# MANAGING THE QUALITY OF COST-PER-CLICK TRAFFIC

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Bobji Mungamuru
October 2010

This dissertation is online at: http://purl.stanford.edu/th614zh7388

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Hector Garcia-Molina, Primary Adviser**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Amin Saberi**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Jeffrey Ullman**

Approved for the Stanford University Committee on Graduate Studies.

**Patricia J. Gumport, Vice Provost Graduate Education**

# Abstract

Traffic quality is critical to the security and viability of an online advertising network's business. As such, how might an advertising network detect and avoid paying for low-quality traffic? The online advertising market can be modeled as a large, multiplayer game with three classes of players: publishers, advertisers and advertising networks. We focus on cost-per-click (CPC) traffic, and identify two distinct-but-related aspects of click quality: *validity* and *targetedness*. Validity refers to whether a click-through is legitimate, whereas targetedness refers to the likelihood that valid clicks become conversions. We study three techniques that influence traffic quality on a network, namely *filtering*, *predictive pricing* and *revenue sharing*. We begin by asking whether it is in an advertising network's interest to filter (i.e., ensure traffic validity) in the first place. This question has been a topic of intense debate in the industry. Our analysis shows definitively that networks do, indeed, have a strong incentive to aggressively filter out invalid traffic. We then consider how a network might use predictive pricing and revenue sharing to attract targeted, high-conversion-rate traffic. We show that effective usage of predictive pricing can yield a competitive edge for a network, and that targetedness has a quantifiable impact on profits. Finally, we study how validity and targetedness can be managed together. What is the relation between filtering, predictive pricing and revenue sharing? We derive efficient, tractable algorithms for computing near-optimal traffic management policies, and also propose strategies to combat publisher click inflation. Perhaps the most important lessons learned from our work are that a) in the online advertising business, traffic quality has a direct impact on profits, and b) traffic quality is often more important than quantity.

# Acknowledgements

Hector Garcia-Molina, you are a most unconventional advisor. Thank you so, so much for supporting me through my unconventional journey through the Ph.D. program. To this day, I am stunned by just how good my life was as a Hector advisee, and just what a nice you guy are. One of the many things I have gained from you is the stubborn confidence that, if I am baffled by the contents of a research paper, it is probably the author's fault and not mine.

Jeff Ullman, Amin Saberi and Kincho Law, thank you for being accommodating even though you had no real reason to be. Jeff and Amin, your exacting standards for this thesis were invaluable in ensuring the quality of the final product. Any remaining errors, of course, are entirely my fault. My respects to the late Rajeev Motwani, who was originally going to be on my reading committee, and who posed the research question that lead to my first published research paper as a Stanford student.

Many thanks to the entire Infolab family for welcoming a lowly electrical engineer into your midst. To the elder statesmen – Zoltan, Paul and David – who am I supposed to make snarky comments to, now that you guys aren't around? Marianne Siroker, I don't know how I would have survived without you.

My sincere gratitude to the Stanford CS department, where I found a home after the trauma that was the EE quals. The view from Gates across Serra Mall was a stark reminder of how good my graduate student life was, and how it could have otherwise turned out. Thank you NSERC, TRUST and PORTIA for paying my bills and keeping me off the streets. Thank you to the Google Security team – "Do Know Evil" – with whom I first worked on the ideas that eventually became this thesis.

I lived a double life during my Stanford years, splitting my time between Palo Alto and the Mission. Crothers Memorial #530, 1025B Hampshire St., 2678 21st St., and 3039 Alma St., each of you occupy a special place in my memories. Julia, Tim, Tony, Anja, Kat, Scott, Hieu and Abalone, life with you at the Shire was a dream. These days, when people ask me where I'm from, I often catch my replying "San Francisco". I'm not through with you yet. Tokyo, I want to blame you for causing me to take so long to write up this thesis. But you know me, I can't stay mad at you for very long.

My deepest gratitude and respect are reserved for my parents and sister – I know how difficult it is for you that I live so far away, which makes your enduring support so much more valuable to me.

Finally, to my legions of adoring fans, there are some of you that I clearly ought to have mentioned here, but didn't. Any such omission is purely a result of my carelessness. So, if you've actually taken the time to read through these acknowledgements, be sure to call me out – I will buy you a drink. Over and out.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Advertisers have been moving online in increasing numbers over the past decade. The online medium has gained popularity because it can potentially reach a very targeted audience, often at a lower cost than traditional print or broadcast media, yielding a higher return on investment.

In any advertising medium, advertisers will be willing to pay more for higher-quality traffic, where "quality" is measured by the likelihood that a member of the target audience will eventually buy the product or service being advertised. The notion of "traffic," of course, depends on the medium – traffic could be television viewers, radio listeners, magazine readers or web surfers. The definition of a "customer," too, depends on the advertiser and medium. For example, if a consumer product is being advertised, a customer is someone who actually purchases the product. If the advertiser is a political campaign, a customer could be defined either as someone who signs up for a mailing list, or someone who makes a campaign contribution. And just as in traditional print and broadcast media, online traffic quality can vary greatly depending on the source of the traffic. In the online context, a "source" is simply a website from which traffic originates.

Broadly speaking, there are three techniques that online advertising networks (e.g., Google, Facebook, AdChina) can use to influence the quality of traffic on their network: filtering, predictive pricing and revenue sharing. *Filtering* refers to distinguishing between valid and invalid traffic, the implication being that the advertiser should not be billed (and the publisher of the website should not be paid) for traffic that is deemed invalid. Examples of invalid traffic include unintentional click-throughs, web crawler traffic and, most famously, click fraud. Click fraud is the act of maliciously clicking on advertisements without having any interest in the product or service being advertised, in order to earn undeserved commissions or manipulate click-through rates. Click fraud has been a hot button issue in the industry. Advertising networks have devoted large amounts of engineering resources (and public relations, too) toward dealing with fraudulent traffic, mainly to appease angry advertisers who feel they are being been "ripped off" by the networks.

*Predictive pricing* is concerned with a different aspect of traffic quality, namely, how targeted it is. For example, it may be more effective for a business school to advertise its executive MBA program in *Barron's* magazine rather than *Wired*. Although both magazines have comparable readerships, the latter is less targeted toward potential students than is the former – leads generated by the latter (presumably) end up enrolling in business school less frequently than leads generated by the former. As such, the business school ought to be charged less for advertising space in *Wired* than *Barron's*. Examples of predictive pricing programs in online advertising include Google's "Smart Pricing" and Yahoo's "Quality-Based Pricing." Finally, *revenue sharing* is the practice of paying out to publishers a fraction of revenue earned from advertisers. Revenue sharing is ostensibly the reason why publishers display advertisements in the first place.

In this thesis, we study how an advertising network can use these three techniques together to optimally manage traffic quality – that is, in a way that maximizes the profits of the advertising network. Our focus will be on *cost-per-click* (CPC) advertising markets, where advertisers are billed (and publishers are paid) only when a user clicks on an advertisement. CPC currently accounts for the vast majority of revenues generated by the online advertising industry.

### A Lemons Market

The CPC market has the two signature characteristics of a "lemons market" [2] – *asymmetric information* and *quality uncertainty*. Asymmetric information means that publishers have more information about the product they are selling than the networks or advertisers do. Moreover, the publishers can secretly engage in click fraud. Quality uncertainty refers to the fact that the products being sold (i.e., click-throughs) are of unequal quality, and that customers cannot be certain about the product's quality at the time of purchase.

In the absence of a reputable, long-lived intermediary to counteract an advertiser's uncertainty about traffic quality, the CPC market would be at risk of a classic lemons-market-style collapse: high-quality publishers would not receive fair value for their high-quality traffic, and consequently advertisers would get inundated with mostly low-quality traffic. Advertising networks act as this intermediary, by filtering and pricing traffic based on its quality, and they extract handsome profits for providing this service.

Of course, for a network to remain in business, it must consistently deliver sufficiently high-quality traffic to advertisers. If substantially higher quality is offered by a competing network, advertisers would quite readily switch over. So, although filtering, predictive pricing and revenue sharing can help advertising networks attract and retain lucrative traffic, applying these tools suboptimally can mean that a network is "leaving money on the table." In a market that, by most estimates (e.g., [12, 6]), is worth tens of billions of dollars per year, the losses due to a suboptimal policy can be tremendous. Managing traffic quality is a real problem of immense practical significance.

### Overview

This thesis is organized as follows. In the remainder of this chapter, we present an overview of the CPC market and a survey of related research. In Chapter 2, we study the validity aspect of click quality, while holding targetedness fixed. Is it in the best interest of advertising networks to ensure the validity of their traffic? Particularly in the context of click fraud, this question has been the subject of much debate in the online-advertising community. We analyze a model of the CPC market in which competing networks are distinguished by their skill at filtering out invalid traffic. The networks are faced with a single decision, namely, how aggressively to filter. We show that networks do, indeed, have a strong incentive to fight fraud.

In Chapter 3, we study the targetedness aspect of click quality, while holding validity fixed. In the presence of competition, how might an advertising network apply predictive pricing and revenue sharing, to most effectively manage traffic of varying targetedness? Here, networks are distinguished by their skill at matching publishers with advertisers, and delivering relevant, targeted ads to users. We derive an algorithm that prescribes an optimal predictive pricing and revenue sharing policy for a network, in response to the policies used by other competing networks.

In Chapter 4, we study the validity and targetedness aspects simultaneously, tying together the results of Chapter 2 and Chapter 3. How might a network use filtering, predictive pricing and revenue sharing together, towards the effective, holistic management of traffic quality? We derive an efficient algorithm that jointly prescribes an optimal policy for filtering, predictive pricing and revenue sharing. Targetedness and validity are closely related, as they both have a direct impact on the advertisers' return on investment (ROI) as well as the revenues of publishers and ad networks. As such, in certain cases, it turns out that predictive pricing can act as a substitute for filtering. We address the publisher's incentive for click inflation, and recast cost-per-acquisition pricing schemes within this context. Finally, we also show how a network can estimate the input parameters necessary for our algorithms, using readily observable traffic data.

In Chapter 5 we present our conclusions, and suggest some avenues for future work in this area.

## 1.1 Cost-Per-Click Advertising

We model the CPC advertising market as a large, multiplayer game between three classes of players: publishers, advertising networks and advertisers. *Publishers* create online content and display advertisements alongside their content. *Advertisers* design advertisements (or, "ads"), as well as bid on *queries* that summarize what their target market might be interested in. *Advertising networks* (or, "ad networks") act as intermediaries between publishers and advertisers by first judging which queries best describe each publisher's content, and then delivering ads to the publisher from the advertisers that have bid on those queries. For example, an ad network might deduce that the query "foreign automobile" is relevant to an online article about cars, and serve an ad for used car inspection reports.

Figure 1.1 is an illustration of the steps in the lifecycle of a click. Suppose a user visits a publisher's webpage (1). The publisher submits a request to an ad network of its choice, for advertisements that are well suited to the content on that page (2). If the user then happens to click on one of the ads, they are redirected to the corresponding advertiser's website (3) – we say that a *click-through* (or, "click") has occurred on that ad. Assuming that the click is deemed valid, the advertiser pays a small amount to the ad network that delivered the ad (5), and a fraction of this amount is in turn paid out to the publisher who displayed the ad (6). The exact amounts paid out to each party depend on several factors including the advertiser's bid, the auction mechanism being used and the revenue sharing agreement between the ad network and publisher.

Advertisers are willing to pay for clicks because some of those clicks may turn into *conversions* (or, "customer acquisitions" or "sales") (4). The exact definition of a conversion depends on the agreement between the advertiser and the ad network, varying from an online purchase to joining a mailing list. In general, a conversion is some action taken by a user. The publishers and ad networks, of course, hope that

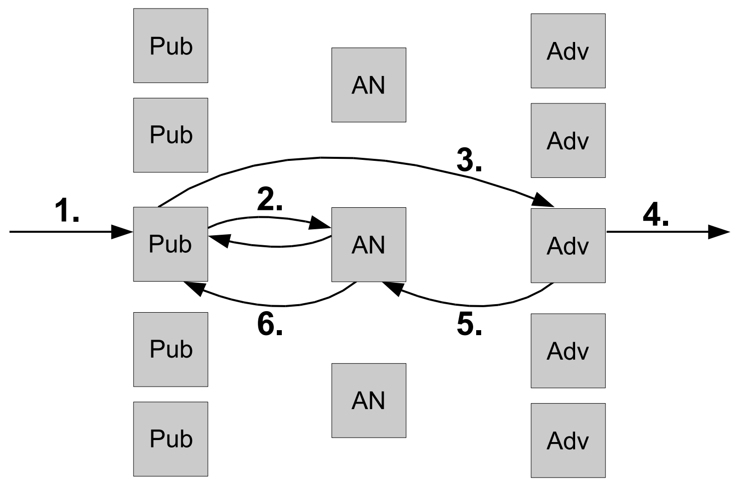


Figure 1.1: Lifecycle of a click-through in CPC advertising.

users will click on ads because of the payment they (i.e., the publishers and ad networks) would receive from the advertiser.

In some cases, a publisher and an ad network are owned by the same business entity. For example, search engines and social networking sites typically display ads from their own network alongside their content. Similarly, a publisher and an advertiser can be owned by the same entity. Online newspapers are a common example. In our model, even if two parties are owned by the same real-world entity, we assume that they will nevertheless both act independently. Recent anecdotal evidence [31] suggests that this may indeed be a reasonable assumption.

### Traffic Quality

As suggested earlier, advertising networks serve a second important function apart from ad delivery, namely, trying to ensure that traffic is of sufficiently high quality. We model two key aspects of traffic quality – validity and targetedness.

*Valid* clicks can be (informally) defined as click-throughs that have nonzero probability of leading to a conversion (it is still a topic of some debate what the exact definition of an valid click should be). Clicks that are not valid are considered *invalid*. Invalid clicks include fraudulent clicks as well as various sources of unintentional clicks. For example, if a user unintentionally double-clicks on an ad, only one of the two clicks has a chance at becoming a conversion, so the second click is considered invalid. As such, we will typically speak of valid and invalid clicks, rather than “legitimate” and “fraudulent” clicks. Advertisers are not billed for clicks that ad networks detect as invalid, although the user is still forwarded to the advertiser’s site.

*Targetedness* refers to the likelihood that a valid click becomes a conversion. Over the course of daily operations, ad networks collect data in order to estimate the conversion rates of various flows of click traffic. Using these estimates, ad networks have the ability to price discriminate based on quality – in particular, advertisers can be offered discounts on poorly targeted traffic. Note that targetedness can vary greatly depending on the query. For example, traffic from the *Bloomberg* financial news pages might have a low conversion rate on advertisements returned by the query “digital camera.” On the other hand, it is probably much better targeted to queries for “home mortgages” or “mutual funds.” Indeed, the ad network’s skill at matching publishers’ content with advertisers’ ads, as well as the quality of an advertiser’s ads themselves, can have a significant impact on the conversion rates resulting from the flow of click-through traffic.

### Decision Variables

Each class of players in our model is faced with a different decision problem. Publishers must choose which ad networks to sell their click-throughs on, and how to allocate traffic across these networks. Advertisers must decide what value to place on clicks coming from each ad network, and how much to bid for the clicks at auction.

Both the publishers and the advertisers conduct business with the ad networks. However, there is a key conceptual difference between the decisions they face. Publishers, essentially, are choosing between the ad networks. They allocate their finite inventory of click-throughs across the ad networks. Advertisers, on the other hand, do not choose between ad networks. They are willing to buy an unlimited number of click-throughs from any and all of the ad networks, as long as the ROI on each ad network is sufficiently high. Moreover, only publishers have users – advertisers and ad networks do not. For this reason, we distinguish between publishers, advertisers and ad networks in our model, even when they are owned by the same business entity.

Ad networks are faced with the decision of a) how aggressively to filter for invalid clicks, and b) how aggressively to price discriminate based on targetedness. In the chapters that follow, we will show how to quantify an ad network’s “aggressiveness.” An ad network’s aggressiveness impacts the profits of both publishers and advertisers. If an ad network is more aggressive, the publisher is effectively paid for fewer clicks. However, the advertiser is also billed for fewer clicks, causing its ROI to increase and thus its valuation to increase as well. Conversely, when an ad network is less aggressive, publishers receive higher payments for their clicks. However, advertisers are billed for more potentially nonconverting clicks, causing valuations to drop. As mentioned earlier, we are interested in the strategic interactions between ad networks, particularly as they relate to the task of selecting filtering, predictive pricing and revenue sharing policies.

## 1.2 Related Work

To gain a thorough understanding of the online advertising market, we must consider the viewpoints of all three classes of market participants, namely, the publishers, advertisers and advertising networks.

Much of the public discussion on online advertising is from the publishers’ and advertisers’ viewpoint. The threat of fraud to the advertising business model and the technical challenge of detecting fraud have been topics of great concern in the industry (e.g., [14]). There have been many informal conjectures in online forums and the media asking whether advertising networks have any incentive at all to fight fraud (e.g., [32, 28, 29, 30, 31]). The arguments, while sometimes intuitive, typically are not backed by a sound economic analysis. Thus, the conclusions arrived at differ widely. In Chapter 2, we attempt to fill this gap by performing just such an analysis. We focus mainly on click fraud (more generally, validity) in pay-per-click advertising systems. Click fraud refers to the act of clicking on advertisements, either by a human or a computer, in an attempt to gain value (e.g., deplete a competitor's budget or earn commissions) without having any actual interest in the advertiser's website. Manipulation of the click-through rate is one incentive [21]. Click fraud is probably the most prevalent form of online advertising fraud today [10, 14]. See [9] for a discussion of the various types of ad fraud. We do not analyze the reasons why fraud occurs (see [9] for such an analysis). Instead, given that fraud does occur in practice, we ask whether advertising networks have an incentive to aggressively combat it.

There is some existing research that considers the advertisers' incentives in isolation. For example, there is a large body of work on auction design (e.g., see [13]) that is closely related, but orthogonal, to our work. Although we include in our model each advertiser's private valuation of click-throughs (which are subsequently translated to auction bids), we never specify or restrict the exact rules of the auction employed by each ad network. Instead we treat each network's auction as a black box, and only make some weak assumptions on the properties of the auction. Optimal competitive bidding, from the advertiser's perspective, has also been studied extensively. For example, see [8] and [22].

One of the main outcomes of our work is an optimization algorithm to be applied by advertising networks (although we explicitly model the incentives of all parties). We ask how a network can apply filtering, predictive pricing and revenue sharing together, towards maximizing revenues. The practice of predictive pricing in the cost-per-click advertising market is relatively new. Google's "Smart Pricing" [16] and Yahoo's "Quality-Based Pricing" [34] are the most well-known examples of such schemes. At the same time, publishers often lament "being smart-priced" (e.g., [20]). To the authors' knowledge, there has been no formal analysis thus far of how to apply predictive pricing and share revenue optimally, apart from our work in Chapters 3 and 4. We suspect that it is currently being applied in an ad hoc manner. Perhaps there has been no need for a principled approach – the fact that any predictive pricing was being done at all may have been sufficient to satisfy publishers and advertisers. However, as more networks adopt such programs, we feel that a principled approach will become necessary. Revenue sharing as a tool for incentivizing supplier behaviour has been studied before, in the supply chain literature (e.g., [7]). Although applicable to the CPC market (i.e., using revenue sharing to attract click inventory from publishers), this idea is not well explored in the online advertising context.

Our model of the online advertising market considers a single keyword query at a time. For a treatment on how an advertising network can be more effective at matching publishers and advertisers, see [24]. Although our focus is on CPC, cost-per-acquisition (CPA) is a pricing scheme that has gained some traction recently. In Chapter 4, we demonstrate how some of the main benefits of CPA can be achieved using an existing CPC infrastructure. See [27] for a study on CPA pricing mechanisms. The classic work on "lemons markets" [2] is a theoretical justification for why intermediaries (i.e., advertising networks) should exist between publishers and advertisers, in the first place.

The computational tools we leverage in Chapters 3 and 4 have been implemented in software only relatively recently, which may explain why there has been a lack of formal analysis applied to this problem. In particular, the key computational steps involve numerically solving a sequence of geometric programming problems [5]. Geometric programming is a convex optimization technique that typically finds application in engineering fields. Its application to online advertising and game theory, to our knowledge, is new. The implementation of our algorithm in our experiments uses the CVX [18] library in MATLAB.

# Chapter 2

# Should Ad Networks Bother Fighting Fraud?

[0]An abbreviated version of this chapter appears in [26].

## 2.1 Introduction

Advertising fraud, particularly click fraud, has been a growing concern to the online advertising industry. Recall that the online advertising market is comprised of three classes of parties: content publishers (e.g., blogs, cnn.com), advertising networks (e.g., Google, MSN, Facebook, AdBrite), and advertisers. Users engage in the market indirectly by clicking on advertisements on publishers' content pages.

At first glance, the incentives regarding fighting fraud seem somewhat perverse: If an advertiser is billed for clicks that are fraudulent (or more generally, invalid), the advertising network's revenues increase. As such, is it even in an advertising network's interest to fight fraud at all? Perhaps it makes more sense for an advertising network to just "let it happen" [3, 17]. If not, can an advertising network actually gain a market advantage by aggressively combating fraud?

In this chapter, we address these and other questions by studying the economic incentives related to filtering, and how these economic incentives might translate into behavior. An economic analysis of online advertising fraud (or, "ad fraud") is appropriate because, unlike many other online security threats, ad fraud is primarily motivated by financial gain. Successfully committing ad fraud yields immediate monetary gains for attackers at the expense of the victims (although it may not seem immediately clear who the actual "victims" of ad fraud are).

Our goal, then, is to construct and analyze a simplified economic model that hones in on the market effects of filtering. However, conducting such an analysis is difficult because faithful models of the market can quickly become very complex. A complete specification of the players' types, decision variables and signals would be intractable. For example, a publisher's type includes, among other things, the volume of traffic they receive, the quality of their content, and their user demographics and interests. Advertisers can be differentiated by the size of their advertising budgets, their valuation of traffic that they receive through online ads, the quality of their campaign, and their relevance to particular demographics. Ad networks can differ in their ability to detect invalid clicks, as well as the quality and relevance of their ad serving mechanisms. The challenge is to retain just those aspects that are relevant to filtering out invalid traffic.

#### Invalid Clicks

One of the most important services provided by an advertising network is the detection and filtering out of invalid clicks. Conceptually, each ad network has an algorithm that they apply on a click-by-click basis, to decide whether each click is valid or invalid. Recall that invalid clicks can be defined as click-throughs that have zero probability of leading to a conversion. Examples of invalid clicks include click fraud, unintentional clicks, double clicks and web crawler traffic. Advertisers are not billed for clicks marked invalid, although the user is still forwarded to the advertiser's site.

The algorithms used by ad networks to detect invalid clicks are, of course, prone to error. In particular, their algorithms may produce *false negatives* by identifying invalid clicks as valid, and *false positives* by identifying valid clicks as invalid. A false negative implies that an advertiser has been unfairly billed for a click that could not lead to a conversion. A false positive, on the other hand, is a valid click that the advertiser has received for free.

We assume that if an ad network detects an invalid click, it will then actually mark the click invalid (and therefore not bill the advertiser for it). The reader (and indeed, many advertisers) may wonder whether an ad network might detect a click as invalid, but then choose to mark it valid anyways. However, for our purposes, doing so is equivalent to simply not detecting the click as invalid, since the user is forwarded to the advertiser's site irrespective of the ad network's decision. Therefore, without loss of generality, we will use the phrases “detect as invalid” and “mark as invalid” synonymously.

In our model, we make an important distinction between how effective and how aggressive an ad network is at detecting invalid clicks. Effectiveness refers to the quality of the algorithms used to detect invalid clicks, whereas aggressiveness refers to the manner in which the algorithms are applied. For example, suppose an ad network uses an algorithm that takes as input a single click, and outputs a score between 0 and 1, where a higher score indicates a higher confidence that the click is invalid. The ad network might then run the algorithm on each click, and apply a threshold to the output score to decide whether to mark the click invalid e.g., they might mark clicks invalid if the score is 0.5 or higher. In this example, effectiveness refers to the accuracy of the scoring algorithm – a good algorithm should output a high score when invalid clicks are input and a low score when valid clicks are input. Aggressiveness, on the other hand, is related to the threshold value chosen – a more aggressive ad network would choose a lower threshold, even though doing so may result in more false positives.

Each ad network in our model uses an algorithm whose effectiveness is known (i.e., it is an inherent property of each ad network). The aggressiveness, on the other hand, is a decision variable that is subject to change over time. Our goal here is to study how aggressive a rational ad network would be in equilibrium:

- Is it in an ad network's interest to aggressively detect invalid clicks?
- If an ad network is more effective at detecting invalid clicks, do they gain an edge in the market?

**Overview**

The remainder of this chapter is structured as follows. In Section 2.2 we analyze a simplified single-period model of the online advertising market (a sequential game), and demonstrate our results in this model. Although the model in Section 2.2 is simplified, it captures the essential intuitions behind our results. In Section 2.3, we present a more realistic (and more complex) multi-period model of the market, as a sequence of interactions between publishers, ad networks and advertisers over a long horizon. Much of the intuition and the proof techniques from the single-period model turn out to be directly applicable in the multi-period setting. Under some additional assumptions, we establish essentially the same result in the multi-period setting as we do for the single-period game. We provide a thorough discussion in Section 2.4 of the impact of various assumptions in our model, and conclude in Section 2.6. Detailed deriviations of our results are presented in the appendices.

The main results in this chapter are equilibrium predictions about how high- and low-quality publishers would respond to various decisions by ad networks. Informally, “high-quality” publishers are those whose traffic contains a relatively low amount of invalid clicks, whereas “low-quality” publishers have a relatively high proportion of invalid clicks (precise definitions will be presented shortly). Our results can be summarized as follows:

1. In equilibrium, all publishers (i.e., even the low-quality ones) will prefer to display ads from the ad network that promises advertisers the highest quality traffic. Why? Because higher-quality click traffic leads to increased ROI for advertisers, which in turn leads to higher bids and more money for publishers. Stated differently, the ad network that promises the highest-quality traffic will not only attract the high-quality publishers, but also the low-quality ones.

2. All other factors remaining equal, the ad network that is most effective at detecting invalid clicks is able to provide advertisers the highest quality traffic. Why? Because filtering effectively ensures that advertisers are billed for fewer invalid (and hence, nonconverting) clicks.

3. However, the ad network that is most effective at detecting invalid clicks can only deliver high-quality traffic if it is sufficiently aggressive. In particular, it is suboptimal to simply mark all clicks valid. As its lead in effectiveness narrows, it is forced to be increasingly aggressive. As ad networks become more aggressive, the high-quality publishers benefit disproportionately.

The first conclusion – that all publishers will choose the same ad network – is perhaps the most surprising, and certainly is the most difficult to prove. It is easy to believe that high-quality publishers will be attracted to the ad network with the highest quality traffic, due to the increased bids on that network. However, we predict that even the low-quality ones will prefer this network. This prediction is surprising because, intuitively, we would expect a low-quality publisher to send traffic to a less-aggressive network, hoping that fewer clicks will be rejected as invalid.

The key phenomenon behind this prediction is that the algorithms used by ad networks are prone to error – no ad network is able to detect 100% of the invalid clicks coming from a publisher (unless, of course, every single click is marked invalid). Even though an effective, aggressive ad network will detect most invalid clicks, a few will always "slip through the cracks." The advertisers' bids will be high enough, in equilibrium, that the few invalid clicks that are mistaken for high-quality traffic generate sufficient revenues to attract even the low-quality publishers.

## 2.2 Single-Period Model

In this section, we model the online advertising market as a single-period, two-step sequential game between publishers, ad networks and advertisers. The two steps of our game are as follows:

1. In the *first step*, ad networks decide how aggressively to filter for invalid clicks.

2. In the *second step* (or, the *subgame*), the publishers and advertisers react to the ad networks' decisions. Publishers decide which ad networks to sell their clicks on, and advertisers decide how much they are willing to pay for clicks from each ad network.

After the second step, profits are realized: a) publishers sell clicks (i.e., display ads) on their chosen ad networks, b) ad networks mark a subset of these clicks invalid and c) advertisers pay the ad networks for the clicks that are marked valid. The game in its entirety (i.e., the first step followed by the second) is referred to as the *supergame*. Our goal is to predict how aggressive ad networks would be in equilibrium, and how the market of publishers and advertisers would react. Our single-period model is simplified in four important ways:

1. It is a single-period game.
2. Ad networks act first, followed by the publishers and advertisers.
3. There is no private information.
4. There are no auctions for click-throughs.

In Section 2.3, we relax the above simplifications to arrive at a more realistic model of the market. However, our main results and intuitions will still apply.

### 2.2.1 Model

Table 2.1 is a summary of the notation used in Section 2.2. Consider a market for click-throughs on a single query. The market is comprised of $I$ publishers, $J$ ad networks, and $K$ advertisers. Click-throughs originate on publishers' pages, and are forwarded on to advertisers via the ad networks.

Suppose each publisher $i$ receives a total volume $V_i$ of click-throughs on its websites. Of these $V_i$ clicks, suppose a fraction $r_i$ are valid. That is, publisher $i$ receives $r_iV_i$ valid clicks and $(1 - r_i)V_i$ invalid clicks. We define the *average publisher quality*, $\bar{r}$, as follows:

$$\bar{r} \equiv \frac{\sum_i r_i V_i}{\sum_i V_i} \tag{2.1}$$

We say that a publisher $i$ is *high-quality* if $r_i \geq \bar{r}$. Conversely, a publisher $i$ is *low-quality* if $r_i < \bar{r}$.

**Ad Networks** We define $x_j \in [0, 1]$ as a measure of network $j$'s aggressiveness in filtering. Concretely, $x_j$ is the fraction of valid clicks that ad network $j$ mistakenly marks invalid (i.e., false positives, where a "positive" is a click that is marked invalid). Intuitively, a higher $x_j$ means $j$ is giving away more clicks for free. $x_j$ is a decision variable for network $j$.

We also define a parameter $\alpha_j \in [0, 1]$ to measure ad network $j$'s filtering effectiveness. For a given $x_j$ chosen by network $j$, $x_j^{\alpha_j}$ is the fraction of invalid clicks that $j$ correctly marks invalid (i.e., true positives). $\alpha_j$ is a fixed, intrinsic parameter that describes network $j$ – that is, it cannot be controlled by network $j$.

A lower value of $\alpha_j$ indicates that network $j$ is more effective at filtering invalid clicks. If $\alpha_j = 0$, it means network $j$'s filtering algorithm can perfectly distinguish between valid and invalid clicks, without ever making any mistakes i.e., the true positive rate is 1 even when the false rate is arbitrarily small. If $\alpha_j = 1$,

**Player Parameters**

| | |
|---|---|
| $V_i$ | Volume of clicks on publisher $i$'s site |
| $r_i$ | Fraction of publisher $i$'s clicks that are valid |
| $\alpha_j$ | Effectiveness of ad network $j$ |
| $y_k$ | Advertiser $k$'s revenue per conversion |
| $R_k$ | Advertiser $k$'s target ROI |

**Decision Variables**

| | |
|---|---|
| $c_{ij}$ | Fraction of publisher $i$'s clicks sent to ad network $j$ |
| $x_j$ | Aggressiveness (false positive rate) of ad network $j$ |
| $v_{kj}$ | Advertiser $k$ valuation of ad network $j$'s clicks |

**Derived Parameters**

| | |
|---|---|
| $N_{ij}$ | Fraction of $i$'s clicks marked valid by $j$ |
| $\pi_{ijk}$ | Publisher $i$'s revenue from clicks sold to advertiser $k$ via ad network $j$ |
| $\pi_i$ | Publisher $i$'s total revenue |
| $Y_{kj}$ | Advertiser $k$'s revenue from ad network $j$'s clicks |
| $Z_{kj}$ | Number of clicks advertiser $k$'s is billed for by ad network $j$ |
| $R_{kj}$ | Advertiser $k$'s ROI on ad network $j$ |
| $\eta_j$ | Ad network $j$'s total revenue |
| $M_j$ | Ad network $j$'s multiplier |

**Constants**

| | |
|---|---|
| $I$ | Number of publishers |
| $J$ | Number of ad networks |
| $K$ | Number of advertisers |
| $h$ | Fraction of revenue paid by ad networks to publishers |
| $\beta$ | Fraction of clicks that become conversions |

Table 2.1: Notation used in Section 2.2.

the false positive rate and true positive rate are both equal to $x_j$, i.e., network $j$ is randomly marking clicks invalid with probability $x_j$.

The functional relationship, $f_j$, between network $j$'s true positive rate and false positive rate is generally referred to as a *receiver operating characteristic*, or "ROC curve." The functional form of $f_j$ that we have chosen in our model, $f_j(x_j;\alpha_j) \equiv x_j^{\alpha_j}$, gives a simple single-parameter family of curves, while capturing the concave shape of typical real-world ROC curves.

A simple calculation then shows that

$$N_{ij} \equiv (1 - x_j)r_i + (1 - x_j^{\alpha_j})(1 - r_i) \tag{2.2}$$

is the fraction of publisher $i$'s clicks marked valid by $j$. $N_{ij}$ is simply a weighted average of the true negative rate and the false negative rate on network $j$. As we demonstrate in Section 2.2.4, the ratio of true negatives to false negatives, $\frac{1-x_j}{1-x_j^{\alpha_j}}$ in our model, is a key measure of "ad network quality" in our model.

**Publishers** Next, we define $c_{ij}$ as the fraction of publisher $i$'s click-throughs that are on advertisements from ad network $j$. In other words, $c_{ij}$ is $i$'s "allocation" of traffic to network $j$. $c_{ij}$ is a decision variable for publisher $i$.

Then, $V_i c_{ij}$ will be the total volume of clicks publisher $i$ sends to network $j$, and from Equation (2.2),

$$N_{ij} V_i c_{ij} \tag{2.3}$$

is the total number of $i$'s clicks that ad network $j$ marks valid. Note that all clicks are forwarded onto advertisers, not just those marked valid. Marking clicks valid or invalid only affects how much advertisers are billed and how much publishers and ad networks are paid.

Some advertising networks sell ad impressions (or "page views" or "eyeballs"), rather than click-throughs. However, under certain assumptions (detailed in Section 2.4), it is equivalent to think of click-throughs as the items that are bought and sold in the market.

**Advertisers** For simplicity, let us assume each advertiser $k$ receives an equal fraction of the clicks from each ad network (or equivalently, ad networks choose an advertiser uniformly at random on each ad impression). That is, there is no competition amongst advertisers for click-throughs, and each advertiser $k$ receives a fraction $\frac{1}{K}$ of each ad network $j$'s clicks (recall that there are $K$ advertisers in total). From Equation (2.3), we get that

$$\frac{1}{K} N_{ij} V_i c_{ij} \tag{2.4}$$

is the number of publisher $i$'s clicks that network $j$ marks valid, and forwards on to each advertiser $k$. As we will see, the manner in which an ad network allocates clicks across advertisers does not impact our results. Thus, in Section 2.3, we will replace our simplifying assumption by a more realistic one (i.e., keyword auctions), without changing our conclusions.

**Payments** We define $v_{kj}$ as advertiser $k$'s valuation of network $j$'s clicks. $v_{kj}$ is a decision variable for advertiser $k$.

In the current simplified setting, we assume that the amount $k$ pays $j$ per click is equal to $v_{kj}$ i.e., we assume that the amount paid per click does not vary across clicks (we will drop this assumption in Section 2.3, once we introduce auctions into our model). In this setting, advertiser $k$ chooses $v_{kj}$ using a simplified optimization problem, described in Equation (2.14).

From Equation (2.4),

$$\frac{1}{K} N_{ij} V_i c_{ij} v_{kj} \tag{2.5}$$

is the total amount paid by advertiser $k$ to network $j$, for publisher $i$'s traffic. A fraction $h$ of each dollar of revenue is in turn paid out by each ad network to the publisher from which the click originated. Therefore, from Equation (2.5),

$$\pi_{ijk} \equiv \frac{1}{K} N_{ij} V_i c_{ij} v_{kj} h \tag{2.6}$$

is the amount ad network $j$ pays publisher $i$ for clicks that are sent to advertiser $k$.

