DOI: https://doi.org/10.35065/ZHOI8576

Revenue Management in Online Advertising

A Dissertation
Submitted to London Business School
in Candidacy for the Degree of
Doctor of Philosophy

by
Sami Najafi Asadolahi

Dissertation Director: Dr. Kristin Fridgeirsdottir

December 2010

Abstract

Revenue Management in Online Advertising

Sami Najafi Asadolahi

2011

Online advertising is a multibillion-dollar business with a promising revenue increase for the coming years. Web publishers that generate revenues from online advertising face several challenging decisions. They need to decide on how many advertising slots to have on their website, whether to hire a sales force to attract advertisers to post ads on their website or rely on advertising networks, how many impressions to promise to deliver, and how much to charge, etc. Revenue management, in particular pricing, is considered one of the most challenging tasks and currently ad-hoc approaches are frequently used. In this dissertation, we provide systematic approaches for managing revenues in online display advertising.

In the first chapter, we consider a web publisher facing uncertain demand from advertisers requesting space on its website, and an uncertain supply of impressions from viewers visiting the website. Formulating the problem as a novel queuing system we show, for example, that the optimal cost-per-impression (CPM) can increase in the number of ads rotated in a slot, which goes against the intuition of supply and demand. In the second chapter, we consider a different pricing scheme, the so-called cost-per-click scheme. Formulating the problem as another novel queuing system, we show that the general heuristic applied by practitioners to convert between the CPC and CPM pricing schemes using the so-called click-through rate (CTR), can be misleading. In the third chapter, we explore the interactions of two web publishers in a competitive setting and provide various interesting insights about their strategic

pricing behavior at equilibrium. Lastly, In the fourth chapter, we obtain the optimality conditions for the advertisers' demand process when the demand distribution, instead of being Poisson, follows an arbitrary continuous distribution.

Contents

List of Figures

List of Tables

Dedication

This work is dedicated to my mother, Diana Lorita Darbin, who never gave up on me, and to the many others, though unnamed, who helped me in the completion of this task.

Acknowledgement

From the initial stages of this thesis, to the final draft, I owe an immense debt of gratitude to my supervisor, Professor Kristin Fridgeirsdottir for her careful and invaluable guidance throughout the stages of this thesis.

To her, I extend my deepest appreciations.

Introduction

The research conducted in this dissertation lies in the interface of operations, marketing, information technology, and economics. Much of this work focuses on providing new tools for understanding and resolving complex operational challenges faced by web publishers in online advertising, using a variety of management science techniques such as stochastic models, game theory, and optimization methods. These sophisticated operational tools assist web publishers in pricing and the revenue management of display ads while operating in risky and uncertain environments. The models discussed in this Ph.D. dissertation reflect, as extensively as possible, the reality of online display advertising by considering some of the most important aspects of a web publisher's advertising operations. The research in this thesis is one of the very first steps to bridge the gap between much of the academic literature on pricing in online advertising, which mainly focuses on deterministic models and the much more complex online advertising settings encountered in practice. The models developed and discussed in this Ph.D. dissertation provide significant contributions to the currently developing management science literature on online advertising, and help to distance from the commonly made assumptions of deterministic systems in the marketing literature. Beyond all, some of the introduced models in various sections of this thesis (particularly, those introduced in Chapters 1 and 2) can serve as decision making tools for web publishers running advertising operations, for instance, by providing extra layers of intelligence on top of their pricing engine software.

Any web publisher that generates revenues from display advertising faces several challenging decisions. They need to decide how many advertising slots to have on its website, whether to hire a sales force to attract advertisers to post ads on its website or rely on advertising networks, how many impressions to promise to deliver, and how much to charge, etc. Revenue management, in particular pricing, is considered to be one of the most challenging decisions that publishers face and currently ad-hoc approaches are frequently used. In this Ph.D. dissertation, we provide systematic approaches to bridge this pricing gap. The chapters that constitute the remainder of this dissertation deal with four different aspects of this problem, all of which are of significant importance to the online advertising industry.

In Chapter 1, we consider a web publisher that generates revenues from displaying advertisements on its website and charges according to a cost-per-impression (CPM) pricing scheme. The advertisers request their ad to be displayed to a certain number of visitors to the website. We focus on the main operational challenge of matching uncertain demand from advertisers requesting advertising space, to uncertain supply from viewers. The publisher faces challenging decisions of determining the price per impression, number of advertising slots, number of advertisements that share each advertising slot and others. Our stylized model is a new queuing system with no waiting space (loss system), where advertising slots correspond to servers. What sets this novel system apart from known multi-server queuing systems is its service mechanism; the advertising slots act as synchronized servers. We derive a closed-form solution for the system's steady-state probabilities and determine the optimal price. Using this solution, we analyze the publisher's optimal decisions and for example show that the optimal price increases in the number of impressions made of each ad, which goes against the quantity-discount commonly offered in practice. In addition, we conduct an empirical analysis of advertisers' and viewer's arrival processes at a large Scandinavian web publisher and link its implications to our model's assumptions. Finally, using extensive simulations for more general operational settings, we demonstrate that the proposed results are promising.

In Chapter 2, we consider a different pricing scheme, the so-called cost-per-click (CPC). We formulate this problem using a different queuing system, where the slots correspond to serving channels. The resulting queuing system is quite complex and different from the system considered in Chapter 1, owing to its multidimensional state space and the fact that the service rate of each server has an inverse relation to the number of active servers. We derive the closed-form solution for the steady-state probabilities of the number of advertisers in this system. Using this solution, we show that the behavior of the two pricing schemes at the optimal level can be considerably different. As described in Chapter 1, for instance, the optimal cost-per-impression (CPM) prices decrease in the number of advertising slots, while in Chapter 2, we show that the optimal CPC prices may increase with the number of slots. A more important result which we show in this chapter is that the common tendency among practitioners to convert the prices between the two schemes using the click-through rate (CTR) can be misleading.

In Chapter 3, we explore the interactions of two web publishers in a competitive setting and provide various interesting insights about their strategic behavior at equilibrium. We focus on the steady-state equilibriums (SSE), which tends to be significantly more reliable in the players' behavior predictions than the equilibriums obtained merely in a one-stage game. The reason for this is that by considering SSE, we study the strategic behavior of the publishers in the limit when the game is played many times. As a result, the players learn from the past and become more sophisticated decision makers. One of the insights that we demonstrate is that in the competition setting more web traffic (visitors) may not mean more revenue for a publisher, as a substantial traffic increase for one of the publishers may lead all publishers to lose profits as a result of the consequent price wars. In addition, we consider the zero-sum repeated competition of incomplete information on one side between the two web publishers. We find that the publisher having private information about the market can always guarantee reaching a higher payoff by adopting a proper partially revealing strategy.

Lastly, in Chapter 4, we obtain the optimality conditions for the advertisers' demand process when the demand follows an arbitrary continuous distribution rather than being Poisson. The results in this chapter go beyond online advertising as they are related to any setting based on customers' demand distributions. More specifically, in this chapter, we consider the optimality condition introduced by Gallego and van Ryzin (1994) for Poisson customers' demands with finite time horizon as well as the optimality condition introduced by Araman and Caldenty (2009) for Poisson demands with stopping (or infinite) time horizon. We extend these two demand optimality conditions from Poisson to an arbitrary continuous demand distribution.

Chapter 1

Revenue Management for a Web Publisher Using Advertising Networks

1.1 Introduction

The Internet has been a fast growing advertising medium. It provides access to a large consumer base and companies are constantly increasing the portion of their marketing budget allocated to online advertising (IAB 2010). Online advertising is a $23 billion business (IAB 2010). It can be divided into two domains: sponsored search advertising, involving advertisers paying a fee to appear next to search results for particular search words (e.g., Google) and display advertising where publishers display banner ads on their website (e.g., CNN.com). Sponsored search advertising involves well established payment procedures based on auctions, while pricing display ads lacks systematic approaches. In this chapter, we focus on display advertising.

Web publishers that generate revenues from display advertising face several challenging decisions. They need to decide on how many advertising slots to have on their website, whether to hire a sales force to attract advertisers to post ads on their website or rely on advertising networks, how many impressions to promise to deliver, and how much to charge, etc. Revenue management, in particular pricing, is considered to be one of the most challenging tasks and currently ad-hoc approaches are frequently used. In this chapter, we provide systematic approaches for managing revenues in online display advertising.

When modeling the advertising operation of a web publisher we focus on the main operational challenge of matching uncertain demand from advertisers requesting advertising space to uncertain supply from viewers visiting the website (often referred to as impressions). We consider a common setting for small and medium publishers where the publishers do not have their own sales force but use advertising networks to provide them with advertisers (see the Appendix for more details on ad networks). The advertisers are charged based on the "cost-per-impression" (CPM) pricing scheme. This setting captures around 25% of the $23 billion online advertising market (IAB 2010, Business Week 2009, and Media Banker 2009).

We consider the demand faced by the web publisher as "arrivals" of advertisers and the supply as "arrivals" of viewers. Advertisers complete their service when the agreed number of viewers has visited the website while the ad is displayed. A web publisher that uses ad networks is only visible to advertisers approaching the ad network when it has advertising slots available. Therefore, if all advertising slots are taken the web publisher does not have advertisers queueing up for their ad to be displayed. Similar dynamics occur with direct sale channels in the case when all advertising slots are full and advertisers are not willing to wait for a slot to become available. Our model captures both settings, but we will refer to the ad network setting for the remainder of the thesis.

The main contributions of this chapter are:

1. We construct a modeling framework capturing the main trade-offs in the operation of a web publisher dealing with an ad network that comes from matching supply with demand. We consider a general setting of multiple webpages, multiple types of ads (e.g. based on location and size) with different prices, and allow ads to share an advertising slot. This model can serve as a building block for studying more complicated operational issues of a web publisher such as competition for which we provide some initial but promising results. (See Sections 1.3 and 3.3.)

2. We derive a closed-form solution of the probability distribution of the number of advertisers in the system. This enables us to determine the optimal price for the web publisher to charge advertisers and analyze the publisher's system in detail. (See Sections 1.3 and 1.5.)

3. We show that the optimal price increases in the number of impressions. While this can be explained based on operational insights, all web publishers we approached offer either fixed prices or quantity discount, except for Yahoo! that recently started to charge a higher price per impression for contracts delivering a large number of impressions[1]. We provide further insights on how the price is affected by web traffic, number of ad slots, number advertisers that share a slot, and other decision factors and design parameters of the website. (See Sections 1.5 and 1.6).

[1] Confirmed by Prof. Preston McAfee, VP and Research Fellow at Yahoo!

4. We provide an analysis based on a real data set from a Scandinavian web publisher and use it to support our assumptions along with a simulation analysis. (Section 1.7.)

The chapter is organized as follows. In the next section, the relevant literature is reviewed. Section 1.3 presents the model developed for the web publisher's operation. The optimal price to charge advertisers is derived in Section 1.5. Section 1.6 provides numerical examples for further insights and Section 1.7 presents several extensions to the model including competition. Finally, we conclude in Section 1.8 and present directions for future research.

1.2 Literature Review

The literature on online advertising within the marketing area is quite extensive. Ha (2008) gives an overview of research on online advertising published in advertising journals and Evans (2008) summarizes the economics of the online advertising industry. Novak and Hoffman (2000) provide an overview of advertising pricing schemes for the internet. However, there is limited literature on analytical models for optimal pricing and other decision making for a web publisher with an advertising operation. (For issues faced by advertisers such as predicting audience for advertising campaigns see, e.g., Danaher (2007) and papers referenced therein.)

Research on online advertising within the operations research and operations management areas is fairly limited and there are few papers on pricing in online advertising. Mangàni (2003) compares the expected revenues from the cost-per-click (CPC) and the CPM schemes using a simple deterministic model. Unlike our paper, he does not consider the uncertainties involved with the advertisers' demand and viewers' supply. Chickering and Heckerman (2003) develop a delivery system that maximizes the click-through rate given inventory-management constraints in the form of advertisement quotas. Both of these papers assume the prices are fixed. Najafi-Asadolahi and Fridgeirsdottir (2010) focus on pricing for a CPC pricing scheme. The common misconception exists in the industry that CPC prices are simply CPM prices scaled by the click-through rate. This paper addresses that issue and shows that the simple scaling has flaws as the actual click-through rate depends on how many ads are on display. The paper develops a novel model for the CPC pricing scheme, which is different from the CPM model as the CPC system has a service rate that depends on the state of the system.

There has been some recent literature on online search and sponsored search advertising, the other section of the online advertising market. Johnson et al. (2004) conduct an empirical study to examine the dynamics of online search behavior. Ghose and Yang (2009) provide an empirical analysis of search engine advertising for sponsored searches on the internet. The nature of search advertising is fundamentally different from display advertising, as its pricing is mainly based on using auctions.

Some researchers have focused on the problem of pricing of goods and services on the internet. Brynjolfsson and Smith (2000) and Clemons et al. (2002) conduct empirical evaluations of price dispersions and price differentiations on the internet. Bakos and Brynjolfsson (1999, 2000) study the optimal strategies of product bundling for a retailer selling products through the internet. Dewan et al. (2000) and (2003) examine the problem of optimal product customization and price strategy both in monopoly and in competition. Jain and Kannan (2002) and Sundararajan (2004) analyze the optimal pricing of information goods. Although all of these papers consider a variety of online pricing problems, none are applicable to the web publisher's setting.

Sometimes web publishers do not only generate revenues from advertising but also from subscriptions. Baye and Morgan (2000) develop a simple economic model of online advertising and subscription fees. Prasad et al. (2003) model two offerings to viewers of a website: a lower fee with more ads and a higher fee with fewer ads. Kumar and Sethi (2008) study the problem of dynamically determining the subscription fee and the size of advertising space on a website. They use optimal control theory to solve the problem and obtain the optimal subscription fee and the optimal advertisement level over time. Unlike our paper, all these papers are focused on capacity management problems not pricing decisions with the price assumed to be fixed.

Scheduling the delivery of ads on a website has recently become a popular topic. Kumar et al. (2008) develop a model that determines how ads on a website should be scheduled in a planning horizon to maximize revenue. They consider geometry and display frequency as the two most important factors specifying the ads. Their problem belongs to the class of NP-hard problems and they develop a heuristic to solve it. They also provided a good overview of other related papers on scheduling. Our paper does not consider the details of scheduling individual ads rather we approach the problem at a higher level.

In this chapter of the dissertation, we develop a novel queuing system to characterize the web publisher's system. Relatively few papers in the queuing literature consider systems with similar characteristics. Green (1980), Brill and Green (1984), Courcoubetis and Reiman (1987), and Hong and Ott (1989) study systems with simultaneous service requirements with a concept of similar nature as the synchronization feature of the publisher's system that is discussed in Section 1.3. Nevertheless, the approaches in these papers do not prove useful when analyzing the publisher's system as the problem structures and the dynamics are significantly different.

We end this section by a short review of related work in revenue management. For a comprehensive reference of traditional revenue management models, we refer the reader to the book by Talluri and van Ryzin (2004a). However, the book does not cover the online setting. Savin et al. (2005) consider revenue management for rental businesses with two customer classes. Although considering a different problem, they have assumed uncertainty in the customers' demand in their model, which has some similarity to our model. Araman and Popescu (2009) also study revenue management for traditional media, specifically broadcasting. Their model is concerned with how to allocate limited advertising space between up-front contracts and the so-called scatter market (i.e., a spot market) in order to maximize profits and meet contractual commitments. Unlike our paper, both of these papers are mainly concerned with the capacity decisions.

We approach the web publisher's operation in a similar manner as Araman and Fridgeirsdottir (2010) with arrivals of advertisers and viewers. They focus on pricing and capacity management for a system where advertisers are willing to wait. This leads to intractable formulations. However, they solve a scaled version of their system and determine asymptotic optimal solutions and show that the fluid solution derived from a system with no uncertainties is asymptotically optimal. They provide insights on the impact of uncertainty on the pricing and capacity management. In contrast, this chapter of the dissertation derives a closed form solution of the web publisher's operation with advertisers that are not willing to wait or request advertising campaigns through advertising networks. This enables us to provide well supported managerial insights and analyze the system in detail. The well established

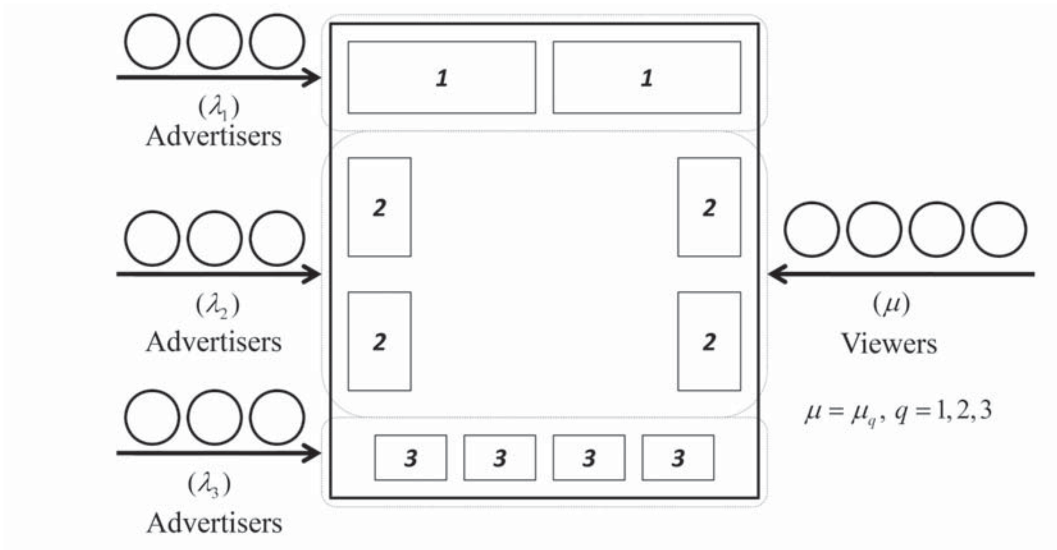


Figure 1.1: An illustrative example demonstrating a webpage with three different types of slots. Advertisers' arrivals tend to be independent as each advertiser determines why type of slot he looks for in advance.

characterization of the publisher's operation based on the closed-form solution can serve as a building block for more complex setting such as competition (see Section 3.3).

1.3 The Model

We consider a web publisher facing uncertain demand from advertisers requesting advertising space and uncertain supply of impressions from viewers[2]. The advertisers request their impressions through an ad network. The ad network supplies the web publisher with advertisers as long as the publisher has space available. If no space is available the network does not assign ads to that publisher. This implies that the publisher's website is a *loss* system (see the Appendix for details on ad networks). Our models also apply to the setting where direct sales channels are used

[2]The ad impressions are often referred to as ad inventory, i.e., the impressions are the items that satisfy the advertisers' demand.

with advertisers not willing to wait for space to become available.

A web publisher often charges different prices based on the size of the ad, the page on which the ad is posted, and the ad's allocated position on the page; e.g., the leaderboard (the horizontal banner at the top) on the homepage of a news site is more expensive than a small square at the bottom of the lifestyle page. Hence, when a web publisher registers with an ad network it classifies similar advertising slots that are charged the same price and registers each group with a separate tracking code.

We assume the web publisher's website (the system) contains J pages labeled from 1 to J. For example, for a news site these pages could correspond to the business page, travel page, etc. Each page can have several groups of ads that are priced equally. For instance (see Figure 1.1), the top of the page can display two equally sized ads. Likewise, similar ads can be positioned along the left and right sides of the page (skyscrapers), while several small ads can be placed at the bottom (rectangles). This leads to a total of three ad groups. More formally, for each page j we group the ads into M^j groups (the subsystems) of equivalent slots, where each subsystem m, $1 \leq m \leq M^j$, contains $n^{j,m}$ equivalent slots. (In Figure 1.1 we have $M^j = 3$, $n^{j,1} = 2$, $n^{j,2} = 4$, and $n^{j,3} = 4$.) We denote by $\lambda^{j,m}$ the rate with which the advertisers arrive requesting space in the subsystem (j, m). An advertiser requesting a slot in group m on page j requires his ad to be posted on the website until displayed $X^{j,m}$ times to viewers visiting the system. $X^{j,m}$ is a random variable. We denote the traffic rate of viewers to a page j by μ^j.

The publisher can often serve more advertisers than there are slots. For example, two ads could share the same slot with each ad displayed to every other viewer. We can experience this kind of rotation of ads into slots when we reload a webpage and see a new set of ads. Then if we continue reloading the page we come back to the first set of ads. We denote $s^{j,m}$ as the number of sets of ads being served in subsystem (j,m), i.e., we need to refresh the page $s^{j,m}$ times to see the same set of ads being displayed.

Furthermore, we note that in practice, the publisher does not usually leave a slot empty; rather it places a "default" ad in there. A default ad (or a filler ad) is often the publisher's own ad that does not generate any revenue. Then, when a revenue generating ad is sent to the publisher it would immediately free up this slot.[3]

The publisher's goal is to maximize its total revenue rate by determining the right prices to charge. The revenue rate for each subsystem consists of the payments made by an advertiser multiplied by the "effective" demand rate for that subsystem. Each payment consists of the price per impression, denoted by $p^{j,m}$, multiplied by the number of impressions requested, $X^{j,m}$. We capture the price-sensitivity of the advertisers with the price-demand function, $p^j_m(\lambda^{j,m})$, which is assumed to be continuous and decreasing in the arrival rate of the advertisers. (In Sections 1.7.1 and 1.7.4 we consider the price also to depend on the number of impressions.) Even though it might not be trivial for the publisher to determine this function, we assume it can do so with trial and error. (Ad networks often encourage publishers to start by offering low prices and then gradually increase them to the appropriate value.) The process of advertisers being matched to web publishers based on type preference and willingness-to-pay can be modeled specifically. However, ultimately it will lead to a price-demand relationship. We will not model the process in detail here but provide in the Appendix a description, from one of the ad networks, of the matching process.

[3] Sometimes the ad network provides the publisher with filler ads. We show in the Online Supplement that charging a (fixed and usually low) price for the filler ads does not affect our pricing results.

Note that an advertiser chooses his desired subsystem in advance when registering with the ad network. For instance, he may request a right hand side banner on the sport page. If that subsystem is fully occupied at the publisher's site then the network does not offer slots in this subsystem. Given that the publisher registers each subsystem separately with the ad network, we consider the demand for each subsystem to be independent.

Now, only a part of the advertisers' demand per time unit can usually be met by the publisher. That is, the demand rate for each subsystem is scaled down by the probability that there are advertising slots available. We denote the probability of having i advertisers in subsystem (j,m) by $\mathbb{P}_i^{j,m}$, $i \in \{0, ..., s^{j,m}n^{j,m}\}$. Note that a total of $s^{j,m}n^{j,m}$ advertisers can be served with $s^{j,m}$ advertisers sharing the same slot.

As we have a one-to-one relationship between the prices and the arrival rates of the advertisers, we will optimize the revenue rate with respect to the arrival rates and then determine the prices from the price-demand functions, $p_m^j(\lambda^{j,m})$. The optimization problem of the publisher of maximizing its expected revenue rate can be formulated as follows:

$$\max_{\mathbf{\Lambda}_1,...,\mathbf{\Lambda}_J} R(\mathbf{\Lambda}_1, ..., \mathbf{\Lambda}_J) = \sum_{j=1}^{J}\sum_{m=1}^{M^j} \lambda^{j,m}(1 - \mathbb{P}^{j,m}_{s^{j,m}n^{j,m}}(\lambda^{j,m}; X^{j,m}, n^{j,m}, s^{j,m}, \mu^j))p^{j,m}(\lambda^{j,m})\mathbb{E}(X^{j,m})$$

$$\mathbf{\Lambda}_j = \left(\lambda^{j,1}, ..., \lambda^{j,M^j}\right)^t \in [0, +\infty)^{M^j}, \quad j = 1, ..., J. \tag{1.1}$$

In this formula $\mathbb{P}^{j,m}_{s^{j,m}n^{j,m}}$ is the probability that the subsystem (j,m) is full. Therefore, $\lambda^{j,m}(1 - \mathbb{P}^{j,m}_{s^{j,m}n^{j,m}}(\lambda^{j,m}; X^{j,m}, n^{j,m}, s^{j,m}, \mu^j))$ is the effective advertisers' arrival rate to subsystem (j,m). (We are slightly abusing the notation by writing $\mathbb{P}^{j,m}_{s^{j,m}n^{j,m}}$ as a function of X.) Note that since the demand processes of the subsystems are independent and the publisher is considered a risk neutral decision maker, we can write the publisher's problem as the sum over all the subsystems. (More on this later.)