The fraction $h$ in Equation (2.6) is referred to as the *revenue share*. In this chapter, we assume that all networks offer the same revenue share, since our focus is on the validity aspect of traffic quality. In Chapters

3 and 4, we model the revenue share as a decision variable for each network $j$.

Each publisher $i$'s total profit, $\pi_i$, is simply $\pi_{ijk}$ summed over all networks and advertisers:

$$\pi_i \equiv \sum_j \sum_k \pi_{ijk} \tag{2.7}$$

The remainder of the revenue is retained by the ad network as profit. In other words, ad network $j$'s profit is simply the total revenue received from advertisers, less the payments made to publishers – a fraction $h$ of each dollar of revenue from the advertisers is paid out, and the other $1 - h$ is retained. Therefore, using (2.6), ad network $j$'s total profit, $\eta_j$, is simply:

$$\eta_j \equiv \frac{1}{K}(1-h) \sum_k \sum_i N_{ij} V_i c_{ij} v_{kj} = \frac{1-h}{h} \sum_k \sum_i \pi_{ijk} \tag{2.8}$$

Now, ad network $j$ receives $\sum_i r_i V_i c_{ij}$ valid clicks in total. Suppose each valid click becomes a conversion with probability $\beta$. Then,

$$\beta \sum_i r_i V_i c_{ij} \tag{2.9}$$

will be the expected number of converted clicks on network $j$. Recall that a fraction $\frac{1}{K}$ of these clicks are sent to advertiser $k$. This means, assuming advertiser $k$ earns $y_k$ dollars from each conversion, $k$ will earn

$$Y_{kj} \equiv \left( \sum_i r_i V_i c_{ij} \right) \frac{1}{K} \beta y_k \tag{2.10}$$

dollars of total revenue from conversions of clicks coming from ad network $j$.

Recall from (2.3), that ad network $j$ marks $N_{ij} V_i c_{ij}$ of publisher $i$'s clicks valid. Therefore,

$$Z_{kj} \equiv \frac{1}{K} \sum_i N_{ij} V_i c_{ij} \tag{2.11}$$

is the total number of clicks that ad network $j$ marks valid, and bills advertiser $k$ for. In the current simplified setting, advertiser $k$ will pay $v_{kj} Z_{kj}$ dollars in total (i.e., exactly $v_{kj}$ per click) to ad network $j$ for these clicks. Combining Equations (2.10) and (2.11), we conclude that advertiser $k$'s *return on investment* (ROI) on ad network $j$ is:

$$R_{kj} \equiv \frac{Y_{kj}}{v_{kj} Z_{kj}} \tag{2.12}$$

### 2.2.2 Player Objectives

In our single-period model, each publisher $i$ is described by the ordered pair $(r_i, V_i)$, which measures the quality and volume of $i$'s click-through traffic. The allocations $\{c_{ij}\ \forall j\}$, on the other hand, are $i$'s decision variables. Publisher $i$'s objective is to choose $\{c_{ij}\ \forall j\}$ so that its total profit across all ad networks, $\pi_i$, is maximized. That is, publisher $i$ chooses $\{c_{ij}\ \forall j\}$ by solving the following optimization problem:

$$\max_{\{c_{ij}\}} \pi_i \ \text{ s.t. } \sum_j c_{ij} = 1 \ \text{ and } \ c_{ij} \geq 0 \tag{2.13}$$

In most cases, publisher $i$ will choose $c_{ij} = 1$, where $j$ is the most profitable network for publisher $i$. But, in the event that $i$ is indifferent between two or more networks, fractional allocations of traffic are possible.

Advertiser $k$'s decision variables are the valuations $\{v_{kj}\ \forall j\}$ i.e., the amount that $k$ pays ad network $j$ per click. Each advertiser $k$'s type is the ordered pair $(y_k, R_k)$, where $R_k$ is $k$'s target ROI. Recall that $y_k$ is the amount $k$ earns per converted click. Intuitively, $R_k$ represents the returns $k$ can realize by advertising offline, or through other alternative media.

In our simplified single-period model, advertiser $k$'s objective is not to maximize profits – if $k$ were a profit maximizer it would simply set $v_{kj} = 0$, since it is guaranteed a $\frac{1}{K}$ fraction of the traffic. Instead, we assume for simplicity that $k$'s sole objective is to achieve an ROI of exactly $R_k$ on every single ad network. That is, $k$ selects $v_{kj}$ such that:

$$R_{kj} = R_k\ \forall j \tag{2.14}$$

In Section 2.3, where advertisers compete for traffic via auctions, we model advertisers are maximizing profits from online advertising, subject to the constraint that $R_{kj} \geq R_k$. In our current setting, however, advertisers don't need to compete for traffic – rather, each advertiser receives a fixed fraction $\frac{1}{K}$ of the traffic, and simply tries to attain and ROI of $R_k$ on each network $j$. The simplification in Equation (2.14) is reasonable because, in the more realistic setting of Section 2.3, the constraint $R_{kj} \geq R_k$ turns out to be binding at the optimum.

Finally, ad network $j$'s type is its effectiveness $\alpha_j$, whereas $j$'s decision variable is the level of aggressiveness $x_j$. Ad network $j$ chooses $x_j$ so that their profit $\eta_j$, in Equation (2.8), is maximized.

### 2.2.3 Numerical Example

To gain some intuition for the single-period model, we present a small numerical example. Note that this example is not an equilibrium scenario.

Consider a market with just $I = 2$ publishers, $J = 2$ ad networks and $K = 1$ advertiser. Assume $r_1 = 0.9$, $r_2 = 0.1$, and $V_1 = V_2 = 100$ i.e., the publishers receive the same volume of clicks but publisher 1's traffic is of much higher quality. Let $\alpha_1 = 0.322$ and $\alpha_2 = 0.555$ i.e., ad network 1 is more effective than ad network 2 at detecting invalid clicks.

Now, suppose $x_1 = 0.5$ and $x_2 = 0.4$ i.e., ad network 1 is more aggressively detecting invalid clicks than ad network 2. Using Equation (2.2), we can compute $N_{ij}$ for $i = 1, 2$ and $j = 1, 2$. The results are shown in Table 2.2. From Table 2.2, we see that ad network 2 would mark more clicks valid than ad network 1,

| | Ad Network 1 | Ad Network 2 |
|---|---|---|
| Publisher 1 | 47% | 58% |
| Publisher 2 | 23% | 42% |

Table 2.2: Fraction $N_{ij}$ of publisher $i$'s clicks marked valid by ad network $j$.

for both publishers 1 and 2. Note that even though only 10% of publisher 2's clicks are valid, ad network 2 marks 42% of them valid. Thus, ad network 2 will be billing advertisers for many invalid clicks that do not generate revenue.

Suppose that $v_{11} = 15$ and $v_{12} = 10$ i.e., the advertiser is willing to pay $15 per click on ad network 1 and $10 per click on ad network 2. The advertiser is willing to pay more per click on ad network 1 than 2 because ad network 1 is giving more clicks away for free (i.e., marking fewer clicks valid), leading to a higher ROI. Again, this is just a numerical example – we are not claiming that these are equilibrium valuations. Assume $h = 0.4$ i.e., both ad networks pay out 40% of their revenue to publishers.

Define $\pi_{ij}$ as the total profit to publisher $i$ if it decides to sell all of its clicks to ad network $j$ exclusively i.e., $c_{ij} = 1$. From Equation (2.6) we get that $\pi_{ij} = N_{ij}V_i v_{1j} h$ (recall that $K = 1$ in this example). Table 2.3 shows $\pi_{ij}$ for $i = 1, 2$ and $j = 1, 2$, assuming $v_{11} = 15$ and $v_{12} = 10$. We see that publisher 1 would earn

| | Ad network $j = 1$ | Ad network $j = 2$ |
|---|---|---|
| Advertiser's valuation, $v_{1,j}$ | $15.00 | $10.00 |
| Publisher 1's profit, $\pi_{1,j}$ | $282 | $232 |
| Publisher 2's profit, $\pi_{2,j}$ | $138 | $168 |

Table 2.3: Publisher profits $\pi_{ij}$ assuming $v_{11} = 15.00$ and $v_{12} = 10.00$.

$282 from ad network 1 for all of its traffic compared to only $232 from ad network 2. Publisher 2 would earn $168 from ad network 2 compared to only $138 from ad network 1. Therefore, based on the valuations $v_{11} = 15$ and $v_{12} = 10$, publisher 1 would send traffic to ad network 1 and publisher 2 would send traffic to ad network 2.

The important features of this numerical example are as follows. Ad network 2 marks more clicks valid than ad network 1, for both publishers 1 and 2. However, the advertiser is willing to pay more for ad network 1's clicks than ad network 2's clicks, since per-click ROI is higher on ad network 1. Based on total profits, publisher 1 would prefer to send its traffic to ad network 1 (i.e., since $\pi_{11} > \pi_{12}$), whereas publisher 2 would prefer to send its traffic to ad network 2 (i.e., since $\pi_{22} > \pi_{21}$).

### A Subgame Equilibrium

We can carry this numerical example a bit further to predict what would happen in equilibrium.

Suppose publisher 1 does indeed send all of its clicks to ad network 1 and publisher 2 sends its clicks to ad network 2. The publishers' profits computed in Table 2.3 assumed $v_{11} = 15$ and $v_{12} = 10$. However, the advertiser had not yet accounted for the difference in traffic quality between the ad networks. Ad network 1's clicks (which come from publisher 1) are much higher quality than ad network 2's clicks (which come from publisher 2). The advertiser, therefore, will adjust $v_{11}$ and $v_{12}$ to reflect this quality difference.

Using Equations (2.10), (2.11), (2.12) and (2.14), the advertiser decides that it is now willing to pay $v_{11} = 19.15$ per click to ad network 1, and only $v_{12} = 2.38$ per click to ad network 2. Using these new valuations, we recalculate the publishers' profits. The results are shown in Table 2.4. From Table 2.4, we see that both publisher 1 and 2 make higher profits with ad network 1 than 2, so they both would now send all their traffic to ad network 1.

| | Ad Network 1 | Ad Network 2 |
|---|---|---|
| Advertiser's valuation, $v_{1,j}$ | \$19.15 | \$2.38 |
| Publisher 1's profit, $\pi_{1,j}$ | \$360 | \$55 |
| Publisher 2's profit, $\pi_{2,j}$ | \$176 | \$40 |

Table 2.4: Publisher profits $\pi_{ij}$ assuming $v_{11} = 19.15$ and $v_{12} = 2.38$.

However, publisher 2's decision to use ad network 1 instead of 2 will lower ad network 1's traffic quality. Thus, the advertiser will again re-evaluate what it is willing to pay per click on each network. It now sets $v_{11} = 14.29$ and $v_{12} = 5$. Re-calculating the values in Table 2.4, we get Table 2.5. Observe that under the updated valuations, both publishers' choices remain unchanged – they both would still display ad network 1's ads. As such, the advertiser no longer needs to recompute its valuations. Therefore, $v_{11} = 14.29$ and $v_{12} = 5$, with both publishers choosing ad network 1 (i.e., $c_{11} = c_{21} = 1$) is an *equilibrium* for the subgame when $x_1 = 0.5$ and $x_2 = 0.4$.

| | Ad Network 1 | Ad Network 2 |
|---|---|---|
| Advertiser's valuation, $v_{1,j}$ | \$14.29 | \$5.00 |
| Publisher 1's profit, $\pi_{1,j}$ | \$269 | \$116 |
| Publisher 2's profit, $\pi_{2,j}$ | \$131 | \$84 |

Table 2.5: Publisher profits $\pi_{ij}$ assuming $v_{11} = 14.29$ and $v_{12} = 5.00$.

The interesting feature of this equilibrium is that both publishers agree on their choice of ad network. As we show next, this effect is not at all specific to this numerical example – it applies irrespective of the numbers of publishers, ad networks and advertisers, and irrespective of the distribution of their types.

### 2.2.4 Equilibria

We will now characterize the equilibria of our single-period sequential game. Recall that the ad networks act first i.e., in the first step, each ad network $j$ must choose $x_j$ without knowing the publishers' and advertisers' decisions. In the subgame, publishers and advertisers simultaneously decide on their allocations and valuations, respectively (recall that we refer to the second step as the *subgame*, while the complete sequential game including both the first and second steps is referred to as the *supergame*). Publishers and advertisers know $\{x_j \; \forall j\}$ (i.e., the outcome of the first step) before making their decisions in the second step. In this sense, our single-period sequential game captures how market participants react to an ad network's choice of aggressiveness level.

**Assumption 1.** *Publishers are not homogeneous with respect to quality of clicks:*

$$\exists m, n \text{ s.t. } r_m \neq r_n \tag{2.15}$$

Assumption 1 will hold in any nontrivial problem instance. Define ad network $j$'s *multiplier*, $M_j$, as the ratio of the total number of converted clicks to the total number of clicks $j$ marks valid, assuming all publishers send all their traffic to $j$ (i.e., $c_{ij} = 1 \ \forall i$, keeping $j$ fixed). Combining Equations (2.3) and (2.9) with $c_{ij} = 1 \ \forall i$, we get

$$M_j = \frac{\beta \sum_i r_i V_i}{\sum_i N_{ij} V_i}. \tag{2.16}$$

Under Assumption 1, in equilibrium, all publishers will choose the same ad network in the subgame:

**Theorem 1.** *Suppose Assumption 1 holds. Then, for any first-step outcome, there exist pure-strategy Nash equilibria (NE) in the subgame. Moreover, in any pure-strategy subgame NE, there exists a* $j^* \in 1, \ldots, J$ *such that*

$$c_{ij^*} = 1 \ \forall i \tag{2.17}$$

$$v_{kj^*} = \frac{y_k}{R_k} M_{j^*} \ \forall k, \tag{2.18}$$

*whereas for all* $j \neq j^*$,

$$c_{ij} = 0 \ \forall i \tag{2.19}$$

$$v_{kj} \leq \frac{y_k}{R_k} M_j \ \forall k. \tag{2.20}$$

*We say that ad network* $j^*$ *has been "chosen" by publishers and advertisers in that equilibrium.*

*Proof.* See Appendix A.2. □

Theorem 1 is a very strong statement. It holds for any number of publishers, ad networks and advertisers, and irrespective of the distribution of player types. Equation (2.17) says that all publishers will choose the same ad network $j^*$ in the subgame – like the example in Section 2.2.3, publishers will never disagree (in equilibrium) on which ad network they choose. Or, from the ad networks' perspective, a single network will win 100% market share, and the rest will get no business at all from any of the publishers. From Equation (2.18), advertiser $k$'s equilibrium valuation $v_{kj^*}$ will be a product of two factors: an advertiser-specific quantity (i.e., $\frac{y_k}{R_k}$) and an ad-network-specific quantity (i.e., $M_{j^*}$). For ad networks $j \neq j^*$, advertiser $k$'s valuation is given by the inequality (2.20).

Theorem 1 does not claim that there is a unique equilibrium in the subgame. Many equilibria exist, and they differ both in which ad network is chosen and the advertiser valuations on those ad networks that aren't chosen. In fact, for any ad network $j'$ and any first-step outcome $\{x_j \ \forall j\}$, there exists an equilibrium for the subgame in which $j^* = j'$ in Theorem 1. In words, irrespective of what the ad networks do in the first step, any ad network can be chosen by publishers in the second step. Theorem 1 only says that all publishers will agree on their choice. For example, in the numerical example in Section 2.2.3, we studied an equilibrium in which ad network 1 was chosen by both publishers. Thus, Theorem 1 alone cannot predict which ad network will be chosen, or how aggressive ad networks will be.

To make a stronger prediction than Theorem 1, we use the notion of average publisher quality, $\bar{r}$, given in Equation (2.1). Recall that a publisher $i$ is high-quality if $r_i \geq \bar{r}$. Conversely, a publisher $i$ is low-quality if $r_i < \bar{r}$. Although any ad network $j$ can be chosen in equilibrium, a given publisher will find some equilibria to be more profitable than others. For a fixed first-step outcome $\{x_j \; \forall j\}$, define $\bar{\pi}_{ij}$ as the profit to publisher $i$ in an equilibrium in which ad network $j$ is chosen by all publishers. We say that publisher $i$ "prefers" ad network $j$ if $\bar{\pi}_{ij} \geq \bar{\pi}_{ij'} \; \forall j' \neq j$. In words, publisher $i$ will prefer ad network $j$ if, amongst the set of all equilibria, publisher $i$'s revenues are highest in equilibria where ad network $j$ is chosen.

The following lemma says that all high-quality publishers will agree on which equilibria they prefer:

**Lemma 1.** *High-quality publishers will prefer ad network $j$ if*

$$\frac{1-x_j}{1-x_j^{\alpha_j}} \geq \frac{1-x_n}{1-x_n^{\alpha_n}} \;\; \forall n \neq j \tag{2.21}$$

*Proof.* See Appendix A.3. □

That is, high-quality publishers will prefer the ad network $j$ that delivers the highest ratio of true negatives (i.e., $1 - x_j$) to false negatives (i.e., $1 - x_j^{\alpha_j}$). As such, we can interpret the ratio $\frac{1-x_j}{1-x_j^{\alpha_j}}$ as a measure of "ad network quality." Lemma 1 simply says that high-quality publishers prefer high-quality ad networks.

**Assumption 2.** *The ad network chosen in the subgame is the one preferred by high-quality publishers.*

Under assumptions 1 and 2, publishers will choose the ad network that is most effective at detecting invalid clicks, as long as the ad network is sufficiently aggressive:

**Theorem 2.** *Suppose Assumptions 1 and 2 hold, $J \geq 2$, and that ad network 1 is the most effective at detecting invalid clicks i.e., $\alpha_1 < \alpha_j \; \forall j \neq 1$. Then,*

1. *There exists $x^* \in (0,1)$ such that if $x_1 > x^*$, then $c_{i1} = 1 \; \forall i$, irrespective of what the other ad networks do (i.e., irrespective of $\{x_j \; \forall j \neq 1\}$). Therefore, it is a dominant strategy for ad network 1 to choose $x_1 > x^*$ in the first step.*

2. *Suppose ad network 2 is the second-most effective at detecting invalid clicks i.e., $\alpha_2 \leq \alpha_j \; \forall j \geq 2$. Then, as $\alpha_2 - \alpha_1 \to 0$, we get $x^* \to 1$.*

3. *Ad network 1's total revenues are the same for all $x^* < x_1 < 1$. As $x_1 \to 1$, high-quality publishers earn an increasingly large share of the total publisher revenues.*

*Proof.* See Appendix A.4. □

The intuition behind Theorem 2 is as follows. By choosing an $x_1$ that is sufficiently large, ad network 1 is able to deliver the highest-quality traffic (Lemma 1) to its advertisers. Since high-quality publishers decide which equilibrium is played (Assumption 2), ad network 1's high-quality traffic is enough to win the market

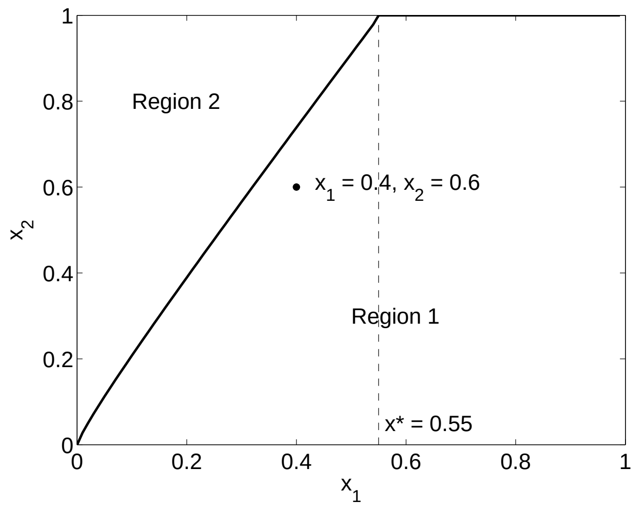


Figure 2.1: Decision regions for $\alpha_1 = 0.20$, $\alpha_2 = 0.25$. Along the boundary (2.21) holds with equality.

over. In particular, it is suboptimal to simply ignore invalid clicks (i.e., set $x_j = 0$) – high-quality publishers would react to an insufficiently aggressive ad network by choosing a more aggressive competitor instead. If ad network 1's detection algorithms are only slightly better than its competitors', it is forced to mark a large fraction of clicks invalid in order to win over the market. High-quality publishers prefer ad networks that filter aggressively, since they would disproportionately benefit compared to low-quality publishers.

Figure 2.1 gives some graphical intuition for Theorems 1 and 2. Consider a market with only $J = 2$ ad networks, where $\alpha_1 = 0.20$ and $\alpha_2 = 0.25$. The set of possible outcomes $(x_1, x_2)$ for the first-step is depicted in Figure 2.1. According to Theorem 1, depending on the first-step outcome, all publishers will choose either ad network 1 or 2 in the subgame. Using Assumption 2 and the inequality (2.21), "Region 1" in Figure 2.1 is the set of points $(x_1, x_2)$ where ad network 1 is chosen, whereas "Region 2" is the set of outcomes where ad network 2 is chosen. The boundary between the two regions is the set of points where (2.21) holds with equality. For example, if $x_1 = 0.4$ and $x_2 = 0.6$, publishers will prefer ad network 1, which means the point $(0.4, 0.6)$ lies in Region 1.

Observe that for any $x_1 \leq 0.55$, there is an $x_2 \in [0, 1]$ that lies in Region 2. On the other hand, if $x_1 > 0.55$, no such $x_2$ exists. Therefore, $x^* = 0.55$ (see Theorem 2). If ad network 1 chooses $x_1 > x^*$, then there is no $x_2$ that ad network 2 can choose that will cause publishers to prefer it over ad network 1. It can be shown that ad network 1's total profit is the same for all points in Region 1. As such, it is a dominant strategy for ad network 1 to choose any $x_1 \in (x^*, 1)$.

Figure 2.2 illustrates the dependence of $x^*$ on $\alpha_2$, for various values of $\alpha_1$. Each curve is obtained by letting $x_n \to 1$ in the right-hand side of (2.21), with $j = 1$ and $n = 2$. Then, $x^*$ is the solution of the resulting polynomial in $x_1$, solved for various $(\alpha_1, \alpha_2)$. On each curve, as $\alpha_2 \to \alpha_1$, note that $x^* \to 1$. As the gap between $\alpha_1$ and $\alpha_2$ increases, $x^*$ decreases i.e., ad network 1 can get away with being less aggressive. Note that the plots in Figure 2.2 do not depend on the numbers of publishers and advertisers, or their types.

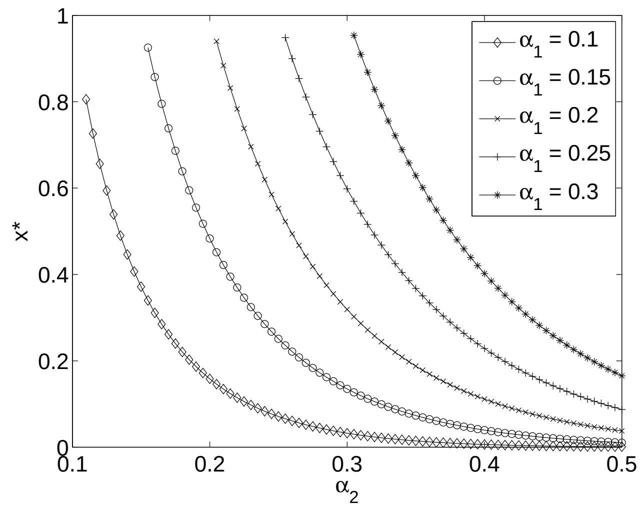


Figure 2.2: Dependence of $x^*$ on $\alpha_1$ and $\alpha_2$.

Observe that an analogue to Theorem 2 would not hold for low-quality publishers. If low-quality publishers were to decide which equilibrium is played, the inequality in (2.21) would be reversed. However any ad network can achieve the minimum by setting $x_j = 0$. That is, only an effective, aggressive ad network can satisfy high-quality publishers, but any ad network can appease the low-quality publishers by simply not filtering at all.

## 2.3 Multi-Period Model

The main intuition of the single-period model in Section 2.2 was that, if all other factors remain equal, the ad network that is most effective at detecting invalid clicks would win over 100% of publishers, provided it is sufficiently aggressive. In particular, the ad network that could deliver the highest ratio of true negatives to false negatives wins.

Recall that the single-period model made the following simplifying assumptions:

1. It is a single-period game.
2. Ad networks act first, followed by the publishers and advertisers.
3. There is no private information.
4. There are no auctions for click-throughs.

In this section, we will relax all of these assumptions and arrive at a more realistic multi-period model of the market. Fortunately, the results, intuitions and proof techniques developed for the single-period model are mostly still applicable in this new setting. Rather than presenting the model in complete detail, we will focus on the differences between the multi-period and single-period models.

### 2.3.1 Model

Consider once again a market comprised $I$ publishers, $J$ ad networks and $K$ advertisers. We model the online advertising market as a multi-period dynamic game, played over a sequence of discrete time periods, indexed by $t \in 1, \ldots, \infty$. We model each period $t$ as a two-step sequential game, which we refer to as the *stage game*. The two steps in the stage game are as follows:

1. At the beginning of period $t$, publishers decide which ad networks to sell their clicks on in period $t$, and advertisers decide how much they are willing to pay for clicks from each ad network in period $t$. Publishers and advertisers make their decision without knowing how aggressive each ad network will be.

2. At the end of period $t$, ad networks decide how aggressively to filter for invalid clicks from their period $t$ traffic.

Each period $t$ represents a "billing cycle" for the ad networks (e.g., one month). At the beginning of each billing cycle, advertisers design ads and enter bids on queries, and publishers publish code snippets on their sites to request ads from their chosen ad networks. In the middle of the cycle, users visit websites, click on ads and sometimes buy products. At the end of each billing cycle, ad networks examine their logs and try to detect which click-throughs were invalid, advertisers pay for ones marked valid, and publishers receive their share of the revenue. Observe that the two-step stage game in the multi-period model is very similar to the two-step sequential game in Section 2.2. However, there is one key difference: in the multi-period model, the publishers and advertisers act first, whereas in Section 2.2 the ad networks acted first. The multi-period model is more realistic, in view of the billing cycle interpretation of each period $t$.

Table 2.6 is a summary of the new notation we will introduce in Section 2.3. Publisher $i$'s type is $(r_i, V_i)$, just like in the single-period model. It receives $V_i$ clicks per period, of which a fraction $r_i$ are valid. Ad network $j$'s period-$t$ aggressiveness is $x_{jt}$, and $j$'s effectiveness (i.e., its game-theoretic type) is $\alpha_j$. The fraction of $i$'s clicks marked valid by $j$ in period $t$ is, then:

$$N_{ijt} \equiv (1 - x_{jt})r_i + (1 - x_{jt}^{\alpha_j})(1 - r_i) \tag{2.22}$$

Publisher $i$ sends a fraction $c_{ijt}$ of its clicks to ad network $j$ in period $t$. The total number of $i$'s clicks marked valid by $j$ is therefore:

$$N_{ijt} V_i c_{ijt} \tag{2.23}$$

### 2.3.2 Advertisers

Advertiser $k$'s type is still $(y_k, R_k)$. However, in the multi-period model, $y_k$ and $R_k$ are private information to advertiser $k$. In particular, $y_k$ and $R_k$ are not known to ad networks or publishers. As we discuss in Section 2.3.3, ad networks use auctions as a mechanism to extract this information from each advertiser $k$. On the other hand, the joint probability distribution (denoted $F_{y,R}$) of $y$ and $R$ over the population of advertisers

| | |
|---|---|
| **Decision Variables** | |
| $c_{ijt}$ | Fraction of publisher $i$'s clicks sent to ad network $j$ in period $t$ |
| $x_{jt}$ | Aggressiveness (false positive rate) of ad network $j$ in period $t$ |
| $v_{kjt}$ | Advertiser $k$ valuation of ad network $j$'s clicks in period $t$ |
| **Forecasts** | |
| $c_{ijt}^{(k)}$ | Advertiser $k$'s forecast of $c_{ijt}$ |
| $v_{kjt}^{(i)}$ | Publisher $i$'s forecast of $v_{kjt}$ |
| $x_{jt}^{(k)}$ | Advertiser $k$'s forecast of $x_{jt}$ |
| $x_{jt}^{(i)}$ | Publisher $i$'s forecast of $x_{jt}$ |
| $\hat{x}_{jt}$ | Publishers' and advertisers' common forecast of $x_{jt}$ |
| **Derived Parameters** | |
| $N_{ijt}$ | Fraction of $i$'s clicks marked valid by $j$ in period $t$ |
| $\pi_{ijt}$ | Publisher $i$'s revenue from clicks sold via ad network $j$ in period $t$ |
| $\pi_{it}$ | Publisher $i$'s total revenue in period $t$ |
| $\xi_{kjt}$ | Fraction of ad network $j$'s clicks send to advertiser $k$ in period $t$ |
| $Y_{kjt}$ | Advertiser $k$'s revenue from ad network $j$'s clicks in period $t$ |
| $Z_{kjt}$ | Number of clicks advertiser $k$'s is billed for by ad network $j$ in period $t$ |
| $R_{kjt}$ | Advertiser $k$'s ROI on ad network $j$ in period $t$ |
| $\theta_{jt}$ | Ad network $j$'s expected revenue per click in period $t$ |
| $\eta_{jt}$ | Ad network $j$'s total revenue in period $t$ |
| **Constants** | |
| $F_{y,R}$ | Joint probability distribution over advertisers' type |

Table 2.6: New notation introduced in Section 2.3.

is public information. Clearly, modeling the advertisers' conversion revenues and ROI targets as private information is more realistic.

Let $\xi_{kjt}$ be the fraction of ad network $j$'s clicks sent to advertiser $k$ in period $t$ (we will never need to compute the exact value of $\xi_{kjt}$ since it "cancels out" in Equation (2.26)). Ad network $j$ receives $\sum_i r_i V_i c_{ijt}$ clicks in total in period $t$. Assuming a conversion rate of $\beta$, advertiser $k$ earns

$$Y_{kjt} \equiv \left( \sum_i r_i V_i c_{ijt} \right) \xi_{kjt} \beta y_k \tag{2.24}$$

dollars of revenue in period $t$ from clicks coming from ad network $j$. From (2.23), the total number of clicks that $k$ is billed for by $j$ in period $t$ is:

$$Z_{kjt} \equiv \left( \sum_i N_{ijt} V_i c_{ijt} \right) \xi_{kjt} \tag{2.25}$$

Suppose $v_{kjt}$ is the amount $k$ is willing to pay $j$ per click in period $t$. Then, $k$'s period-$t$ ROI would be:

$$R_{kjt} \equiv \frac{Y_{kjt}}{v_{kjt} Z_{kjt}} \tag{2.26}$$

### 2.3.3 Auctions

In the single-period model, $v_{kj}$ was the actual dollar amount that $k$ paid $j$ per click. Since advertiser $k$'s objective was simply to achieve an ROI each period of exactly $R_k$ – no more, no less – it had no reason to falsely report its valuation $v_{kj}$ to the ad networks. Each advertiser also received a fixed fraction $\frac{1}{K}$ of each ad network's clicks, so there was no connection between $v_{kj}$ and the number of clicks $k$ was allocated by $j$.

In the multi-period model, we are in a more realistic setting where advertisers aim to maximize profits (see (2.32)). If advertisers knew they would receive a $\frac{1}{K}$ fraction of clicks irrespective of $v_{kjt}$, then every advertiser would simply pick $v_{kjt} = 0$ every period – profits are clearly maximized if clicks are free, and ROI is infinite! Advertisers would have no incentive to report how much they are actually willing to pay for clicks.

Ad network $j$ might try charging advertiser $k$ a fixed amount $\frac{y_k}{R_k} M_j$ per click (as in Theorem 1). However, doing so is impossible because $y_k$ and $R_k$ are private information, and not known to $j$. The solution is for ad networks to create a link between the valuation $v_{kjt}$ and the allocation $\xi_{kjt}$. Advertisers who are willing to pay more should receive a larger fraction of clicks. However, as it stands, ad network $j$ does not know $v_{kjt}$.

To encourage advertisers to report their valuations, we assume that each ad network runs an auction to determine which advertiser receives each click. On each click, a fixed number of advertisers are selected uniformly-at-random from the pool of advertisers who have entered a bid on a given query. Auction participants are chosen randomly to ensure that the top bidder does not receive 100% of the clicks (this practice is referred to as "throttling"). The click is then awarded to the highest bidder, and the amount billed is related to the randomly-selected advertisers' bids. The benefit of using auctions is that advertisers have an incentive to report $v_{kjt}$. The downside for ad networks is that, in most cases, advertiser $k$ ends up paying less than the true valuation, $v_{kjt}$, per click.

Let $\theta_{jt}$ be ad network $j$'s expected revenue per click in period $t$ i.e., if $j$ marks $Z$ clicks valid in period $t$, then its period-$t$ gross revenue (before paying publishers) would be $Z\theta_{jt}$. Our focus here is not on the specifics of the auction mechanism. For our purposes, we only assume the following:

1. The expected revenue per click is some function of advertisers' valuations i.e., $\theta_{jt} = g_j\left(\{v_{kjt}\}\right)$
2. All ad networks use the same auction mechanism i.e., $g_j = g\ \forall j$
3. The auction mechanism is *linear* in the following sense: if every single advertiser's valuation is scaled by a constant factor $\gamma$, then the expected revenue from the auction is also scaled by $\gamma$ i.e., if $v_{kjt} \leftarrow \gamma v_{kjt}\ \forall(k,j)$, then $\theta_{jt} \leftarrow \gamma\theta_{jt}$

The first and third assumptions are quite weak, and hold for most auction mechanisms used in practice. We make the second assumption (i.e., $g_j = g\ \forall j$) because our focus here is on the validity of traffic, rather than comparing the efficiency of ad networks' auction mechanisms. In particular, we do not assume that the auction mechanism is truthful, or any other property about the mapping from advertisers' valuations to bids.

From these assumptions alone, it is shown in Chapter 3 that if advertisers compute $v_{kjt}$ optimally, then:

$$\theta_{jt} = \kappa a_{jt} \tag{2.27}$$

where $\kappa$ is a constant across all ad networks, and $a_{jt}$ is ad network $j$'s *adjustment factor* in period $t$:

$$a_{jt} \equiv \frac{\left(\sum_i r_i V_i c_{ijt}\right)\beta}{\sum_i N_{ijt} V_i c_{ijt}} \tag{2.28}$$

Recall that $\beta$ is the fraction of clicks that become conversions. Intuitively, $a_{jt}$ is the ratio of the total number of converted clicks on ad network $j$ to the total number of clicks $j$ marks valid, in period $t$. Changing the (common) auction mechanism only affects the value of $\kappa$. In Section 2.4, we discuss the conditions under which we can think of click-throughs, rather than impressions, as the items that are auctioned off.

### 2.3.4 Player Objectives

Publisher $i$'s expected revenue from ad network $j$ is:

$$\pi_{ijt} \equiv \left(N_{ijt} V_i c_{ijt}\right) h\theta_{jt} \tag{2.29}$$

Publisher $i$'s total period-$t$ expected revenues are:

$$\pi_{it} \equiv \sum_j \pi_{ijt} \tag{2.30}$$

Therefore, in each period $t$, $i$ chooses $\{c_{ijt}\ \forall j\}$ by solving the following optimization problem:

$$\max_{\{c_{ijt}\}} \pi_{it} \text{ s.t. } \sum_j c_{ijt} = 1 \text{ and } c_{ijt} \geq 0 \tag{2.31}$$

Advertiser $k$ selects its valuation $v_{kjt}$ so that its revenues from ad network $j$ are maximized, subject to a lower bound $R_k$ on ROI:

$$\max_{v_{kjt}} Y_{kjt} \text{ s.t. } R_{kjt} \geq R_k \tag{2.32}$$

At the optimum, the constraint in (2.32) will be binding:

$$R_{kjt} = R_k\ \forall (j,t) \tag{2.33}$$

To understand why (2.33) holds, assume for a moment that network $j$'s auction is truthful i.e., that advertiser $k$'s period-$t$ bid is $v_{kjt}$. Recall that $R_k$ is the ROI that advertiser $k$ can achieve through channels other than CPC advertising. So, if $R_{kjt} > R_k$, advertiser $k$ will want to spend more money on network $j$ i.e., it will want to buy more clicks from network $j$. To receive more clicks from network $j$, however, it must increase its bid $v_{kjt}$. From (2.26), we know that $R_{kjt}$ is decreasing in $v_{kjt}$. Therefore, advertiser $k$ will keep increasing $v_{kjt}$ as long as $R_{kjt} > R_k$, meaning (2.33) will hold at the optimum. This argument is informal, but is made rigourous in Appendix A.1. Essentially, even if the auction mechanism used is not truthful, a similar argument would apply as long as the number of clicks allocated to $k$ is increasing in $k$'s bid.