For the next sections we make the following assumptions. However, in Section 1.7 we illustrate how our results hold without them.

Assumption 1 *Advertisers' demand follows a Poisson process.* The demand for the publisher's slots belonging to the subsystem (j, m) comes through an ad network. We do not attempt to model the operation of the ad network rather assume that the publisher receives a certain rate of demand that depends on the price it offers for a certain number of impressions. For tractability, we assume that the demand for subsystem (j, m) is stationary and follows a Poisson process with rate $\lambda^{j,m}$. A Poisson assumption of this type is common in the service literature (see e.g. Savin et al. 2005). Our goodness-of-fit tests in Section 1.7.3 based on real data from a Scandinavian web publisher indicate that this is a restrictive assumption. However, our extensive simulation analysis in Section 1.7.3 illustrates that the Poisson assumption minimally affects the optimal revenues.

Assumption 2 *Advertisers are offered the same number of impressions.* We assume the web publisher offers a single number of impressions $x^{j,m}$, i.e., the ad will be shown to $x^{j,m}$ viewers. This assumption is restrictive, as the advertisers may choose to request different numbers of impressions. Nevertheless, in Sections 1.7.1 and 1.7.4 we consider different generalizations of this assumption and show through numerical simulations that even if the advertisers choose different numbers of impressions according to the random variable $X^{j,m}$ and charge a price depending on $X^{j,m}$, the problem can be well approximated by assuming that all advertisers request $x^{j,m} = \mathbb{E}(X^{j,m})$ with a single price charged. On a different note, some ad networks allow the advertisers to request a certain advertising campaign length instead of the number of impressions. We consider that case in Section 1.7.2.

Assumption 3 *Viewers' visits follow a Poisson process.* The viewers are assumed to visit a webpage containing a subsystem (j, m) according to a Poisson process with rate μ^j. This assumption could be considered restrictive as some research supports that web traffic shows self similarity, long range dependence and heavy tailed distribution (see Gong et al. 2005), which are not properties of the Poisson process. However, other studies recognize that a Poisson distribution is a reasonable assumption (see Cao et al. 2002). Our goodness-of-fit tests in Section 1.7.3 based on real data from a Scandinavian web publisher indicate that this is an appropriate assumption. Furthermore, our extensive simulation analysis in Section 1.7.3 illustrates that the Poisson assumption minimally affects the optimal revenues.

When an advertiser requests a particular type of slot that is available in one of the publisher's subsystem, the ad is displayed. More specifically, when a viewer arrives at that page all the advertisers *whose ads are displayed* are served together. That is, the remaining numbers of impressions for all the displayed ads decrease by one at the same time. We refer to this phenomenon as *synchronization* or the synchronized service. Synchronization differentiates the publisher's system from classic multi-server systems, where servers are independent.

If we consider the case with no rotation of ads into slots, i.e., $s^{j,m} = 1$, then it takes $x^{j,m}$ viewers with exponential interarrival times with rate μ^j to serve one advertiser in subsystem (j, m). Therefore, the service time of an advertiser follows

$Erlang(x^{j,m}, \mu^{j,m})$ distribution. The fact that the displayed slots operate in a *synchronized* manner makes this system different from the Erlang Loss System, denoted by $M/E_{x^{j,m}}/n^{j,m}/n^{j,m}$ in our context, where servers operate independently. Note that with only one slot (i.e. $n^{j,m} = 1$) there is no notion of *synchronized* servers and the web publisher's system is equivalent to the Erlang Loss System. For this case, the probability of having the subsystem (j, m) full is $\mathbb{P}_1 = \frac{r^{j,m}x^{j,m}}{1+r^{j,m}x^{j,m}}$ (see Gross and Harris (1998)) where $r^{j,m} = \frac{\lambda^{j,m}}{\mu^j}$.

Let us now consider the service structure for the case with rotation of ads, i.e., $s^{j,m} > 1$. We assume that each slot can display up to $s^{j,m}$ different ads, one at a time sequentially, i.e., up to $s^{j,m}$ ads can share the same slot. Hence, every time the webpage is loaded one of the ads sharing the slot is displayed to the viewer. Since each ad is only shown to every $s^{j,m}$ viewer, $s^{j,m}x$ viewers need to visit the website to complete the service for each ad. In addition, the ad might need to wait for its turn among other ads in the same slot. We denote by H the initial position of the ad among the ads that share the slot, with $1 \leq H \leq s^{j,m}$. For example, $H = 1$ means that the ad is going to be immediately displayed to the next viewer, while $H = h$ indicates that the ad is going to be displayed to the h^{th} viewer. Hence, a total of $s^{j,m}x + H - 1$ viewers need to visit the website to complete the service of an ad. As $H << s^{j,m}x$ we assume that the waiting time of starting display is negligible and the number of impressions needed to complete each advertiser's service can be well approximated with $s^{j,m}x$. As a result, the publisher's revenue function can be re-expressed as:

$$\max_{\mathbf{\Lambda}_1,...,\mathbf{\Lambda}_J} R(\mathbf{\Lambda}_1, ..., \mathbf{\Lambda}_J) = \sum_{j=1}^{J} \sum_{m=1}^{M^j} \lambda^{j,m}(1 - \mathbb{P}^{j,m}_{s^{j,m}n^{j,m}}(\lambda^{j,m}; s^{j,m}x^{j,m}, s^{j,m}n^{j,m}, \mu^j))p^{j,m}(\lambda^{j,m})x^j_m$$

$$\mathbf{\Lambda}_j = \left(\lambda^{j,1}, ..., \lambda^{j,M^j}\right)^t \in [0, +\infty)^{M^j}, \quad j = 1, ..., J. \tag{1.2}$$

Therefore, each subsystem m on page j, which has $n^{j,m}$ slots, delivers $x^{j,m}$ impressions, and rotates among $s^{j,m}$ ads, is approximately equivalent to a system with $s^{j,m}n^{j,m}$ slots, that delivers $s^{j,m}x^{j,m}$ impressions, without ad rotation.

Note that the revenue function in (1.2) is separable in the decision variables $\lambda^{j,m}$. The reason is that the advertiser specifies his target class (i.e., subsystem) in advance with the ad network. For instance, he may specify that his ad should be displayed in the first page of a sport website as a side rectangle. The network then matches this demand with the publisher's listed specifications[4]. This means that the demand processes for the subsystems are independent. Now, the service processes of the subsystems are dependent as the number of ads occupying each subsystem depends on the common arrival stream of viewers. However, as the publisher maximizes its *expected* revenue rate the objective function depends on the *sum of the expected* number of ads in each subsystem, which allows for separation of the revenues from each subsystem. That said, instead of maximizing the whole revenue function, the publisher simply maximizes each subsystem separately and for convenience we drop the indices (j, m):

$$\begin{aligned} \max_{\lambda}\ & R(\lambda) = \lambda(1 - \mathbb{P}_{sn}(\lambda; sx, sn, \mu))p(\lambda)x \\ & \lambda \in [0, +\infty). \end{aligned} \tag{1.3}$$

In order to solve the optimization problem above the full-state probability, $\mathbb{P}_{sn}(\lambda; sx, sn, \mu)$, should be characterized. We derive its closed-form solution in the next section.

[4]Note that most ad networks are blind, which means that the advertiser does not know on which website his ad will be placed.

1.3.1 The Probability Distribution

Having Markovian arrival and service processes we can now model a subsystem using Markov chains. For convenience, in this section we refer to this subsystem as a *system*. Without loss of generality we set $s = 1$ and consider later how the rotation affects our results. Note that even though we are ultimately interested in keeping track of the number of advertisers in the system, in order to set up a Markov chain we need to keep track of the system at a more detailed level; i.e., of the number of impressions left to be delivered for each slot. When an advertiser arrives, he is randomly assigned to one of the available slots with equal probability as they are equivalent. This random ad-to-slot allocation means that we can keep track of the dynamics of the system without distinguishing between the slots.

We formulate the problem as a queuing model with the state vector $\mathbf{k} = (k_1, ..., k_i, 0, ..., 0)$ with $1 \leq k_j \leq x$ for $j = 1, 2, ..., i \leq n$. This indicates that in the system, there is one slot with k_1 impressions remaining (i.e., impressions left to be satisfied), another slot with k_2 impressions remaining, etc. Then there are $n-i$ slots empty. As the slots are considered to be identical, we do not distinguish between them. Consequently, any rearrangement of vector $\mathbf{k}$'s components does not lead to a new state. For example, $(5, 2, 0, 7)$, $(7, 2, 5, 0)$, and $(0, 7, 5, 2)$, all refer to the same state of the system. In order to see the nature of the state transitions we consider the following example.

Consider the state of the system $\mathbf{k} = (k_1, ..., k_i, 0, ..., 0)$ with $1 \leq k_j \leq x$ for $j = 1, 2, ..., i \leq n$. When a viewer arrives at the system, since all the ads are displayed, the state of the system goes to $\mathbf{k}^- = (k_1^-, ..., k_i^-, 0, ..., 0)$ with rate μ, where $k_j^- = k_j - 1$, i.e., all the positive components' values reduce by one at the same time (the synchronization), while the zero components do not change. Note that one important difference between the publisher's system and the more traditional loss systems such as Erlang is that in those systems we do not need to define the *n-tuple* vector $\mathbf{k}$ to characterize the system as we need here. As a result, the characterization of the Erlang Loss System is significantly easier, which is due to the independence among the servers (or lack of synchronization). Now, when an advertiser arrives, the publisher assigns one of the empty slots to him with x impressions to be displayed. Hence, the state of the system will do a transition from $\mathbf{k}$ to $(k_1, ..., k_i, x, ..., 0)$, with rate λ.

In order to find $\pi_{\mathbf{k}}$, the probability of finding the system in state $\mathbf{k}$, we characterize all possible states and transitions of the system and solve the flow balance equations. The following proposition states the closed-form solution of the probability distribution of the web publisher's system. The proof of the proposition can be found in the Appendix A. All other proofs can be found in the Appendix B.

Proposition 1 *The probability of a web publisher's system with n slots being in state $\mathbf{k} = (k_1, ..., k_i, 0, ..., 0)$, where $k_1, k_2, ..., k_i$ impressions are left in i slots and $n - i$ are empty is:*

$$\pi_{\mathbf{k}}(r, x, n) = \frac{r^i (1+r)^{n-i-1}}{\sum\limits_{j=0}^{n} \binom{x+n-1}{j} r^j}, \ \ i < n, \tag{1.4}$$

$$\pi_{\mathbf{k}}(r, x, n) = \frac{r^n}{\sum\limits_{j=0}^{n} \binom{x+n-1}{j} r^j}, \ \ i = n, \tag{1.5}$$

where $r = \lambda/\mu$. Moreover, the steady-state probability of having i advertisers in the system is:

$$\mathbb{P}_i(r, x, n) = \frac{\binom{x+i-1}{i} r^i (1+r)^{n-i-1}}{\sum\limits_{j=0}^{n} \binom{x+n-1}{j} r^j}, \ \ i < n, \tag{1.6}$$

$$\mathbb{P}_n(r, x, n) = \frac{\binom{x+n-1}{n} r^n}{\sum\limits_{j=0}^{n} \binom{x+n-1}{j} r^j}, \ \ i = n. \tag{1.7}$$

Note that $\pi_{\mathbf{k}}(r, x, n)$ does not depend on the actual number of impressions left in each slot, it only depends on the number of filled slots.

Let us consider how the interaction between the empty and the occupied slots comes through in the publisher's system. In the formula for $\mathbb{P}_i(r, x, n)$, r^i plays the role of the i occupied slots while $(1+r)^{n-i-1}$ plays the role of the $n-i$ empty slots. The multiplication of those two terms captures the effect of the interaction of i occupied slots with $n - i$ empty slots. Since in $\mathbb{P}_n(r, x, n)$ all the n slots are occupied there is no interaction between the empty and the occupied slots. Therefore, $\mathbb{P}_n(r, x, n)$ does not have a term of the form $(1 + r)$.

With the proposition above we have fully characterized the probabilistic properties of the web publisher's system with a closed-form solution of the steady-state probabilities. Next we provide some structural properties for $\mathbb{P}_n(r,x,n)$, the probability that the system is full and

$$L(r,x,n) = \lambda(1-\mathbb{P}_n(r,x,n))x/\mu, \tag{1.8}$$

the average number of advertisers in the system (based on Little's law). Those are useful when proving properties of the optimal price for the web publisher to charge in the next section.

Proposition 2 *$\forall x,n,r$ the full-state probability of the system, $\mathbb{P}_n(r,x,n)$, defined by (1.7) satisfies*:

(i) $\frac{\partial \mathbb{P}_n(r,x,n)}{\partial r} \geq 0$,

(ii) $\mathbb{P}_n(r,x+1,n) - \mathbb{P}_n(r,x,n) \geq 0$,

(iii) $\mathbb{P}_{n+1}(r,x,n+1) \leq \mathbb{P}_n(r,x,n)$.

This proposition confirms the intuition that the web publisher is busier if there is more demand, less traffic, more impressions, and fewer slots. Numerical analysis indicates that $\mathbb{P}_n$ is not necessarily concave in the number of impressions.

Proposition 3 *$\forall x,n,r$ the average number of advertisers, $L(r,x,n)$, defined by (1.8) and its increment $\Delta L_x(r,x,n) = L(r,x+1,n) - L(r,x,n)$ satisfy:*

(i) $\Delta L_x(r,x,n) \geq 0$, $\Delta L_x(r,x+1,n) \leq \Delta L_x(r,x,n)$,

(ii) $\frac{\partial L(r,x,n)}{\partial r} \geq 0$, $\frac{\partial^2 L(r,x,n)}{\partial r^2} \leq 0$,

(iii) $L(r,x,n) \leq L(r,x,n+1)$.

Part (i) implies that the average number of advertisers in the web publisher's system is increasing and concave in the number of impressions. Hence, the publisher is busier the larger number of impressions it offers. However, the impact levels off. Part (ii) implies that the average number of advertisers in the system is increasing and concave in the intensity, r. Hence, the publisher is busier with more demand, less traffic, or higher demand-traffic ratio. However, the impact levels off. Finally, Part (iii) indicates that the average number of advertisers in the web publisher's system increases in the number of slots. More slots on the website imply that fewer advertisers are being rejected and more can be served.

1.4 Comparison With Known Queuing Models

As the synchronization of the publisher's advertising slots leads to a novel queueing model, in this section we briefly compare it to related models from a queueing theory stand point before we move to the pricing section[5].

1.4.1 Erlang's Loss System

We first compare the web publisher's model with the $M/E_x/n/n$ queue, the so-called Erlang's loss system. As in our system this system does not have any waiting space and the only jobs in the system are the ones being served by one of the n servers. The difference comes from the operation of the servers.

In Erlang's loss system the servers operate independently, while in our system the slots are synchronized, i.e., the advertisers receive service simultaneously. Erlang's

[5] The reader who is interested in pricing only can conveniently skip this part and move directly to the next section.

loss formula that represents the probability distribution of the number of jobs in the system is the following:

$$\mathbb{P}_i^E = \frac{\frac{(xr)^i}{n!}}{\sum_{j=0}^{n} \frac{(xr)^j}{j!}}, \; 0 \leq i \leq n,$$

which we can compare to the distribution for the web publisher's system:

$$\mathbb{P}_i = \frac{\binom{x+i-1}{i} r^i (1+r)^{n-i-1}}{\sum_{j=0}^{n} \binom{x+n-1}{j} r^j}, \; i < n,$$
$$\mathbb{P}_n = \frac{\binom{x+n-1}{n} r^n}{\sum_{j=0}^{n} \binom{x+n-1}{j} r^j}.$$

If $n = 1$ the two formulas yield the same results as expected.

As we discussed before r^i plays the role of the i occupied slots while $(1+r)^{n-i-1}$ plays the role of the $n-i$ empty slots. The multiplication of those two terms captures the effect of the interaction of i occupied slots with $n-i$ empty slots. Since in $\mathbb{P}_n$ all the n slots are occupied there is no interaction between the empty and the occupied slots. Therefore, $\mathbb{P}_n$ does not have a term of the form $(1+r)$. This is different from the $M/E_x/n/n$ model with its independent servers, where the formula for $\mathbb{P}_i$, $0 \leq i \leq n$, has the same format even though there are empty servers.

In the following proposition we compare the probability of the system being full for Erlang's loss system and the web publisher's system.

Proposition 4 *The probability of a fully occupied system is higher for the web publisher than for the Erlang's loss system, i.e.,* $\mathbb{P}_n \geq \mathbb{P}_n^E$. *In addition, the average number of jobs in the web publisher's system is less than the average number of jobs in the Erlang's loss system,* $L \leq L_E$.

This proposition shows that the online system is less efficient than Erlang's loss system with independent servers. This is intuitive as the synchronization of servers imposes a restriction compared to independence.

1.4.2 Bulk Service

A bulk service system, denoted $M/M^{[n]}/1$, has arrivals that are Poisson and the service time is exponential. There are n slots for service and an infinite waiting space. When n or less jobs are in the system they are all served at the same time and if a job arrives during the service and a slot is empty that job is also served and finishes at the same time as the others (memoryless service property). If there are more than n jobs in the system only n are served simultaneously and the rest wait.

This system with the additional assumption of no waiting space is the same as the online system with *one impression*. We can denote it by $M/M^{[n]}/1/n$. Since having one impression is not realistic for the online setting there does not seem to be much to gain for us from the bulk service literature. However, next we illustrate how we can use the results from the web publisher's system to learn more about the bulk service system. First, the solution of the system $M/M^{[n]}/1$ (Gross and Harris (1998)) is the following:

$$
\begin{aligned}
\mathbb{P}_0^B &= (1 - x_0), \\
\mathbb{P}_i^B &= (1 - x_0)x_0^i, \quad i = 1, 2, 3, ...,
\end{aligned}
$$

where x_0 is the unique solution (between zero and one) of the characteristic equation: $\mu x^{n+1} - (\lambda + \mu)x + \lambda = 0$ with λ as the customers arrival rate and μ as the service rate (corresponds to the arrival rate of viewers, λ_v, in the online setting). The

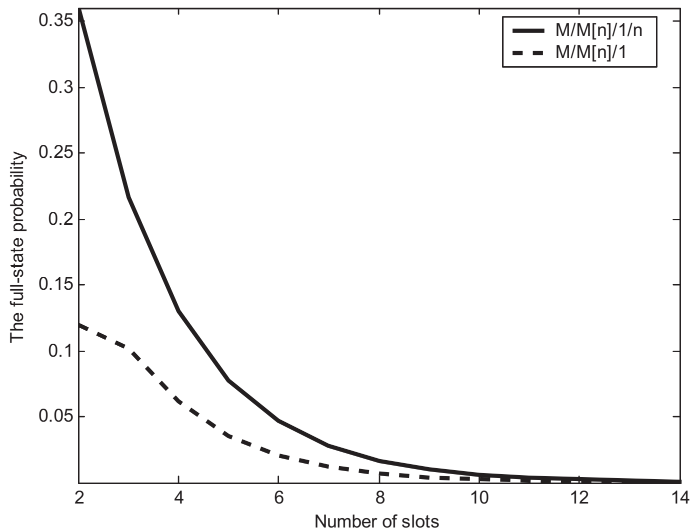


Figure 1.2: Comparison of the full-state probabilities

characteristic equation is of order $(n+1)$ and has at most $(n+1)$ roots but in most applications there is one real root. The drawback of this formula is when n is large or approaches infinity, i.e., the bulk service system has very large or infinite capacity, the characteristic equation will be hard to solve. However, in our model when $n \to \infty$ the assumption of having no waiting space does not play a role anymore and the result from the web publisher's model can be used to approximate the bulk service solution. This is formalized in the following proposition.

Proposition 5 *As the number of service slots in the bulk service system approach infinity, $n \to \infty$, the probability distribution of the number of jobs, $\mathbb{P}_i^B$ has the following property:*

$$\mathbb{P}_i^B \to \frac{r^i}{(1+r)^{i+1}} \quad for\ n \to \infty. \tag{1.9}$$

In addition, the average number and the variance of jobs in the system are $L^B = r$ and $Var^B = r(1+r)$, respectively, where $r = \lambda/\mu$.

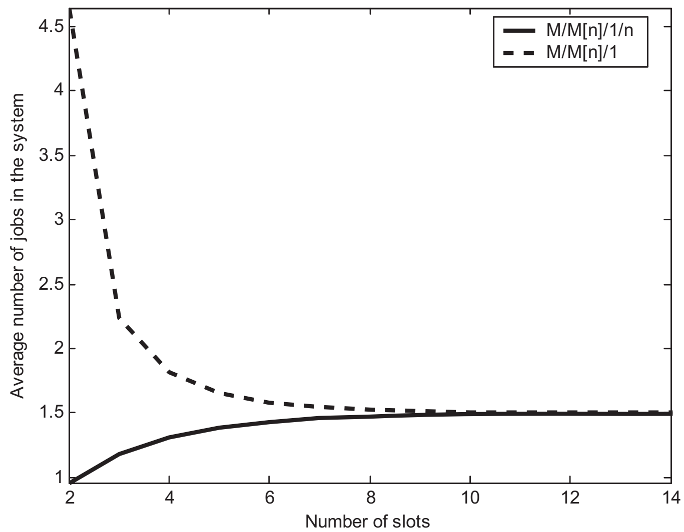


Figure 1.3: Comparison of the average number of jobs in the system

Let us explore this in a numerical example where we consider two systems $M/M^{[n]}/1$ and $M/M^{[n]}/1/n$ (which is the same as the online system with n slots and one impression) with $\lambda = 15$ and $\mu = 10$ and different values for n. When calculating accurately the average number of jobs in $M/M^{[n]}/1$ (using the characteristic equations above from Gross and Harris (1998)) and in $M/M^{[n]}/1/n$ (using Equations (1.7) and (1.8) with $x = 1$) we obtain a difference in L, the average number of jobs in the system, of less than 1.2% with $n \geq 10$. Figure 1.3 illustrates this difference.

Note that L for the $M/M^{[n]}/1$ system is higher as could be expected since there can be jobs waiting in a queue ready to go into service while the $M/M^{[n]}/1/n$ system needs to wait for the next arrival.

The full state probability is illustrated in Figure 1.2. The convergence of $\mathbb{P}_n$ is a bit slower and there is less than 1.4% difference for $n \geq 10$.

The bulk service system with Erlang service time (instead of exponential) is not the same as the web publisher's system. In the publisher's system the "jobs" can leave and enter the “bulk”; i.e., the jobs being served simultaneously are not necessarily in the same phase of Erlang distribution. However, in the bulk service system all jobs belonging to the same bulk have the same service time.

1.5 The Optimal Price

The web publisher's objective is to determine the price to charge per impression in order to maximize the revenue rate, which we can write as $R(\lambda) = \lambda(1-\mathbb{P}_n(\lambda, \mu, n, x))px = L(\lambda)\mu p$, based on Equation (1.8).

As we have a one-to-one relationship between the price and the arrival rate of the advertisers, λ, we will optimize the revenue rate with respect to λ and then determine the price from the price-demand function, $p(\lambda)$. The optimization problem of the web publisher can now be expressed as:

$$\max_{\lambda} R(\lambda) = \lambda(1 - \mathbb{P}_n(\lambda; \mu, n, x))p(\lambda)x. \tag{1.10}$$

$$\lambda \in [0, +\infty).$$

The following proposition ensures the existence of the optimal solution and gives an implicit equation for the optimal price.