Let $\eta_{jt}$ be ad network $j$'s total expected revenue in period-$t$:

$$\eta_{jt} = (1-h)\left(\sum_i N_{ijt} V_i c_{ijt}\right)\theta_{jt} \tag{2.34}$$

Each ad network $j$ has a *discount factor*, $\delta_j \in [0,1)$, which represents how much network $j$ values future earnings relative to the present. $\delta_j$ is a fixed parameter for each network $j$ i.e., it is not a decision variable. Using $\delta_j$, we can compute network $j$'s total expected long-term revenue, $\eta_j$:

$$\eta_j = \sum_t (\delta_j)^t \, \eta_{jt} \tag{2.35}$$

In the multi-period model, network $j$'s objective is to choose $\{x_{jt} \ \forall (j,t)\}$ such that $\eta_j$ is maximized. Note that in our multi-period model, publishers and advertisers are "greedy," in the sense that their objective is to maximize period-$t$ profits. Ad networks, on the other hand, aim to maximize long-term profits. We make this simplifying assumption in order to focus on the interactions between ad networks, rather than possible collusions between publishers and/or advertisers.

### 2.3.5 Forecasts

Publishers and advertisers base their decisions on *forecasts* of what other players will do in period $t$. An example of a forecast is to assume that each player simply makes the same decision as the previous period. Another example is to assume that publishers will allocate traffic randomly and that advertiser valuations are random. The notion of forecasting is reflective of how publishers and advertisers must make their decisions each period, without knowledge of other players' decisions (how the forecasts are actually computed is not important here).

At the start of each period, each publisher $i$ computes forecasts $\{v_{kjt}^{(i)} \ \forall (k,j)\}$ of each advertiser $k$'s bid on each ad network $j$, as well as forecasts $\{x_{jt}^{(i)} \ \forall j\}$ of the ad networks' actions. Publisher $i$ then chooses allocations that are the best responses to $\{v_{kjt}^{(i)} \ \forall (k,j)\}$ and $\{x_{jt}^{(i)} \ \forall j\}$ i.e., $i$ solves (2.31) assuming its forecasts are accurate.

Similarly, advertiser $k$ computes forecasts $\{c_{ijt}^{(k)} \ \forall (i,j)\}$ of each publisher $i$'s allocation to each ad network $j$, as well as forecasts $\{x_{jt}^{(k)} \ \forall j\}$ of the ad networks' actions. Advertiser $k$'s period-$t$ valuations are then best responses to $\{c_{ijt}^{(k)} \ \forall (i,j)\}$ and $\{x_{jt}^{(k)} \ \forall j\}$ i.e., $k$ solves (2.32) under the assumption that its forecasts are accurate.

### 2.3.6 Equilibria

Thus far in Section 2.3, we have relaxed some key assumptions made in the single-period model in Section 2.2. We can now state generalized versions of Theorems 1 and 2 for the multi-period case.

**Assumption 3.** *Publishers and advertisers agree on their forecast of ad networks' actions:*

$$x_{jt}^{(i)} = x_{jt}^{(k)} \equiv \hat{x}_{jt} \;\; \forall (i,j,k,t) \tag{2.36}$$

Informally, Assumption 3 says that nobody has "insider information" regarding the ad networks. This assumption is analogous to the single-period model (where publishers and advertisers observed $x_j$ before making their decisions) in the sense that all publishers and advertisers will use the same value of $x_{jt}$ in their calculations. However, Assumption 3 does not say that the forecasts are accurate i.e., it certainly is possible that $\hat{x}_{jt} \neq x_{jt}$.

**Assumption 4.** *Publishers and advertisers are able to forecast each others actions accurately:*

$$c_{ijt}^{(k)} = c_{ijt} \;\; \forall (i,j,k,t) \tag{2.37}$$
$$v_{kjt}^{(i)} = v_{kjt} \;\; \forall (i,j,k,t) \tag{2.38}$$

If publishers and advertisers optimize their decisions conditional on their forecasts, and all forecasts were indeed accurate as in Assumption 4, then no player would change their decision (or, feel "regret") upon observing the other players' actual decisions. Hence, Assumption 4 is equivalent to simply assuming that a "Nash equilibrium" is played each period between publishers and advertisers in the first step of each stage game. Assumptions 3 and 4 provide enough structure to allow us to treat publishers' and advertisers' actions in each period in the same way as the single-period model – publishers and advertisers arrive at an equilibrium, with the networks' actions fixed.

Recall Assumption 1, which was that $r_m \neq r_n$ for some $m$ and $n$. Under Assumptions 1, 3 and 4, we obtain the following analogue of Theorem 1, which says that in each period all publishers will choose the same ad network in equilibrium:

**Theorem 3.** *Suppose Assumptions 1, 3 and 4 hold. Then, there exist subgame perfect Nash equilibria (SPNE) for the multi-period game with the following property – In every period t, there exists a $j^* \equiv j^*(t)$ such that*

$$c_{ij^*t} = 1 \;\; \forall i \tag{2.39}$$
$$v_{kj^*t} = \frac{y_k}{R_k} M_{j^*} \;\; \forall k, \tag{2.40}$$

*whereas for all $j \neq j^*$,*

$$c_{ijt} = 0 \;\; \forall i \tag{2.41}$$
$$v_{kjt} \leq \frac{y_k}{R_k} M_j \;\; \forall k. \tag{2.42}$$

*We say that ad network $j^*$ has been "chosen" by publishers and advertisers in period t.*

*Proof.* Showing that all publishers agree on their choice in each stage game is similar to Theorem 1 (see Appendix A.2). Then, we apply the result that playing an NE in each stage game is also an SPNE for the repeated game (e.g., [13]). □

Observe that $j^*$ can vary between periods i.e., it is possible that $c_{ijt} = 1$ and $c_{ijs} = 0$ for $t \neq s$. In words, publishers may choose a different ad network each period. However, Theorem 3 assures us that all publishers will agree on their choice each period i.e., $c_{ijt} = c_{mjt}\ \forall(i, m)$.

To make a stronger statement analogous to Theorem 2, we now consider a steady-state equilibrium.

**Assumption 5.** *Publishers and advertisers are able to forecast ad networks' actions accurately:*

$$\hat{x}_{jt} = x_{jt}\ \ \forall(j, t) \tag{2.43}$$

Stated differently, Assumption 5 says that ad networks behave "predictably." For example, Assumption 5 would hold in a "steady state" where $x_{jt} = x_{j,t-1}$, so that $\hat{x}_{jt} = x_{j,t-1} = x_{jt}$. Assumption 5 provides us with the same structure as the single-period setting, in the sense that acting last is no longer an advantage for the networks. The networks now behave predictably, giving publishers and advertisers the chance to anticipate and react.

**Assumption 6.** *In each period, the ad network chosen is the one preferred by high-quality publishers.*

Assumption 6 is an analogue of Assumption 2, extended to the multiperiod setting. Under assumptions 1, 3, 4, 5 and 6 we obtain a multiperiod analogue to Theorem 2. In equilibrium, publishers will always choose the ad network that is most effective at detecting invalid clicks, as long as the ad network is sufficiently aggressive:

**Theorem 4.** *Suppose Assumptions 1, 3, 4, 5 and 6 hold, $J \geq 2$, $\delta_1 > 0$, and that ad network 1 is the most effective at detecting invalid clicks i.e., $\alpha_1 < \alpha_j\ \forall j \neq 1$. Then,*

1. *There exists $x^* \in (0, 1)$ such that if $x_{1,t} > x^*$, then $c_{i1t} = 1\ \forall(i, t)$, irrespective of what the other ad networks do (i.e., irrespective of $\{x_{jt}\ \forall(j, t), j \neq 1\}$). Therefore, it is a dominant strategy for ad network 1 to choose $x_{1t} > x^*$ in every period $t$.*

2. *Suppose ad network 2 is the second-most effective at detecting invalid clicks i.e., $\alpha_2 \leq \alpha_j\ \forall j \geq 2$. Then, as $\alpha_2 - \alpha_1 \to 0$, we get $x^* \to 1$.*

3. *Ad network 1's total revenues are the same for all $x^* < x_{1t} < 1$. As $x_{1t} \to 1$, high-quality publishers earn an increasingly large share of the total period-$t$ publisher revenues.*

*Proof.* Simply apply the proof techniques for Theorem 2 on a period-by-period basis. □

Notice that $\delta_1 > 0$ implies that network 1 does indeed care about future earnings, so it's strategy will not be to, say, quickly exit the market after "stealing" huge revenues in a single period.

## 2.4 Discussion

We devote this section to a brief discussion of some modeling decisions and simplifying assumptions used in our models.

### 2.4.1 Modeling Decisions

**Single query market** Our models consider the market for a single search query. In real markets involving an entire collection of queries, we envision an entire "collection" of independently functioning markets. For example, ad network 1 might earn 100% market share for the query "digital camera," while ad network 2 might earn 100% share for the query "mortgage."

**Pricing model** We did not consider *cost-per-mille* (CPM) or *cost-per-acquisition* (CPA) pricing models in our analysis. In these pricing models, fraud occurs in different ways (e.g., [11]). We refer the reader to [9] to learn about other forms of online advertising fraud.

**No direct deals** We did not explicitly model "direct deals" between publishers and advertisers. These direct agreements were more prevalent in the earlier days of the Internet, before the emergence of the larger ad networks and ad aggregators.

Suppose a popular content publisher with many users decides not to deal with any ad networks, and opts for direct deals with advertisers only. Advertisers might indeed be interested in advertising on this publisher's site because of its large user base – of course, an advertiser $k$ would do so only if the ROI was larger than $R_k$. In effect, this publisher would be a) starting up its own ad network, b) sending its traffic to this ad network irrespective of profitability, and then c) rejecting business from all publishers other than itself. Thus, direct deals violate our earlier assumption that an entity that owns both a publisher and an ad network (e.g., Google) will allow them to operate independently.

**Innocent publishers** We have assumed that publishers are "innocent" – they are unaware whether a given click is valid, and are therefore unable to send all their valid clicks to one ad network and all their invalid clicks to another. The innocent-publishers assumption is without loss of generality, since a publisher that can distinguish between valid and invalid traffic (but chooses to ignore the invalid traffic) can simply be modeled as two separate publishers – one that receives valid traffic, and one that generates a stream of fraudulent click-throughs. On the other hand, such a publisher might choose to voluntarily "filter out" invalid clicks (that is, report them to the ad network), raising the overall quality of the traffic it sends to ad networks.

Similarly, publishers are assumed to be profit maximizers who do not collude with each other. Those that do collude can simply be combined into a single publisher.

**No investing in technology** Ad networks cannot invest in improving their invalid click detection algorithms. For example, an ad network $j$ cannot hire more engineers in order to lower their $\alpha_j$. Theorems 2 and 4 imply that if such investment is feasible, it is absolutely in an ad network's interest to invest heavily –

we have seen that the ad network whose is most effective at detecting invalid clicks has a dramatic market advantage.

### 2.4.2 Impressions vs. Clicks

In practice, each publisher $i$ has $P_i$ ad impressions (or, "page views") to allocate across the ad networks. On each ad impression, the publisher decides which ad network to allocate the impression to. An auction is then run by the network, between $D_i$ randomly chosen advertisers, and the impression is sold to the $A_i$ advertisers with the highest valuations (i.e., there are $A_i$ ad slots on the webpage). Then, at most one advertiser (but usually none) will have its ad impression become a click-through (although some click-throughs will be invalid). Payment is collected from the advertiser only if the ad impression is clicked on (some advertisers elect to pay on a per-impression basis, however). So, to summarize, publishers allocate impressions across networks, but clicks are the objects that are sold to advertisers.

Our model of publishers, however, is somewhat different – in particular, each publisher $i$ has an inventory of $V_i$ clicks-throughs to allocate across the ad networks. This modeling decision is very convenient, as it allows us to sidestep the issues of click-through rates (i.e., the ratio of clicks to impressions) and impression spam (the goal of which is to manipulate click-through rates).

Under the following conditions, it is equivalent to model clicks (rather than impressions) as the object being bought and sold:

- Selection of the ad network on each impression is independent and identical across impressions, with probabilities given by $c_{ij}$, for a fixed publisher $i$.
- $D_i$, $A_i$ and the auction mechanism used don't vary across impressions, for a fixed publisher $i$.
- Selection of the $D_i$ advertisers in each auction is independent and identical across impressions, for a fixed publisher $i$.

The first three conditions imply that the list of ads shown on each impression is an independent, identically-distributed (IID) random vector.

- All impressions for a given publisher have equal *a priori* probability of being clicked on.
- For any impression that has been clicked on, the probability of being valid is exactly $r_i$.
- The click-through rate for a given advertiser does not vary across ad networks.

The final three conditions mean that, for a fixed publisher, the revenue generated from each impression (possibly zero) is an IID random variable. Therefore, from the publisher's perspective, we can model the $P_i$ impressions instead as a fixed volume $V_i$ of click-throughs, of which a fraction $r_i$ are valid.

### 2.4.3 Simplifying Assumptions

The model we presented here made a few simplifying assumptions. These assumptions can be divided into two groups: "essential" and "nonessential." Essential assumptions are those without which our results would not hold (or would hold only approximately). Viewed differently, deviations in the real world from our models' predictions can be explained by the extents to which our essential assumptions are violated. Non-essential assumptions are those that we can relax easily, and have our key results still hold.

#### Essential Assumptions

**Equal revenue shares** All ad networks are assumed to offer the same revenue share, $h$, to their publishers. It is intuitive why this assumption is essential – if an ad network does not have the most effective invalid click detection algorithm, one way to draw publishers over to its network is to simply offer them a larger fraction of the revenue. It can be shown analytically that such a tactic can, in fact, attract publishers (albeit low-quality ones) to a lower-quality ad network. In such cases, *market segmentation* amongst the publishers is observed, contrary to the predictions of Theorems 1 and 3.

Notice that if a lower-quality ad network decided to offer a slightly-higher revenue share, low-quality publishers would be the first to be enticed. High-quality publishers may not want to switch networks immediately, since a small increase in revenue share would not compensate for much lower revenues earned per-click.

**No discounts** Similarly, we disallow the practice of predictive pricing. In practice, such discounts would alter the bids submitted by advertisers, and would therefore impact the attractiveness of various ad networks to market participants. Since we assume that conversion rates are constant across all publishers, advertisers and ad networks, such a tactic would be ineffective in our model. In Chapters 3 and 4, we study the use of predictive pricing and revenue shares as tools that ad networks can use to attract and retain traffic on their networks.

#### Non-Essential Assumptions

**ROC curve** Our models assumed, for simplicity, that if an ad network $j$'s aggressiveness is $x_j$ and its effectiveness is $\alpha_j$, then its true positive rate is described by a function $f(x_j;\alpha_j) \equiv x_j^{\alpha_j}\ \forall j$ known as a receiver operating characteristic (or, "ROC curve").

However, our results apply for any function $f(x_j;\alpha_j)$ that is continuous in $x_j$ and $\alpha_j$, nondecreasing over $x_j \in [0,1]$ and nonincreasing over $\alpha_j \in [0,1]$. These conditions ensure that, for a large enough choice of $x_j$, the value of $\frac{1-x_j}{1-f(x_j;\alpha_j)}$ attained by network $j$ is unattainable by all other competing networks, assuming $j$ is the most-effective network. In the most general case, an ad network's effectiveness would be described by an entire curve, $f_j(x_j)$.

**No sticky publishers** There are no "sticky" publishers in our model. A sticky publisher is an "irrational" publisher who decides to display ads from a particular ad network, even though it may not be the most profitable one. A common example is a search engine company that is both a publisher and an ad network. Another example is a publisher that colludes with a particular ad network.

If publisher $i$ is sticky in its choice of ad network $j$, we can simply fix $c_{ij} = 1$. Unless publisher $i$ has a very large volume of traffic, our results would remain qualitatively unchanged, in the sense that a single high-quality ad network would still attract traffic from most of the publishers.

**Time-invariant types** We assumed in the multiperiod model that the players' types are time-invariant. That is, $r_i, V_i, \alpha_j, \delta_j, y_k$ and $R_k$ do not vary across time periods (e.g., $r_{it} = r_i \; \forall t$). However, the theorems in Section 2.3 can easily be generalized to account for the time-varying case. Essentially, we apply Theorem 1 on a period-by-period basis, and recognize that playing an Nash equilibrium in every stage game is a subgame perfect Nash equilibrium for the supergame.

**Non-zero true-negative rates** Only to simplify the exposition, we assume in Section 2.2 that $x_j < 1 \; \forall j$ i.e., no ad network gives away 100% of its clicks for free. It can be shown that $x_j = 1$ will never occur in equilibrium. Similarly, in Section 2.3, we assume $x_{jt} < 1 \; \forall (j,t)$.

## 2.5 Managerial Implications

The results in this chapter make it clear how ad networks ought to think about publishers and advertisers. The decision to model advertisers as agents that maximize profits subject to a lower bound on ROI merits special attention. Advertisers enter bids on each ad network in order to attain their target ROI on every ad network (we refer to this strategy as "ROI targeting"), and are willing to pay for an unlimited number of clicks as long as the return on their investment in these clicks is sufficiently high. Therefore, in principle, advertisers have infinite budgets.

It is, of course, possible (and seemingly more natural) to model advertisers as allocating a finite advertising budget across the ad networks. However, in practice, ROI targeting is a more common behaviour amongst advertisers. It is actually quite rare that advertising budgets are ever exhausted. Advertisers typically are willing to pay for more click-throughs if they are available, since doing so simply means they are making more money from clicks becoming conversions. In Appendix A.1, we describe conditions under which ROI targeting is actually equivalent to profit maximization under a large-but-finite budget constraint. In fact, in can be shown via a simple economic argument that in a budget-allocation model, if an advertiser's budget for click-throughs is being exhausted, it is rational for the advertiser to simply lower their bid. Doing so would result in strictly more click-throughs for the exact same cost.

Publishers and advertisers should, of course, feel more at ease in light of our results. We have our first clear demonstration that traffic quality has an impact on profits, not just for publishers and advertisers, but for ad networks as well. Indeed, in the setting of this chapter, the impact of quality is dramatic – it makes the difference between zero profit and complete market dominance.

## 2.6 Summary

We have presented an economic model of the cost-per-click advertising market, focusing on the effects of detecting invalid clicks (and click fraud, in particular). Our results suggest that, indeed, letting fraud go unchecked (i.e., choosing $x_j = 0$) is suboptimal for ad networks – a network that sets $x_j = 0$ would lose business to a more aggressive competitor. In particular, under reasonable assumptions on how the market functions, the ad network that can filter most effectively has a significant competitive advantage.

In practice, obviously no ad network is earning 100% market share. On the other hand, publishers in the real world do often choose the most profitable ad network, and do switch ad networks when revenue prospects seem higher. So, to the extent that publishers and advertisers act rationally, we conjecture that our predictions hold true in practice, at least approximately i.e., ad networks that filter effectively and aggressively are rewarded by the market. As we shall see Chapters 3 and 4, accounting for differences between ad networks in revenue sharing, predictive pricing, and ad targeting will help paint a more complete picture.

# Chapter 3

# Predictive Pricing and Revenue Sharing

[0]An abbreviated version of this chapter appears in [25].

## 3.1 Introduction

Google's "Smart Pricing" [16] and Yahoo's "Quality-Based Pricing" [34] are examples of a practice we refer to as *predictive pricing*. The idea behind predictive pricing in cost-per-click advertising is to charge the same advertiser different prices for click-throughs, depending on which publisher the click-through originated from. For example, an advertiser who bid on the keyword "camera" might be charged less for a click-through from a travel website than one from a photographer's blog, since the latter would (ostensibly) be more targeted to potential camera purchasers than the former. Advertising networks use predictive pricing to attract publishers and advertisers to their network.

*Revenue sharing*, which is the practice of paying out a fraction the ad network's revenues to the publishers where click-throughs originate, is another tool used by advertising networks to attract traffic. Revenue sharing is the reason publishers display advertisements alongside their content in the first place. In this chapter, we study how an online advertising network can apply predictive pricing and revenue sharing "optimally" – that is, in a manner that maximizes the advertising network's profits.

The sheer size of the online advertising market makes this problem interesting and important. Although predictive pricing and revenue sharing can help advertising networks attract and retain lucrative traffic, applying these tools sub-optimally can mean that a network is "leaving money on the table" (either by paying out an unnecessarily large revenue share, or by attracting less- or lower-quality traffic than they could be). Advertising networks that currently do not apply predictive pricing should feel compelled to start – our results suggest that they are yielding a significant advantage to their competitors.

Finding an optimal pricing policy is challenging because faithful models of the online advertising market can spiral in complexity. The challenge is to capture just those aspects that impact pricing policy decisions. Our model is game-theoretic, necessitating the computation of equilibria (a task that is known to be difficult). Worse yet, the optimization problems involved in computing equilibria are highly nonconvex and thus computationally intractable. Thus, we face significant challenges from both modeling and computational perspectives.

#### Overview

We begin by constructing a model of the online advertising market as a game between content publishers, advertising networks and advertisers. The model is a simplification of what happens in practice – our intent is to hone in on the market effects of predictive pricing and revenue sharing decisions. Our model is similar to Chapter 2 – however, the focus there was on validity, so predictive pricing was disallowed and all networks were assumed to offer the same fixed revenue share. In this chapter, we instead ignore validity and choose to focus on predictive pricing and revenue sharing.

We derive an expression for an advertising network's *best-response function*. That is, if an advertising network knows the predictive pricing and revenue sharing policies of its competitors, what policy should the network choose in response, in order to maximize its profits? The expression we derive for the best-response function is implicit – it is the solution to a difficult optimization problem. We then present an algorithm, PRICINGPOLICY, for solving this optimization problem, yielding a best-response predictive pricing and revenue sharing policy.

Finally, we apply PRICINGPOLICY toward answering some qualitative questions about predictive pricing:

- Is it always better to charge less for lower-quality traffic? (Yes.)
- Should an advertising network always try to attract as much traffic as it can, regardless of traffic quality? (No.)
- If a network is better at targeting, can it offer a lower revenue share? (Yes.)
- Does predictive pricing harm publishers, as has been conjectured in online forums? (Yes and no – it can harm low-quality publishers while helping high-quality publishers.)

In principle, the best-response function can be used as a “subroutine” for computing equilibrium policies for advertising networks (an equilibrium is, by definition, a fixed point of the networks’ best-response functions). However, we believe that the practical value of our algorithm lies in computing best responses, rather than equilibria. It prescribes actions that networks can take “today” in response to their competitors, rather than waiting for equilibria to unfold. Thus, our focus will be on computing and studying the properties of the best-response functions.

## 3.2 Model

As in Chapter 2, we model the *cost-per-click* (CPC) advertising market as a one-shot dynamic game between three classes of players: content publishers, advertising networks and advertisers. Content publishers publish websites and display advertisements alongside their content. Advertisers design advertisements and bid on keywords that describe the interests of their target market. Advertising networks act as intermediaries, auctioning off click-throughs to advertisers and delivering relevant ads to publishers upon request.

If a user visits a publisher’s site and clicks on an ad, the advertiser pays the network a small amount. The network then pays out a fraction of this amount to the publisher where the click originated. Predictive pricing affects how much the advertiser is billed by the network, whereas the revenue share determines what fraction of this revenue the network will pay out to the publisher. A small fraction of clicks become conversions e.g., a purchase, or a signup to an email list. The advertiser earns some revenue each time a click becomes a conversion.

Our dynamic game is comprised of two steps:

1. In the *first step*, networks select and announce their predictive pricing and revenue sharing policies.
2. In the *second step*, publishers decide which networks to sell their clicks on, and advertisers simultaneously decide how much they are willing to pay for clicks from each network.

Recall that, in Chapter 2, networks decided how aggressively to filter for invalid clicks. In this chapter, we assume that all clicks are valid, so filtering is unnecessary. After the second step, payoffs are realized: a)

| | |
|---|---|
| **Player Parameters** | |
| $V_i$ | Volume of clicks on publisher $i$'s site |
| $\beta_i^{\text{Pub}}$ | Quality of publisher $i$'s traffic |
| $\beta_j^{\text{Net}}$ | Network $j$'s skill at matching publishers and advertisers |
| $y_k$ | Advertiser $k$'s revenue per conversion |
| $R_k$ | Advertiser $k$'s target ROI |
| $\beta_k^{\text{Adv}}$ | Effectiveness of advertiser $k$'s ads |
| **Decision Variables** | |
| $c_{ij}$ | Fraction of publisher $i$'s clicks sent to ad network $j$ |
| $h_j$ | Revenue share paid out by network $j$ |
| $g_{ij}$ | Predictive pricing factor applied to publisher $i$'s traffic by network $j$ |
| $v_{kj}$ | Advertiser $k$ valuation of ad network $j$'s clicks |
| **Derived Parameters** | |
| $\pi_{ij}$ | Publisher $i$'s revenue from clicks sent to network $j$ |
| $\pi_i$ | Publisher $i$'s total revenue |
| $\beta_{ijk}$ | Conversion rate of clicks going from $i$ to $j$ to $k$ |
| $Y_{kj}$ | Advertiser $k$'s revenue from network $j$'s clicks |
| $Z_{kj}$ | Number of clicks advertiser $k$ is billed for by network $j$ |
| $R_{kj}$ | Advertiser $k$'s ROI on ad network $j$ |
| $\bar{v}_k$ | Advertiser $k$'s nominal valuation |
| $\theta_j$ | Network $j$'s expected auction revenue per click |
| $\kappa_j$ | Network $j$'s expected auction revenue per click when $a_j = 1$ |
| $a_j$ | Network $j$'s adjustment factor |
| $\eta_j$ | Network $j$'s total revenue |
| $\eta_j^{\max}$ | Network $j$'s maximum possible revenue |
| **Constants** | |
| $I$ | Number of publishers |
| $J$ | Number of ad networks |
| $K$ | Number of advertisers |

Table 3.1: Notation used in Chapter 3.

publishers sell clicks (i.e., display ads) on their chosen networks, and b) advertisers pay the networks, who then pay the publishers. We consider a one-shot game although, as in Chapter 2, the extension to a multi-period model is straightforward. Table 3.1 is a summary of the notation used in this chapter. We remain notationally consistent with Chapter 2 as much as possible.

Consider the market for click-throughs on a single keyword. There are $I$ publishers whose content is relevant to the keyword, $K$ advertisers interested in buying clicks on this keyword, and $J$ networks.

**Publishers** Each publisher in our model has an inventory of clicks to allocate across the networks. That is, each publisher $i$ receives $V_i$ clicks on his website. In practice, however, publishers allocate impressions (or, "page views"), not clicks. In Section 2.4, we include a detailed discussion of the conditions under which it is equivalent to model clicks (rather than impressions) as the objects being bought and sold.

Let $c_{ij}$ be the fraction of these clicks that publisher $i$ sends to network $j$. As in Chapter 2, $c_{ij}$ is a decision variable for publisher $i$. Then,

$$V_i c_{ij} \tag{3.1}$$

is the total number of clicks that publisher $i$ sends to network $j$. Recall that, in this chapter, we ignore validity and focus on targetedness. That is, we assume that all clicks are valid and that networks mark all clicks valid ($r_i = 1\ \forall i$ and $N_{ij} = 1\ \forall(i,j)$, using the notation of Chapter 2). Our results are in no way dependent on this assumption. For example, if we were to account for validity, Equation (3.1) would be $N_{ij}V_i c_{ij}$, where $N_{ij}$ is the fraction of publisher $i$'s clicks marked valid by network $j$ (see Equation (2.3)).

**Ad Networks** For each click coming from publisher $i$, network $j$ bills advertisers for only a fraction $g_{ij}$ of a click i.e., advertisers receive a $(1 - g_{ij})$ discount. The fraction $g_{ij}$ is the *predictive pricing factor* that network $j$ applies to publisher $i$'s traffic. The term "predictive pricing" alludes to network $j$'s prediction about the quality of publisher $i$'s traffic (i.e., accounting for click-through rates, click fraud and conversion rates). The *effective* number of clicks publisher $i$ is paid for by network $j$ is then:

$$V_i c_{ij} g_{ij} \tag{3.2}$$

Of each dollar of revenue from advertisers, network $j$ pays out a fraction $h_j$ to publishers. The fraction $h_j$ is referred to as the *revenue share*.

We refer to $\{g_{ij}\ \forall i\}$ and $h_j$ together as network $j$'s *pricing policy*. Each ad network in our model decides on a pricing policy, with the goal of maximizing their profits.

Let $\theta_j$ be the expected auction revenue per click on network $j$. That is, if network $j$ were to auction off $Z$ clicks, its total expected revenue would be $Z\theta_j$. The value of $\theta_j$ depends on the auction mechanism used by network $j$, as well as all the advertisers' bids. Then, from Equation (3.2), the revenue to publisher $i$ from network $j$ is:

$$\pi_{ij} \equiv V_i c_{ij} g_{ij} h_j \theta_j \tag{3.3}$$

Publisher $i$'s total revenue, across all networks, is then:

$$\pi_i \equiv \sum_j \pi_{ij} = \sum_j V_i c_{ij} g_{ij} h_j \theta_j \tag{3.4}$$

Network $j$'s total profit, $\eta_j$, is the amount collected from advertisers less the amount paid out to publishers. Therefore:

$$\eta_j \equiv \left( \sum_i V_i c_{ij} g_{ij} \right) (1 - h_j)\theta_j = \frac{1 - h_j}{h_j} \sum_i \pi_{ij} \tag{3.5}$$

**Advertisers** Of all the clicks sent to network $j$, a fraction $\xi_{jk}$ is sent on to advertiser $k$. So,

$$V_i c_{ij} \xi_{jk} \tag{3.6}$$

is the number of clicks sent from publisher $i$ to advertiser $k$ via network $j$. Note that $\xi_{jk}$ is not the same for all $k$. The fraction $\xi_{jk}$ depends on network $j$'s auction mechanism, as well as all of the other advertisers' bids. Fortunately, as we demonstrate later, we will never need to explicitly compute the advertisers' bids or the value of $\xi_{jk}$ in our model.

Of the clicks going from publisher $i$ to network $j$ to advertiser $k$, let $\beta_{ijk}$ be the fraction that become conversions i.e., the *conversion rate*. From Equation (3.6), the number of clicks converted by advertiser $k$ that came from publisher $i$ via network $j$ is then:

$$V_i c_{ij} \xi_{jk} \beta_{ijk} \tag{3.7}$$

Let $y_k$ be the revenue that advertiser $k$ earns from each conversion. The total revenue to advertiser $k$ from conversions of clicks from network $j$, across all publishers, is then:

$$Y_{kj} \equiv \left( \sum_i V_i c_{ij} \xi_{jk} \beta_{ijk} \right) y_k = \left( \sum_i V_i c_{ij} \beta_{ijk} \right) \xi_{jk} y_k \tag{3.8}$$

Using (3.2), the effective number of clicks (that is, after predictive pricing is applied) originating from publisher $i$ that advertiser $k$ is billed for by network $j$ is:

$$V_i c_{ij} g_{ij} \xi_{jk} \tag{3.9}$$

The total number of clicks advertiser $k$ is billed for by network $j$ is then:

$$Z_{kj} \equiv \left( \sum_i V_i c_{ij} g_{ij} \right) \xi_{jk} \tag{3.10}$$

Let $v_{kj}$ be advertiser $k$'s valuation for network $j$'s clicks i.e., $v_{kj}$ is what advertiser $k$ is willing to pay network $j$ per click. The total amount that advertiser $k$ is willing to pay network $j$ is then:

$$Z_{kj} v_{kj} \tag{3.11}$$

Advertiser $k$'s *return on investment* (ROI) on clicks purchased from network $j$ would therefore be:

$$R_{kj} \equiv \frac{Y_{kj}}{Z_{kj} v_{kj}} = \frac{\left( \sum_i V_i c_{ij} \beta_{ijk} \right) y_k}{\left( \sum_i V_i c_{ij} g_{ij} \right) v_{kj}} \tag{3.12}$$

Note that we are differentiating between bids and valuations here. Network $j$ does not know $v_{kj}$, so it runs auctions to extract this information. Advertiser $k$'s bid at auction does not necessarily have to be $v_{kj}$. Network $j$'s expected per-click auction revenue, $\theta_j$, is a function of $\{v_{kj}\ \forall k\}$ and the auction mechanism used by $j$.

### 3.2.1 Assumptions

**Separable Conversion Rates** We assume that conversion rates are *separable* i.e., that each $\beta_{ijk}$ is a product of three factors:

$$\beta_{ijk} = \beta_i^{\text{Pub}} \beta_j^{\text{Net}} \beta_k^{\text{Adv}} \ \ \forall (i, j, k) \tag{3.13}$$

Each factor in (3.13) has a different interpretation. $\beta_i^{\text{Pub}}$ measures how targeted publisher $i$'s traffic is with respect to the keyword in question. $\beta_j^{\text{Net}}$ measures how good network $j$ is at matching publishers' content with advertisers' ads. $\beta_k^{\text{Adv}}$ measures the quality and effectiveness of advertiser $k$'s ads. From (3.8), separability of conversion rates implies:

$$Y_{kj} = \beta_j^{\text{Net}} \left( \sum_i V_i c_{ij} \beta_i^{\text{Pub}} \right) \xi_{jk} \beta_k^{\text{Adv}} y_k \tag{3.14}$$

A related "separability" assumption is made in [1] and [33], where the click-through rate (CTR) for an advertisement is assumed to be the product of an advertiser-specific factor and a position/"slot"-specific factor.

**Linear Auctions** We assume that every network uses an auction that is *linear* in the sense that we discussed in Chapter 2: If all agents' valuations are scaled by a factor $\gamma$, then the expected revenue from the auction is also scaled by $\gamma$. First-price, second-price, Dutch and English auctions can all be shown to have this property. The maximal and minimal equilibrium revenues for the position auction in [33] and the generalized second-price auction in [12] are also linear in this sense.

We are not assuming that all networks use the same auction mechanism, or even that the mechanisms are truthful – only that each auction is linear. The linearity assumption will allow us to derive an explicit expression for network $j$'s expected per-click revenue, $\theta_j$ (see (3.20)).

Auction participants are chosen randomly on each click, to ensure that the same advertiser is not awarded every click. This practice is referred to as "throttling." Ad networks throttle ad placement to ensure that advertisers do not exhaust their budgets too early in a billing cycle, and that a variety of ads are displayed to the user. Thus, when computing expected auction revenues, the expectation is taken over the valuation of the randomly-chosen participants, as well as the possibly-randomized bidding and pricing strategies employed by advertisers and networks, respectively, in the auction.

**Sticky publishers** We assume that every network $j$ has some *sticky* publishers i.e., publishers who will sell their clicks to $j$ irrespective of how profitable it may be to do so. Search engines are perhaps the largest instances of sticky publishers e.g., only ads from the Google network are likely to appear on the website `www.google.com`, irrespective of how profitable other networks' ads may seem.

We combine all of network $j$'s sticky publishers into one, and assume that publisher $j$ always sets $c_{jj} = 1 \ \forall j \in 1, \ldots, J$ (recall that $J \ll I$). Thus, $V_j$ is the total volume across network $j$'s sticky publishers, and $j$ is guaranteed to receive at least $V_j$ clicks. The existence of sticky publishers implies that $Y_{kj}$ and $Z_{kj}$ are never zero, and therefore that advertiser $k$'s ROI on network $j$, $R_{kj}$, is always well-defined (see (3.8), (3.10) and (3.12)). Assuming the existence of sticky publishers assumption, though not essential for our results, helps to simplify our notation and analysis.

### 3.2.2 Publisher and Advertiser Objectives

In the first step, network $j$ chooses its pricing policy (i.e., $h_j$ and $\{g_{ij}\ \forall i\}$) such that its profit, $\eta_j$, is maximized. We discuss network $j$'s optimization problem in Section 3.3.