Proposition 6 *If the price-demand function, $p(\lambda)$, is concave decreasing in the advertisers' arrival rate, λ, then $R(\lambda)$ is concave in λ. Furthermore, at the optimal advertisers' arrival rate, λ^*, the following condition is satisfied:*

$$\left.\frac{\partial L(\lambda)}{\partial \lambda}\right|_{\lambda^*} p(\lambda^*) + \left.\frac{\partial p(\lambda)}{\partial \lambda}\right|_{\lambda^*} L(\lambda^*) = 0. \tag{1.11}$$

Note that in order to ensure concavity of the objective function, $p(\lambda)$ needs to be concave. This includes a linear price, which is widely used in the economics literature. In Section 1.6 we show that even some convex pricing functions give a unimodal revenue function. (Other "weaker" conditions such as assuming concave payment rate $\lambda p(\lambda)$ or monotonicity of the price elasticity $-\frac{\partial \lambda}{dp}\frac{p}{\lambda}$ are not sufficient.) Furthermore, Equation (1.11) implies that at the optimal point, λ^*, the proportional change in the average number of advertisers, L, equals the negative proportional change in the optimal price, $p(\lambda^*)$.

The proposition below confirms the intuitive results that the web publisher increases its revenue by having more slots, offering higher numbers of impressions, and having more traffic to its website. We denote $R(\lambda^*)$ by $R_{n,x}(\lambda^*(n,x);\mu)$ to emphasize the dependence on n, x, and μ.

Proposition 7 *The optimal revenue rate, $R_{n,x}(\lambda^*(n,x);\mu)$, defined by (1.10) satisfies:*

(i) $R_{n,x}(\lambda^*(n,x);\mu) \leq R_{n+1,x}(\lambda^*(n+1,x);\mu)$,

(ii) $R_{n,x}(\lambda^*(n,x);\mu) \leq R_{n,x+1}(\lambda^*(n,x+1);\mu)$,

(iii) $R_{n,x}(\lambda^*(n,x);\mu^1) \leq R_{n,x}(\lambda^*(n,x);\mu^2),\ \ \mu^1 \leq \mu^2$.

Note that although some of the results of Propositions 2, 3, and 7 are intuitive, we will see in Section 3.3 that a few of them are overturned in competitive settings. For instance, part (ii) of Proposition 7 mentions that the optimal revenue increases with the number of slots in the publisher's system. However, we note that in the competition setting more slots may no longer mean more revenue. Furthermore, we observe a similar interesting behavior with respect to the web traffic μ indicating that in the competition setting more traffic to the publisher's system may not mean more revenue (see Section 3.3).

The following proposition states the counter-intuitive result from a marketing point of view that it is optimal to charge more per impression if the advertisers are offered a higher number of impressions.

Proposition 8 *If the price-demand function, $p(\lambda)$, is concave decreasing in the advertisers' arrival rate λ, then $p(\lambda^*)$ is increasing in x.*

This proposition is interesting as in practice web publisher's usually offer quantity discounts. However, from an operational point of view when more impressions are requested per advertiser, the advertisers provide more workload to the system and fewer advertisers are needed, which means a higher price can be charged. Practically speaking, the web publisher should not offer quantity discounts from an operational point of view. All publishers we had a conversation with offer quantity discounts except Yahoo!. Prof. Preston McAfee, a vice president and senior research fellow at Yahoo!, confirmed that they now increase the CPM price for large contracts instead of giving a discount. However, they did not have any theoretical underpinnings for doing so, rather they had come to this pricing approach through a series of trials and errors over time. We were pleased to offer a theoretical explanation. It is interesting to mention that our analysis of the competitive setting indicates that larger contracts impact not only the publisher offering them but also its competitor. Hence, both will charge a higher price if one offers more impressions.

We note that in the price-demand function, $p(\lambda)$, we have not yet considered the fact that advertisers might not be willing to pay as much for their ad to be posted on a website with many ads compared to a website with few. To capture this feature, we set the price to depend not only on the advertisers' arrival rate, λ, but also on the number of slots n. Proposition 9 describes the solution of the publisher's objective revenue function. We use the notation $\lambda^*(n)$ to emphasize the implicit dependence on n at the optimal value.

Proposition 9 *Let the price function $p(\lambda, n)$ be decreasing in the advertisers' arrival rate, λ, and the number of slots, n. In addition, let $n_c \in \mathbb{R}^+$ be the continuous version of n. Given the following property is satisfied:*

$$\frac{p(\lambda, n_c)}{L(\lambda, n_c)} \leq \min\left(-\frac{p''_{\lambda n_c}}{L''_{\lambda n_c}}, -\frac{p''_{\lambda\lambda}}{L''_{\lambda\lambda}}\right) \tag{1.12}$$

then:

(i) $\left.\frac{\partial L(\lambda,n)}{\partial\lambda}\right|_{\lambda^*} \geq 0$,

(ii) $\lambda^*(n+1) \geq \lambda^*(n)$,

(iii) $p(\lambda^*(n+1), n+1) \leq p(\lambda^*(n), n)$,

(iv) $L(\lambda^*(n+1), n+1) \geq L(\lambda^*(n), n)$.

Proposition 9, Part (iv) indicates that with an upper bound on the ratio of the price and the number of advertisers in the system, the publisher has more ads on display with a larger number of advertising slots, even though advertisers are discouraged by a large number of advertising slots.

1.6 Numerical Analysis

In Section 1.5, we derived the optimal price for the publisher to charge and several structural properties. In this section, we show numerically that those structural properties hold for more general price-demand functions than concave ones. Moreover, we provide further insights on how the number of slots on a website, the web traffic, the offered impressions, and the number of ads that share a slot affect the publisher's system.

1.6.1 Advertising Slots

We first explore the properties of the optimal revenue and prices with respect to the number of advertising slots. The viewers' arrival rate at the publisher's website is $\mu = 2,000$. Each advertiser is offered $x = 100,000$ impressions through the ad network. The price-demand relationship (per impression) for the advertisers is set $p(\lambda) = 0.02 - 0.2\lambda^c$, where $c = 0.8$, 1, or 1.2, i.e., the price function is convex, linear, or concave.

An extra slot on the website means additional capacity to serve advertisers. Hence, the publisher charges a lower price to attract more advertisers as indicated in Figure 1.5 and by doing so it increases the revenues (see Figure 1.4). However, this effect levels off as indicated for $c = 0.8$ in Figures 1.4 and 1.5.

As in Proposition 9 we set the price to depend not only on the arrival rate of advertisers, λ, but also the number of slots n. We consider the following price function:

$$p(\lambda, n) = 0.02 - 0.2\lambda^c - 0.001n.$$

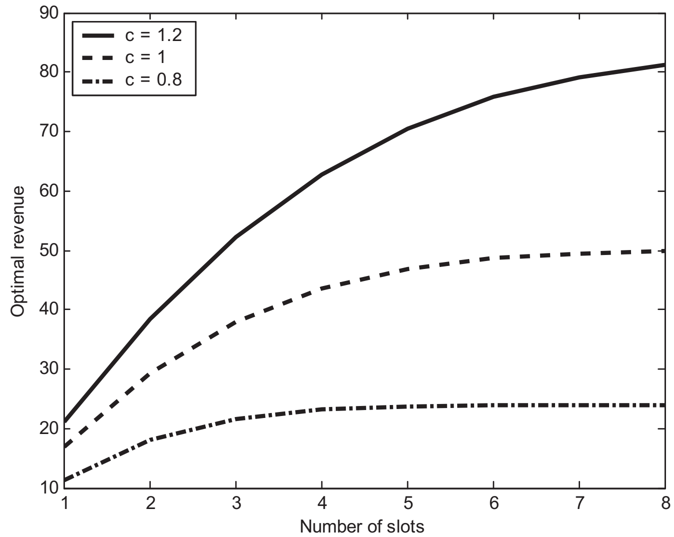


Figure 1.4: Optimal revenue vs. slots

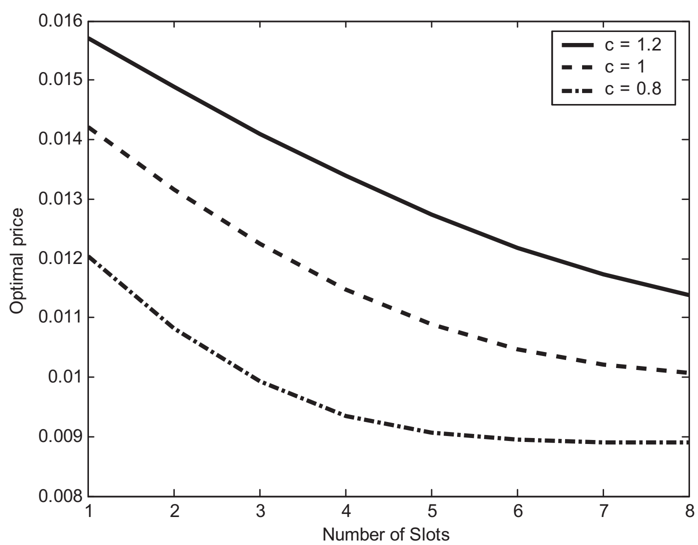


Figure 1.5: Optimal price vs. slots

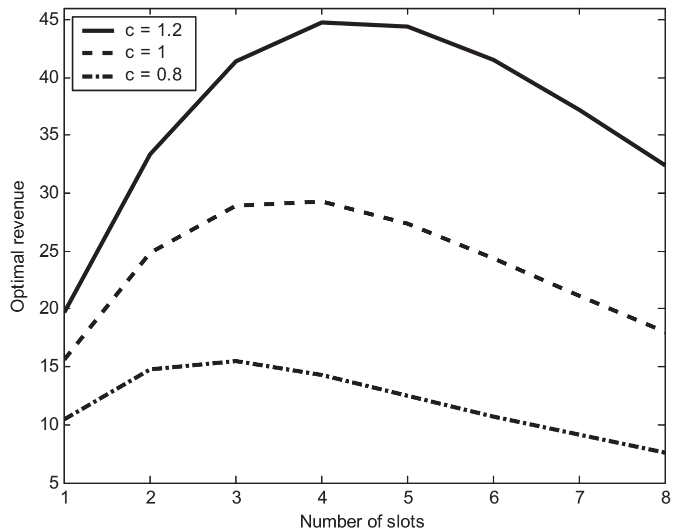


Figure 1.6: Optimal revenue vs. slots with price depending on number of slots

Figures 1.6 and 1.7 show the optimal revenue and the optimal price vs. number of advertising slots taking into account that ads can jeopardize each other.

Comparing Figures 1.4 and 1.6, we can see that the optimal revenue does not continue to increase with the number of slots as before. Instead, after a certain number of slots the impact of the price sensitivity with respect to the number of slots starts playing a role and the revenue starts decreasing. Here, the optimal number of slots to choose varies from three to four slots depending on the price-demand relationship. On the price side, Figure 1.7 indicates that the optimal price decreases more sharply in the number of slots than before. This means that the web publisher has to lower the price faster to attract the customers lost due to the impact of the increased number of slots.

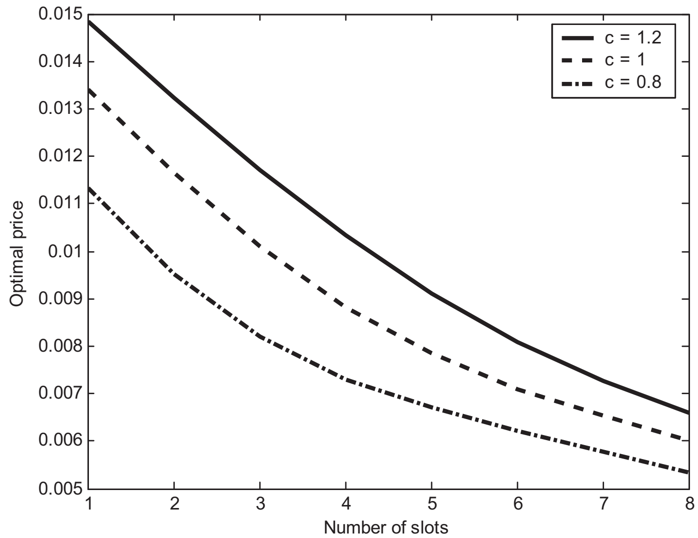


Figure 1.7: Optimal price vs. slots with price depending on number of slots

1.6.2 Impressions

Next, we consider the sensitivity of the optimal price and revenue with respect to the number of impressions. We assume that there are two slots on the website, $n = 2$, and the price-demand function for the advertisers is chosen as before to be $p(\lambda) = 0.02 - 0.2\lambda^c$, where $c = 0.8$, 1 or 1.2.