Publishers and advertisers know the networks' pricing policies when they make their decisions in the second step. Publisher $i$ chooses allocations $\{c_{ij}\ \forall j\}$ such that the total revenue generated from its sites is maximized:

$$\text{maximize}\ \ \pi_i\ \ \text{subject to}\ \ \sum_j c_{ij} = 1 \tag{3.15}$$

At the same time, each advertiser $k$ chooses valuations $v_{kj}$ that maximize its revenue from each network $j$, subject to a lower bound $R_k$ on ROI:

$$\text{maximize}\ \ Y_{kj}\ \ \text{subject to}\ \ R_{kj} \geq R_k \tag{3.16}$$

Here, $R_k$ is advertiser $k$'s *target ROI*. Intuitively, $R_k$ is the ROI that advertiser $k$ can achieve by advertising through channels other than CPC. As we demonstrate in Appendix A.1 solving (3.16) is equivalent to maximizing advertiser $k$'s combined profits from both online and "offline" advertising.

Publisher $i$'s type is $\left(V_i, \beta_i^{\text{Pub}}\right)$, network $j$'s type is $\beta_j^{\text{Net}}$ and advertiser $k$'s type is $\left(y_k, R_k, \beta_k^{\text{Adv}}\right)$. We assume that publishers' and networks' types are common knowledge, whereas each advertiser $k$'s type is known only to $k$. Assuming knowledge of the publishers' types may not seem realistic – however, one of the topics explored in Chapter 4 is how a network can, indeed, estimate the necessary parameters from readily observable click traffic data.

### 3.2.3 Publisher and Advertiser Best Responses

At the optimum, the constraint in (3.16) will be binding for each network $j$:

$$R_{kj} = R_k\ \forall j \tag{3.17}$$

To understand intuitively why (3.17) holds, assume for a moment that network $j$'s auction is truthful i.e., that advertiser $k$'s bid equal to it's valuation, $v_{kj}$. Recall that $R_k$ is the ROI that advertiser $k$ can achieve through channels other than CPC advertising. So, if $R_{kj} > R_k$, advertiser $k$ will want to spend more money on network $j$ i.e., it will want to buy more clicks from network $j$. To receive more clicks from network $j$, it must increase its bid, and hence it's valuation $v_{kj}$. However, from (3.12), we know that $R_{kj}$ is decreasing in $v_{kj}$. Therefore, advertiser $k$ will keep increasing $v_{kj}$ as long as $R_{kj} > R_k$, meaning (3.17) will hold at the optimum. This argument is informal – a more detailed discussion can be found in Appendix A.1.

From (3.12) and (3.17), advertiser $k$'s optimal valuation (i.e., its best response) is:

$$v_{kj} = \beta_j^{\text{Net}} \frac{\left(\sum_i V_i c_{ij} \beta_i^{\text{Pub}}\right)}{\left(\sum_i V_i c_{ij} g_{ij}\right)} \frac{\beta_k^{\text{Adv}} y_k}{R_k} = \bar{v}_k a_j \tag{3.18}$$

where $\bar{v}_k$ is defined as advertiser $k$'s *nominal valuation*, and $a_j$ is an *adjustment factor* applied to network $j$:

$$\bar{v}_k \equiv \frac{\beta_k^{\text{Adv}} y_k}{R_k} \qquad a_j \equiv \beta_j^{\text{Net}} \frac{\left(\sum_i V_i c_{ij} \beta_i^{\text{Pub}}\right)}{\left(\sum_i V_i c_{ij} g_{ij}\right)} \tag{3.19}$$

Intuitively, $a_j$ is proportional to the ratio between the number of clicks converted on network $j$ and the effective number of clicks billed for, after predictive pricing. The key point here is that publishers' and networks' decisions affect $v_{kj}$ only through the multiplicative factor $a_j$. Moreover, $v_{kj}$ depends on the actions of all publishers, but does not depend on the actions of any other advertisers.

Let $\kappa_j$ be network $j$'s expected revenue per click assuming $v_{kj} = \bar{v}_k \ \forall k$ i.e., $\theta_j = \kappa_j$ when $v_{kj} = \bar{v}_k$. Note that $\bar{v}_k$ does not depend on $j$. Therefore, if $\kappa_1 > \kappa_2$, it would mean network 1 is extracting more revenue per click than network 2, from the same set of valuations i.e., network 1's auction mechanism is more "efficient" than network 2's auction mechanism.

Now, suppose each advertiser $k$ chooses his valuation optimally i.e., (3.18) holds for all $k$. Compared to the scenario where $v_{kj} = \bar{v}_k \ \forall k$, each advertiser $k$'s valuation has been scaled up by a factor $a_j$. The assumption that network $j$'s auction is linear would therefore imply that:

$$\theta_j = \kappa_j a_j \tag{3.20}$$

From (3.3) and (3.4), $\pi_i$ is linear in publisher $i$'s allocations $\{c_{ij} \ \forall j\}$. The lone constraint in (3.15) is also linear in $c_{ij}$. Thus, solutions to (3.15) have a simple and intuitive form. Let $X_{ij}$ be publisher $i$'s revenue assuming it sends all of its traffic to network $j$ (i.e., $c_{ij} = 1$). From (3.3), we get:

$$X_{ij} = V_i g_{ij} h_j \theta_j \tag{3.21}$$

The optimal allocations $\{c'_{ij} \ \forall j\}$ for publisher $i$ (i.e., its best response) satisfy:

$$\sum_j c'_{ij} X_{ij} = \max_j X_{ij} \ \text{ and } \ \sum_j c'_{ij} = 1 \tag{3.22}$$

In words, it is optimal for publisher $i$ to send all its traffic to the single network whose $X_{ij}$ value is highest. If there is a tie between two or more networks, publisher $i$ can split its traffic arbitrarily between these networks.

We emphasize that the networks act first and publishers and advertisers second. So, when publishers compute their optimal allocations and advertisers compute their optimal valuations, the networks' actions (i.e., their pricing policies) are known. Therefore, $\{g_{ij} \ \forall (i,j)\}$ and $\{h_j \ \forall j\}$ are treated as constants and not variables in (3.15) and (3.16).

For a given first-step outcome $\{g_{ij} \ \forall (i,j)\}$ and $\{h_j \ \forall j\}$, an equilibrium in the second step is defined as a scenario where every advertiser $k$ chooses its valuations $\{v_{kj}\ \forall j\}$ optimally and every publisher $i$ chooses its allocations $\{c_{ij}\ \forall j\}$ optimally i.e., (3.18), (3.20) and (3.22) hold simultaneously for all $(i, j, k)$. Therefore, if an equilibrium is played in the second step, we can substitute (3.18) and (3.20) into (3.5), and simplify:

$$\eta_j = \left(\sum_i V_i c_{ij} g_{ij}\right)(1 - h_j)\theta_j \tag{3.23}$$

$$= \beta_j^{\text{Net}} \left(\sum_i V_i \beta_i^{\text{Pub}} c_{ij}\right)(1 - h_j)\kappa_j \tag{3.24}$$

The structure of (3.23) and (3.24) is quite intuitive. In order for network $j$ to maximize profits, it must simultaneously: a) attract a large volume of billable traffic from publishers (i.e., high $\sum_i V_i c_{ij} g_{ij}$), b) deliver high-conversion-rate traffic to advertisers (i.e., high $\theta_j$), and c) retain a large share of revenues (i.e., low $h_j$). Interestingly, the predictive pricing factors $g_{ij}$ do not appear anywhere in (3.24). However, they clearly have a strong influence on publisher allocations $c_{ij}$ and advertiser valuations $v_{kj}$ (see (3.18), (3.21) and (3.22)), and thus on network profits.

## 3.3 Optimal Pricing Policies

Network $j$'s goal is to maximize its own profit, $\eta_j$. From (3.24), $\eta_j$ depends on the decisions made by publishers and advertisers. However, the networks act first in our game. Publishers and advertisers observe the networks' decisions in the first step before deciding on their allocations and valuations in the second step. In other words, the outcome in the second step (i.e., allocations and valuations) is the market's reaction to first-step outcome (i.e., networks' pricing policies). Therefore, to maximize revenue, each network $j$ will: a) assume that an equilibrium will be played in the second step, and b) choose a pricing policy that induces the most profitable equilibrium in the second step.

The second-step outcome depends not only on network $j$'s pricing policy, but also on the pricing policies chosen by competing networks in the first step. For example, if the revenue share $h_j$ offered by network $j$ is too low, then very few publishers may send traffic to $j$ (i.e., $c_{ij} = 0$ for most $i$), leading to a low $\eta_j$. If $h_j$ were too high, more publishers may send traffic to network $j$, but $\eta_j$ might be low again since $j$ would be paying out a large fraction of revenues to publishers. Therefore, network $j$ must account for the actions of all other networks when choosing its own pricing policy.

We can now express the *best response* of network 1, holding the actions of all other networks fixed, and assuming an equilibrium is played in the second step. Our choice of network 1 is without loss of generality – obviously, we can compute the best response for any network $j$ in a similar manner, holding the actions of all other networks fixed. Combining Equations (3.5), (3.19), (3.20), (3.21) and (3.22), network 1's best response is a solution to the following optimization problem:

$$\begin{aligned}
\text{maximize} \quad & \eta_1 \equiv \beta_1^{\text{Net}} \left( \sum_i V_i \beta_i^{\text{Pub}} c_{i1} \right) (1 - h_1)\kappa_1 \\
\text{subject to} \quad & X_{ij} = V_i g_{ij} h_j \theta_j \;\; \forall (i,j) \\
& \sum_j c_{ij} X_{ij} = \max_j X_{ij} \;\; \forall i \\
& \sum_j c_{ij} = 1 \;\; \forall i \\
& \theta_j = \kappa_j a_j \;\; \forall j \\
& a_j = \beta_j^{\text{Net}} \frac{\left(\sum_i V_i c_{ij} \beta_i^{\text{Pub}}\right)}{\left(\sum_i V_i c_{ij} g_{ij}\right)} \;\; \forall j \\
& 0 \le g_{i1}, h_1, c_{ij} \le 1 \;\; \forall (i,j)
\end{aligned} \tag{3.25}$$

The objective in (3.25) is an expression for network 1's profit (see (3.24)). The first three constraints encode the assumption that each publisher chooses allocations optimally in the second step (see (3.21) and (3.22)). The fourth and fifth constraints say that advertisers also choose valuations optimally (see (3.19) and (3.20)) i.e., that there is an equilibrium in the second step between publishers and advertisers. The final constraint gives ranges for the various decision variables we are interested in.

Perhaps the most striking feature of (3.25) is that the advertisers' valuations $v_{kj}$ do not appear anywhere. Neither do the advertisers' types, $y_k$, $R_k$ and $\beta_k^{\text{Adv}}$. The constants $\{\kappa_j \; \forall j\}$ act as sufficient statistics for the distribution of advertiser types in our problem. The fact that $v_{kj}$ does not appear greatly reduces the complexity of solving (3.25), since the number of decision variables is reduced by a factor $K$.

### 3.3.1 PRICINGPOLICY

Network 1's optimization problem (3.25) is highly nonconvex, which makes even feasible points difficult to find. One of our main contributions is an iterative algorithm, which we call PRICINGPOLICY, for finding near-optimal solutions to (3.25). We describe the key concepts here, and present a detailed derivation of PRICINGPOLICY in Appendix A.5.

Define $\mathbf{g}_1 \equiv \{g_{i1}\}_{I\times 1}$ i.e., $\mathbf{g}_1$ is an $I$-by-1 matrix (or, a length-$I$ vector) whose $i^{\text{th}}$ element is $g_{i1}$. Similarly, let $\mathbf{C} \equiv \{c_{ij}\}_{I\times J}$ be an $I$-by-$J$ matrix of publisher allocations, and let $\mathbf{G}_{-1} \equiv \{g_{ij} \; \forall j \neq 1\}_{I\times(J-1)}$ and $\mathbf{h}_{-1} \equiv \{h_j \; \forall j \neq 1\}_{(J-1)\times 1}$ denote the actions of the other networks. Recall that in (3.25), $\mathbf{G}_{-1}$ and $\mathbf{h}_{-1}$ are given as inputs – we are finding network 1's best response to $\mathbf{G}_{-1}$ and $\mathbf{h}_{-1}$, so they are not variables.

We refer to triples $(h_1, \mathbf{g}_1, \mathbf{C})$ as *points*. We say that the point $(h_1, \mathbf{g}_1, \mathbf{C})$ is *feasible* if it satisfies the constraints in (3.25). If $(h_1, \mathbf{g}_1, \mathbf{C})$ is feasible, it means that if network 1 plays $(h_1, \mathbf{g}_1)$ and the other networks play $(\mathbf{h}_{-1}, \mathbf{G}_{-1})$ in the first step, then the publishers' equilibrium allocations in the second step will be $\mathbf{C}$ (recall that the corresponding advertiser valuations can be computed from (3.18)). We say that $(h_1^*, \mathbf{g}_1^*, \mathbf{C}^*)$ is *optimal* if it is feasible and network 1's profit is (weakly) the highest when $(h_1^*, \mathbf{g}_1^*, \mathbf{C}^*)$ is played, compared to any other feasible point i.e., it is a best response to $(\mathbf{h}_{-1}, \mathbf{G}_{-1})$.

Starting from (3.25), we can make a sequence of transformations (described in Appendix A.5), that yield a *geometric programming (GP) relaxation* of (3.25) around a given point $\left(h_1^{(t)}, \mathbf{g}_1^{(t)}, \mathbf{C}^{(t)}\right)$. That is, we approximate (3.25) by a GP in the vicinity of the point $\left(h_1^{(t)}, \mathbf{g}_1^{(t)}, \mathbf{C}^{(t)}\right)$. GPs are log-convex [4], and therefore can be solved globally and efficiently. PRICINGPOLICY works by solving a sequence of these GPs. It outputs a sequence of feasible (but not-necessarily optimal) points, where each point yields weakly higher profits for network 1 than the previous point. The sequence of solutions converge to an approximate solution (i.e., a local optimum) to (3.25).

**Algorithm 1** PRICINGPOLICY

**Require:** $\mathbf{G}_{-1}, \mathbf{h}_{-1}, T$

1: Select arbitrary initializations $h_1^{(0)}$ and $\mathbf{g}_1^{(0)}$
2: Compute second-step equilibrium, $\mathbf{C}^{(0)}$, assuming other networks play $(\mathbf{h}_{-1}, \mathbf{G}_{-1})$ and network 1 plays $\left(h_1^{(0)}, \mathbf{g}_1^{(0)}\right)$
3: **for** $t \in 1, \ldots, T$ **do**
4: Solve GP-relaxation of (3.25) to find an optimal point $(h_1', \mathbf{g}_1', \mathbf{C}')$ that is "close to" $\left(h_1^{(t-1)}, \mathbf{g}_1^{(t-1)}, \mathbf{C}^{(t-1)}\right)$, assuming other networks play $(\mathbf{h}_{-1}, \mathbf{G}_{-1})$
5: $\left(h_1^{(t)}, \mathbf{g}_1^{(t)}, \mathbf{C}^{(t)}\right) \leftarrow (h_1', \mathbf{g}_1', \mathbf{C}')$
6: **end for**
7: Recompute second-step equilibrium, $\mathbf{C}^{(T)}$, assuming other networks play $(\mathbf{h}_{-1}, \mathbf{G}_{-1})$ and network 1 plays $\left(h_1^{(T)}, \mathbf{g}_1^{(T)}\right)$
8: **return** $\left(h_1^{(T)}, \mathbf{g}_1^{(T)}, \mathbf{C}^{(T)}\right)$

Given $\left(h_1^{(0)}, \mathbf{g}_1^{(0)}\right)$ and $(\mathbf{h}_{-1}, \mathbf{G}_{-1})$ (i.e., the first-step outcome), the second-step equilibrium allocations $\mathbf{C}^{(0)}$ in line 2 of PRICINGPOLICY can be computed using fixed-point iteration on Equations (3.20), (3.21) and (3.22). The purpose of line 2 is to provide a feasible starting point $\left(h_1^{(0)}, \mathbf{g}_1^{(0)}, \mathbf{C}^{(0)}\right)$ for the inner loop of PRICINGPOLICY (if the solution is not unique, we select one at random). Similarly, line 7 ensures that the final output $\left(h_1^{(T)}, \mathbf{g}_1^{(T)}, \mathbf{C}^{(T)}\right)$ is feasible for (3.25), since the solutions of the relaxed problem may be infeasible for the original optimization problem (3.25).

Different initializations $\mathbf{g}_1^{(0)}$ and $h_1^{(0)}$ may lead to different local optima, so PRICINGPOLICY should be executed several times with different initializations, keeping the best result. There are also a number of minor tweaks needed to ensure that PRICINGPOLICY works well in practice. Some of these tweaks are related to "regularization," to ensure the numerical stability of the algorithm. It is also prudent to verify that the sequence of approximate solutions output by PRICINGPOLICY corresponds to progress being made on the original problem.

In principle, PRICINGPOLICY could be used as a subroutine to compute an equilibrium for the first step i.e., for computing a subgame-perfect equilibrium for our one-shot dynamic game. We would run the algorithm (i.e., compute a best response) for each network $j$ holding all other networks' actions fixed, and iterate until convergence. However, we feel that computing the best-response function is more useful in practice – what pricing policy should network $j$ use in order to maximize its own profits?

## 3.4 Experiments

Using PRICINGPOLICY, we can gain some interesting insights into the structure of best-response pricing policies, and the effects of various parameters in our model.

### 3.4.1 Predictive Pricing

Our first experiment examines whether networks that apply predictive pricing gain a competitive edge, compared to networks that do not. Consider a market with $J = 2$ networks and $I = 20$ publishers. Each publisher $i$ receives 100 clicks (i.e., $V_i = 100\ \forall i$), and the quality of $i$'s traffic, $\beta_i^{\text{Pub}}$, is linear in $i$ with values ranging from 0.25% to 5% (i.e., $\beta_i^{\text{Pub}} = 0.0025i$). Such a range is realistic – 5% would be considered a very high conversion rate in practice. The networks are equally effective at matching up publishers and advertisers i.e., $\beta_1^{\text{Net}} = \beta_2^{\text{Net}} = 1.0$. We assume $\kappa_1 = \kappa_2 = 10$, which means the auction mechanisms used by each network are also equally efficient.

We used PRICINGPOLICY to compute the best-response pricing policy for network 1, assuming network 2 does not use predictive pricing (i.e., $g_{i2} = 1\ \forall i$) and offers publishers a revenue share of 50% (i.e., $h_2 = 0.5$). To solve the GP-relaxation of (3.25) in line 4 of PRICINGPOLICY, we used CVX, a software package for specifying and solving convex programs [18, 19]. We initialized the algorithm with random choices of $\mathbf{g}_1$ and $h_1$ for network 1.

Figure 3.1 shows the revenue share $h_1^{(t)}$ output at each iteration $t$, as well as the market share $\frac{1}{I}\sum_i c_{i1}^{(t)}$, estimated profit $\hat{\eta}_1^{(t)}$ and actual profit $\eta_1^{(t)}$ at each iteration. From (3.24), note that $\eta_1 \leq \left(\sum_i V_i \beta_i^{\text{Pub}}\right) \kappa_1 \beta_1^{\text{Net}} \equiv \eta_1^{\text{max}}$, which is the maximum possible profit network 1 can attain in any outcome. Thus, in Figure 3.1 (and later in Figure 3.7), we normalize profits by $\eta_1^{\text{max}}$. The "estimated profit" is computed using the estimated allocations $\mathbf{C}^{(t)}$ output by iteration $t$ of PRICINGPOLICY, whereas the "actual profit" is computed using the actual second-step equilibrium allocations that would result if network 1 played $\left(h_1^{(t)}, \mathbf{g}_1^{(t)}\right)$. Observe that the estimated profit tracks the actual profit reasonably well – in this case it is an underestimate of the actual profit, but in other experiments we ran it was an overestimate.

From Figure 3.1, we see that the algorithm converges after roughly $T = 50$ iterations. PRICINGPOLICY, by its nature (i.e., solving a sequence of locally-smoothed approximations), always converges to a local optimum of (3.25). In each iteration, a GP with $O(IJ)$ constraints is solved, so the computational effort required in each iteration scales with the product $IJ$. However, the constraint set is "sparse" (i.e., each constraint involves a small number of variables). There may be an opportunity for computational savings in cleverly exploiting this sparsity (we leave this problem for future work). In the experiments that we ran, the number of iterations required for convergence did not vary greatly with the size of market.

As iterations progress, the revenue share $h_1^{(t)}$ in Figure 3.1 steadily decreases. PRICINGPOLICY is trying to minimize the share that network 1 pays out (i.e., so that network 1 can retain a larger share of total revenues) by recommending progressively better predictive prices $\mathbf{g}_1^{(t)}$ (not shown). It may seem surprising that the

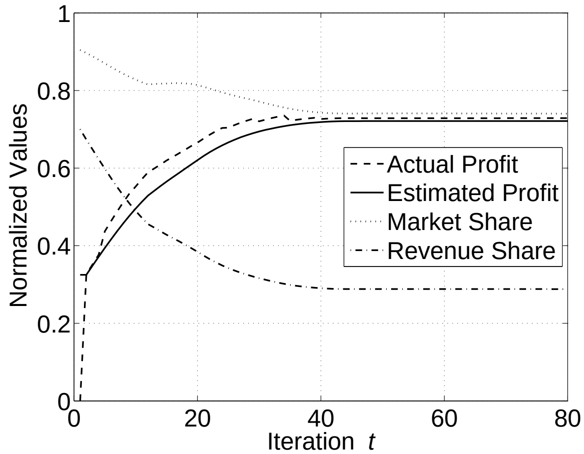


Figure 3.1: Progress of PRICINGPOLICY.

market share in Figure 3.1 is also falling across iterations. This decrease reflects the idea that attracting traffic from every single publisher is not necessarily optimal – quality is just as important quantity.

Observe that the algorithm converges to a revenue share of 29%, which is much lower than the 50% being offered by network 2. But despite offering a lower revenue share, network 1 manages to attract a 74% market share of publishers (recall that advertisers enter bids on all networks). The lowest-quality (i.e., lowest $\beta_i^{\text{Pub}}$) publishers are being dissuaded from sending traffic to network 1, and so they choose network 2 instead.

Figure 3.2 shows $\beta_i^{\text{Pub}}$ for each publisher $i$, and the final set of predictive pricing factors $g_{i1}^{(T)}$ for each publisher. Figure 3.2 suggests why the low-quality publishers choose network 2 over network 1. Advertisers are being charged very low prices (i.e., low $g_{i1}$) for traffic from low-quality publishers (i.e., low $\beta_i^{\text{Pub}}$). Consequently, network 1 offers to pay these low-quality publishers very little for their traffic (see (3.3)), causing them to choose network 2 instead. Thus, the use of predictive pricing is giving network 1 a significant advantage.

We may now ask what is the effect of network 1's predictive prices on publishers' profits? Consider a scenario where, initially, network 2 is the only network operating in the market. That is, network 2 initially has a "monopoly." Network 2 does not use predictive pricing and offers a revenue share of $h_2 = 50\%$. Then, network 1 enters the market, and uses the revenue share and predictive pricing factors shown in Figures 3.1 and 3.2. We now have a "duopoly."

Figure 3.3 is a plot of each publisher $i$'s profits in the monopoly and duopoly cases. In the monopoly case, all publishers earn the same amount of revenue because predictive pricing is not being used and their click volumes are equal (as such, we normalize the profits in Figure 3.3 by the publishers' profits in the monopoly case). In the duopoly case, we see a rather surprising effect – all publishers (both high- and low-quality) make less money than in the monopoly case, although high-quality publishers make more than low-quality ones. In both cases, observe that each publisher's revenue is proportional to the predictive pricing factor applied by the network that they send traffic to (see (3.3) and Figure 3.2).

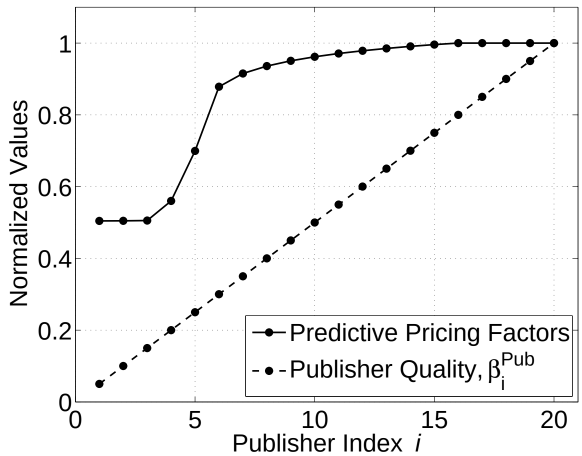


Figure 3.2: Best-response predictive prices, $g_{i1}^{(T)}$.

Why do all publishers earn less in the duopoly case? Network 1's entry is creating a bad equilibrium for the publishers. When network 1 enters the market (and creates a duopoly), predictive pricing causes advertisers to bid more on network 1 than network 2. Attracted by the higher bids, high-quality publishers leave network 2 and send their traffic to network 1 instead. Thus, the average quality on network 2 decreases, leading to lower bids on network 2, and lower revenues for publishers that remain on network 2. Publishers on network 1 are willing to accept a lower revenue share (leading to lower publisher revenues) in the second-step equilibrium, since advertisers' bids are increased.

Essentially, a "lemons market" effect is avoided on network 1 as a result of predictive pricing. The lack of low-quality publishers on network 1 raises the average quality of network 1's traffic (recall that quality is measured by conversion rate), causing advertisers' bids to increase. Unfortunately for the publishers, the high-quality publishers get paid more per click (i.e., $X_{ij}$ is higher), but have to settle for a lower revenue share as a result.

In this sense, the use of predictive pricing is allowing network 1 to "bully" publishers into accepting lower revenues. It is worth noting then, that if high-quality publishers were to collude (that is, collaborate on their choice of network), they could avoid yielding such an advantage. This situation explains the popularity of traffic aggregator services, amongst smaller content publishers.

### 3.4.2 Publisher Types

As we have seen in the experiments thus far, networks are able to extract value by differentiating between publishers of different traffic qualities. The objective is to increase the average traffic quality on the network, causing publishers to accept a lower revenue share in return. It would seem, then, that networks can only profit and extract value from markets where there is a wide variation in publisher quality.

To test this intuition, we can manipulate the variance in publisher quality, and study the impact of doing so

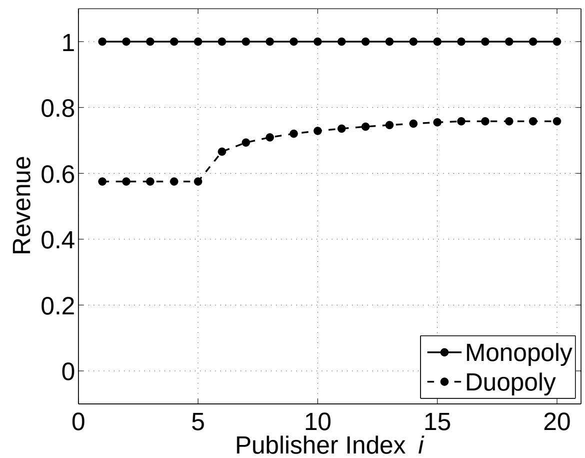


Figure 3.3: Effects of predictive pricing on publishers' profits.

on the market outcome. Suppose publisher $i$'s conversion rate, $\beta_i^{\text{Pub}}$, is given by the following linear equation:

$$\beta_i^{\text{Pub}} = \beta_0 + \gamma \frac{i}{I} \quad \text{where} \quad \beta_0 \equiv \frac{I+1}{2I}(\beta^{\max} - \gamma) \tag{3.26}$$

In (3.26), the parameter $\beta^{\max} = 5\%$ is the conversion rate of the highest-quality publisher, and the parameter $\gamma$ is the "slope." Larger $\gamma$ means that there is greater variation between the conversion rates of the publishers in the market. Note that the value of $\sum_i \beta_i^{\text{Pub}}$ due to (3.26) remains constant, irrespective of $\gamma$ i.e., the total number of conversions (and, consequently, the total amount of money to be divided amongst the publishers and networks) is fixed – see (3.24). Setting $\gamma = 5\%$ gives us the scenario considered in Section 3.4.1.

We varied the value of $\gamma$ between 1% and 5%, and observed the resulting division of profits between the publishers and networks. Figure 3.4 shows the distribution of publisher quality for the various values of $\gamma$. For each value, we used PRICINGPOLICY to compute the best response for network 1, and computed network 1's resulting profits and the publishers' total revenues. As in the earlier example, $J = 2$, $I = 20$, and network 1 does not use predictive pricing and offers a revenue share of $h_2 = 50\%$.

Figure 3.5 shows network 1's normalized profit, best-response revenue share and the total publisher profits for each value of $\gamma$ (in Figure 3.5, note that the total normalized profits of publishers and network 1 add up to 1). As predicted, we see that larger $\gamma$ means that network 1 can apply predictive pricing more aggressively, leading to a lower offered revenue share $h_1$, and thus increased profits for network 1. In general, predictive pricing is used to attract high-quality traffic, and the payoff to network 1 comes via a lower revenue share.

So far, we have focused on the effects of $\beta_i^{\text{Pub}}$ on market outcomes. The other element of publisher $i$'s type, of course, is the volume $V_i$. In our next experiment, we examine the effect of volume on the predictive prices chosen by networks. For example, if a low-quality network attracts a very high volume of traffic, network 1 may not want to apply such a low predictive pricing factor. That is, the aggregate number of conversions might be substantial despite the low conversion rate, making the low-quality publisher's traffic

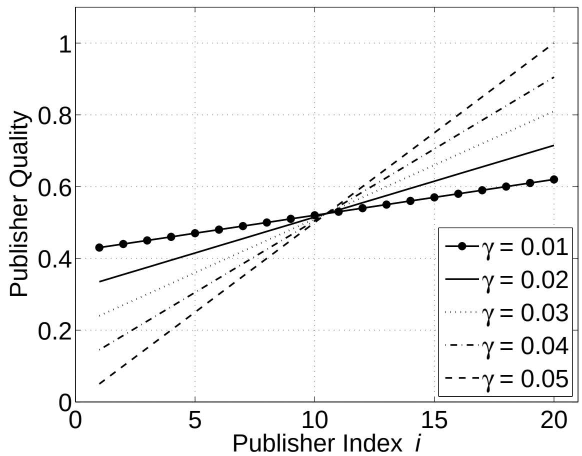


Figure 3.4: Publishers' conversion rates for various $\gamma$. Larger slopes correspond to larger $\gamma$.

attractive to network 1.

To study the effect of volume, we considered the same example as in Section 3.4.1, except that we assumed $V_i$ was decreasing linearly in $i$:

$$V_i = (I + 1 - i)V^{\text{min}} \tag{3.27}$$

Here, $V^{\text{min}}$ is the click volume of the lowest-volume publisher. So, in this example, the highest-quality publishers receive the lowest volume of traffic, whereas the lowest-quality publishers receive the highest volume (recall that $\beta_i^{\text{Pub}} = 0.0025i$). We computed the best-response pricing policy assuming (3.27). Figure 3.6 is a plot of the predictive prices in this case ("decreasing volumes") compared to the earlier case where $V_i$ was a equal across all $i$ ("equal volumes").

We see that volume does have a small effect on the network's best-response predictive prices. Qualitatively, however, the profiles are similar – the lowest-quality publisher still receives a very low predictive pricing factor, even though it actually generates the same number of conversions as the highest quality publisher (observe that $\beta_1^{\text{Pub}} V_1 = \beta_{20}^{\text{Pub}} V_{20}$).

Recall that advertisers adjust their valuations so that they attain their target ROI on each network (see (3.18)). As such, publishers with higher aggregate numbers of conversions can receive low predictive prices based on a low conversion rate. A high volume of nonconverting clicks will result in advertisers being billed for a lot of "junk" (see (3.10)), leading to lower valuations and bids. Thus, in our model, conversion rates are the primary determinant of predictive prices.

To summarize, we have learned the following from Sections 3.4.1 and 3.4.2:

- Even though two networks can be equals in all other respects, the use of predictive pricing alone can give one network a significant market edge.

- It is not necessarily optimal to attract as much traffic as possible – traffic quality on a network is

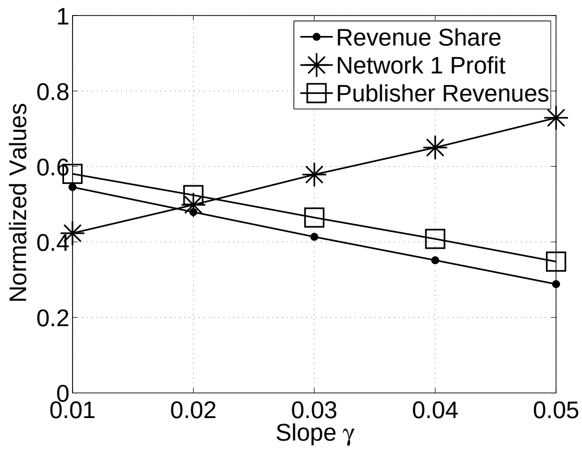


Figure 3.5: The effect of variation in publishers' conversion rates.

sometimes more important than traffic volume.

- Predictive pricing disproportionately helps high-quality publishers and hurts low-quality publishers, although it can lead to uniformly lower profits for publishers.

### 3.4.3 Targeting

Our final experiment considers the impact of targeting (i.e., $\beta_j^{\text{Net}}$) on market outcomes. In particular, if a network is more effective than its competitors at matching publishers with advertisers (i.e., its $\beta_j^{\text{Net}}$ is higher), does it translate to higher profits for that network?

Consider a market with $J = 3$ networks and $I = 20$ publishers. For variety, we will assume $\beta_i^{\text{Pub}} = 0.000125i^2$. That is, $\beta_i^{\text{Pub}}$ is quadratic in $i$ (the previous experiments assumed $\beta_i^{\text{Pub}}$ was linear in $i$), with values ranging from 0.0125% to 5% (there are many low-quality publishers and a few high-quality ones).

The three networks each behave differently:

- Network 1 is the "entrant" into the market. It observes the pricing policies being used by networks 2 and 3, and chooses a best-response for itself.
- Network 2 is a "quasi-CPA" network. It uses a predictive pricing rule that is linear in publisher $i$'s conversion rate (i.e., $g_{i2} = 20\beta_i^{\text{Pub}}$), and offers a revenue share of $h_2 = 0.5$.
- Network 3 is a "traditional" network. It does not use predictive pricing (i.e., $g_{i3} = 1\ \forall i$). To compensate, it offers a higher revenue share to publishers than network 2, i.e., $h_3 = 0.6$.

We refer to network 2 as "quasi-CPA" since, as demonstrated in Chapter 4, choosing $g_{ij}$ proportional to $\beta_i^{\text{Pub}}$ is closely related to *cost-per-action* (CPA) pricing schemes (see [23] for a detailed discussion of CPA).

Networks 2 and 3 are equally skilled at matching i.e., $\beta_2^{\text{Net}} = \beta_3^{\text{Net}} = 1.0$. We will select different values for $\beta_1^{\text{Net}}$ and observe the market outcome in each instance. We also assume that $\kappa_1 = \kappa_2 = \kappa_3 = 10$, so no

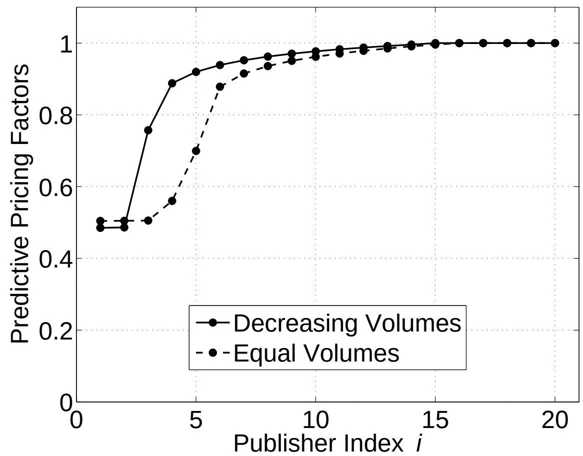


Figure 3.6: Effect of traffic volume on predictive prices.

network has an edge due to the auction mechanism they use. It can be shown that varying the (relative) value of $\kappa_j$ has a similar market effect as varying $\beta_j^{\text{Net}}$.