As shown in Proposition 8 and illustrated in Figures 1.8 and 1.9, the optimal revenue and the optimal price increase with the number of impressions. From a marketing point of view, one might expect quantity discounts, i.e., that the price per impression would decrease with the number of impressions. However, from an operational point of view the opposite is optimal as more impressions mean the web publisher needs fewer advertisers and thus can charge higher price (based on the decreasing price-demand curve).

By considering the advertisers' expectations for quantity discounts, we incorporate this in a simple way in the price-demand function, $p(\lambda) = 0.02 - 0.2\lambda^c - 10^{-7}x$.

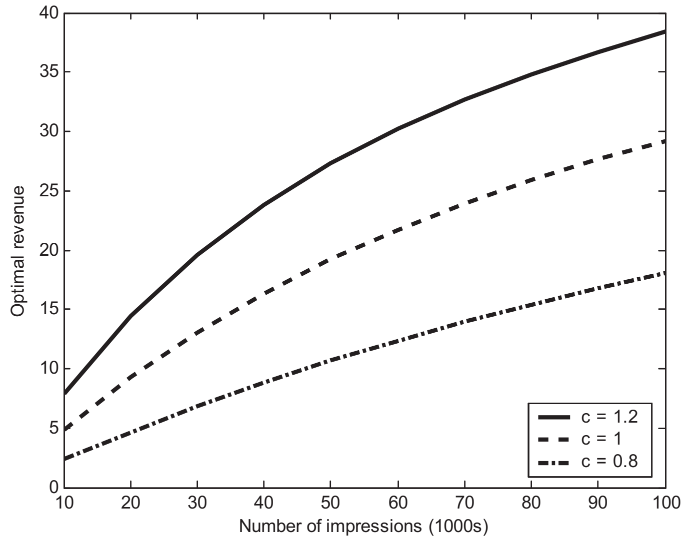


Figure 1.8: Optimal revenue vs. impressions

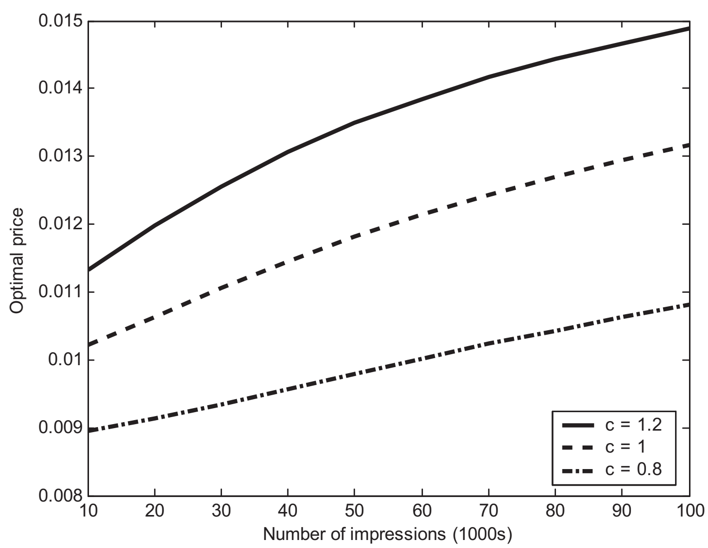


Figure 1.9: Optimal price vs. impressions

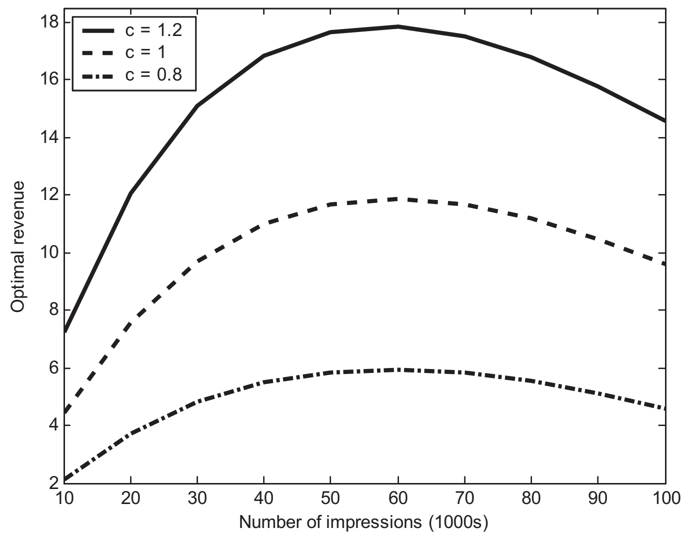


Figure 1.10: Optimal revenue vs. impressions with price depending on number of impressions

Using this function we explore how the optimal price and revenues change with the number of impressions. By incorporating quantity discounts in the pricing, the optimal revenue does not continue to increase as before, instead it starts decreasing, indicating an optimal value for the number of impressions to offer.

1.6.3 Ad Rotation

We consider the impact of serving more advertisers than there are slots by rotating the ads into slots. As defined before, s is the number of ads that share a slot. We let the number of slots on the website be $n = 4$ and as before we assume the price-demand function to be $p(\lambda) = 0.02 - 0.2\lambda^c$, where $c = 0.8$, 1 or, 1.2.

In Figure 1.12, in which the number of impressions is set to be $x = 100,000$, we observe that the optimal price is increasing in the number of rotating ads, while this relationship is overturned in Figure 1.13 where we assume $x = 1,000,000$. Hence, the

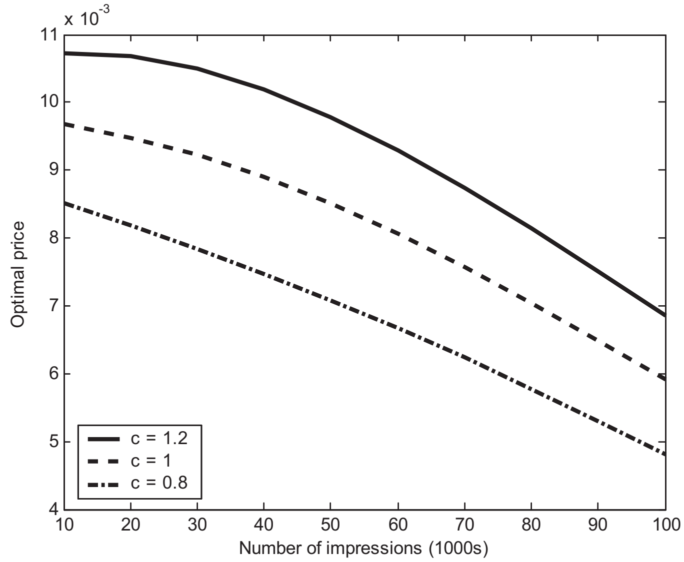


Figure 1.11: Optimal price vs. impressions with price depending on number of impressions

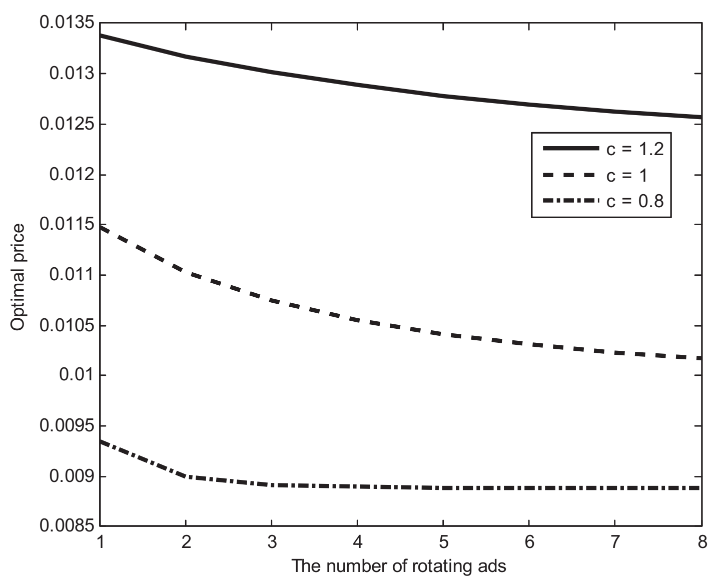


Figure 1.12: Optimal price vs. the number of rotating ads sharing the same slot. In this graph: $x = 100,000$.

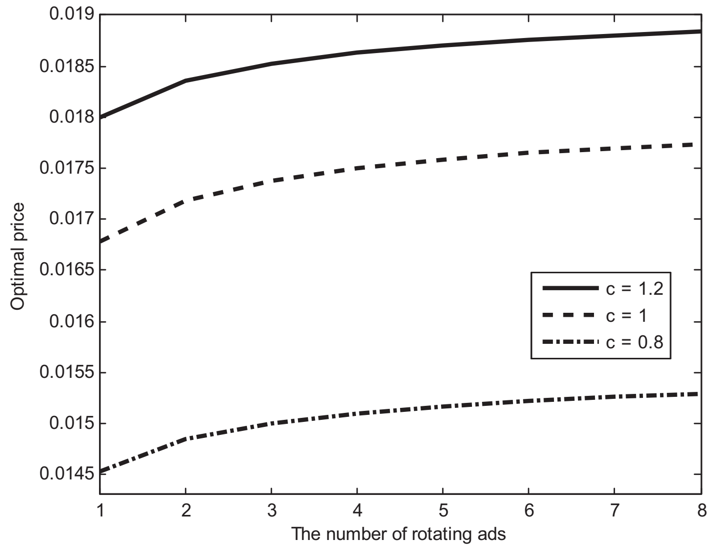


Figure 1.13: Optimal price vs. the number of rotating ads sharing the same slot. In this graph: $x = 1,000,000$.

rotation does not have an obvious impact on the publisher's optimal decisions. First, increasing the number of rotating ads leads to an increase in the system's capacity as more advertisers can be served at the same time. However, rotating a larger number of ads means that advertisers take longer to be served and occupy the capacity of the system for longer. Hence, fewer advertisers are needed. Depending on the system parameters (such as the number of impressions) one of these two impacts dominates.

1.7 Extensions

Our model provides a web publisher with insights on how to manage its revenues based on the price to charge and the operational characteristics of its website. We have focused on the fundamental trade-off of matching supply with demand under the assumptions listed in Section 1.3. The web publisher's advertising operation is quite complex and we do not attempt to capture every element of it. However, we have provided significant steps towards a systematic approach. In the following sections, we illustrate how our model can be extended and used as a building block for more complicated settings than considered in the previous sections. In addition, we show how our results hold without the assumptions listed in Section 1.3.

1.7.1 Different Numbers of Impressions

Publishers usually display any number of impressions requested by advertisers or allow them to choose among several listed quantities. As mentioned in Section 1.3, we can model this choice by defining X as a random variable representing the number of impressions chosen by advertisers. In this section, we compare the simulated values of two system quantities when random numbers of impressions are requested, with their corresponding analytical values when all advertisers request the same number of impressions, $\mathbb{E}(X)$ (Assumption 2). The random variable X can either have a discrete distribution representing a list of numbers offered, or it can be assumed to be continuous (as X is usually large) representing that any number can be chosen. We denote this system by Stochastic Request system (SR) and the system where Assumption 2 applies by Deterministic Request system (DR). Solving this SR system analytically does not appear to be tractable but to gain further insights, we perform a simulation study and simulate the system quantities of interest; L, the average number of advertisers in the system and $\mathbb{P}_n$, the probability that the system is full. In our simulation study, we let the advertisers' arrival rate be equal to 0.1 per time unit, $\lambda = 0.1$, and the viewers' arrival rate be equal to 10 per time unit, $\mu = 10$. These numbers are chosen for illustration purposes. The number of slots is chosen to be, $n = 4$. Each arriving advertiser requests $X = Y \cdot 1_{\{Y \geq 0\}}$ impressions, where $Y \sim N(\mu, \vartheta\mu)$, i.e., X is a truncated normal random variable. We compare the SR

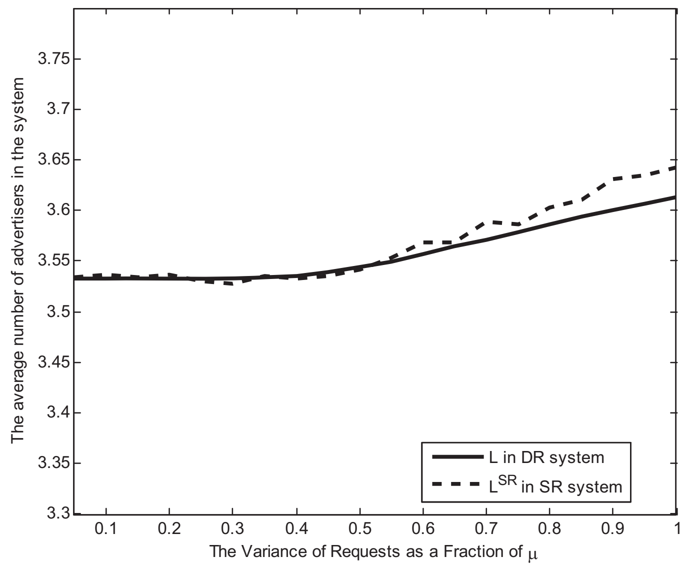


Figure 1.14: Comparison of L vs. the simulated L^{SR}

system with the DR system, in which all advertisers request $\mathbb{E}(X) = \mu/(1 - \Phi(\frac{-1}{\vartheta}))$ impressions (the mean of a truncated normal random variable), wherein $\Phi(\cdot)$ is the standard normal distribution and $(1 - \Phi(\frac{-1}{\vartheta}))$ is the probability of the event $\{X \geq 0\}$. We run each simulation for $100,000$ time units varying ϑ from $\vartheta = 0.05$ to $\vartheta = 1$. Figures 1.14 and 1.15 compare the values of L^{SR} and $\mathbb{P}_n^{SR}$ obtained through simulations with the corresponding values L and $\mathbb{P}_n$ calculated using the Equations (1)-(1.8) for the DR system with $x = \mathbb{E}(X) = \mu/(1 - \Phi(\frac{-1}{\vartheta}))$.

Based on Figures 1.14 and 1.15 we can see that the performance measures considered for the SR system are very similar to the ones of the DR system with an increasing difference when the variance increases, as can be expected. Other simulation results using different distributions for X confirm this result. These results indicate that the DR system seems to be an accurate estimator for the SR system's behavior even for low numbers of impressions.

Note here we have explored two quantities characterizing the operation of the

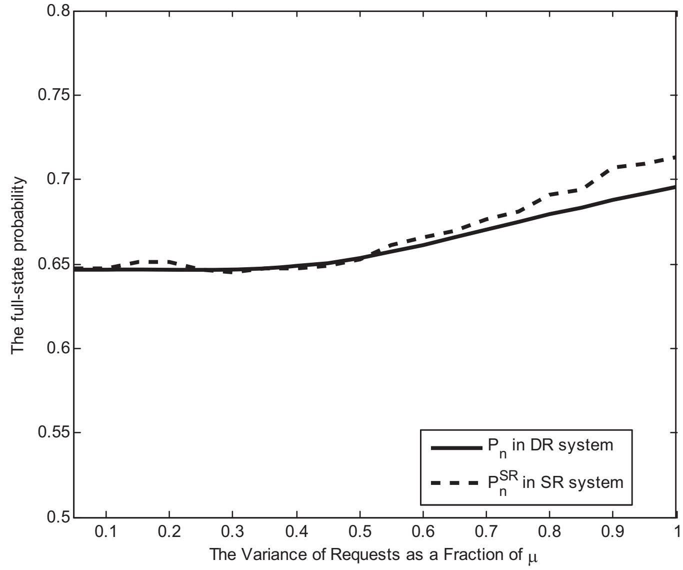


Figure 1.15: Comparison of $\mathbb{P}_n$ vs. the simulated $\mathbb{P}_n^{SR}$

system, L and $\mathbb{P}_n$. In Section 1.7.4 we explicitly explore the impact of the random number of impressions on the revenues and consider the case where the price depends on the number of impressions, X. Hence, the publisher can adjust the optimal price based on the requested impressions.

1.7.2 Fixed Advertising Campaign Length

Some ad networks allow the advertisers to request a certain advertising campaign length instead of the number of impressions. The publisher might then give some estimates on how many impressions the advertiser can expect to receive during the campaign. This system is a special case of the SR system since the number of impressions received by each advertiser during a horizon T, is a random variable $X_T \sim Poisson(\mu T)$, where μ is the viewers' arrival rate, based on the fact that the interarrival times of the viewers are exponential. Following the approach of the last section, we can approximate this system of fixed campaign length by setting the single impressions' number to be $\mathbb{E}(X_T) = \mu T$ in the DR system.

We can extend the fixed campaign length system to incorporate not a single horizon T but multiple horizon values that the advertisers can choose from. We define the choice set as $\Omega = \{T_1, ..., T_m\}$. We can argue that this system is equivalent to the SR system. We let $\tau_i \in [0, 1]$ be the percentage of the advertisers preferring to stay in the system for $T_i \in \Omega$ time units. Since the viewers' interarrival times are exponential each advertiser choosing T_i is served with $X_{T_i} \sim Poisson(\mu T_i)$ impressions. This system of multiple campaign lengths can be approximated with a DR system with $x = \sum_{i=1}^{m} \tau_i \mu T_i$ impressions. The continuous version of the multiple campaign lengths system, in which the service time T is a continuous random variable can be approximated by a DR system with $x = \mathbb{E}(X) = \int_0^\infty \mu t h(t) dt$ impressions, where $T \sim h(T)$.

1.7.3 Non-Poisson Arrivals

In Section 1.3 we assumed that the advertisers' arrivals at the web publisher from the ad network follow a Poisson process (Assumption 1), which might not be the case in reality. In addition, the viewers' arrival process might not be Poisson either (Assumption 3). In this section, we explore other distributions for both the demand and supply sides. Figures 1.16 and 1.17 show the empirical distributions, based on data from a large Scandinavian web publisher, for the advertisers' and viewers' arrivals as well as other fitted distributions. For the arrival distribution of the viewers, the Poisson, Weibull, and Normal distributions pass the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) goodness-of-fit tests at the 5% significance level. For the arrival distribution of the advertisers only the Uniform and Normal distributions pass the tests, not the Poisson. However, even though assuming Poisson arrivals of

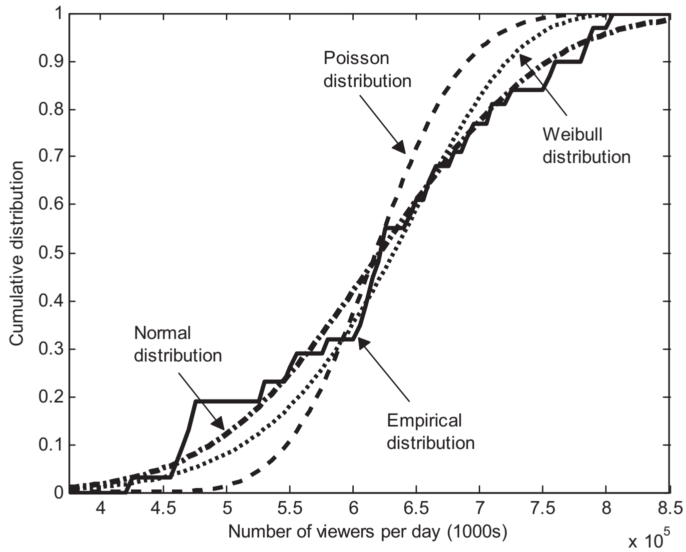


Figure 1.16: The empirical cumulative distribution of the viewers' arrivals obtained from a Scandinavian publisher based on daily data, and other fitted distributions.

advertisers might not be a realistic assumption, our simulation study illustrates that the revenues of the web publisher are only slightly affected.

In our simulation study, we specifically examine the amount of revenue a publisher can lose by using the base model's solution obtained in Section 1.5, while both the advertisers' and the viewers' arrivals, in reality, do not follow a Poisson process. We let the viewers' arrival rate be $\mu = 1$. For the advertisers' interarrival time distributions, we consider the following distributions: Normal with mean $1/\lambda$ and standard deviation $1/\lambda$, Erlang-2 with mean $1/\lambda$ and standard deviation $1/\sqrt{2}\lambda$, Erlang-4 with mean $1/\lambda$ and standard deviation $1/2\lambda$, uniform with the two parameters 0 and $2/\lambda$, exponential with rate λ, and finally deterministic arrivals. For the viewers' inter arrival time distributions, we consider the same distributions with λ replaced with $\mu = 1$.

The number of slots is $n = 4$. We choose the pricing function to be $p(\lambda) = 0.02 - 0.2\lambda^{0.8}$. Moreover, we let the number of impressions be $x = 1000$. The steps

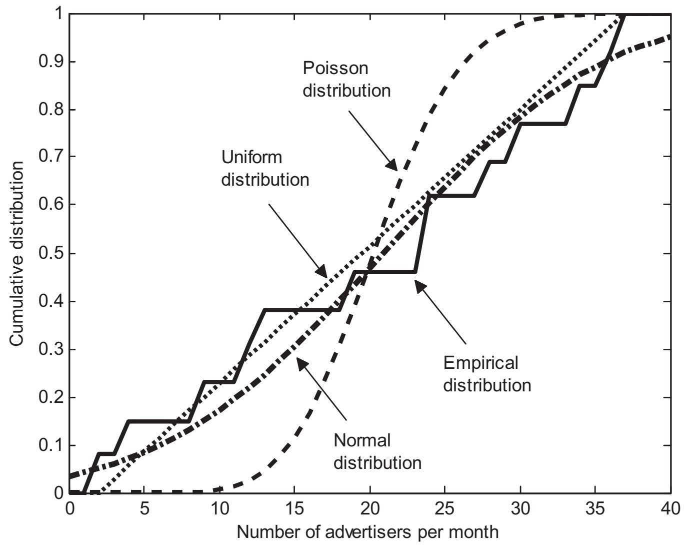


Figure 1.17: The empirical cumulative distribution of the advertisers' arrivals obtained from a Scandinavian publisher based on monthly data, and other fitted distributions.

of each simulation process are as follows:

First, we obtain the advertisers' optimal arrival rate, $\lambda^*_{D_1}$, when the advertisers' interarrival times follow the generic distribution D_1, and the viewers' interarrival times follow D_2. This includes simulating the publisher's system for multiple values of λ and then selecting $\lambda^*_{D_1,D_2}$, the rate that gives the highest revenue. We represent the revenue related to $\lambda^*_{D_1,D_2}$ with $R_{D_1,D_2}(\lambda^*_{D_1,D_2})$.

Next, we compute the optimal value for λ using the solution provided in Equation (1.11). We represent this with λ^*_{Exp}. If the web publisher used our analytical solution for a system that does not have Poisson arrivals on either side, its revenue would be $R_{D_1,D_2}(\lambda^*_{Exp})$, where $R_{D_1,D_2}(\lambda^*_{Exp})$ is the simulated revenue given the advertisers' arrival rate is $\lambda_{D_1,D_2} = \lambda^*_{Exp}$.

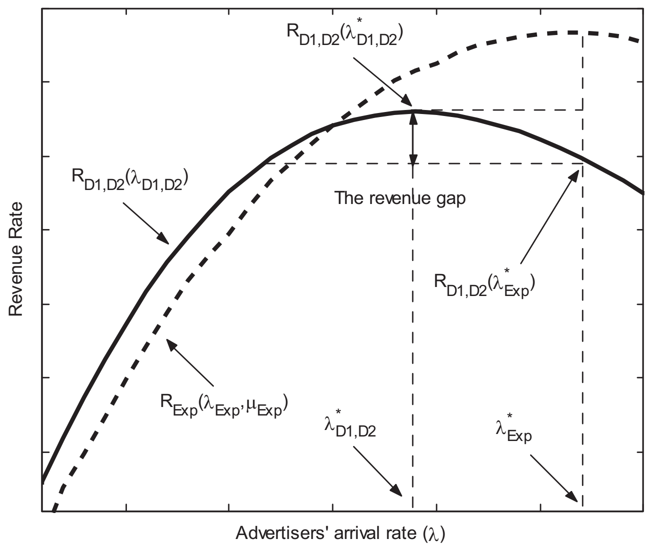


Figure 1.18: A schematic presentation of how the revenue gap is computed through the mentioned steps.