We computed optimal pricing policies for network 1, for values of $\beta_1^{\text{Net}}$ ranging from 0.7 to 1.3. Recall that $\beta_1^{\text{Net}}$ greater than (less than) 1.0 means that network 1 is better (resp., worse) at targeting than networks 2 and 3. Figure 3.7 shows network 1's optimal revenue share $h_1^*$ and its resulting profits (normalized by $\eta_1^{\max}$). As we might expect, network 1 earns higher (lower) profits when $\beta_1^{\text{Net}}$ is higher (resp., lower). From Figure 3.7, we see that network 1 is able to offer a lower revenue share when $\beta_1^{\text{Net}}$ is higher, since network 1 is generating more conversions for advertisers, causing bids (and consequently publishers' revenues) to increase.

Finally, the values that we chose for $\mathbf{g}_2$, $\mathbf{g}_3$, $h_2$ and $h_3$ in this experiment were such that, for smaller $\beta_1^{\text{Net}}$, all three networks captured a positive market share. Such outcomes suggest that in real-world markets with many networks, it is possible for each network to select its pricing policy to target a particular "niche" or "stratum" of the publisher population.

## 3.5 Managerial Implications

Our results have clear implications for those engaged in the online advertising business. We saw that it is in a network's best interest to judiciously apply predictive pricing and revenue sharing, towards attracting high-quality traffic. Pricing policies are relatively simple to implement – it amounts to simply offering discounts – and networks that do not use predictive pricing yield a significant advantage to their competitors.

Advertisers really ought to be thinking in terms of conversion rates and traffic quality when bidding for click-throughs, rather than just trying to maximize the number of click-throughs. It is a frequently-recommended strategy on search-engine optimization (SEO) sites to hunt for keywords that will result in a high click-through rate. There exists prior work on optimal keyword bidding strategies (e.g., [22]), but

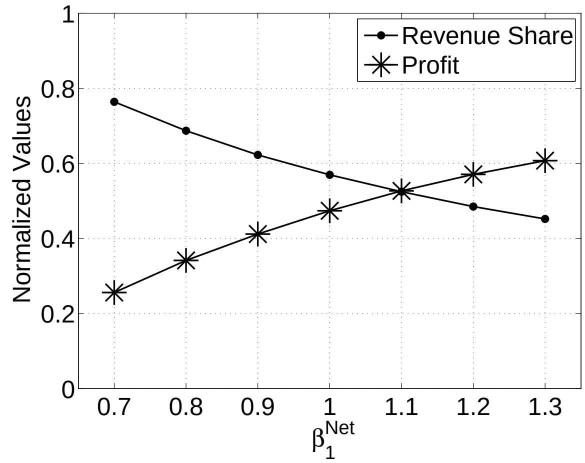


Figure 3.7: The effect of network 1's skill at matching publishers and advertisers (i.e., $\beta_1^{\mathrm{Net}}$).

not much focus on the issue of click quality. Our model also reinforces the role of ROI as the primary criterion for advertisers' decision-making (see Appendix A.1). As for publishers, we observed that high-quality publishers can benefit disproportionately from predictive pricing. So, rather than attempting to attract an increased volume of traffic, it may indeed be more profitable to improve conversion rates (perhaps, by publishing better-targeted content). On the other hand, lower-quality publishers may actually be better off targeting networks that do not use predictive pricing (or, ones that use sub-optimal policies).

At an industry level, we believe further adoption of predictive pricing policies is inevitable. Probably, more networks are currently using such pricing schemes than would publicly admit. The provision of detailed traffic quality data to prospective publishers and advertisers is, itself, a competitive strategy that is perhaps worthy of analysis by the networks. Finally, we do not believe CPA schemes are necessarily the right direction for the industry, mainly due to the inherently unequal distribution of risk amongst publishers, advertisers and networks. Many of the benefits of CPA can be achieved using CPC and predictive pricing, and more equitably so, in terms of risk.

## 3.6 Summary

We have presented an economic model of the CPC advertising market that captures the effects of predictive pricing and revenue sharing. The model is simple, yet flexible enough to account not only for conversion rates, but also click-through rates and click fraud (although we did not discuss them in this chapter). We derived an implicit expression for the optimal pricing policy for a network, as the solution to a difficult optimization problem. We then presented an iterative algorithm, PRICINGPOLICY, which finds near-optimal solutions to this problem.

Through experiments, we found that predictive pricing and revenue sharing can be very effective tools for advertising networks to attract publishers and advertisers, especially if their competitors are not using predictive pricing. It is not necessarily optimal to attract as much traffic as possible – quality can be just as important as quantity. Being more effective at matching publishers and advertisers can increase a network's profits, so improving their matching algorithms may be a worthwhile investment for networks.

# Chapter 4

# Managing the Quality of CPC Traffic

## 4.1 Introduction

As we saw in Chapters 2 and 3, there are two distinct-but-related aspects of traffic quality: *validity* and *targetedness*. Validity refers to whether a click is valid or invalid. Valid clicks have a strictly positive probability of becoming a conversion, whereas invalid clicks have zero probability. Targetedness refers to the likelihood that a valid click will become a conversion, and is measured by the conversion rate.

In this chapter, we show how these two aspects of click quality can be managed together, using filtering, predictive pricing and revenue sharing.

#### Overview

We begin by revisiting our economic model of the CPC advertising market (Sections 4.2.1 and 4.2.2). Our intent is to hone in on the decisions that affect traffic quality. Although our focus is on CPC, an analogous model can be developed for the *cost-per-mille* (CPM) and *cost-per-action* (CPA) pricing schemes.

The main difference between this chapter and earlier ones is that we apply filtering, predictive pricing and revenue sharing simultaneously. In Chapters 2 and 3, we considered targetedness and validity in isolation; now, we analyze them together. In particular, we provide an expression for an advertising network's *best-response* function (Section 4.2.3). That is, if a network knows the filtering, predictive pricing and revenue sharing policies of its competitors, what policy should the network choose in response?

We then study the properties of this best-response function, as they relate to the joint management of the validity and targetedness aspects of traffic quality. Looking at filtering, predictive pricing and revenue sharing together, we gain insight into the holistic management of CPC traffic quality.

Our main conclusions can be summarized as follows:

- Due to the underlying incentives, it is important to distinguish between the two sources of click traffic: *organic* traffic and publisher-initiated *click inflation*. All valid traffic is organic, by definition, although organic traffic can also be invalid. All click inflation is, by definition, invalid (it is a form of click fraud).
- To manage the quality of organic traffic, it is unnecessary (and, in some cases, suboptimal) to use both predictive pricing and filtering simultaneously. Predictive pricing alone is enough. (Section 4.3)
- Filtering can, however, be (indirectly) useful in fighting click inflation, as long as the performance of the filtering algorithm can be accurately characterized. Otherwise, predictive pricing can be used to fight click inflation. (Sections 4.4.2 and 4.4.3)

## 4.2 CPC Advertising

In this section, we review our model of the CPC advertising market. Our model is a generalization of Chapters 2 and 3, so we remain notationally consistent whenever possible.

We model the CPC market as a one-shot dynamic game between three classes of players: content publishers, advertising networks and advertisers. Content publishers publish websites and display advertisements alongside their content. Advertisers design advertisements and bid at auction on keyword queries that best describe the interests of their target market. Advertising networks act as intermediaries between publishers and advertisers, by first judging which keywords best describe each publisher's content, and then delivering ads to the publisher from advertisers that have bid on those keywords.

When a user visits a publisher's site and clicks on an ad related to a given keyword, we say that a click-through has occurred on that ad. Networks apply both filtering and predictive pricing. If the click is deemed valid by the network, the advertiser pays the network a small amount. The network then pays out a fraction of this amount to the publisher where the click originated. Filtering is the process of detecting invalid clicks. Recall that predictive pricing affects how much the advertiser is billed by the network. The revenue share determines what fraction of this billed amount the network will pay out to the publisher.

A small fraction of valid clicks become conversions for the advertiser e.g., a product purchase, or a signup to an e-mail list. The advertiser earns some revenue each time a click becomes a conversion.

### 4.2.1 Notation

Table 4.1 is a summary of the notation used in this chapter. To a large extent, this section is an integration of Chapters 2 and 3, so we remain notationally consistent whenever possible.

Consider a single keyword. Suppose there are $I$ publishers whose content is relevant to the keyword, $K$ advertisers interested in buying clicks on this keyword, and $J$ networks.

**Traffic.** Each publisher $i$ receives $V_i$ clicks on his website, of which only a fraction $r_i$ are valid. That is, publisher $i$ receives $r_iV_i$ valid clicks and $(1 - r_i)V_i$ invalid clicks in total. For now, we assume that $r_i$ and $V_i$ are fixed parameters that describe the validity of publisher $i$'s traffic (we will relax this assumption in Section 4.4, when we discuss click inflation).

In our model, each publisher $i$ decides how to allocate its volume, $V_i$, of clicks across the $J$ networks. Note that in practice, publishers allocate ad impressions (or, "page views"), rather than clicks. Under some reasonable assumptions (see Chapter 2 for a discussion), however, it is equivalent to model clicks (rather than impressions) as the objects being bought and sold. Let $c_{ij}$ be the fraction of publisher $i$'s clicks that are sent to network $j$. Then, $V_ic_{ij}$ is the total number of clicks that publisher $i$ sends network $j$, of which

$$r_iV_ic_{ij} \tag{4.1}$$

are valid clicks.

**Filtering.** The algorithms used by networks to filter out invalid clicks are prone to error. In particular, the algorithms may produce false positives ("Type I" errors) by marking valid clicks invalid. They may also produce false negatives ("Type II" errors) by marking invalid clicks valid.

Let $u_j$ be the fraction of valid clicks that network $j$'s filtering algorithm correctly identifies as valid (i.e., *true negatives*). We assume in our model that $f_j(u_j) \equiv u_j^{\gamma_j}$ will be the fraction of invalid clicks that network $j$ mistakenly marks valid (i.e., *false negatives*). $\gamma_j \in [1, \infty)$ is a fixed, intrinsic parameter for $j$ i.e., it cannot

| | |
|---|---|
| **Player Parameters** | |
| $r_i$ | Fraction of publisher $i$'s traffic that is valid |
| $V_i$ | Volume of clicks on publisher $i$'s site |
| $\beta_i^{\text{Pub}}$ | Targetedness of publisher $i$'s traffic |
| $\gamma_j$ | Network $j$'s effectiveness at filtering |
| $\beta_j^{\text{Net}}$ | Network $j$'s skill at matching |
| $y_k$ | Revenue earned by advertiser $k$ on each conversion |
| $R_k$ | Advertiser $k$'s target ROI |
| $\beta_k^{\text{Adv}}$ | Effectiveness of advertiser $k$'s ads |
| **Decision Variables** | |
| $c_{ij}$ | Fraction of publisher $i$'s clicks sent to network $j$ |
| $u_j$ | Network $j$'s aggressiveness at filtering |
| $g_{ij}$ | Predictive price applied by $j$ to publisher $i$'s clicks |
| $\mathbf{g}_j$ | $= \{g_{ij}\ \forall i\}$, predictive prices chosen by network $j$ |
| $h_j$ | Revenue share paid out by network $j$ |
| $v_{kj}$ | Advertiser $k$'s valuation of network $j$'s clicks |
| **Derived Parameters** | |
| $B_i$ | $= (1 - r_i)V_i$, volume of invalid ("bad") clicks |
| $G_i$ | $= r_i V_i$, volume of valid ("good") clicks |
| $\beta_{ijk}$ | Conversion rate of clicks going from $i$ to $j$ to $k$ |
| $A_i$ | Nominal number of conversions due to publisher $i$ |
| $A_{ijk}$ | Number of converted clicks going from $i$ to $j$ to $k$ |
| $\xi_{ijk}$ | Fraction of $i$'s traffic that $j$ forwards on to $k$ |
| $\theta_j$ | Network $j$'s auction revenue per click |
| $\kappa_j$ | Network $j$'s nominal auction revenue per click |
| $N_{ij}$ | Fraction of $i$'s clicks that are marked valid by $j$ |
| $E_{ij}$ | Effective number of clicks for which $j$ pays $i$ |
| $\pi_{ij}$ | Publisher $i$'s revenue from clicks sent to network $j$ |
| $X_{ij}$ | Publisher $i$'s revenue from sending all its clicks to $j$ |
| $\eta_j$ | Network $j$'s total revenue |

Table 4.1: Notation used in Chapter 4.

be changed or controlled by network $j$.

The fraction $u_j$ is a measure of how aggressively network $j$ is filtering for invalid clicks (lower $u_j$ means more aggressive). The parameter $\gamma_j$ is a measure how effective network $j$ is at filtering (higher $\gamma_j$ means more effective, since it leads to fewer false negatives for a given $u_j$).

Notice that the formulation in this chapter differs slightly from Chapter 2, but is entirely equivalent. Whereas earlier we modeled the false positive rate, $x_j$, as the control variable, here we use the true negative rate – these quantities are related simply by $u_j = 1 - x_j$. Taking the true negative rate as the control variable will be convenient for the analysis in this chapter. Our choice of $f_j$ reasonably approximates the relationship between the true-negative rate and false-negative rate in many real-world binary-decision tasks. In particular, $f_j$ corresponds to a concave *receiver operating characteristic*, or "ROC curve."

Of all the clicks that publisher $i$ sends to network $j$,

$$N_{ij} \equiv u_j r_i + u_j^{\gamma_j}(1 - r_i) \tag{4.2}$$

is the fraction that is marked valid. Thus,

$$N_{ij} V_i c_{ij} \tag{4.3}$$

is the number of publisher $i$'s clicks that network $j$ marks valid. Marking a click invalid only means that the network will not charge the advertiser for the click, and will consequently not pay the publisher. The user is forwarded to the advertiser's site, irrespective of whether the click is marked valid or invalid.

**Predictive Pricing and Revenue Sharing.** For each click coming from publisher $i$ that is marked valid, network $j$ bills advertisers for only a fraction $g_{ij}$ of a click i.e., advertisers receive a $(1 - g_{ij})$ discount. The fraction $g_{ij}$ is the *predictive price* that network $j$ applies to publisher $i$'s traffic. The *effective* number of clicks for which network $j$ pays publisher $i$ is then:

$$E_{ij} \equiv N_{ij} V_i c_{ij} g_{ij} \tag{4.4}$$

Suppose $\theta_j$ is the expected auction revenue per click on network $j$. Of each dollar of revenue from advertisers, network $j$ pays out a fraction $h_j$ to publishers. The fraction $h_j$ is referred to as the *revenue share*. Then, the total revenue to publisher $i$ from network $j$ is:

$$\pi_{ij} \equiv N_{ij} V_i c_{ij} g_{ij} h_j \theta_j \tag{4.5}$$

Let $\mathbf{g}_j \equiv \{g_{ij}\ \forall i\}$. We refer to the pair $(\mathbf{g}_j, h_j)$ together as network $j$'s *pricing policy*. We refer to the triple $(u_j, \mathbf{g}_j, h_j)$ together as network $j$'s *traffic policy*.

**Conversion Rates.** The clicks sent by the publishers to the networks are, in turn, distributed amongst the $K$ advertisers (in proportions related to the advertisers' bids). Of all the clicks sent from publisher $i$ to advertiser $k$ via network $j$, let $\beta_{ijk}$ be the fraction that become conversions. The fraction $\beta_{ijk}$ is referred to as a *conversion rate*. From each conversion, advertiser $k$ earns an amount $y_k$.

As in Chapter 3, we assume that conversion rates are *separable* i.e., that each $\beta_{ijk}$ is a product of three factors:

$$\beta_{ijk} = \beta_i^{\text{Pub}} \beta_j^{\text{Net}} \beta_k^{\text{Adv}} \ \ \forall (i, j, k) \tag{4.6}$$

Each factor in (4.6) has a different interpretation. $\beta_i^{\text{Pub}}$ measures how targeted publisher $i$'s traffic is with respect to the keyword in question. $\beta_j^{\text{Net}}$ measures how good network $j$ is at matching publishers' content with advertisers' ads. $\beta_k^{\text{Adv}}$ measures the quality and effectiveness of advertiser $k$'s ads. See Section 4.5 for further discussion on separability.

### 4.2.2 Sequence of Events

Our one-shot dynamic game is comprised of two stages:

1. In the *first stage*, each network $j$ selects and announces its traffic policy (i.e., its filtering aggressiveness, $u_j$, predictive prices, $\mathbf{g}_j$, and revenue share, $h_j$).

2. In the *second stage* (or, the *subgame*), each publisher $i$ decides on which networks to sell traffic (i.e., its *allocations*, $\{c_{ij} \; \forall j\}$). Simultaneously, each advertiser $k$ decides how much it is willing to pay for clicks from each network $j$ (i.e., its *valuations*, $\{v_{kj} \; \forall j\}$).

Recall that in Chapter 2 networks only selected an aggressiveness level, and in Chapter 3 networks only selected a pricing policy. In this chapter, networks make both of these decisions simultaneously.

All players aim to maximize revenues. After the second stage, payoffs are realized: a) publishers sell clicks (i.e., display ads) on their chosen networks, b) networks mark a subset of these clicks valid, c) advertisers pay the networks (possibly a discounted price) for the clicks marked valid, and d) networks pay out a fraction of earned revenues to publishers. Recall that users are forwarded to the advertiser's site even if a click is marked invalid.

Network $j$'s revenues are a function of the decisions taken by advertisers and publishers in the second stage. Advertisers' and publishers' decisions, in turn, are a reaction to the traffic policies chosen by all the networks in the first stage. Thus, each network's profits will depend on the traffic policies of all of its competing networks.

### 4.2.3 Best-Response Traffic Policy

We can now write down an expression for the best-response traffic policy of network 1, holding the policies of all other networks fixed, and assuming equilibrium in the second stage. First, define:

$$X_{ij} \equiv N_{ij} V_i g_{ij} h_j \theta_j \tag{4.7}$$

$X_{ij}$ is simply the revenue publisher $i$ would earn if it sent all its traffic to network $j$ (see (4.5)). Then, define $A_i$, the *nominal number of conversions* for publisher $i$, as follows:

$$A_i \equiv r_i \beta_i^{\text{Pub}} V_i \tag{4.8}$$

The best-response traffic policy for network 1 can then be found as a solution to the following optimization problem:

$$\begin{aligned}
\text{maximize} \quad & \eta_1 \equiv \left(\sum_i E_{i1}\right)(1-h_1)\theta_1 \\
\text{subject to} \quad & \sum_j c_{ij} X_{ij} = \max_j X_{ij} \;\; \forall i \\
& \sum_j c_{ij} = 1 \;\; \forall i \\
& \theta_j = \kappa_j \left(\frac{\sum_i A_i c_{ij}}{\sum_i E_{ij}}\right) \beta_j^{\text{Net}} \;\; \forall j \\
& N_{ij} = u_j r_i + u_j^{\gamma_j}(1-r_i) \;\; \forall (i,j) \\
& E_{ij} = N_{ij} V_i c_{ij} g_{ij} \;\; \forall (i,j) \\
& X_{ij} = N_{ij} V_i g_{ij} h_j \theta_j \;\; \forall (i,j) \\
& A_i = r_i \beta_i^{\text{Pub}} V_i \;\; \forall i \\
& 0 \le u_1, g_{i1}, h_1, c_{ij} \le 1 \;\; \forall (i,j)
\end{aligned} \tag{4.9}$$

The objective in (4.9) is an expression for network 1's profit. The first two constraints encode the assumption that each publisher chooses allocations optimally in the second stage. The third constraint (discussed below) implies that advertisers choose their valuations and bids optimally. Thus, the first three constraints together imply that there is an equilibrium in the second stage between publishers and advertisers. The next four constraints simply enforce the definitions given earlier (see (4.2), (4.4), (4.7) and (4.8)), while the final constraint gives ranges for the decision variables we are interested in.

The most striking feature of (4.9) is that the advertisers' valuations $v_{kj}$ do not appear anywhere. In fact, there are no $k$-subscripts at all. The fact that $v_{kj}$ does not appear greatly reduces the complexity of solving (4.9), since the number of decision variables is reduced by a factor $K$.

The advertisers' equilibrium behaviour (i.e., their private valuations and corresponding bids at auction) is captured implicitly in (4.9), via the constraint:

$$\theta_j = \kappa_j \left(\frac{\sum_i A_i c_{ij}}{\sum_i E_{ij}}\right) \beta_j^{\text{Net}} \;\; \forall j \tag{4.10}$$

The derivation of (4.10) is nontrivial – we refer the reader to Section 3.2.3 for complete details. We will instead highlight the key intuitions here. We assume that each advertiser $k$ chooses its private valuations, $\{v_{kj} \; \forall j\}$, such that its total revenues from online advertising are maximized, subject to a lower-bound, $R_k$, on its *return on investment* (ROI). Advertisers then strategically bid at auction for click-throughs on each network (without collusion). We also assume that each network $j$'s auction mechanism is *linear*, in the sense that if all advertisers were to simultaneously scale their valuation by a factor $\delta$ (i.e., $v_{kj} \leftarrow \delta v_{kj}$), then network $j$'s expected auction revenue would also be scaled by $\delta$ (i.e., $\theta_j \leftarrow \delta\theta_j$).

Under just these (weak) assumptions, we can deduce Equation (4.10), which says that that the expected per-click auction revenue $\theta_j$ will be proportional to $\left(\frac{\sum_i A_i c_{ij}}{\sum_i E_{ij}}\right)\beta_j^{\text{Net}}$, which (roughly speaking) is the ratio of converted clicks to billed-for clicks on network $j$. The proportionality constant $\kappa_j$ can be computed from the distribution of advertiser types $\left\{\left(y_k, R_k, \beta_k^{\text{Adv}}\right)\ \forall k\right\}$, and the auction mechanism used by network $j$. In particular, $\kappa_j$ does not depend on the actions of publishers or networks. In this sense, the constants $\{\kappa_j\ \forall j\}$ act as *sufficient statistics* for the distribution of advertiser types, leading to (4.10).

Seen in this light, the structure of the objective function in (4.9) is quite intuitive. In order for network 1 to maximize profits, it must simultaneously: a) attract a large volume of billable traffic from publishers (i.e., high $\sum_i E_{i1}$), b) deliver high-conversion-rate traffic to advertisers (i.e., high $\theta_1$), and c) retain a large share of revenues (i.e., low $h_1$).

### 4.2.4 TRAFFICQUALITY

The optimization problem (4.9) is highly nonconvex, making it difficult to even find feasible points. In Appendix A.6 we describe an iterative algorithm, TRAFFICQUALITY, for finding approximate solutions to (4.9). The key step is to solve a sequence of geometric programs, each of which is a relaxation of (4.9). TRAFFICQUALITY is a generalization of the PRICINGPOLICY algorithm derived in Chapter 3.

In Section 4.3, we study the relationship between filtering and predictive pricing. Our findings will allow us to reduce the number of input parameters required to run TRAFFICQUALITY, and estimate the rest from readily observable data. However, in practice, the incentive for click inflation means that some of the observable data may be manipulated by the publishers. Section 4.4 is thus devoted to dealing with the incentive for click inflation.

Finally, TRAFFICQUALITY computes a best-response for network 1, given the policies of competing networks, and assuming an equilibrium is played in the second-stage. We refer the reader to Section 4.5 for a brief discussion of equilibria in the first-stage of our game.

## 4.3 Filtering And Predictive Pricing

In this section, we study the relationship between filtering and predictive pricing. Given the option, should a network choose one over the other, or use both together?

The main result of Chapter 2 was that, if predictive pricing is disallowed and all networks offer the same revenue share, then it is optimal for network $j$ to filter aggressively. In particular, if network 1 is the most skilled at filtering (i.e., $\gamma_1 > \gamma_j\ \forall j \neq 1$), all publishers (even low-quality ones) will prefer to send their traffic to network 1, provided it is filtering aggressively enough (i.e., $u_1 \leq u^*$ for some known threshold value $u^*$).

As a concrete example, consider a market with $I = 20$ publishers and $J = 2$ networks. Both networks offer publishers a revenue share of $h_1 = h_2 = 50\%$ and set $g_{ij} = 1\ \forall i$ (i.e., no predictive pricing). The networks are equally skilled at selecting ads ($\beta_1^{\text{Net}} = \beta_2^{\text{Net}} = 1$), but network 1 is more skilled at filtering ($\gamma_1 = 10$ and $\gamma_2 = 8$). Network 2 is marking valid $u_2 = 80\%$ of valid clicks (thus, it also marks valid $u_2^{\gamma_2} = 17\%$ of invalid clicks). To simulate a wide variation in traffic quality, we assume that $r_i = 0.05i$ and $\beta_i^{\text{Pub}} = 0.0025i$ i.e., publishers are sorted in increasing order of traffic validity and targetedness, with $r_i$

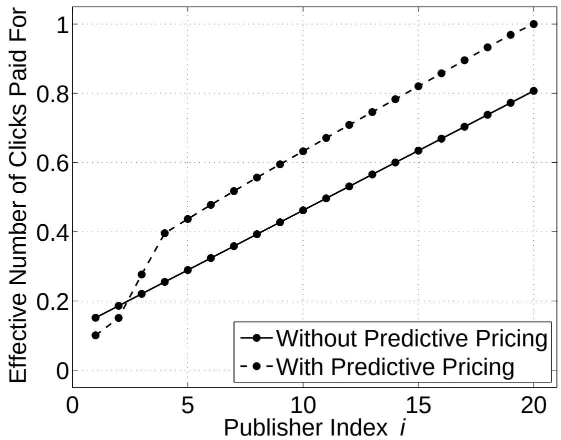


Figure 4.1: Effective number of clicks paid for, $E_{i1}$.

ranging between $5\%$ and $100\%$, and $\beta_i^{\text{Pub}}$ ranging between $0.25\%$ and $5\%$ (a $5\%$ conversion rate would be considered very high in practice).

For this scenario, we used TRAFFICQUALITY (with $\mathbf{g}_1$ and $h_1$ disabled) to compute the best-response traffic policy for network 1. As predicted, TRAFFICQUALITY recommends that network 1 filter aggressively (i.e., set $u_1 \to 0$), causing all publishers (including the low-quality ones) to send their traffic to network 1 (in this example, $u_1 \leq u^* = 81\%$ was sufficient for network 1 to win over the market).

Now, consider the same scenario, except that network 1 is allowed to use predictive pricing. That is, we allow $\mathbf{g}_1$ to be a decision variable for network 1 (note that the revenue share, $h_1$, is still fixed at $50\%$). Using TRAFFICQUALITY with $\mathbf{g}_1$ enabled this time, we get a very interesting outcome: the optimum is now $u_1 = 100\%$. That is, TRAFFICQUALITY is recommending that network 1 stop filtering altogether, and just use predictive pricing instead.

Figure 4.1 shows the effective number of clicks, $E_{i1}$, that each publisher is paid for (as a fraction of $V_i c_{i1}$), with and without predictive pricing (see (4.4)). With filtering alone, network 1 is restricted to a linear $E_{i1}$ profile (recall that $N_{i1}$ is linear in $r_i$, and in this example $r_i$ is linear in $i$), and all publishers choose to send traffic to network 1. With both filtering and predictive pricing enabled, TRAFFICQUALITY outputs a nonlinear profile of predictive prices that achieves exactly the same market outcome as the filtering-only case (i.e., all publishers choose network 1), but it does so without needing to using filtering at all.

This outcome is not peculiar to the scenario described above – it is actually a general phenomenon:

**Theorem 5.** *Suppose* $(\tilde{u}_1, \tilde{\mathbf{g}}_1, h_1^*)$ *is a solution to (4.9), with* $\tilde{u}_1 < 1$. *Then, there exists another solution* $(u_1^*, \mathbf{g}_1^*, h_1^*)$ *to (4.9) where* $u_1^* = 1$. *The converse is not necessarily true.*

*Proof.* In (4.9), $u_j$ only appears in the fourth constraint (i.e., the definition of $N_{ij}$), and in turn $N_{ij}$ only appears in the product $N_{ij} g_{ij}$. Also, note that if $u_j = 1$, then $N_{ij} = 1$ irrespective of $r_i$ (i.e., all clicks are being marked valid).

Therefore, if $(\tilde{u}_1, \tilde{\mathbf{g}}_1, h_1^*)$ is feasible for (4.9), then $(1, \mathbf{g}_1^*, h_1^*)$ will also be feasible, where $g_{i1}^* = \tilde{g}_{i1} \tilde{N}_{i1}$ and $\tilde{N}_{i1} = \tilde{u}_1 r_i + \tilde{u}_1^{\gamma_1}(1 - r_i)$. Moreover, all other variables are left unchanged. Neither $N_{ij}$ nor $g_{ij}$ appear in the objective of (4.9). Therefore, if $(\tilde{u}_1, \tilde{\mathbf{g}}_1, h_1^*)$ is optimal, then $(1, \mathbf{g}_1^*, h_1^*)$ will also be optimal.

Conversely, suppose $(1, \mathbf{g}_1^*, h_1^*)$ is a solution to (4.9), and let $\tilde{u}_j < 1$. If there is an $i$ such that $r_i < 1$ and $g_{i1}^* = 1$, then $\tilde{g}_{i1} = \frac{g_{i1}^*}{\tilde{N}_{i1}} > 1$, which is infeasible. Therefore, an optimal $(\tilde{u}_1, \tilde{\mathbf{g}}_1, h_1^*)$ where $\tilde{u}_j < 1$ does not necessarily exist. □

Theorem 5 says that there is always a best-response for network 1 (i.e., a solution to (4.9)) that involves using predictive pricing alone (i.e., setting $u_j = 1$). Therefore, in most settings, predictive pricing can and should be used for managing traffic quality instead of filtering. Although this result may not seem surprising at first, it has important practical implications, which we discuss in Section 4.3.1.

Theorem 5 holds irrespective of the numbers of publishers and advertisers and their traffic qualities. An immediate implication is that a network seemingly gains no competitive advantage from having superior algorithms for filtering, since competing networks can simply respond by implementing predictive pricing. This conclusion is a partial contradiction of the result established in Chapter 2.

Whereas filtering effectively requires careful development of algorithms for detecting specific traffic patterns, predictive pricing only requires aggregate measurements of click- and acquisition traffic and the selection of a set of coefficients (i.e., predictive prices). Of course, the challenge in predictive pricing is being able to accurately estimate these required traffic statistics (more on this issue in Section 4.3.2).

Combining Theorem 5 with the result from Chapter 2, we can also deduce the following:

**Corollary 1.** *In a subgame perfect equilibrium, each network uses either filtering or predictive pricing, but not both.*

*Proof.* In equilibrium, each network uses a best-response traffic policy. From Theorem 5, if a network uses predictive pricing, it is best not to filter. From Theorem 2 in Chapter 2, if a network is not predictive pricing, its best response is to filter aggressively. □

### 4.3.1 Practical Implications

Theorem 5 may not seem very surprising, since predictive pricing does allow for much finer-grained publisher-level control (i.e., a factor $g_{ij}$ for each publisher $i$) compared to filtering, which provides a single control $u_j$ for the entire population of publishers. However, Theorem 5 has very significant practical implications.

#### No need to measure or control $u_j$

In most practical settings, $u_j$, the true-negative rate (i.e., aggressiveness) of network $j$'s filtering algorithm, is difficult to measure since such a measurement typically requires a "ground truth" data set, which may not be available. In our context, ground truth would be a "representative" sample of traffic where clicks are labelled valid and invalid. Even if $u_j$ could be measured, network $j$ may not be able to select arbitrary values for $u_j$ strictly between 0 and 1.

On the other hand, it is always possible to set $u_j = 0$ or $u_j = 1$ by simply marking all clicks invalid or valid, respectively (of course, with $u_j = 1$, the false-negative rate $u_j^{\gamma_j}$ would also be 1 i.e., all invalid clicks would also be marked valid). Theorem 1 guarantees that there is always a best-response where $u_j = 1$.

#### No need to know $r_i$ and $\beta_i^{\text{Pub}}$

Even if accurate measurement and control of $u_j$ were possible, there is a more serious practical issue concerning parameter estimation. In order for a network to use TRAFFICQUALITY to find an optimal traffic policy, it needs to know $r_i$ and $\beta_i^{\text{Pub}}$ for each publisher $i$ (see (4.9)). However, in practice, these quantities are not observed – if $r_i$ could be observed, filtering would be unnecessary! Typically, networks are only able to measure the volume of clicks (e.g., using click logs) and numbers of conversions (e.g., using conversion tracking code installed on the advertisers' sites).

Fortunately, Theorem 1 allows a network to run TRAFFICQUALITY even without knowing $r_i$ and $\beta_i^{\text{Pub}}$. We know that it is sufficient for network 1 to search for a traffic policy where $u_j = 1$. From (4.2), setting $u_j = 1$ implies that $N_{ij} = 1\ \forall i$, irrespective of $r_i$ and $\gamma_j$. Upon substituting $u_j = N_{ij} = 1$ into fourth constraint of (4.9), we observe that the only other place that $r_i$ or $\beta_i^{\text{Pub}}$ appear in (4.9) is as the product $r_i\beta_i^{\text{Pub}}$ in the seventh constraint. As we describe in Section 4.3.2, it is possible to estimate the product $r_i\beta_i^{\text{Pub}}$ by observing the volumes of clicks and conversions alone.

Therefore, network $j$ can run TRAFFICQUALITY even without knowing $r_i$ and $\beta_i^{\text{Pub}}$, by simply setting $u_j = 1$ and estimating the product $r_i\beta_i^{\text{Pub}}$ instead. Theorem 1 guarantees that there is no loss in profits from calculating a best response in this way.

### 4.3.2 Parameter Estimation

We now demonstrate how network $j$ can estimate the product $r_i\beta_i^{\text{Pub}}$ for each publisher $i$, as well as other parameters required to run TRAFFICQUALITY, using data that can be readily observed in practice.

#### Estimating $r_i\beta_i^{\text{Pub}}$

Suppose a user visits publisher $i$'s website, on which advertiser $k$'s ad is displayed, and the ad is delivered by network $j$. If the user then clicks on the ad, network $j$ redirects the user from publisher $i$'s site to advertiser $k$'s site, and records the click-through in its click logs. So, by simply analysing its click logs, network $j$ can compute $V_i c_{ij}\ \forall i$, which is the total number of clicks sent to network $j$ by publisher $i$. Network $j$ can also compute $\xi_{ijk}\ \forall (i,k)$, where $\xi_{ijk}$ is the fraction of publisher $i$'s traffic sent to advertiser $k$ by network $j$.

Let $A_{ijk}$ denote the number of clicks originating on publisher $i$'s site that eventually become conversions for advertiser $k$, where the ad is delivered by network $j$ (the letter $A$ stands for "actions" or "acquisitions"). Conversion tracking software installed on the advertiser's site typically gives the total number of clicks that advertiser $k$ converts, along with the times and dates of those clicks and conversions (alternatively, the advertiser can "self-report" the number of conversions, although the advertiser may not self-report truthfully – see [23] and [27] for a discussion). Cross-referencing this data with the click logs, a network can infer which publisher each converted click originated from. Thus, as long as the required conversion-tracking infrastructure is in place, each network $j$ can observe $A_{ijk}\ \forall (i,k)$.

Using the quantities defined in Section 4.2, $A_{ijk}$ can be decomposed into a product as follows:

$$A_{ijk} = V_i c_{ij} r_i \xi_{ijk} \beta_{ijk} \tag{4.11}$$

Equation (4.11) holds since $V_i c_{ij} r_i \xi_{ijk}$ is the number of valid clicks sent from publisher $i$ to advertiser $k$ via network $j$, and $\beta_{ijk}$ is the fraction of these clicks that become conversions.

Therefore, using its observations of $V_i c_{ij}$, $\xi_{ijk}$ and $A_{ijk}$, network $j$ can compute the product $r_i \beta_{ijk}\ \forall(i,k)$ as follows:

$$r_i \beta_{ijk} = \frac{A_{ijk}}{V_i c_{ij} \xi_{ijk}} \tag{4.12}$$

Then, applying the separability assumption (4.6), we get:

$$\left(r_i \beta_i^{\text{Pub}}\right) \beta_k^{\text{Adv}} = \frac{A_{ijk}}{V_i c_{ij} \xi_{ijk} \beta_j^{\text{Net}}} \tag{4.13}$$

Observe that the right-hand side of (4.13) can be computed directly from available data for every $(i,k)$, whereas the left-hand side is comprised of parameters that network $j$ needs to estimate. In particular, network $j$ has $IK$ (possibly noisy) data points with which to estimate the $I+K$ parameters $\{r_i \beta_i^{\text{Pub}}\ \forall i\}$ and $\{\beta_k^{\text{Adv}}\ \forall k\}$. Since there will typically be many more data points than parameters (i.e., $IK \gg I + K$), a data-fitting technique (e.g., least-squares) can be used to do the estimation.

As mentioned in Section 4.3.1, when $u_j = 1$ the parameters $r_i$ and $\beta_i^{\text{Pub}}$ only appear in (4.9) as the product $r_i \beta_i^{\text{Pub}}$. Thus, once network $j$ computes estimates of $r_i \beta_i^{\text{Pub}}$, it can "naively" assume that $r_i = 1$ and use $\hat{\beta}_i \equiv r_i \beta_i^{\text{Pub}}$ as an estimate of $\beta_i^{\text{Pub}}$. Clearly, $\hat{\beta}_i$ will be an over-estimate of $\beta_i^{\text{Pub}}$ since $r_i \leq 1$, but this estimation error does not adversely affect network $j$'s profits since Theorem 5 says $u_j = 1$ is optimal anyways. Setting $\hat{\beta}_i = r_i \beta_i^{\text{Pub}}$ will become useful in Section 4.4, when we discuss click inflation.

In practice, some advertisers may not be willing to install conversion tracking software on their sites. However, network $j$ is most interested in estimating the product $r_i \beta_i^{\text{Pub}}$ (for use in TRAFFICQUALITY), as opposed to $\beta_k^{\text{Adv}}$ for specific $k$. So, as long as "enough" advertisers do install conversion tracking software, data-fitting techniques can be used to compute good estimates of $r_i \beta_i^{\text{Pub}}$. Real-world deviations from, say, the separability assumption can also be (partially) compensated-for using fitting techniques.

**Other parameters**

In many cases, network $j$ can assume $V_i = V_i c_{ij}$ whenever publisher $i$ sends it any traffic at all (i.e., that $c_{ij} = 1$ whenever $c_{ij} > 0$). Publisher $i$'s optimization problem (not discussed here) is such that, in most cases, its optimal allocation $c_{ij}$ will be either 0 or 1. That is, if publisher $i$ sends network $j$ any traffic at all, it will send network $j$ all of its traffic. Fractional allocations of traffic across multiple networks (i.e., $c_{ij} \in (0,1)$) are only optimal for publisher $i$ when there a "tie" between those networks in terms of profitability, which happens infrequently in practice.

In such cases, network $j$ can use (4.13) to also infer $A_i$, the nominal number of conversions for publisher $i$, as follows:

$$A_i = r_i \beta_i^{\text{Pub}} V_i = \frac{A_{ijk}}{\xi_{ijk} \beta_j^{\text{Net}} \beta_k^{\text{Adv}}} \tag{4.14}$$

Intuitively, $A_i$ measures the "potential" number of conversions that can result from publisher $i$'s traffic, before adjusting for the matching algorithms of network $j$ (i.e., $\beta_j^{\text{Net}}$) and the ad quality of advertiser $k$ (i.e., $\beta_k^{\text{Adv}}$).

In the case where $\beta_j^{\text{Net}} = \beta_k^{\text{Adv}} = 1 \; \forall (j,k)$, we get $A_i c_{ij} \xi_{ijk} = A_{ijk}$ and $\sum_{j,k} A_{ijk} = A_i$, both of which are consistent with the interpretations of $c_{ij}$ and $\xi_{ijk}$ as fractional traffic allocations.

Finally, we assume in (4.13) and (4.14) that $\beta_j^{\text{Net}}$ is known to each network $j$. More generally, $\beta_j^{\text{Net}}$ should be known for each network $j$ to within a constant factor. Relative values of $\beta_j^{\text{Net}}$ can be estimated from historical competitive data, or by other comparative means.

## 4.4 Click Inflation

Theorem 5 states that, for most sources of traffic, predictive pricing can and should be used for managing traffic quality instead of filtering. In fact, there are cases where filtering gives lower profits. However, Theorem 5 paints an incomplete picture. An underlying assumption in the proof was that $r_i$ and $V_i$ are fixed parameters that describe publisher $i$'s traffic. In particular, they were not considered decision variables for publisher $i$. Stated differently, we assumed that all of the traffic on publisher $i$'s site is *organic*, in the sense that the traffic is not generated or caused by the publisher itself. All valid traffic, by definition, is organic. Most forms of invalid traffic can also be considered organic, including the various forms of non-click-fraud invalid traffic (e.g., double-clicks, unintentional clicks, web crawlers), as well as click fraud due to "competitor clicking" [9].

Unfortunately, $r_i$ and $V_i$ can indeed be manipulated by publisher $i$ in practice. In particular, a publisher can inject a stream of fraudulent clicks into the traffic it sends to network $j$ – this practice is known as *click inflation* [9]. Engaging in click inflation increases the total volume of clicks while leaving the amount of valid clicks and the number of resulting conversions unchanged (since none of the fraudulent clicks become conversions).

In Section 4.4.1, we discuss why click inflation might occur, and how to account for it in our model. In Sections 4.4.2 and 4.4.3, we present a pair of approaches that networks can use to compensate for click inflation. Roughly, one approach is to use filtering to estimate $r_i$, and then pay publisher $i$ for its valid, organic traffic only. Another is to find predictive prices such that the incentive for click inflation is eliminated altogether. In Section 4.4.4, we compare these approaches.

### 4.4.1 Why Click Inflation Occurs

It is easy to see why publishers might have an incentive to engage in click inflation. Define, for convenience, $G_i \equiv r_i V_i$ and $B_i \equiv (1 - r_i)V_i$ (the letters $G$ and $B$ stand for "good" and "bad"). Click inflation causes $B_i$ to increase without a change in $G_i$. For example, publisher $i$ might pay users to visit its site and click on ads, even though the users are uninterested in the product being advertised. The problem for network $j$ is that it pays for $N_{ij} V_i c_{ij} = \left( u_j G_i + u_j^{\gamma_j} B_i \right) c_{ij}$ clicks, so for any $u_j > 0$, the number of clicks that publisher $i$ is paid for will be increasing in $B_i$.

As an illustration, we computed the best-response traffic policy for network 1 in a scenario where $r_i = 1 \; \forall i$ (i.e., $G_i = V_i$) and $\beta_i^{\text{Pub}}$ is linear in $i$. We then chose the highest-quality publisher, and computed by how much its revenues would increase if it inflated its click volume by various amounts. The results are shown in

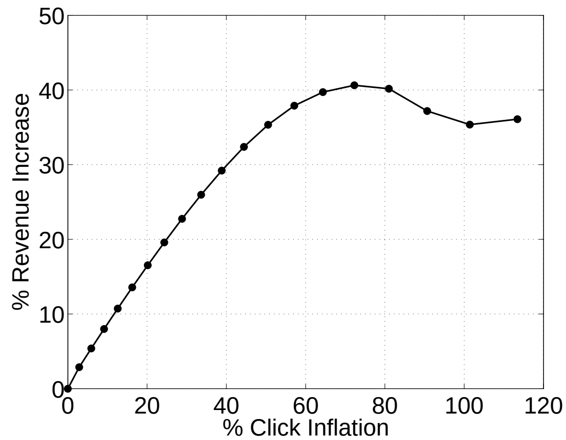


Figure 4.2: Increase in revenues due to click inflation.

Figure 4.2. For example, inflating click volumes by $24.3\%$ (i.e., injecting $B_i = 0.243G_i$ fraudulent clicks) results in a $19.6\%$ increase in revenues.

There are two features that stand out in Figure 4.2. First, publishers can increase their revenues significantly by generating fraudulent traffic. Second, it not optimal for a publishers to generate an arbitrarily large amount of fraudulent traffic. In our example, the publisher's revenues are maximized (a gain of $41\%$) when it inflates its traffic by $72.3\%$. Any further inflation causes its revenues to decrease, since network 1 would apply a very low predictive price to its traffic. Intuitively, this example suggests that even relatively high-quality publishers may try to slip some fraudulent traffic through the networks' filters.

To model click inflation, we slightly modify the dynamic game described in Section 4.2.2 as follows:

- In the first stage, networks select traffic policies based on forecasts of organic traffic quality.
- In the second stage, publishers decide on allocations as well as whether (and by how much) to inflate $B_i$.
- After the second stage (i.e., after receiving the publishers' traffic, but before any payments are made), networks adjust their traffic policies to account for any perceived click inflation in each publisher's traffic.

In other words, $r_i$ and $V_i$ are treated as decision variables for publisher $i$. Recall that click inflation causes an increase in total volume $V_i$, but leaves $G_i$ unchanged. In Section 4.2, the fixed set of parameters that described publisher $i$'s traffic (i.e., its "type") was the triple $(r_i, \beta_i^{\text{Pub}}, V_i)$. In this section, publisher $i$'s organic traffic is described by $(G_i, \beta_i^{\text{Pub}}, A_i)$.

### 4.4.2 Solution 1: Estimate $r_i$

One approach for network $j$ to fight click inflation is to somehow estimate $r_i$, so that publisher $i$ is (effectively) paid for only $r_i V_i c_{ij} = G_i c_{ij}$ valid clicks.

For example, one possible way to derive an estimate of $r_i$ is to: a) run the clicks through a filtering algorithm to observe $N_{ij}$, b) use labeled "synthetic" traffic to measure $u_j$ and $\gamma_j$, and then c) use $N_{ij}$, $u_j$ and $\gamma_j$ to invert (4.2). The accuracy of the resulting estimate, of course, will depend on how closely the synthetic traffic resembles actual invalid traffic seen by the filtering algorithm in practice – perhaps the most unbiased approach would be to hire a third-party firm or research lab to act as an adversary for testing purposes. We note here that accurately estimating $r_i$ is not the same as accurately distinguishing between valid and invalid clicks (i.e., filtering). The latter involves making accurate decisions on a click-by-click basis, whereas the former requires deriving only a single *aggregate* estimate.

In this section, rather than detailing an estimation procedure, we study the effect of estimation errors on a network's profits (since network $j$'s estimation procedure may be highly dependent on the specifics of its ad-serving mechanism). As described in Section 4.4.1, the incentive for click inflation arises because network $j$ usually pays each publisher $i$ for $N_{ij} V_i c_{ij} = \left(u_j G_i + u_j^{\gamma_j} B_i\right) c_{ij}$ clicks. The expression for $N_{ij}$ captures the fact that $r_i$ is unknown to network $j$, and that its filtering algorithms are prone to error. Hypothetically, suppose network $j$ knew the exact value of $r_i$. It could then simply set $N_{ij} = r_i \; \forall(i,j)$ i.e., pay each publisher $i$ for exactly $r_i V_i c_{ij}$ clicks. Then, (4.4) would be $E_{ij} = G_i g_{ij} c_{ij}$, which is independent of $B_i$, implying that publisher $i$ would gain nothing from engaging in click inflation.

Operationally, network $j$ would then compute its best-response traffic policy by simply replacing the fourth constraint in (4.9) with:

$$N_{ij} = r_i \; \forall(i,j) \tag{4.15}$$

With this motivation in mind, network $j$ can try to estimate $r_i$ for each publisher $i$.

The extent to which network $j$ can deter click inflation will depend on how accurately it can estimate $r_i$. Inaccurate estimates mean that network $j$ solves (4.9) with incorrect coefficients, and so the policy output by TRAFFICQUALITY may be suboptimal. We ran an experiment to quantify the sensitivity of network 1's profits to errors in estimating $r_i$. For concreteness, suppose network 1 uses filtering to derive a noisy estimate, $\hat{r}_i$, of $r_i$ for each publisher $i$:

$$\hat{r}_i = r_i + \sigma Z_i \tag{4.16}$$

Each $Z_i$ is an independent, zero-mean, unit-variance normal random variable ($\hat{r}_i$ is truncated so that it is between 0 and 1). Smaller (larger) values of the standard error, $\sigma$, mean that network 1 is more (less) accurate at estimating $r_i$.

The value of $\sigma$ was varied between $0\%$ and $40\%$, and several trials were run at each value. In each trial, network 1 first uses filtering to derive a set of estimates $\{\hat{r}_i \; \forall i\}$ (the estimation error is given by (4.16)). Network 1 then sets $N_{ij} = \hat{r}_i$ in (4.9), and uses TRAFFICQUALITY to compute a best-response pricing policy $(\hat{\mathbf{g}}_1, \hat{h}_1)$. Of course, $(\hat{\mathbf{g}}_1, \hat{h}_1)$ may be suboptimal since $\hat{r}_i \neq r_i$ (although $r_i \beta_i^{\text{Pub}}$ can still be estimated accurately, as discussed in Section 4.3.2). Network 1 assumes that network 2 is setting $g_{i2} \propto r_i \beta_i^{\text{Pub}}$. We then compute the actual profit to network 1 resulting from using $(\hat{\mathbf{g}}_1, \hat{h}_1)$ and $N_{ij} = \hat{r}_i$. When computing actual profits, network 2 is assumed to know $r_i$ and $\beta_i^{\text{Pub}}$ exactly.

Figure 4.3 shows network 1's profits in each trial, as well as the average profit across the trials for each value of $\sigma$. Profits have been normalized by network 1's profit in the $\sigma = 0\%$ case. As expected, from Figure

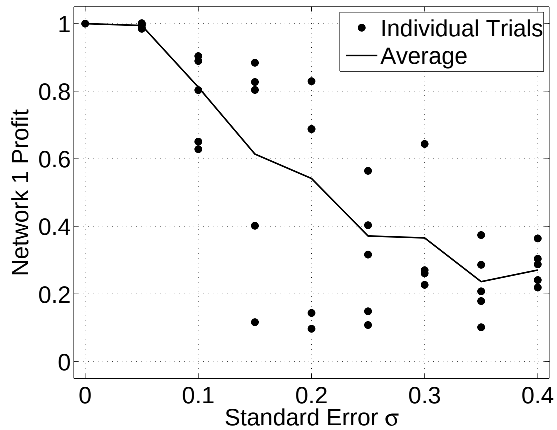


Figure 4.3: Sensitivity of profits to accuracy of $\hat{r}_i$.

4.3 we see that lower values of $\sigma$ (i.e., higher estimation accuracy) result in higher profits for network 1. For $\sigma \leq 10\%$, the losses are relatively small (less than $10\%$ loss in some trials). For $\sigma > 10\%$, however, the variance in the outcome increases greatly, due to estimation errors. The average losses also steadily increase, exceeding $60\%$ at $\sigma = 30\%$.

We conclude, in this example, that the estimates $\hat{r}_i$ are useful only when $\sigma \leq 10\%$ or so. There is one more issue: in practice, networks can only refuse to pay publishers for clicks that their algorithm marks invalid. That is, networks cannot refuse payment for a given click-through, without providing the publisher a justification for doing so. To utilize its estimates of $r_i$, network $j$ would take the following steps:

1. Use filtering to derive estimates $\hat{r}_i$, but mark all clicks valid irrespective of what the filter decides.

2. Use TRAFFICQUALITY with $N_{ij} \leftarrow \hat{r}_i$ to compute predictive prices $\mathbf{g}_j^*$ and revenue share $h_j$. Effectively, network $j$ is computing a pricing policy for a setting where all clicks are valid.

3. Apply $g_{ij} \leftarrow g_{ij}^* \hat{r}_i$ to each publisher $i$'s traffic.

Observe that network $j$ does not actually mark any clicks invalid. The desired effect of paying for only $G_i c_{ij}$ clicks is achieved indirectly using predictive prices $g_{ij} = g_{ij}^* \hat{r}_i$.

To summarize, if network $j$ could determine $r_i$ exactly, it would use TRAFFICQUALITY with $N_{ij} \leftarrow r_i$ to compute its traffic policy. There is zero loss in network $j$'s profits due to click inflation since no new restrictions are placed on $\mathbf{g}_j$ (compare this situation to Section 4.4.3). Therefore, when fighting click inflation using estimates $\hat{r}_i$ of $r_i$, any and all losses are purely due to inaccurate estimates.

### 4.4.3 Solution 2: Quasi-CPA

An alternate approach to fighting click inflation is to directly constrain the search for predictive prices in a way that eliminates the incentive, as we describe in this section. The main advantage of doing so is that no estimates of $r_i$ are needed. As we will demonstrate, this approach is closely related to *cost-per-action*

(CPA) pricing schemes. It is widely agreed-upon that CPA schemes are resistant to click inflation. They are susceptible to other forms of fraud, however – we refer the reader to [9, 23, 27] for more details.

Suppose network $j$ simply assumes that all clicks are valid i.e., that $r_i = 1 \; \forall i$. The only other measure of publisher quality would then be the conversion rate, $\beta_i^{\text{Pub}}$. Using the approach discussed in Section 4.3.2, network $j$ could simply compute an estimate $\hat{\beta}_i$ of $\beta_i^{\text{Pub}}$ as $\hat{\beta}_i = r_i \beta_i^{\text{Pub}}$. Clearly, $\hat{\beta}_i$ would be an underestimate of $\beta_i^{\text{Pub}}$, since $r_i \in [0, 1]$. Let us assume in this section that the publishers are sorted by $\hat{\beta}_i$ (i.e., $\hat{\beta}_i$ is increasing in $i$).

We can interpret a vector of predictive prices $\mathbf{g}_j$ as samples of a continuous function $g_j(\beta)$. More specifically, $\mathbf{g}_j$ is simply the continuous function $g_j(\beta)$ sampled at the $I$ points $\{\hat{\beta}_i, i \in 1, \ldots, I\}$. These $I$ points could then be interpolated smoothly to reconstruct the function $g_j(\beta)$ over the domain $[0, 1]$. Figure 4.1 in Section 4.3 is an example.

Let $u_j = 1$ (due to Theorem 5), and define $e_{ij} \equiv c_{ij} h_j \theta_j$ for convenience. Interpreting $g_{ij}$ as the function $g_j(\beta)$ evaluated at the point $\hat{\beta}_i$, we can then rewrite (4.5) as follows:

$$\begin{aligned} \pi_{ij} &= N_{ij} V_i c_{ij} h_j \theta_j g_{ij} \\ &= e_{ij} V_i g_j(\hat{\beta}_i) \end{aligned} \tag{4.17}$$

Recall that $\pi_{ij}$ is the revenue earned by publisher $i$ from traffic sent to network $j$.

To simplify our discussion, let us assume in this section that $\beta_j^{\text{Net}} = \beta_k^{\text{Net}} = 1 \; \forall (j, k)$ (the results that follow do not depend on this assumption). Then, as discussed in Section 4.3.2, network $j$ can compute $\hat{\beta}_i = r_i \beta_i^{\text{Pub}}$ by just counting up the number of conversions, $A_i$, and dividing by the volume of clicks, $V_i$ (see (4.8)). Therefore, we can write (4.17) as:

$$\pi_{ij} = e_{ij} \frac{A_i}{\hat{\beta}_i} g_j(\hat{\beta}_i) \tag{4.18}$$

Clearly, the incentive for click inflation would be eliminated if $\pi_{ij}$ were nonincreasing in $B_i$. Since $V_i = G_i + B_i$ and $\hat{\beta}_i$ is inversely proportional to $V_i$, it is sufficient that $\pi_{ij}$ is nondecreasing in $\hat{\beta}_i$. We can therefore differentiate (4.18) and impose a nonnegativity condition:

$$\frac{\partial \pi_{ij}}{\partial \hat{\beta}_i} = e_{ij} A_i \left( \frac{g_j'(\hat{\beta}_i)}{\hat{\beta}_i} - \frac{g_j(\hat{\beta}_i)}{\hat{\beta}_i^2} \right) \geq 0 \tag{4.19}$$

Simplifying, we get:

$$g_j'(\hat{\beta}_i) \geq \frac{g_j(\hat{\beta}_i)}{\hat{\beta}_i} \tag{4.20}$$

To eliminate the incentive for click inflation, network $j$ should select predictive prices $g_{ij}$ such that (4.20) holds for all $\hat{\beta}_i$. Intuitively, a high-quality publisher $i$ should not feel tempted to "masquerade" (i.e., by engaging in click inflation) as any other lower-quality publisher just because the latter is "getting a better deal" than the former. We can achieve this effect by approximating the derivative in (4.20) with a backward difference between publishers $i$ and $i-1$:

$$g'_j(\hat{\beta}_i) \approx \frac{g_{ij} - g_{i-1,j}}{\hat{\beta}_i - \hat{\beta}_{i-1}} \tag{4.21}$$

Simplifying again, we arrive at:

$$\frac{g_{i-1,j}}{g_{ij}} \leq \frac{\hat{\beta}_{i-1}}{\hat{\beta}_i} \tag{4.22}$$

Therefore, we can simply add the constraint (4.22) to the optimization problem (4.9), for each pair of publishers $i$ and $i-1$. It is very convenient that (4.22) can be added as-is to a geometric program (i.e., without using an approximation), meaning we can enforce these constraints exactly using the TRAFFICQUALITY algorithm.

The constraint (4.22) has a very interesting form. At equality, it forces $g_{ij}$ to be proportional to $\hat{\beta}_i$ i.e., $g_{ij} = \delta\hat{\beta}_i$ for some constant $\delta$. Substituting $g_{ij} = \delta\hat{\beta}_i$ into (4.4), and using (4.8), we find that the effective number of clicks for which network $j$ pays publisher $i$ is simply $E_{ij} = \delta c_{ij} A_i$. That is, $E_{ij}$ is directly proportional to the number of clicks that become conversions. We have, essentially, a CPA pricing scheme. It can even be shown that the expected revenues to each player in the market are equal to those that would be obtained from a CPA pricing scheme. Therefore, what the constraint (4.22) tells us is that CPA is just a special case within the set of traffic policies that eliminate the incentive for click inflation. For this reason, we refer to traffic policies that enforce (4.22) as *quasi-CPA* policies.

Consider a scenario with $I = 20$ publishers and $J = 2$ networks. The fraction of valid traffic and the conversion rate for each publisher $i$ is linear in $i$ ($r_i = 0.05i$ and $\beta_i^{\text{Pub}} = 0.0025i$). This scenario is the same as the one considered in Figure 4.1, except that we will now allow $h_1$ to be a decision variable. Figure 4.4 is a plot of the product $\hat{\beta}_i = r_i\beta_i^{\text{Pub}}$ for each publisher $i$ in this example – there are many low-quality publishers and relatively few high-quality publishers. Figure 4.5 shows the predictive prices that are recommended by TRAFFICQUALITY with and without the quasi-CPA constraint (4.22) included in the optimization problem (4.9). For example, the predictive price chosen for publisher 15 in the quasi-CPA case is $g_{ij} = 0.5625$, compared to $g_{ij} = 0.9843$ for the unconstrained case.

Recall that the relative values of predictive prices are important here, rather than the absolute values – if all predictive prices were halved, advertisers' bids would simply double in response. From Figures 4.4 and 4.5, we see that the quasi-CPA constraint has yielded predictive prices that are (almost exactly) proportional to $\hat{\beta}_i$, whereas the unconstrained profile has a very different shape.

The unconstrained policy tries to dissuade low-quality publishers 1 through 7 by applying drastically lower predictive prices to them compared to medium- and high-quality publishers (8 through 20). The quasi-CPA policy, by comparison, only allows network 1 to capture traffic from publishers 12 through 20. Due to (4.22), the medium-quality publishers 8 through 11 are penalized harshly compared to the unconstrained case, so they choose network 2 instead. In this sense, the potential revenues generated from publishers 8 through 11 is the price network 1 must pay in exchange for using a click-inflation-deterrent traffic policy.

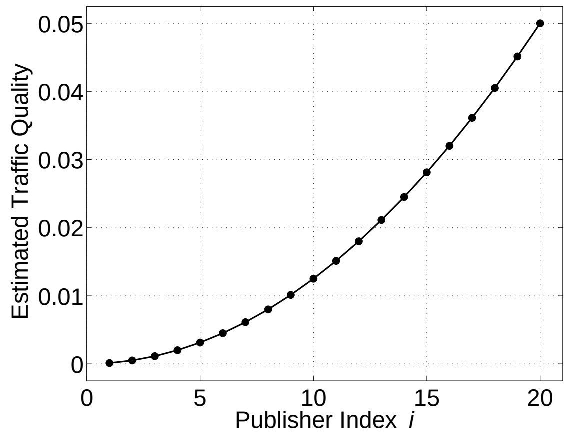


Figure 4.4: Estimated traffic quality, $\hat{\beta}_i = r_i \beta_i^{\text{Pub}}$.

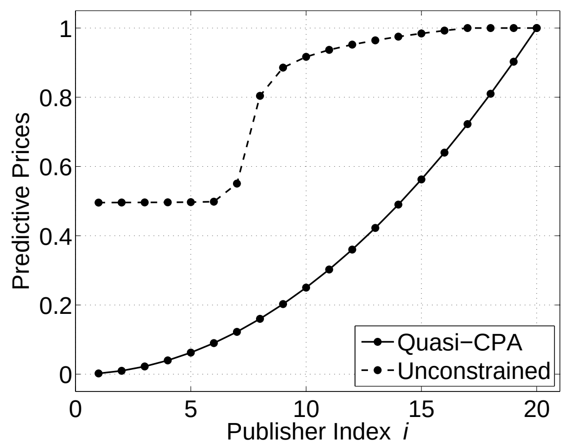


Figure 4.5: Effect of (4.22) on predictive prices.

### 4.4.4 Comparison

In Sections 4.4.2 and 4.4.3, we presented two different approaches for removing the incentive for click inflation. Which approach should a network use?

The answer depends completely on how accurately the network is able to estimate $r_i$. In the example in Section 4.4.3, network 1's profit using a quasi-CPA policy turns out to be $79\%$ of the profit using the unconstrained policy (which could be used if there was no click inflation, or if $r_i$ could be estimated exactly). From the experiment in Section 4.4.2 (see Figure 4.3), we see that average profits exceed $80\%$ when the standard estimation error $\sigma$ is less than $10\%$. Therefore, in this instance, network 1 should use its estimates of $r_i$ if $\sigma \leq 10\%$. Otherwise, a quasi-CPA traffic policy should be used.

In general, there will be a threshold value, say $\sigma^*$, such that if $\sigma \leq \sigma^*$, estimating $r_i$ directly would be more profitable on average for network 1. The value of $\sigma^*$, of course, would need to be identified by each network in practice.

## 4.5 Discussion

**Relation to Chapters 2 and 3**

The model we presented in Section 4.2 is strictly more general than the models in Chapters 2 and 3, so we have remained notationally consistent with earlier chapters whenever possible. Fixing $g_{ij} = 1\ \forall(i,j)$ and $h_j = h\ \forall j$ (i.e., no predictive pricing and equal revenue shares), assuming $\beta_{ijk} = \beta\ \forall(i,j,k)$ (equal conversion rates) and treating $u_j$ as a decision variable places us in the setting of Chapter 2. On the other hand, fixing $u_j = 1$, assuming $r_i = 1$ and treating $\mathbf{g}_j$ and $h_j$ as decision variables gives us the model used in Chapter 3. In this sense, TRAFFICQUALITY is a generalization of the PRICINGPOLICY algorithm presented in Chapter 3, since the optimization is done over the space of traffic policies $(u_j, \mathbf{g}_j, h_j)$, rather than pricing policies $(\mathbf{g}_j, h_j)$.

In Chapter 2, filtering aggressiveness was measured as the rate of false positives, $x_j$ i.e., the fraction of valid clicks that are mistakenly marked invalid. Thus, $x_j = 1 - u_j$. The true-positive rate was then $x_j^{\alpha_j}$, for $\alpha_j \in (0, 1]$. A lower value of $\alpha_j$ indicated that network $j$ was more effective at filtering, as opposed to a higher value of $\gamma_j$ in our current model. The benefit of switching the notation in this chapter is that the resulting expression for $N_{ij}$ (see (4.2)) can be more easily converted to a geometric programming constraint, as discussed in Appendix A.6.

**Yahoo! and Google**

There is anecdotal evidence that Yahoo! and Google employ predictive pricing profiles qualitatively similar to the unconstrained case in Figure 4.5, i.e., a few low-quality publishers being punished severely, and the rest not even noticing the effects of predictive pricing. Just search online for “quality-based pricing” or “smart pricing,” and read about publishers about “getting smart priced.”

**Separability**

We can derive an expression analogous to (4.9) by assuming a “weaker” form of separability, namely:

$$\beta_{ijk} = \beta_{ij}\beta_k^{\text{Adv}}\ \ \forall(i,j,k) \tag{4.23}$$

Here, $\{\beta_{ij}\ \forall(i,j)\}$ would be a matrix of pairwise conversion rates. Separability as assumed in (4.6), however, is most useful when estimating parameters as described in Section 4.3.2. Using a weaker assumption would necessitate collecting more data in order to get accurate parameter estimates.

An important follow-up to this work would be an empirical test of how well conversion-rate separability approximates true market conditions. A related separability assumption is made in [1] and [33], where the click-through rate for an impression is assumed to be the product of an advertiser-specific factor and a position/slot-specific factor, although empirical support has been mixed (e.g., [15]).

### Quasi-CPA vs. CPA

An advantage of using quasi-CPA instead of CPA is that a network that uses the former is not forced to choose predictive prices where (4.22) holds with equality. Allowing (4.22) to be strict (i.e., choosing $g_{ij}$ to be superlinear in $\hat{\beta}_i$) may, for example, allow network $j$ to exploit a suboptimal policy choice by a competitor. In such cases, a network would create not only a disincentive for click-inflation, but also increased profits compared to using CPA.

### Publishers

The results of Section 4.4.2 suggest that networks should devote engineering resources to estimating $r_i$ accurately, rather than trying to mark individual clicks valid or invalid. For example, suppose network $j$'s filtering algorithm marks $u_j = 50\%$ of valid clicks as valid. Then, the other $1 - u_j = 50\%$ of valid clicks are marked invalid. Such a filtering algorithm would be utterly unusable in practice for the task of marking individual clicks valid or invalid – no high-quality publisher would tolerate payment being erroneously denied for half of its valid traffic! Now, suppose network $j$ actually knew that $u_j = 50\%$ for its algorithm and also knew that, say, $\gamma_j = 2$ (i.e., network $j$ is doing a relatively poor job of filtering, but knows exactly how poor a job it is doing). If publisher $i$'s traffic is $r_i = 10\%$ valid, network $j$'s filter will mark $N_{ij} = u_j r_i + u_j^{\gamma_j}(1 - r_i) = 28\%$ of it valid. However, since $u_j$ and $\gamma_j$ are known to $j$, it can simply estimate $r_i$ as $\hat{r}_i = \frac{N_{ij} - u_j^{\gamma_j}}{u_j - u_j^{\gamma_j}}$, and adjust the predictive price $g_{ij}$ so that publisher $i$ is only paid for $\hat{r}_i = 10\%$ of its traffic. The outcome is equivalent, in an economic sense, as if the network were highly skilled at filtering.

Most publishers in practice, especially high-quality ones, are honest. They generate their revenues from legitimate, organic traffic, and do not engage in click inflation. So, perhaps, countermeasures against click inflation should only be activated when it is suspected that a particular publisher is inflating its click volumes, and then only on that publisher's traffic. Doing so would allow a network increased freedom in choosing its traffic policy. Suspicion may be triggered when estimated $\hat{r}_i$ or $\hat{\beta}_i$ values start to increase from their historical values. On the other hand, there may be marketing or public-relations reasons for enabling countermeasures for all publishers at all times, even at the cost of suboptimal profits.

## 4.6 Managerial Implications

In the management of CPC traffic quality, we drew an important distinction between organic traffic and publisher-initiated click inflation. If possible, predictive pricing should be chosen in favour of filtering to manage organic traffic quality. To fight click inflation, either filtering or predictive pricing can be used, depending on how well a network can characterize the performance of its filtering algorithm. In either case, it is important to remember that eliminating the incentive for click inflation does not mean that invalid traffic will disappear. Many forms of invalid traffic are, in fact, organic – indeed, a high amount of invalid traffic from web crawlers may imply that a publisher's site has become quite popular with users!

As for filtering, the key point is that a network's competitive advantage arises from accurately characterizing the aggregate performance of its filtering algorithm (i.e., knowing $u_j$ and $\gamma_j$), rather than correctly deciding on the validity of individual clicks. Even though a network's filtering algorithm makes incorrect decisions on a large number of individual clicks, it can compensate for these errors in aggregate by simply adjusting its predictive prices.

Since much of the online advertising industry (including some of the largest networks) currently operates on a CPC basis, there may be large costs and risks associated with switching over to a CPA pricing scheme. Using a quasi-CPA traffic policy allows a network to reap the benefits of a CPA scheme (i.e., no click inflation) while maintaining a CPC infrastructure. A very important reason publishers and networks shy away from (and advertisers prefer) CPA schemes is that they transfer the risk of organic nonconverting traffic entirely away from the advertisers (and on to publishers and networks). Using a CPC-based quasi-CPA policy enables a more equitable distribution of risk between all players in the market.

## 4.7 Summary

In this chapter, we discussed how filtering, predictive pricing and revenue sharing can be used together to influence the quality of traffic delivered by a CPC advertising network. Managing traffic quality is critical, since quality uncertainty and asymmetric information between publishers and advertisers can destroy value (i.e., the lemons market effect). Useful directions for future work include the scale-up and parallelization of the TRAFFICQUALITY algorithm, testing the separability assumptions (4.6) and (4.23), and running the estimation procedures described in Section 4.3.2 on real data. Ultimately, the applicability of our proposed techniques can only be gauged via a real implementation – we invite and encourage advertising networks to share their experiences in future published work.

# Chapter 5

# Conclusion

As we have seen, CPC advertising markets have the two hallmarks of a lemons market: *asymmetric information*, as publishers and ad networks hold an informational advantage over the advertisers, and *quality uncertainty*, with respect to the validity and targetedness of the traffic being bought and sold. Advertising networks act as reputable, long-lived intermediaries in the CPC market, to prevent prices from collapsing due to low average traffic quality. Fundamentally, networks can take on this role (and earn economic rents) because advertisers are able to "punish" the network for poor quality traffic by denying the network future business. Publishers are often short-lived, and cannot be punished in the same way.

Filtering and predictive pricing are the tools that networks use to ensure that average traffic quality is sufficiently high. The revenue share, on the other hand, is used to extract profits – if the network delivers high quality, advertiser bids will be higher, which means publishers will accept a lower share of revenue. We have presented techniques that allow an advertising network to calculate a near-optimal filtering, predictive pricing and revenue sharing policy to apply to its click-through traffic. We modeled the CPC market as a large, multiplayer game between publishers, advertisers and advertising networks, and then analyzed how publishers and advertisers would respond, in equilibrium, to various policy choices by the networks.

#### Lessons Learned

In retrospect, it seems that much of the fear in the market about click fraud is misguided. Rather than fixating narrowly on the click fraud problem, what an advertising network really ought to care about is increasing its own revenue, and consequently the return on investment (ROI) for its advertisers. Revenues and ROI are determined mainly by traffic quality, and traffic quality is a much broader issue than simply click fraud.

Indeed, in the land of CPC advertising, ROI is king. We worked with a model where advertisers maximize profits with respect to a ROI constraint, as opposed to a seemingly more natural budget allocation model. The ROI targeting model is more reflective of the decision problem facing real advertisers.

Traffic quality was shown to be more important than traffic quantity. Considering validity alone, we saw that low-quality publishers would choose to send traffic to an high-quality ad network, even though it meant fewer clicks being marked valid. As for targetedness, we saw networks trying to turn away low-quality publishers, choosing a lower traffic volume in exchange for higher quality.

The right way to manage CPC traffic quality seems to be a two-pronged approach. For the vast majority of click traffic (that is, organic traffic), influencing quality via predictive pricing and revenue sharing is most effective. That is not to say that filtering is unnecessary. Rather, filtering algorithms are a useful source of side information for networks to use in managing overall traffic quality.

### Future Work

There are a few promising directions in which our work could be extended. For example, certain networks offer advertisers a degree of freedom that we have not modeled here, namely, the ability to block out (or, indeed, request) traffic from specific publishers. Similarly, an extension of our model to multiple periods would give us an grasp of how pricing policies may evolve over time.

Next, we discuss what we believe to be a pair of particularly interesting directions.

**Equilibria** The output of the PRICINGPOLICY and TRAFFICQUALITY algorithms is a best-response policy for a single advertising network. A natural question to ask is, then, what about the equilibria of our game? For example, is it possible to characterize the set of subgame-perfect equilibria in our model?

Due to the richness of the networks' decision spaces in our game, general properties of subgame-perfect equilibria seem difficult to derive. Corollary 1 in Section 4.3 is one of the few general statements we can make. Another partial answer is given in Chapter 2, where it is shown that if no network uses predictive pricing, then in equilibrium all publishers and advertisers will flock to the network that is most skilled at filtering. However, with predictive pricing, such a "crisp" result does not hold. Many equilibria exist, and the structure is highly dependent on the players' types. In particular, in every equilibrium that we were able to compute (i.e., iteratively computing best-response functions until we arrived at a fixed-point), all networks received some positive share of the market. This, of course, is a realistic outcome.

A complete, analytical characterization would be a very interesting extension to our work. On the other hand, perhaps it is more realistic that the industry will only gradually evolve toward an equilibrium, as more networks slowly adopt the practice of predictive pricing. In such a case, our algorithm for computing best-responses is probably of more practical utility to an individual network than any statement on long-run equilibria.

**Stochasticity** As we suggested earlier, we believe that risk management is the most compelling direction for future research in online advertising. The first step in this direction is to introduce stochasticity into our model. Traffic volume, validity, conversion rates and revenue per conversion are quantities that seemingly ought to be modeled as random variables. Revenue for each player would then need to be expressed a function of these random variables. Also, since players can know each others' types only approximately, it may be realistic to model this knowledge as probability distributions over players' types. The structure of optimal pricing and traffic policies may indeed be quite different, under a model that includes stochasticity.

From there, the most interesting question will be how to define, price and manage the risk of nonconverting, low-quality click traffic (i.e., the danger of paying for a large number of clicks, but getting very few conversions as a result). Cost-per-click pricing schemes involve the advertisers bearing some risk of nonconverting traffic (that is, paying for clicks from which no conversions result). Advertisers prefer cost-per-acquisition (CPA) schemes for the simple reason that this notion of risk is completely eliminated, since payment only occurs if a click is converted to an acquisition. And for much the same reason, publishers prefer cost-per-impression schemes.

As such, it seems fair that the price paid per acquisition in a CPA scheme ought to be higher than the expected cost per acquisition in a CPC scheme (i.e., the number of clicks multiplied by the expected conversion rate). And on the flipside, networks and publishers ought to be rewarded for bearing this risk on behalf of the advertisers. Borrowing a term from economics, is the market is pricing in such a "risk premium"? If not, it would be a business opportunity for a clever entrepreneur! How large is this risk premium? Can we measure it? These are fascinating open questions, in our opinion.

## Appendix A

# Proofs

## A.1 Advertisers

We present a simple argument for why Equation (2.14) – repeated below for convenience – will hold at the optimum for each advertiser $k$ and network $j$:

$$R_{kj} = R_k \ \forall j \tag{A.1}$$

Intuitively, (A.1) says that advertiser $k$ will choose its valuations $\{v_{kj} \ \forall j\}$ such that a) its ROI on network $j$'s traffic, $R_{kj}$, is the same across all networks, and b) $R_{kj}$ will be equal to $R_k$, which is the ROI attainable on channels other than CPC.

Let us first consider a setting where an advertiser can buy clicks from just two channels, say, $C_1$ and $C_2$ (we comment on the extension to $J$ networks shortly). Here, $C_1$ meant to be a proxy for channels other than CPC, to which advertisers could allocate their budgets. Traffic from $C_1$ costs $p_1$ per click, irrespective of the quantity purchased, $Q_1$. Traffic from $C_2$, on the other hand, is sold at auction.

For a bidder with a private per-click valuation $v$, let $b(v)$ be the optimal bid, let $p_2(v) \equiv p_2(b(v))$ be the corresponding price per click and let $Q_2(v) \equiv Q_2(b(v))$ be the corresponding quantity of clicks awarded. The advertiser's expected revenue from $C_1$'s traffic would then be $Q_1\beta_1 y$, where $\beta_1$ is the conversion rate and $y$ is the revenue generated per conversion. The cost incurred for this traffic is $Q_1 p_1$. Similarly, for $C_2$'s traffic, the expected revenue is $Q_2(v)\beta_2 y$ and the cost is $Q_2(v)p_2(v)$. The advertiser selects quantities $Q_1$ and $Q_2(v)$ such that revenues are maximized, subject to a (large, but finite) budget constraint, $B$, on the total cost:

$$\begin{aligned} \text{maximize} \quad & Q_1\beta_1 y + Q_2(v)\beta_2 y \\ \text{subject to} \quad & Q_1 p_1 + Q_2(v)p_2(v) \le B \end{aligned} \tag{A.2}$$

So, in this setting, what can we say about the advertiser's equilibrium valuation, $v^*$, for traffic from $C_2$? It is easy to see (for example, from the Lagrangian) that the constraint in (A.2) will hold with equality. Hence, we get following expression for the advertiser's total revenue:

$$\begin{aligned} \pi(v) &\equiv \left(\frac{B - Q_2(v)p_2(v)}{p_1}\right)\beta_1 y + Q_2(v)\beta_2 y \\ &= Q_2(v)\left(\beta_2 - \frac{p_2(v)\beta_1}{p_1}\right) y + \frac{B\beta_1 y}{p_1} \end{aligned} \tag{A.3}$$

The equilibrium valuation, $v^*$, will be consistent with the choices of $Q_1$ and $Q_2(v)$ that maximize the advertiser's revenues, $\pi(v)$. That is, $v^* \in \arg\max_v \pi(v)$.

Let us assume that the quantity $Q_2(v)$ is concave and nondecreasing in $v$, the cost function $Q_2(v)p_2(v)$ is convex in $v$, and that $Q_2(0) = p_2(0) = 0$. These assumptions are sufficient (though not necessary) to ensure that $\pi(v)$ is concave. We can therefore differentiate $\pi(v)$, equate it to zero, and conclude that $v^*$ satisfies:

$$R_1 \equiv \frac{\beta_1 y}{p_1} = \frac{\beta_2 y}{p_2(v^*)} \equiv R_2(v^*) \tag{A.4}$$

The left-hand side of (A.4), $R_1$, is exactly the ROI from $C_1$. The right-hand side, $R_2$, is exactly the ROI from $C_2$. Thus, (A.4) tells us that the equilibrium valuation, $v^*$, and the equilibrium budget allocations are such that the ROI from channels $C_1$ and $C_2$ will be equal. Moreover, since $Q_2(v)$ is increasing in $v$ and $R_2(v)$ is decreasing in $v$, we can rewrite (A.2) as:

$$\begin{aligned} \text{maximize} \quad & Q_1\beta_1 y + Q_2(v)\beta_2 y \\ \text{subject to} \quad & R_2(v) \geq R_1 \end{aligned} \tag{A.5}$$

In words, the advertiser will continue to increase $v$, as long the ROI $R_2(v)$ continues to be higher than the alternative, which is $R_1$. The optimization problem (3.16) (see Section 3.2) and Equation (A.1) are direct generalizations of (A.5) and (A.4).

This argument generalizes directly to $J > 2$ networks – simply modify Equation (A.2) to be an optimization over valuations $v_1, \ldots, v_J$ corresponding to networks $C_1, \ldots, C_J$. Moreover, $C_1$ does not have to be an online medium – we would replace $Q_1 p_1$ with the amount spent offline, and replace $Q_1\beta_1 y$ with the resulting revenue. In general, advertisers allocate their budgets across a diverse set of advertising channels, rather than just CPC (e.g., print, radio, or cost-per-impression schemes). In such a setting, $R_1$ would represent the highest ROI that the advertiser can achieve through channels other than CPC advertising. ROI is the key criterion used by advertisers in deciding how to allocate their budgets.

Finally, notice that the advertiser faces two distinct, orthogonal decision problems. First, the advertiser must decide its valuation, $v^*$, as discussed above. Second, it must translate its valuation into its optimal bid at auction, $b(v^*)$. The first is an individual optimization problem based on ROIs on alternative advertising channels, and is wholly independent of other advertisers. The second problem, on the other hand, is clearly a competitive bidding process versus other advertisers for CPC traffic. We do not need explicit solutions for $b(v)$, the optimal bid, or $Q(v)$, the quantity of clicks awarded. The budget, $B$, is also unimportant. In our derivations, we can simply apply the ROI constraints (3.16) and (A.1), and treat the specifics of each network's auction mechanism as a "black box."

## A.2 Proof of Theorem 1

### A.2.1 Outline

There are three main steps involved in proving Theorem 1:

1. For a fixed first-step outcome, derive the *best-response functions* for publishers and advertisers.

2. Show that the best-response functions have a fixed point, implying the existence of a pure-strategy equilibrium in the subgame for any first-step outcome.

3. Show that for any fixed point of the best-response functions, equations (2.17) - (2.20) hold.

We will use the symbol $\mathbf{C}$ to denote a matrix whose $(i, j)$-element is publisher $i$'s allocation $c_{ij}$. Similarly, $\mathbf{V}$ is a matrix whose $(kj)$-element is advertiser $k$'s valuation $v_{kj}$, and $\mathbf{x}$ is a vector whose element $j$ is $x_j$. We also define $\mathbf{c}_i$ to be the length-$J$ vector whose element $j$ is $c_{ij}$:

$$\mathbf{c}_i \equiv [c_{i1}\ c_{i2}\ \cdots\ c_{ij}]^T \tag{A.6}$$

Let $M_J$ be the $J$-dimensional simplex (i.e., the set of $J$-dimensional vector whose elements sum to 1). Each point in $M_J$ is a feasible allocation of traffic for a single publisher. Then, let $S_{Pub} \equiv M_J^I$ be the $I$-fold cross-product of $M_J$ with itself. $S_{Pub}$ represents the set of possible decisions that the $I$ publishers can collectively make, i.e., $\mathbf{C} \in S_{Pub}$.

Similarly, let $S_{Adv} \equiv \mathbb{R}_+^{KJ}$ (i.e., the set of $K$-by-$J$ matrices whose elements are all nonnegative). Each point in $S_{Adv}$ is a feasible profile of valuations for each advertiser $k$ on each network $j$, i.e., $\mathbf{V} \in S_{Adv}$.

Finally, let $S_{Net} \equiv [0, 1)^J$ (i.e., the open unit-hypercube of dimension-$J$). Each point in $S_{Net}$ is a profile of filtering decisions for the $J$ networks, i.e., $\mathbf{x} \in S_{Net}$.

Define $S_1 \equiv S_{Pub} \times S_{Adv} \times S_{Net}$ and $S_2 \equiv S_{Pub} \times S_{Adv}$. The symbol "$\times$" denotes a cross-product of sets. We refer to $S_1$ and $S_2$ as *decision spaces*, and elements of $S_1$ and $S_2$ as *decision profiles*. Each decision profile $(\mathbf{C}, \mathbf{V}, \mathbf{x}) \in S_1$ is a possible outcome of the supergame (i.e., first- and second-steps), whereas each decision profile $(\mathbf{C}, \mathbf{V}) \in S_2$ is a possible outcome of the subgame (i.e., the second step). Each $\mathbf{x} \in S_{Net}$ is a possible outcome of the first step.

We assume throughout the proof that $x_j \in [0, 1)\ \forall j$ i.e., that $x_j = 1$ is not allowed. In words, no ad network marks 100% of its clicks invalid. This assumption allows us to sidestep some "corner cases" in our analysis. However, assuming $x_j < 1$ is by no means necessary for our results, since it can be shown that $x_j = 1$ never occurs in equilibrium.

Table A.1 is a list of notation we introduce for the proof of Theorem 1.

### A.2.2 Advertiser $k$'s Best Response

Fix a first-stage outcome, $\mathbf{x} \in S_{Net}$.

Let $BR_{kj}^{Adv} : S_2 \rightarrow 2^{S_{Adv}}$ be advertiser $k$'s *best-response function* on ad network $j$. That is, for any decision profile $(\mathbf{C}, \mathbf{V}) \in S_2$, each element $v'_{kj}$ of the set $BR_{kj}^{Adv}(\mathbf{C}, \mathbf{V}; \mathbf{x})$ is an optimal choice of $v_{kj}$ for advertiser $k$ on ad network $j$, assuming $\mathbf{x}$ was played in the first stage and other players' choices in the subgame are given by $(\mathbf{C}, \mathbf{V})$. Notice that $BR_{kj}^{Adv}$ is a set-valued function because $k$'s optimal valuation $v'_{kj}$ on ad network $j$ need not be unique.

We will now derive an expression for $BR_{kj}^{Adv}$. Fix a decision profile $(\mathbf{C}, \mathbf{V}) \in S_2$ and a particular ad network $j$. There are two possible cases: 1) $c_{ij} \neq 0$ for some $i$, and 2) $c_{ij} = 0\ \forall i$.

1. $c_{ij} \neq 0$ for some $i$.

| Symbol | Description |
|---|---|
| $\mathbf{C}$ | $I$-by-$J$ matrix whose $(i,j)$-element is $c_{ij}$ |
| $\mathbf{V}$ | $K$-by-$J$ matrix whose $(k,j)$-element is $v_{kj}$ |
| $\mathbf{x}$ | Length-$J$ vector whose element $j$ is $x_j$ |
| $\mathbf{c}_i$ | Length-$J$ vector whose element $j$ is $c_{ij}$ |
| $S_{Pub}$ | Set of possible allocations, $\mathbf{C}$, for the publishers |
| $S_{Adv}$ | Set of possible valuations, $\mathbf{V}$, for the advertisers |
| $S_{Net}$ | Set of possible filtering decisions, $\mathbf{x}$, for the networks |
| $S_1$ | Set of possible outcomes $(\mathbf{C},\mathbf{V})$ for the subgame |
| $S_2$ | Set of possible outcomes $(\mathbf{C},\mathbf{V},\mathbf{x})$ for the supergame |
| $BR_{kj}^{Adv}$ | Advertiser $k$'s best response function on ad network $j$ |
| $BR_{ij}^{Pub}$ | Publisher $i$'s best response function on ad network $j$ |
| $BR$ | Best response function for all publishers and advertisers |
| $a_j$ | Advertisers' adjustment factor for ad network $j$ |
| $X_{ij}$ | Publisher $i$'s expected revenue for sending a single click to ad network $j$ |
| $X_i^*$ | Publisher $i$'s maximal per-click expected revenue |
| $\Phi\left(\{X_{ij}\}\right)$ | Set of solutions $\mathbf{c}_i'$ to equations (A.14) - (A.16) |
| $(\mathbf{C}^*,\mathbf{V}^*)$ | A fixed point of the best-response function, $BR$ |

Table A.1: Notation used in Appendix A.2.

Advertiser $k$ chooses $v_{kj}$ so that $R_{kj} = R_k$. From (2.10), (2.11) and (2.12):

$$R_{kj} = \frac{Y_{kj}}{v_{kj} Z_{kj}} = \frac{\left(\sum_i r_i V_i c_{ij}\right) \frac{1}{K} \beta y_k}{v_{kj} \frac{1}{K} \sum_i N_{ij} V_i c_{ij}} \tag{A.7}$$

Notice that the factor $\frac{1}{K}$ cancels out from the numerator and denominator in (A.7). The factor $\xi_{kjt}$ defined in Section 2.3 cancels out in a similar manner. As such, our theorems are independent of the schemes used by ad networks to distribute clicks. Setting $R_{kj} = R_k$ in (A.7) and solving for the optimal $v_{kj}$, we get:

$$v_{kj}' \equiv \frac{\left(\sum_i r_i V_i c_{ij}\right) \beta y_k}{R_k \sum_i N_{ij} V_i c_{ij}} = \frac{y_k}{R_k} a_j \tag{A.8}$$

where we have defined the *adjustment factor*, $a_j$, as:

$$a_j \equiv \frac{\left(\sum_i r_i V_i c_{ij}\right) \beta}{\sum_i N_{ij} V_i c_{ij}} \tag{A.9}$$

Intuitively, (A.8) says that advertiser $k$'s optimal valuation $v_{kj}'$, in response to $(\mathbf{C},\mathbf{V};\mathbf{x})$, is equal to their "nominal valuation" $\frac{y_k}{R_k}$, scaled by the per-ad network adjustment factor $a_j$. Observe that if $c_{ij} = 1\ \forall i$, then $a_j = M_j$, where $M_j$ was defined in (2.16).

2. $c_{ij} = 0\ \forall i$.

   The ROI $R_{kj}$ is undefined since publishers do not send any clicks to ad network $j$, and so $j$ doesn't have any clicks to sell to advertiser $k$. The optimal valuation is not unique – in fact, any nonnegative value of $v_{kj}$ is reasonable. Therefore, when $c_{ij} = 0\ \forall i$, we define $BR^{Adv}_{kj}(\mathbf{C}, \mathbf{V}; \mathbf{x}) = [0, \infty)$.

Combining these two cases, we get:

$$v'_{kj} \in BR^{Adv}_{kj}(\mathbf{C}, \mathbf{V}; \mathbf{x}) = \begin{cases} [0, \infty) & \text{if } c_{ij} = 0\ \forall i \\ \frac{y_k}{R_k} a_j & \text{if } c_{ij} > 0 \text{ for some } i \end{cases} \tag{A.10}$$

### A.2.3 Publisher $i$'s Best Response

Fix a first-stage outcome, $\mathbf{x} \in S_{Net}$.

Let $BR^{Pub}_i : S_2 \to 2^{S_{Pub}}$ be publisher $i$'s *best-response function* on ad network $j$. That is, for any given decision profile $(\mathbf{C}, \mathbf{V}, \mathbf{x}) \in S_1$, each element $\mathbf{c}'_i$ of the set $BR^{Pub}_{kj}(\mathbf{C}, \mathbf{V}; \mathbf{x})$ is an optimal vector of allocations $\mathbf{c}_i$ for publisher $i$, assuming $\mathbf{x}$ was played in the first stage and other players' choices in the subgame are given by $(\mathbf{C}, \mathbf{V})$. Notice that $BR^{Pub}_i$ is a set-valued function because $i$'s optimal allocation $c'_{ij}$ on each ad network $j$ need not be unique.

We will now derive an expression for $BR^{Pub}_i$. Fix a decision profile $(\mathbf{C}, \mathbf{V}, \mathbf{x}) \in S_1$. From (2.13), publisher $i$ chooses $\mathbf{c}_i$ to maximize its total profits, as follows:

$$\max_{\mathbf{c}_i} \sum_j \sum_k \pi_{ijk} \ \text{ s.t. } \ \sum_j c_{ij} = 1 \ \text{ and } \ c_{ij} \geq 0\ \forall j \tag{A.11}$$

From (2.6), $\sum_k \pi_{ijk}$ is linear in $c_{ij}$:

$$\sum_k \pi_{ijk} = N_{ij} V_i c_{ij} \frac{1}{K} h \sum_k v_{kj} \tag{A.12}$$

Therefore, (A.11) is very easy to solve – publisher $i$ simply sends all its clicks to the single ad network (or multiple ad networks, if there is a tie) that will generate the highest expected revenue per click.

Unfortunately, formally expressing the solution to (A.11) is somewhat cumbersome notationally. Define $X_{ij}$ as follows:

$$X_{ij} \equiv N_{ij} \frac{1}{K} h \sum_k v_{kj} \tag{A.13}$$

$X_{ij}$ is a key quantity in the proof of Theorem 1 – it is the criterion used by publisher $i$ to decide what fraction of clicks to allocate to $j$. Intuitively, $X_{ij}$ is $i$'s expected revenue for sending a single click to $j$. $X_{ij}$ is equal to $\sum_k \pi_{ijk}$ when $c_{ij} = 1$ and $V_i = 1$.

Let $X^*_i \equiv \max_j X_{ij}$. Any $\mathbf{c}_i$ that satisfies:

$$c_{ij} \geq 0 \text{ if } X_{ij} = X^*_i \tag{A.14}$$

$$c_{ij} = 0 \text{ if } X_{ij} < X^*_i \tag{A.15}$$

$$\sum_j c_{ij} = 1 \tag{A.16}$$

is a best response to $(\mathbf{C}, \mathbf{V}, \mathbf{x})$ for publisher $i$.

Let $\Phi(\{X_{ij}\})$ denote the set of vectors $\mathbf{c}_i$ that satisfy equations (A.14) - (A.16). Using this notation, we can write the solution to (A.11) as:

$$\mathbf{c}_i' \in BR_i^{Pub}(\mathbf{C}, \mathbf{V}; \mathbf{x}) = \Phi(\{X_{ij}\}) \tag{A.17}$$

### A.2.4 Existence of Fixed Points

In this section, we show that the best-response functions always have a fixed point. Define $BR : S_2 \to 2^{S_2}$ as follows:

$$\begin{aligned} BR(\mathbf{C}, \mathbf{V}; \mathbf{x}) \quad \equiv \quad & BR_1^{Pub}(\mathbf{C}, \mathbf{V}; \mathbf{x}) \times BR_2^{Pub}(\mathbf{C}, \mathbf{V}; \mathbf{x}) \times \cdots \times BR_I^{Pub}(\mathbf{C}, \mathbf{V}; \mathbf{x}) \times \\ & BR_{11}^{Adv}(\mathbf{C}, \mathbf{V}; \mathbf{x}) \times BR_{12}^{Adv}(\mathbf{C}, \mathbf{V}; \mathbf{x}) \times \cdots \times BR_{1,J}^{Adv}(\mathbf{C}, \mathbf{V}; \mathbf{x}) \times \\ & \cdots \times \\ & BR_{K,1}^{Adv}(\mathbf{C}, \mathbf{V}; \mathbf{x}) \times BR_{K,2}^{Adv}(\mathbf{C}, \mathbf{V}; \mathbf{x}) \times \cdots \times BR_{kj}^{Adv}(\mathbf{C}, \mathbf{V}; \mathbf{x}) \end{aligned} \tag{A.18}$$

Each element $(\mathbf{C}', \mathbf{V}') \in BR(\mathbf{C}, \mathbf{V}; \mathbf{x})$ is a pair of matrices of optimal allocations $c_{ij}'$ and valuations $v_{kj}'$ in response to the decision profile $(\mathbf{C}, \mathbf{V}) \in S_2$, assuming $\mathbf{x}$ was played in the first step.

We claim that for any first-step outcome $\mathbf{x}$, $BR(\mathbf{C}, \mathbf{V}; \mathbf{x})$ has at least one fixed point i.e., for any $\mathbf{x} \in [0,1)^J$ there exists $(\mathbf{C}^*, \mathbf{V}^*)$ such that:

$$(\mathbf{C}^*, \mathbf{V}^*) \in BR(\mathbf{C}^*, \mathbf{V}^*; \mathbf{x}) \tag{A.19}$$

Our claim can be proved by construction. Pick any integer $n \in 1 \ldots J$. From (A.10) and (A.14) - (A.16), it is easy to see that the following is always a fixed point of $BR$:

$$c_{ij} = \begin{cases} 1 & j = n \\ 0 & j \neq n \end{cases} \qquad v_{kj} = \begin{cases} \frac{y_k}{R_k} M_j & j = n \\ 0 & j \neq n \end{cases} \tag{A.20}$$

A fixed point of the form (A.20) exists for every $n \in 1 \ldots J$. The fact that this construction works for any $n$ is why any ad network can be chosen in equilibrium.

A pure-strategy Nash equilibrium in the subgame is, by definition, a fixed point of $BR$. Therefore, for any first-stage outcome $\mathbf{x} \in [0,1)^J$, there exist pure-strategy Nash equilibria in the subgame.

### A.2.5 Analysis of Fixed Points

In this section, we will show that for any fixed point of $BR$, equations (2.17) - (2.20) must hold, thereby proving Theorem 1.

Fix an $\mathbf{x}$ and let $(\mathbf{C}^*, \mathbf{V}^*)$ be any fixed point of $BR$ i.e., $(\mathbf{C}^*, \mathbf{V}^*) \in BR(\mathbf{C}^*, \mathbf{V}^*; \mathbf{x})$. Let element $(i,j)$ of $\mathbf{C}^*$ be $c_{ij}$, and let element $(k,j)$ of $\mathbf{V}^*$ be $v_{kj}$. Going forward, we will only be interested in analyzing fixed points of $BR$. Unless specified otherwise, an allocation $c_{ij}$ and valuation $v_{kj}$ will always correspond to a fixed point $(\mathbf{C}^*, \mathbf{V}^*)$ of $BR$.

To prove Theorem 1, we only need to show that for any $(\mathbf{C}^*, \mathbf{V}^*)$, equations (2.17) and (2.19) will hold for any fixed point i.e., that $\exists j^*$ such that $c_{ij^*} = 1 \; \forall i$ and $c_{ij} = 0 \; \forall i, j \neq j^*$. If $c_{ij^*} = 1 \; \forall i$ and $c_{ij} = 0 \; \forall i, j \neq j^*$, (A.14) - (A.16) would imply that $X_{ij^*} \geq X_{ij} \; \forall i, j \neq j^*$. Equations (2.18) and (2.20) would then follow directly.

On our way to proving that equations (2.17) and (2.19) hold for any fixed point, we will state and prove Lemmas 2, 3 and 4. But first, we observe from equations (A.14), (A.15) and (A.16) that:

- If $c_{ij} > 0$ then $X_{ij} \geq X_{in} \; \forall n \neq j$
  - If publisher $i$ allocates any clicks to ad network $j$, then $i$'s per click revenue is highest on $j$ (there might be a tie, though)
- If $X_{ij} > X_{in} \; \forall n \neq j$ then $c_{ij} = 1$
  - If publisher $i$'s per click revenue on ad network $j$ is strictly higher on $j$ than any other ad network, $i$ will sends all its clicks to $j$

We will use these observations repeatedly in the proofs of Lemmas 2, 3 and 4.

The assumption that $(\mathbf{C}^*, \mathbf{V}^*)$ is a fixed point of $BR$ means that publishers and advertisers are both playing their best responses. Therefore, we can substitute (A.8) into (A.13). We get:

$$X_{ij} = N_{ij} \frac{1}{K} h \sum_k v_{kj} \tag{A.21}$$

$$= N_{ij} \frac{1}{K} h \frac{\beta \sum_i r_i V_i c_{ij}}{\sum_i N_{ij} V_i c_{ij}} \sum_k \frac{y_k}{R_k} \tag{A.22}$$

$$= N_{ij} \frac{1}{K} h \kappa \frac{\beta \sum_i r_i V_i c_{ij}}{\sum_i N_{ij} V_i c_{ij}} \tag{A.23}$$

where we have defined $\kappa \equiv \sum_k \frac{y_k}{R_k}$.

**Lemma 2.** *If $\exists (i, j, m, n)$ such that $n \neq j$, $c_{ij} > 0$, and $X_{mj} > X_{mn}$, then $c_{pn} = 0 \; \forall p$. It is sufficient but not necessary that $m = i$.*

In words, Lemma 2 says that if ad network $j$ receives any traffic at all, and at least one publisher strictly prefers ad network $j$ over some other ad network $n$, then that ad network $n$ will receive no traffic at all from any publisher.

*Proof.* Without loss of generality, let us arbitrarily choose $i = 1, j = 1, m = 1, n = 2, p = 2$ in the statement of Lemma 2, and demonstrate the result. That is, we will show that if $c_{11} > 0$ and $X_{11} > X_{12}$, then $c_{22} = 0$. Since the choice of $(i, j, m, n, p)$ was arbitrary, we will obtain the desired result.

Assume that $c_{11} > 0$ and $X_{11} > X_{12}$. Suppose $c_{22} > 0$ – we will show that $c_{22} > 0$ leads to a contradiction, which implies $c_{22} = 0$.

By assumption:

$$\begin{aligned} X_{11} &> X_{12} \\ N_{11}\frac{1}{K}h\kappa\frac{\beta\left(\sum_i r_i V_i c_{i1}\right)}{\sum_i N_{i1}V_i c_{i1}} &> N_{12}\frac{1}{K}h\kappa\frac{\beta\left(\sum_i r_i V_i c_{i2}\right)}{\sum_i N_{i2}V_i c_{i2}} \\ N_{11}\frac{\sum_i r_i V_i c_{i1}}{\sum_i N_{i1}V_i c_{i1}} &> N_{12}\frac{\sum_i r_i V_i c_{i2}}{\sum_i N_{i2}V_i c_{i2}} \end{aligned} \quad \text{(A.24)}$$

Define, for convenience:

$$A_1 \equiv \sum_i r_i V_i c_{i1} \quad \text{(A.25)}$$

$$A_2 \equiv \sum_i r_i V_i c_{i2} \quad \text{(A.26)}$$

$$B_1 \equiv \sum_i (1-r_i) V_i c_{i1} \quad \text{(A.27)}$$

$$B_2 \equiv \sum_i (1-r_i) V_i c_{i2} \quad \text{(A.28)}$$

Since $c_{11} > 0$ and $c_{22} > 0$ by assumption, we know $A_1$, $A_2$, $B_1$, and $B_2$ are all strictly positive. Using equation (2.2) along with this notation, we observe that:

$$\sum_i N_{i1} V_i c_{i1} = (1-x_1)A_1 + (1-x_1^{\alpha_1})B_1 \quad \text{(A.29)}$$

$$\sum_i N_{i2} V_i c_{i2} = (1-x_2)A_2 + (1-x_2^{\alpha_2})B_2 \quad \text{(A.30)}$$

Therefore, the inequality (A.24) can be written as:

$$\frac{\left((1-x_1)r_1 + (1-x_1^{\alpha_1})(1-r_1)\right)A_1}{(1-x_1)A_1 + (1-x_1^{\alpha_1})B_1} > \frac{\left((1-x_2)r_1 + (1-x_2^{\alpha_2})(1-r_1)\right)A_2}{(1-x_2)A_2 + (1-x_2^{\alpha_2})B_2} \quad \text{(A.31)}$$

The numerators and denominators in (A.31) are all strictly positive. Therefore, rearranging (A.31) gives:

$$\begin{aligned} (1-x_1)(1-x_2^{\alpha_2})(r_1A_1B_2 - (1-r_1)A_1A_2) &+ \\ (1-x_1^{\alpha_1})(1-x_2)((1-r_1)A_1A_2 - r_1A_2B_1) &+ \\ (1-x_1^{\alpha_1})(1-x_2^{\alpha_2})((1-r_1)A_1B_2 - (1-r_1)A_2B_1) &> 0 \end{aligned} \quad \text{(A.32)}$$

Multiplying (A.32) through by $V_1c_{11}$ gives:

$$\begin{aligned} (1-x_1)(1-x_2^{\alpha_2})(r_1V_1c_{11}A_1B_2 - (1-r_1)V_1c_{11}A_1A_2) &+ \\ (1-x_1^{\alpha_1})(1-x_2)((1-r_1)V_1c_{11}A_1A_2 - r_1V_1c_{11}A_2B_1) &+ \\ (1-x_1^{\alpha_1})(1-x_2^{\alpha_2})((1-r_1)V_1c_{11}A_1B_2 - (1-r_1)V_1c_{11}A_2B_1) &> 0 \end{aligned} \quad \text{(A.33)}$$

Let $\Delta_1 \equiv \{i \mid X_{i1} \geq X_{i2}\}$ i.e., the set of publishers that weakly prefer ad network 1 over ad network 2 – the assumption that $X_{11} > X_{12}$ means $\Delta_1$ is nonempty. Note that $c_{i1} = 0 \; \forall i \notin \Delta_1$, so that:

$$A_1 = \sum_i r_i V_i c_{i1} = \sum_{i \in \Delta_1} r_i V_i c_{i1} \tag{A.34}$$

$$B_1 = \sum_i (1-r_i) V_i c_{i1} = \sum_{i \in \Delta_1} (1-r_i) V_i c_{i1} \tag{A.35}$$

We can write an expression similar to (A.33) for each $i \in \Delta_1$:

$$\begin{aligned} (1-x_1)(1-x_2^{\alpha_2})(r_i V_i c_{i1} A_1 B_2 - (1-r_i) V_i c_{i1} A_1 A_2) &\quad + \\ (1-x_1^{\alpha_1})(1-x_2)((1-r_i) V_i c_{i1} A_1 A_2 - r_i V_i c_{i1} A_2 B_1) &\quad + \\ (1-x_1^{\alpha_1})(1-x_2^{\alpha_2})((1-r_i) V_i c_{i1} A_1 B_2 - (1-r_i) V_i c_{i1} A_2 B_1) &\quad \geq \quad 0 \end{aligned} \tag{A.36}$$

Therefore, the inequality will still hold when we sum over $i \in \Delta_1$:

$$\begin{aligned} \sum_{i \in \Delta_1} (1-x_1)(1-x_2^{\alpha_2})(r_i V_i c_{i1} A_1 B_2 - (1-r_i) V_i c_{i1} A_1 A_2) &\quad + \\ (1-x_1^{\alpha_1})(1-x_2)((1-r_i) V_i c_{i1} A_1 A_2 - r_i V_i c_{i1} A_2 B_1) &\quad + \\ (1-x_1^{\alpha_1})(1-x_2^{\alpha_2})((1-r_i) V_i c_{i1} A_1 B_2 - (1-r_i) V_i c_{i1} A_2 B_1) &\quad > \quad 0 \end{aligned} \tag{A.37}$$

Note that the inequality is strict since it is strict for $i = 1$ in (A.33). Moving the summation in (A.37) inside and simplifying gives:

$$\begin{aligned} 0 \quad < &\quad (1-x_1)(1-x_2^{\alpha_2})(A_1 A_1 B_2 - B_1 A_1 A_2) \\ + &\quad (1-x_1^{\alpha_1})(1-x_2)(B_1 A_1 A_2 - A_1 A_2 B_1) \\ + &\quad (1-x_1^{\alpha_1})(1-x_2^{\alpha_2})(B_1 A_1 B_2 - B_1 A_2 B_1) \\ = &\quad (1-x_2^{\alpha_2})((1-x_1) A_1 + (1-x_1^{\alpha_1}) B_1)(A_1 B_2 - B_1 A_2) \end{aligned} \tag{A.38}$$

We know that the following quantity is strictly positive:

$$(1-x_2^{\alpha_2})((1-x_1) A_1 + (1-x_1^{\alpha_1}) B_1) \tag{A.39}$$

Therefore, so we conclude that:

$$(A_1 B_2 - B_1 A_2) > 0 \tag{A.40}$$

Now, since $c_{22} > 0$ by assumption, we know that $X_{22} \geq X_{21}$. Let $\Delta_2 \equiv \{i \mid X_{i2} \geq X_{i1}\}$ i.e., the set of publishers that weakly prefer ad network 2 over ad network 1 – the assumption that $c_{22} > 0$ means $\Delta_2$ is nonempty. Note that $c_{i2} = 0\ \forall i \notin \Delta_2$, so:

$$A_2 = \sum_i r_i V_i c_{i2} = \sum_{i \in \Delta_2} r_i V_i c_{i2} \tag{A.41}$$

$$B_2 = \sum_i (1 - r_i) V_i c_{i2} = \sum_{i \in \Delta_2} (1 - r_i) V_i c_{i2} \tag{A.42}$$

Using exactly analogous reasoning as above, we can write an expression similar to (A.36) for each $i \in \Delta_2$:

$$\begin{aligned} (1 - x_1)(1 - x_2^{\alpha_2})(r_i V_i c_{i2} A_1 B_2 - (1 - r_i) V_i c_{i2} A_1 A_2) &+ \\ (1 - x_1^{\alpha_1})(1 - x_2)((1 - r_i) V_i c_{i2} A_1 A_2 - r_i V_i c_{i2} A_2 B_1) &+ \\ (1 - x_1^{\alpha_1})(1 - x_2^{\alpha_2})((1 - r_i) V_i c_{i2} A_1 B_2 - (1 - r_i) V_i c_{i2} A_2 B_1) &\leq 0 \end{aligned} \tag{A.43}$$

Note that the direction of the inequality in (A.43) is opposite that of (A.36), since the publishers weakly prefer ad network 2 rather than 1.

Summing over $i \in \Delta_2$ now gives:

$$\begin{aligned} 0 \geq\ & (1 - x_1)(1 - x_2^{\alpha_2})(A_2 A_1 B_2 - B_2 A_1 A_2) \\ +\ & (1 - x_1^{\alpha_1})(1 - x_2)(B_2 A_1 A_2 - A_2 A_2 B_1) \\ +\ & (1 - x_1^{\alpha_1})(1 - x_2^{\alpha_2})(B_2 A_1 B_2 - B_2 A_2 B_1) \\ =\ & (1 - x_1^{\alpha_1})((1 - x_2) A_2 + (1 - x_2^{\alpha_2}) B_2)(A_1 B_2 - B_1 A_2) \end{aligned} \tag{A.44}$$

But (A.44) implies:

$$(A_1 B_2 - B_1 A_2) \leq 0 \tag{A.45}$$

which is a contradiction, due to (A.40). We conclude that $c_{22} > 0$ is impossible, which means $c_{22} = 0$.

□

Repeated application of Lemma 2 will yield Lemma 4, which is at the heart of proving Theorem 1. We will also need Lemma 3 in the proof of Lemma 4:

**Lemma 3.** *$\exists(j, n)$ such that $X_{ij} = X_{in}\ \forall i$ if and only if $r_i = r\ \forall i$ for some constant $r$.*

In words, Lemma 3 says that all publishers will be indifferent between ad networks $j$ and $n$ if and only if all publishers have exactly the same quality, $r$.

*Proof.* If $X_{ij} = X_{in}\ \forall i$, then it is easy to show that $\forall\ i$:

$$\frac{N_{ij}}{N_{in}} = \frac{(x_j^{\alpha_j} - x_j) r_i + 1 - x_j^{\alpha_j}}{x_n^{\alpha_n} - x_n) r_i + 1 - x_n^{\alpha_n}} = \psi \tag{A.46}$$

where $\psi$ is some constant. Therefore:

$$r_i = \frac{\psi(1 - x_n^{\alpha_n}) - (1 - x_j^{\alpha_j})}{(x_j^{\alpha_j} - x_j) - \psi(x_n^{\alpha_n} - x_n)} \equiv r \tag{A.47}$$

Conversely, if $r_i = r\ \forall i$, it is easy to show that $X_{ij} = X_{in}\ \forall(j, n)$.

□

By repeated application of Lemma 2, we arrive at the next lemma:

**Lemma 4.** *If* $\exists(i, j)$ *such that* $c_{ij} = 1$*, then* $c_{mn} = 0\ \forall m \neq i, n \neq j$.

In words, Lemma 4 says that if publisher $i$ sends ad network $j$ all of its traffic, then no other ad network will receive any traffic from any publishers.

*Proof.* As discussed earlier, since $(\mathbf{C}^*, \mathbf{V}^*)$ is a fixed point of $BR$, we know from equations (A.14) - (A.16) that:

- If $c_{ij} > 0$ then $X_{ij} \geq X_{in}\ \forall n \neq j$.
- If $X_{ij} > X_{in}\ \forall n \neq j$ then $c_{ij} = 1$.

WOLOG, assume $i = 1, j = 1$. We will show that if $c_{11} = 1$, then $c_{mn} = 0\ \forall m \neq 1, n \neq 1$. Since the choice of $(i, j)$ was arbitrary, we will obtain the desired result.

Suppose $c_{11} = 1$. There are two possible cases: 1) $X_{11} > X_{1n}\ \forall n \neq 1$; and 2) $X_{11} = X_{1n}$ for some $n \neq 1$.

1. $X_{11} > X_{1n}\ \forall n \neq 1$.

   Applying Lemma 2 with $i = j = m = 1$, we get for each $n \neq 1$ that $c_{pn} = 0\ \forall p$.

2. $X_{11} = X_{1n}$ for some $n \neq 1$.

   Let $\Delta \equiv \{n \mid n \neq 1,\ X_{1n} = X_{11}\}$. Since $c_{11} > 0$, we know $X_{11} \geq X_{1n}\ \forall n$. Therefore, $X_{11} > X_{1n}\ \forall n \notin \Delta$. Applying Lemma 2 with $i = j = m = 1$ to each $n \notin \Delta$, we get $c_{pn} = 0\ \forall p, n \notin \Delta$.

   Since $c_{11} > 0$ and $X_{11} = X_{1n}\ \forall n \in \Delta$, Lemma 2 implies that there do not exist $m \neq 1, n \in \Delta$ such that $c_{mn} > 0$ and $X_{mn} > X_{m1}$. Observe that if $c_{mn} = 0$ and $X_{mn} > X_{m1}$ for some $m \neq 1, n \in \Delta$, then there must be some other $q \notin \{n, 1\}$ such that $c_{mq} > 0$ and $X_{mq} \geq X_{mn} > X_{m1}$, since $c_{mn} = c_{m1} = 0$. Therefore, $X_{mn} \leq X_{m1}\ \forall m \neq 1, n \in \Delta$.

   Fix an $n \in \Delta$. By Lemma 2, if $\exists m$ such that $X_{m1} > X_{mn}$, then $c_{pn} = 0\ \forall p$. Therefore, if $\exists m$ such that $n \in \Delta$ and $c_{mn} > 0$, then $X_{i1} = X_{in}\ \forall i$. However, from Lemma 3, this would imply $r_i = r\ \forall i$ and some constant $r$, violating Assumption 1. Therefore $\forall m, n \in \Delta$ we have $c_{mn} = 0$. We conclude that $c_{mn} = 0\ \forall m, n \neq 1$.

In both Case 1 and Case 2, we get $c_{mn} = 0\ \forall m \neq 1, n \neq 1$, which is the desired result.

□

We can now prove Theorem 1, which says $\exists j^*$ such that $c_{ij^*} = 1\, \forall i$ and $c_{ij} = 0\, \forall i, j \neq j^*$.

*Proof.* The proof involves three simple steps:

1. $\exists (i, j^*)$ such that $c_{ij^*} = 1$.

   Otherwise, $\exists (j, n)$ such that $X_{mj} = X_{mn}\, \forall m$. From Lemma 3 and Assumption 1, this is impossible.

2. If $c_{ij^*} = 1$ then $c_{mn} = 0\, \forall m, n \neq j^*$.

   See Lemma 4.

3. Therefore, we conclude that $c_{ij^*} = 1\, \forall i$.

□

## A.3 Proof of Lemma 1

Let $\mathbf{x} = \{x_1, x_2, \ldots, x_J\}$ be the ad networks' decisions in the first stage, and let $(\mathbf{C}^*, \mathbf{V}^*)$ be a Nash equilibrium in the second stage, where $\mathbf{C}^* = \{c_{ij}\}$ denotes the publishers' allocations and $\mathbf{V}^* = \{v_{kj}\}$ are the advertisers' valuations in this equilibrium. From Theorem 1, we know that there exists some $j^* \in 1, \ldots, J$, such that $c_{ij^*} = 1\, \forall i$.

Recall that $\bar{\pi}_{ij}$ is publisher $i$'s profit in an equilibrium where all publishers choose ad network $j$. Setting $c_{ij} = 1\, \forall i$ in Equation (2.6), we get:

$$\bar{\pi}_{ij} = N_{ij} V_i \frac{1}{K} h \sum_k v_{kj} \tag{A.48}$$

Using Equations (2.16) and (2.20), which describe the advertisers' equilibrium valuations, we also have:

$$v_{kj} = \frac{y_k}{R_k} \frac{\beta \sum_i r_i V_i}{\sum_i N_{ij} r_i V_i} \tag{A.49}$$

Now, suppose publisher $i$ is high-quality and prefers network $j$. That is,

$$\bar{\pi}_{ij} \geq \bar{\pi}_{in}\, \forall n \neq j \tag{A.50}$$

Using (A.48) and (A.49) in (A.50) and simplifying, we get:

$$\left( \frac{1 - x_j}{1 - x_j^{\alpha_j}} - \frac{1 - x_n}{1 - x_n^{\alpha_n}} \right) (r_i - \bar{r}) \geq 0 \;\; \forall n \neq j \tag{A.51}$$

By definition, if publisher $i$ is high-quality, then $r_i - \bar{r} \geq 0$. Therefore, we conclude that:

$$\frac{1 - x_j}{1 - x_j^{\alpha_j}} \geq \frac{1 - x_n}{1 - x_n^{\alpha_n}} \;\; \forall n \neq j \tag{A.52}$$

## A.4 Proof of Theorem 2

From Lemma 1 and Assumption 2, (A.52) will hold for the ad network $j$ that is chosen in equilibrium.

We define the function $f : [0,1] \times (0,1] \to \mathbb{R}$ as follows:

$$f(x,\alpha) \equiv \begin{cases} \frac{1-x}{1-x^{\alpha}} & x < 1 \\ \frac{1}{\alpha} & x = 1 \end{cases} \tag{A.53}$$

Thus, (A.52) says that $c_{ij} = 1 \; \forall i$ if and only if $f(x_j, \alpha_j) \geq f(x_n, \alpha_n) \; \forall n \neq j$. The function $f(x,\alpha)$ has the following properties:

1. $f(x,\alpha)$ is continuous on $x \in [0,1)$.
2. By definition, $f(x,\alpha)$ is also continuous in the limit as $x \to 1$.

$$\lim_{x\to 1} f(x,\alpha) = \lim_{x\to 1} \frac{1-x}{1-x^{\alpha}} = \lim_{x\to 1} \frac{\frac{\partial}{\partial x}(1-x)}{\frac{\partial}{\partial x}(1-x^{\alpha})} = \frac{1}{\alpha} \equiv f(1,\alpha) \tag{A.54}$$

3. $\frac{\partial}{\partial x} f(x,\alpha) = \left(\frac{1}{1-x^{\alpha}}\right)^2 \left(x^{\alpha}(1-\alpha) + (\alpha x^{\alpha-1} - 1)\right) > 0$.

   $\frac{\partial}{\partial x} f(x,\alpha)$ is strictly positive because a) $\lim_{x\to 0} \frac{\partial}{\partial x} f(x,\alpha) = \infty$ b) $\lim_{x\to 1} \frac{\partial}{\partial x} f(x,\alpha) = \frac{1-\alpha}{2\alpha} > 0$ and c) $\frac{\partial^2}{\partial x^2} f(x,\alpha) < 0$. Showing b) and c) is straightforward, but requires some arithmetic.

4. $\frac{\partial}{\partial \alpha} f(x,\alpha) = \frac{1-x}{(1-x^{\alpha})^2} x^{\alpha} \ln x < 0$.

Suppose ad network 1 is (strictly) the most effective at detecting invalid clicks, and ad network 2 is (weakly) the second-most effective i.e., $\alpha_1 < \alpha_2 \leq \alpha_j \; \forall j > 2$. Since $\frac{\partial}{\partial x} f(x,\alpha) > 0$, we get:

$$1 \leq f(x_2,\alpha_2) \leq f(1,\alpha_2) = \frac{1}{\alpha_2} < \frac{1}{\alpha_1} = f(1,\alpha_1) \tag{A.55}$$

$f(x_j,\alpha_1)$ is continuous on $x_j \in [0,1]$ with $f(0,\alpha_1) = 1$ and $f(1,\alpha_1) = \frac{1}{\alpha_1}$. So, by the intermediate value theorem, $\exists x^*$ such that $0 < x^* < 1$ and $f(x^*,\alpha_1) = \frac{1}{\alpha_2}$. Since $\alpha_2 \geq \alpha_j \; \forall j > 2$, a similar statement also applies for any other ad network $j > 2$.

Therefore, as long as $x_1 > x^*$, ad network 1 can guarantee that $f(x_1,\alpha_1) \geq f(x_2,\alpha_2)$, and consequently that $c_{i1} = 1 \; \forall i$. It is therefore a dominant strategy for ad network 1 to choose $x_1 > x^*$ in the first-stage game.

Finally, it is easy to see that $x^* \to 1$ as $\alpha_2 - \alpha_2 \to 0$. For $x \geq x^*$:

$$\frac{1}{\alpha_1} \geq f(x,\alpha_1) \geq f(x^*,\alpha_1) = \frac{1}{\alpha_2} \tag{A.56}$$

As $\alpha_2 - \alpha_1 \to 0$, we also have $\frac{1}{\alpha_2} \to \frac{1}{\alpha_1}$. Therefore, by the squeeze theorem:

$$f(x^*,\alpha_1) \to \frac{1}{\alpha_1} = \lim_{x\to 1} f(x,\alpha_1) \tag{A.57}$$

which, by continuity, implies that $x^* \to 1$.

Observe that:

$$\frac{h}{1-h}\eta_j = \sum_i \bar{\pi}_{ij} = \frac{\beta h}{K}\left(\sum_i r_i V_i\right)\sum_k \frac{y_k}{R_k} \tag{A.58}$$

which is independent of $j$. That is, ad network 1's total profit the total profit across all publishers is the same irrespective of which equilibrium is chosen and how aggressive the ad networks are. However, it can be shown from (A.48) that:

$$\frac{\partial^2 \bar{\pi}_{ij}}{\partial x_j \partial r_i} > 0 \tag{A.59}$$

which means as $x^* \to 1$, higher-quality publishers get a larger fraction of the total profits.

## A.5 Derivation of PRICINGPOLICY

In this section, we show how to form a *geometric programming (GP) relaxation* of (3.25). GPs are log-convex [4], and can therefore be solved globally and efficiently. Instances of the relaxed problem are solved in each iteration of the `for` loop (lines 3 to 6) of PRICINGPOLICY.

Begin by rewriting problem (3.25) as follows:

$$\begin{array}{ll} \text{maximize} & \left(\sum_i A_{i1} c_{i1}\right)(1-h_1) \\ \text{subject to} & u_{ij} = V_i g_{ij} h_j y_j \;\; \forall (i,j) \\ & \sum_j c_{ij} u_{ij} = \max_j u_{ij} \;\; \forall i \\ & \sum_j c_{ij} = 1 \;\; \forall i \\ & y_j = \dfrac{(\sum_i A_{ij} c_{ij})}{(\sum_i V_i c_{ij} g_{ij})} \;\; \forall j \\ & 0 \le g_{i1}, h_1, c_{ij} \le 1 \;\; \forall (i,j) \end{array} \tag{A.60}$$

where we have defined $y_j \equiv \theta_j$, $u_{ij} \equiv X_{ij}$, and $A_{ij} \equiv V_i \beta_i^{\text{Pub}} \kappa_j \beta_j^{\text{Net}}$ for notational convenience. In (A.60), upper case quantities (i.e., $A_{ij}$ and $V_i$) are constants, whereas lower case quantities are variables (except $g_{ij}$ and $h_j$ for $j \neq 1$, since $\mathbf{G}_{-1}$ and $\mathbf{h}_{-1}$ are given as input). The fourth and fifth constraints in (3.25) have been combined into one.

**Replace max with softmax** Replace the second constraint in (A.60) by the following approximation:

$$c_{ij} = \frac{u_{ij}^{\alpha}}{\sum_n u_{in}^{\alpha}} \;\; \forall (i,j) \tag{A.61}$$

where $\alpha$ is some large positive (but finite) constant. We are replacing the "hard" maximum in (A.60) with a "softmax" approximation i.e., if $u_{ij} > u_{in}\ \forall n \neq j$, then $c_{ij}$ will be close to 1 (but not exactly 1). The softmax is continuous and differentiable whereas the maximum operator is not. Larger values of $\alpha$ yield better approximations, although choosing $\alpha$ too large may lead to numerical instability in practice.

**Introduce linear equalities** Define $w_{ij} \equiv A_{ij}c_{ij}$ and $p_j \equiv \sum_i w_{ij}$. Define $z_{ij} \equiv V_i c_{ij} g_{ij}$ and $q_j \equiv \sum_i z_{ij}$. The fourth constraint in (A.60) then becomes $y_j = \frac{p_j}{q_j}\ \forall j$. Define $d_1 \equiv 1 - h_1$. The objective function in (A.60) can then be expressed as $d_1 p_1$.

We can then rewrite (A.60) as:

$$
\begin{aligned}
\text{maximize} \quad & d_1 p_1 \\
\text{subject to} \quad & d_1 = 1 - h_1 \\
& u_{ij} = V_i g_{ij} h_j y_j \ \ \forall (i,j) \\
& c_{ij} = \frac{u_{ij}^{\alpha}}{\sum_m u_{im}^{\alpha}} \ \ \forall (i,j) && \text{(A.62)} \\
& \sum_j c_{ij} = 1 \ \ \forall i && \text{(A.63)} \\
& y_j = \frac{p_j}{q_j} \ \ \forall j \\
& \sum_i w_{ij} = p_j \ \ \forall j && \text{(A.64)} \\
& \sum_i z_{ij} = q_j \ \ \forall j && \text{(A.65)} \\
& w_{ij} = A_{ij} c_{ij} \ \ \forall (i,j) \\
& z_{ij} = V_i c_{ij} g_{ij} \ \ \forall (i,j) \\
& 0 \leq g_{i1}, h_1, c_{ij} \leq 1 \ \ \forall (i,j) && \text{(A.66)}
\end{aligned}
$$

The problem (A.66) is almost a GP, except for the linear equalities (A.63), (A.64) and (A.65), and the softmax equality (A.62), all of which are not GP-compatible (a constraint is said to be *GP-compatible* if it is permitted in a geometric program [4]). Due to (A.63), the softmax equality can simply be replaced by a less-than inequality. To deal with the linear equalities, we replace them with *local monomial approximations* that are GP-compatible, as suggested by [4].

**Aside: Local monomial approximations** Consider any nonnegative length-$L$ vector $\mathbf{x} \equiv \{x_i\}_{L\times 1}$ and a linear equality constraint $\sum_i x_i = 1$. The equality $\sum_i x_i = 1$ is equivalent to the following pair of inequalities:

$$\sum_i x_i \leq 1 \tag{A.67}$$
$$\sum_i x_i \geq 1 \tag{A.68}$$

The less-than constraint (A.67) is GP-compatible, whereas the greater-than constraint (A.68) is not. To make (A.68) GP-compatible, fix a vector $\mathbf{x}_0$ and define $m : \mathbb{R}^L \to \mathbb{R}$ as follows:

$$m(\mathbf{x};\mathbf{x}_0) \equiv d(\mathbf{x}_0) \prod_{i=1}^{L} x_i^{f_i(\mathbf{x}_0)} \tag{A.69}$$

where

$$f_i(\mathbf{x}) \equiv \frac{x_i}{\sum_{m=1}^{L} x_m}$$

and

$$d(\mathbf{x}) \equiv \left( \sum_{i=1}^{L} x_i \right) \prod_{i=1}^{L} x_i^{-f_i(\mathbf{x})}$$

The function $m(\mathbf{x};\mathbf{x}_0)$ is a good monomial approximation of $\sum_i x_i$ around the point $\mathbf{x}_0$ in the sense that $\log m(\mathbf{x};\mathbf{x}_0)$ is the first-order Taylor-series approximation of $\log \sum_i x_i$ about the point $\mathbf{x}_0$. Moreover,

$$m(\mathbf{x};\mathbf{x}_0) \leq \sum_i x_i \ \forall \mathbf{x} \tag{A.70}$$

with equality iff $\mathbf{x} = \mathbf{x}_0$, which means $m$ is a global under-approximation of $\sum_i x_i$.

Monomial greater-than inequalities are GP-compatible. Therefore, to approximate a a linear equality constraint $\sum_i x_i = 1$ in a GP, we replace the equality with the pair of inequalities:

$$\sum_i x_i \leq 1 + \epsilon \tag{A.71}$$
$$m(\mathbf{x};\mathbf{x}_0) \geq 1 - \epsilon \tag{A.72}$$

The slack parameter $\epsilon > 0$ restricts our search space to a small $\epsilon$-neighbourhood around $\mathbf{x}_0$, and ensures a nonsingleton feasible set i.e., with $\epsilon = 0$, the only feasible solution to the inequalities (A.71) and (A.72) would be $\mathbf{x} = \mathbf{x}_0$.

**Introduce monomial approximations** Given $\left(h_1^{(t)}, \mathbf{g}_1^{(t)}, \mathbf{C}^{(t)}\right)$, $(\mathbf{G}_{-1}, \mathbf{h}_{-1})$, $\epsilon > 0$ and $\alpha \gg 0$, define:

$$\mathbf{w}_j^{(t)} \equiv \{w_{ij}^{(t)}\}_{I\times 1} \qquad \text{where } w_{ij}^{(t)} \equiv A_{ij} c_{ij}^{(t)} \tag{A.73}$$

$$\mathbf{z}_j^{(t)} \equiv \{z_{ij}^{(t)}\}_{I\times 1} \qquad \text{where } z_{ij}^{(t)} \equiv \begin{cases} V_i c_{i1}^{(t)} g_{i1}^{(t)} & j = 1 \\ V_i c_{ij}^{(t)} g_{ij} & j \neq 1 \end{cases} \tag{A.74}$$

Also define $\mathbf{c}_i^{(t)} \equiv \{c_{ij}^{(t)}\}_{J\times 1}$. We can now replace each linear equality in (A.66) with a monomial approximation about the point $\left(h_1^{(t)}, \mathbf{g}_1^{(t)}, \mathbf{C}^{(t)}\right)$:

$$
\begin{aligned}
\text{maximize} \quad & d_1 p_1 \\
\text{subject to} \quad & d_1 \leq 1 - h_1 \\
& u_{ij} = V_i g_{ij} h_j y_j \;\; \forall (i,j) \\
& c_{ij} \leq (1+\epsilon) \frac{u_{ij}^{\alpha}}{\sum_m u_{im}^{\alpha}} \;\; \forall (i,j) \\
& 1 - \epsilon \leq m\left(\mathbf{c}_i; \mathbf{c}_i^{(t)}\right) \;\; \forall i \\
& \sum_j c_{ij} \leq 1 + \epsilon \;\; \forall i \\
& y_j = \frac{p_j}{q_j} \;\; \forall j \\
& (1-\epsilon) p_j \leq m\left(\mathbf{w}_j; \mathbf{w}_j^{(t)}\right) \;\; \forall j \\
& \sum_i w_{ij} \leq (1+\epsilon) p_j \;\; \forall j \\
& (1-\epsilon) q_j \leq m\left(\mathbf{z}_j; \mathbf{z}_j^{(t)}\right) \;\; \forall j \\
& \sum_i z_{ij} \leq (1+\epsilon) q_j \;\; \forall j \\
& w_{ij} = A_{ij} c_{ij} \;\; \forall (i,j) \\
& z_{ij} = V_i c_{ij} g_{ij} \;\; \forall (i,j) \\
& g_{i1}, h_1, c_{ij} \leq 1 \;\; \forall (i,j)
\end{aligned}
\tag{A.75}
$$

Problem (A.75) is a relaxed version of (A.66), and it is a GP. We have eliminated all the nonnegativity constraints from (A.66) since such constraints are implicit in any GP. The constraint $d_1 = 1 - h_1$ has been replaced by a less-than inequality – this constraint will be tight at the optimum since we are maximizing $d_1 p_1$. The slack parameter $\epsilon$ allows solutions of (A.75) to be infeasible for the original problem (A.60). That is why we recompute $\mathbf{C}^{(T)}$ in line 7 of PRICINGPOLICY.

To summarize, in each iteration $t$ of PRICINGPOLICY, the point $\left(h_1^{(t)}, \mathbf{g}_1^{(t)}, \mathbf{C}^{(t)}\right)$ is given as input. We then search in the $\epsilon$-vicinity of $\left(h_1^{(t)}, \mathbf{g}_1^{(t)}, \mathbf{C}^{(t)}\right)$ for a point that is more profitable for network 1. The optimum is then labeled $\left(h_1^{(t+1)}, \mathbf{g}_1^{(t+1)}, \mathbf{C}^{(t+1)}\right)$, and we iterate until convergence. Network 1's profit is monotonically increasing across iterations $t$, and is bounded above by $\eta_1^{\max}$, so convergence is guaranteed. Although there are far fewer variables in (A.60) than (A.75), the latter can be solved efficiently due to the log-convexity of GPs. Thus, the increased number of variables in (A.75) is acceptable.

## A.6 Derivation of TRAFFICQUALITY

We now describe the TRAFFICQUALITY algorithm for finding approximate solutions to (4.9).

Let $\mathbf{u}_{-1} \equiv \{u_j \; \forall j \neq 1\}$, $\mathbf{G}_{-1} \equiv \{\mathbf{g}_j \; \forall j \neq 1\}$ and $\mathbf{h}_{-1} \equiv \{h_j \; \forall j \neq 1\}$ denote the actions of network 1's competitors. Let $\mathbf{C} \equiv \{c_{ij} \; \forall (i,j)\}$ denote the publishers' allocations in the second stage, and let $\mathbf{c}_i$ be column $i$ of $\mathbf{C}$.

---

**Algorithm 2** TRAFFICQUALITY

---

**Require:** $\mathbf{u}_{-1}, \mathbf{G}_{-1}, \mathbf{h}_{-1}, T$

1: Select arbitrary initializations $u_1^{(1)}$, $\mathbf{g}_1^{(1)}$ and $h_1^{(1)}$.

2: Use fixed-point iteration to compute the second-stage equilibrium, $\mathbf{C}^{(1)}$, assuming other networks play $(\mathbf{u}_{-1}, \mathbf{G}_{-1}, \mathbf{h}_{-1})$ and network 1 plays $\left(u_1^{(1)}, \mathbf{g}_1^{(1)}, h_1^{(1)}\right)$. Note that $\left(u_1^{(1)}, h_1^{(1)}, \mathbf{g}_1^{(1)}, \mathbf{C}^{(1)}\right)$ is feasible for (4.9).

3: **for** $t \in 1, \ldots, T-1$ **do**

4: Solve a geometric program relaxation of (4.9) to find an optimal point $(u_1', \mathbf{g}_1', h_1', \mathbf{C}')$ that is $\epsilon$-"close to" $\left(u_1^{(t)}, \mathbf{g}_1^{(t)}, h_1^{(t)}, \mathbf{C}^{(t)}\right)$, assuming other networks play $(\mathbf{u}_{-1}, \mathbf{G}_{-1}, \mathbf{h}_{-1})$.

5: $\left(u_1^{(t+1)}, \mathbf{g}_1^{(t+1)}, h_1^{(t+1)}, \mathbf{C}^{(t+1)}\right) \leftarrow (u_1', \mathbf{g}_1', h_1', \mathbf{C}')$

6: **end for**

7: Use fixed-point iteration to recompute the second-stage equilibrium, $\mathbf{C}^{(T)}$, assuming other networks play $(\mathbf{u}_{-1}, \mathbf{G}_{-1}, \mathbf{h}_{-1})$ and network 1 plays $\left(u_1^{(T)}, \mathbf{g}_1^{(T)}, h_1^{(T)}\right)$. Note that $\left(u_1^{(T)}, h_1^{(T)}, \mathbf{g}_1^{(T)}, \mathbf{C}^{(T)}\right)$ is feasible for (4.9).

8: **return** $\left(u_1^{(T)}, \mathbf{g}_1^{(T)}, h_1^{(T)}, \mathbf{C}^{(T)}\right)$

---

Most of the computation in TRAFFICQUALITY is done inside the `for` loop (lines 3 - 6). In each iteration $t$ of the loop, we are given as input $(\mathbf{u}_{-1}, \mathbf{G}_{-1}, \mathbf{h}_{-1})$ and a point $\left(u_1^{(t)}, \mathbf{g}_1^{(t)}, h_1^{(t)}, \mathbf{C}^{(t)}\right)$. We first compute:

$$N_{i1}^{(t)} = u_1^{(t)} r_i + \left(u_1^{(t)}\right)^{\gamma_1} (1 - r_i) \; \forall i$$
$$N_{ij} = u_j r_i + u_j^{\gamma_j} (1 - r_i) \;\; \forall (i, j \neq 1)$$

We then define:

$$\mathbf{w}_j^{(t)} \equiv \{w_{ij}^{(t)}\} \qquad \text{where } \; w_{ij}^{(t)} \equiv r_i V_i \beta_i^{\text{Pub}} \kappa_j \beta_j^{\text{Net}} c_{ij}^{(t)}$$
$$\mathbf{z}_j^{(t)} \equiv \{z_{ij}^{(t)}\} \qquad \text{where } \; z_{ij}^{(t)} \equiv \begin{cases} N_{i1}^{(t)} V_i c_{i1}^{(t)} g_{i1}^{(t)} & j = 1 \\ N_{ij} V_i c_{ij}^{(t)} g_{ij} & j \neq 1 \end{cases}$$

Using these quantities, we solve the following geometric program (GP) [4], which is a relaxation of (4.9) around the input point $\left(u_1^{(t)}, \mathbf{g}_1^{(t)}, h_1^{(t)}, \mathbf{C}^{(t)}\right)$:

$$\begin{aligned}
\text{maximize} \quad & d_1 p_1 \\
\text{subject to} \quad & d_1 \leq 1 - h_1 \\
& X_{ij} = N_{ij} V_i g_{ij} h_j \theta_j \;\; \forall (i,j) \\
& c_{ij} \leq (1+\epsilon) \frac{X_{ij}^{\alpha}}{\sum_m X_{im}^{\alpha}} \;\; \forall (i,j) \\
& 1 - \epsilon \leq m\left(\mathbf{c}_i; \mathbf{c}_i^{(t)}\right) \;\; \forall i \\
& \sum_j c_{ij} \leq 1 + \epsilon \;\; \forall i \\
& \theta_j = \frac{p_j}{q_j} \;\; \forall j \\
& (1-\epsilon) p_j \leq m\left(\mathbf{w}_j; \mathbf{w}_j^{(t)}\right) \;\; \forall j \\
& \sum_i w_{ij} \leq (1+\epsilon) p_j \;\; \forall j \\
& (1-\epsilon) q_j \leq m\left(\mathbf{z}_j; \mathbf{z}_j^{(t)}\right) \;\; \forall j \\
& \sum_i z_{ij} \leq (1+\epsilon) q_j \;\; \forall j \\
& w_{ij} = r_i V_i \beta_i^{\text{Pub}} \kappa_j \beta_j^{\text{Net}} c_{ij} \;\; \forall (i,j) \\
& u_1 r_i + u_1^{\gamma_1} (1 - r_i) \leq (1+\epsilon) N_{i1} \;\; \forall i \qquad \text{(A.76)} \\
& z_{ij} = N_{ij} V_i c_{ij} g_{ij} \;\; \forall (i,j) \\
& (1-\epsilon) N_{i1} \leq m_u\left(u_1; u_1^{(t)}, r_i\right) \;\; \forall i \\
& u_1, g_{i1}, h_1, c_{ij} \leq 1 \;\; \forall (i,j) \qquad \text{(A.77)}
\end{aligned}$$

The derivation of the optimization problem (A.77) closely parallels the discussion in Chapter 3 – the key difference is (A.76), which allows the network to optimize over filtering aggressiveness, $u_j$. In (A.77), $\epsilon > 0$ and $\alpha \gg 0$ are fixed parameters. Given a vector $\mathbf{x}^{(t)} \equiv \{x_i^{(t)}\}$ and a generic point $\mathbf{x} \equiv \{x_i \; \forall i\}$, the function $m(\mathbf{x}; \mathbf{x}^{(t)})$ is the best monomial approximation of $\sum_i x_i$ about the point $\mathbf{x}^{(t)}$. Similarly, $m_u(u; u^{(t)}, r)$ is a monomial approximation of $ur + u^{\gamma_1}(1-r)$ about the point $(u^{(t)}, r)$. To solve the GP (A.75) in our experiments, we used CVX, a package for solving convex programs [18].

# Bibliography


[1] Gagan Aggarwal, Ashish Goel, and Rajeev Motwani. Truthful auctions for pricing search keywords. In *ACM Conference on Electronic Commerce (EC06)*, June 2006.

[2] George Akerlof. The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500, 1970.

[3] Donna Bogatin. Google CEO on click fraud: ’Let it happen’ is perfect economic solution. *ZDNet*, July 2006.

[4] Stephen Boyd, Seung-Jean Kim, Lieven Vandenberghe, and Arash Hassibi. A tutorial on geometric programming. *Optimization and Engineering*, 8(1):67–127, 2007.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, USA, 2004.

[6] Interactive Advertising Bureau. Internet advertising revenues hit $5.9 billion in q1 ́10. *IAB Press Release*, May 2010.

[7] Gerard Cachon and Martin Lariviere. Supply chain coordination with revenue-sharing contracts: Strengths and limitations. *Management Science*, 51(1):30–44, January 2005.

[8] Ali Dasdan, Santanu Kolay, Panagiotis Papadimitriou, and Hector Garcia-Molina. Output bidding: A new search advertising model complementary to keyword bidding. In *Proceedings of the Fifth Workshop on Ad Auctions*, June 2009.

[9] Neil Daswani, Chris Mysen, Vinay Rao, Stephen Weis, Kourosh Gharachorloo, and Shuman Ghosemajumder. *Crimeware: Understanding New Attacks and Defenses*, chapter Online Advertising Fraud. Addison-Wesley Professional, April 2008.

[10] Neil Daswani and Michael Stoppelman. The anatomy of clickbot.A. In *Hot Topics in Understanding Botnets (HotBots)*. Usenix, April 2007.

[11] Benjamin Edelman. CPA advertising fraud: Forced clicks and invisible windows. *www.benedelman.org*, October 2008.

[12] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, March 2007.

[13] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.

[14] Mona Gandhi, Markus Jakobsson, and Jacob Ratkiewicz. Badvertisements: Stealthy click-fraud with unwitting accessories. *Journal of Digital Forensic Practice*, 1(2):131–142, 2006.

[15] Anindya Ghose and Sha Yang. An empirical analysis of sponsored search performance in search engine advertising. In *ACM Conference on Web Search and Data Mining (WSDM)*, 2008.

[16] Google. The facts about smart pricing. *Google AdSense Blog*, October 2005.

[17] Google. "Let click fraud happen"? Uh, no. *Google Blog*, July 2006.

[18] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming (web page and software), June 2008. `http://stanford.edu/~boyd/cvx`.

[19] Michael Grant and Stephen Boyd. *Recent Advances in Learning and Control (a tribute to M. Vidyasagar)*, chapter Graph Implementations for Nonsmooth Convex Programs. Springer, 2008.

[20] Grizzly. Is adsense smart pricing killing your earnings? *Make Money With AdSense*, Feb 2009.

[21] Nicole Immorlica, Kamal Jain, Mohammad Mahdian, and Kunal Talwar. Click fraud resistant methods for learning click-through rates. In *Internet and Network Econonomics*, volume 3828 of *Lecture Notes in Computer Science*, pages 34–45. Springer, November 2005.

[22] Brendan Kitts and Benjamin Leblanc. Optimal bidding on keyword auctions. *Electronic Markets*, 14(3):186–201, 2004.

[23] Mohammad Mahdian and Kerem Tomak. Pay-per-action model for online advertising. In *Internet and Network Econonomics*, volume 4858 of *Lecture Notes in Computer Science*, pages 549–557. Springer, 2007.

[24] Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. *Journal of the ACM*, 54(5), 2007.

[25] Bob Mungamuru and Hector Garcia-Molina. Predictive pricing and revenue sharing. In *Internet and Network Econonomics*, volume 5385 of *Lecture Notes in Computer Science*, pages 53–60. Springer, November 2008.

[26] Bob Mungamuru and Stephen Weis. Competition and fraud in online advertising markets. In *Financial Cryptography and Data Security*, volume 5143 of *Lecture Notes in Computer Science*, pages 187–191. Springer, January 2008.

[27] Hamid Nazerzadeh, Amin Saberi, and Rakesh Vohra. Dynamic CPA mechanisms and applications to online advertising. In *ACM Conf. on the World Wide Web*, April 2008.

[28] Adam L. Penenberg. Click fraud threatens web. *Wired News*, October 2004.

[29] Eric Savitz. Eric Schmidt: Click fraud is self-correcting. Mark Cuban: No, it isn't. *Barron's*, July 2006.

[30] Bruce Schneier. Google's click-fraud crackdown. *Wired News*, July 2006.

[31] Stephen Shankland. Yahoo to try using Google for search ads. *CNET News*, April 2008.

[32] TechCrunch. Facebook click fraud enraging advertisers. *TechCrunch*, July 2009.

[33] Hal Varian. Position auctions. *International Journal of Industrial Organization*, October 2006.

[34] Yahoo! A new pricing model rolls out today. *Yahoo! Search Marketing Blog*, June 2007.