Finally, we compute the revenue gap using the following formula

$$Gap = \frac{R_{D_1,D_2}(\lambda^*_{D_1,D_2}) - R_{D_1,D_2}(\lambda^*_{Exp})}{R_{D_1,D_2}(\lambda^*_{D_1,D_2})} \times 100(\%). \tag{1.13}$$

Figure 1.18 shows a schematic presentation of how the revenue gap is obtained using the above steps.

Table 1.1 shows the relative revenue performance gaps for the different interarrival time distributions considered for advertisers' and viewers' arrivals. We can observe that the computed revenue gaps are usually between $0.1\% - 3\%$. This suggests that the Poisson policy tends to be a relatively accurate estimate for the publisher's model even when both the viewers' and the advertisers' arrivals are non-Poisson. Nevertheless, we notice that the revenue gap is considerably higher (i.e., about 9.4%) when the viewers' and the advertisers' arrival processes are both *deterministic*. This implies that Poisson's policy may not be a good approximate when there is no uncertainty in the model. However, we notice that even when either arrival processes is not deterministic, Poisson tends to perform well. Likewise, other simulation results indicate that the converse also tends to be true. That being said, a purely deterministic system might not be a good approximation for the Poisson system. This emphasizes the fact that uncertainty plays a significant role in the publisher's system.

Interarrival dist.	*Viewers*					
Advertisers	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Det.*	*Exp.*
Erlang-2	1.08%	1.15%	1.03%	0.99%	1.23%	1.34%
Erlang-4	0.40%	0.59%	0.50%	0.44%	0.56%	0.13%
Normal	0.52%	1.06%	1.09%	2.26%	2.48%	0.06%
Uniform	1.03%	1.05%	0.79%	1.23%	0.95%	1.05%
Det.	2.93%	2.89%	3.29%	2.80%	9.42%	2.73%
Exp.	0.10%	0.27%	0.38%	0.60%	0.24%	–

Table 1.1: The relative performance gap $\frac{R_{D_1,D_2}(\lambda^*_{D_1,D_2})-R_{D_1,D_2}(\lambda^*_{Exp})}{R_{D_1,D_2}(\lambda^*_{D_1,D_2})}\times 100(\%)$

1.7.4 Random Price

To validate our results further, we consider the case where each advertiser requests a random number of impressions, X, and the publisher charges a (random) price accordingly. We specifically examine the amount of revenue a publisher can lose by using the base model's solution obtained in Section 1.5 (based on Poisson arrivals, a single number of impressions offered, and a single price charged) to determine the price, while the impressions requested are random and both the advertisers' and the viewers' arrival processes are non-Poisson.

To explore this scenario, we assume the price function to be $p(\lambda, X) = 0.02 - 0.2\lambda^{0.8} - 10^{-7}X$, where X is a random variable following the arbitrary distribution f, with $\mathbb{E}_f(X) = 1000$. The rest of the simulations' settings is the same as in Section 1.7.3. The steps for each simulation process are as follows:

First, we obtain the advertisers' optimal arrival rate, $\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2}$, when the advertisers' interarrival times follow the generic distribution $\mathbf{D}_1$, the viewers' interarrival times follow $\mathbf{D}_2$, and each advertiser requests a different number of impressions according to a random variable X. This includes simulating the publisher's system for multiple values of λ and then selecting $\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2}$, the rate that gives the highest revenue. We represent the revenue related to $\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2}$ with $R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2})$.

Second, we compute the optimal value for λ using the closed-form solution provided in Equation (1.11) by assuming the price function to be $p(\lambda, x) = 0.02 - 0.2\lambda^{0.8} - 10^{-7}x$, where $x = \mathbb{E}_f(X) = 1000$. We represent this optimal value with $\lambda^*_{x,Exp}$. If the web publisher uses our analytical solution with the average demand x, for a system that does not have Poisson arrivals of advertisers and viewers, and each advertiser requests X impressions its "real" revenue would become $R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{x,Exp})$.

Finally, we compute the revenue gap using the following formula

$$Gap = \frac{R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2}) - R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{x,Exp})}{R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2})} \times 100(\%).$$

Figure 1.19 illustrates a schematic presentation of how the revenue gap is computed through the above steps.

Table 1.2 shows the relative revenue performance gaps for the different distributions for the number of impressions resulting in random prices as well as the interarrival time distributions for the advertisers when $\mathbf{D}_2$ is a *normal* distribution. We observe that the computed revenue gaps are on average between $0.08\% - 1.46\%$. This suggests that the optimal revenue when considering the average of the requested number of impressions tends to be very close to the revenue obtained through the real-time price adjusting based on each advertiser's requested impressions, while the

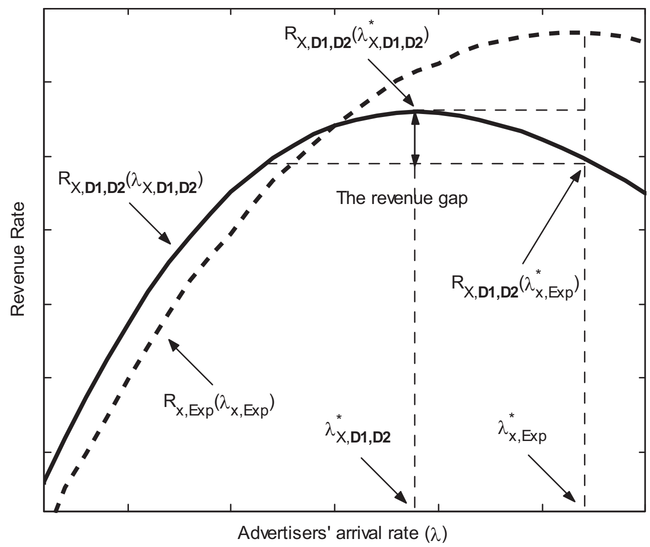


Figure 1.19: A schematic presentation of calculating the revenue gap when the price function depends on the number of impressions X in which X is a random variable.

$f(X)$	*Advertisers* $(\mathbf{D}_1)$				
Impressions	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Exponential*
Erlang-2	0.42%	1.25%	0.28%	1.20%	0.33%
Erlang-4	0.29%	1.46%	0.63%	0.96%	0.08%
Normal	0.67%	0.40%	0.13%	0.95%	0.15%
Uniform	0.29%	0.43%	0.19%	0.57%	0.50%

Table 1.2: The relative gap $\frac{R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2})-R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{x,Exp})}{R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2})}\times 100(\%)$

arrival processes are non-Poisson.

1.8 Conclusion

Online advertising is a promising research area suitable for quantitative approaches. Until recently this area has received little attention from the operations management community. Managing revenues in online advertising is a challenging task and few systematic approaches exist in practice. This chapter is a step towards filling that gap.

We consider a web publisher that generates revenues from displaying advertisements on its website with advertisers being supplied through an advertising network. We focus on the main operational challenge of matching uncertain demand from advertisers requesting advertising space to uncertain supply from viewers visiting the website. The publisher faces challenging decisions of determining the prices to charge, number of ads to display, number of advertisements to share each advertising slots, and others. The publisher's website can consist of multiple pages and each page can have ads of different sizes posted in different locations. We group the website's advertising slots into subsystems that specify the characteristics of the slots such as page, location on a page, size, and price. When a web publisher registers with an ad network it registers each subsystem with a separate tracking code. This enables us to decouple the publisher's website into subsystems and analyze them separately.

We develop a stylized model of a publisher's system with the arrival process corresponding to the advertisers who are supplied by ad networks and are interested in posting their ads and the service process corresponding to the viewers visiting the website. Given the fact that all advertisers on display pay when a single viewer uploads the webpage, the advertisers are served in a "synchronized" manner. These dynamics are different from those of known multi-server systems with *independent* servers corresponding to advertising slots. We derive a closed-form solution of the probability distribution of the number of advertisers in the system, which enables us to characterize the price and other decision variables for the publisher and analyze in detail. For example, we show that the optimal price to charge per impression increases in the number of impressions, contrary to the quantity discount common in practice. Yahoo! is the only publisher we came across that charges higher CPM price for larger contracts instead of giving a discount. We are pleased to offer the theoretical explanation that they were lacking. Furthermore, we show that the increase in the number of ads that share a slot has a non-obvious impact on the optimal price. To illustrate how our model can be used as a building block for a more complicated setting, we extend it to incorporate competition and provide some initial results. We observe that some of our results hold in competitive settings while others are overturned.

The framework for the web publisher's operations can be expanded beyond the results of this chapter. Even though cost-per-impression is a common pricing scheme, others exist such as cost-per-click or even a mixture of the two. Cost-per-click contracts require different models and comparison of them to cost-per-impression contracts can be found in the next chapter. Furthermore, exploring competition between web publishers with different information structures would be an interesting extension. We consider this topic in Chapter 3. In addition, given how easy it is to keep track of viewers' behavior and profile, targeted advertising is very attractive to advertisers as well as to web publishers who can charge a higher price for a more targeted audience. We have analyzed the operation of the web publisher from a steady state point of view. Dynamic pricing would also be interesting and possible to implement as advertisers often buy their advertising space online, which makes it feasible for the web publisher to change the prices dynamically.

In summary, the modeling framework developed in this paper captures the fundamental operational challenges faced by a web publisher and with its closed-form solution it can provide a basis for many promising research directions.

Appendix A

A1. Advertising Networks

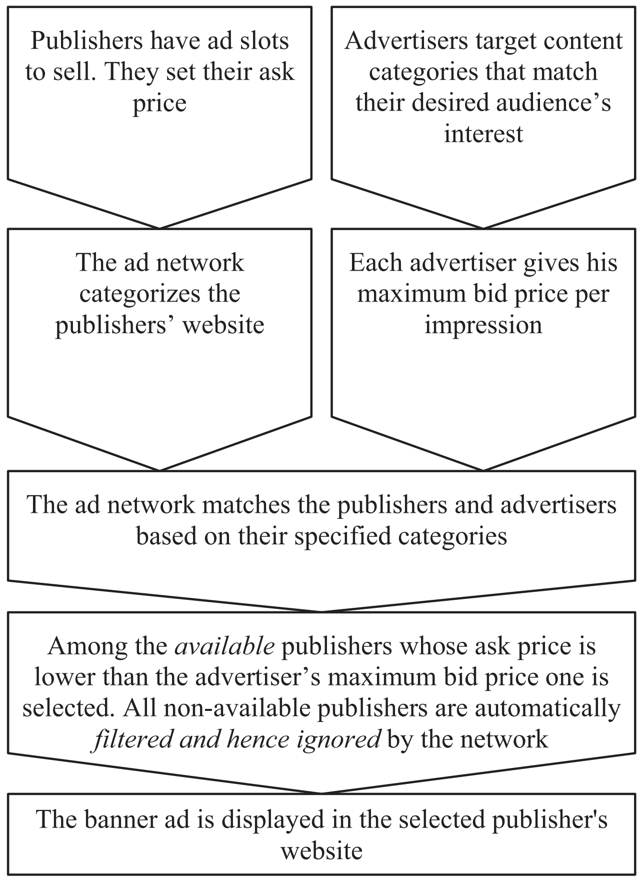


Figure 1.20: The general steps for the transaction between advertisers and web publishers through advertising networks

Ad networks act as brokers between advertisers and web publishers. Ad networks can operate in many different ways. Some ad networks set the price themselves (e.g., for non premium publishers at the ad network Clicksor), but many allow the publisher to set the price to charge advertisers (e.g., Clicksor's premium advertisers and on ADSDAQ). In addition to the price, the number of impressions that the publisher will deliver is also stated. The mechanism of how advertisers are allocated to publishers varies. In some advertising networks the advertisers can choose with which publishers their ad is placed (e.g., Adtoll) while other ad networks define different categories of target websites for advertisers to choose from based on the content of the websites, viewer's locations, etc. (so-called blind ad networks, e.g., Contextweb, Valueclick, and Clicksor). For those ad networks, matching advertisers with suppliers is usually along the following lines: First, the advertiser selects its target website category and his maximum willingness-to-pay (called bid price). Then the network selects the relevant publishers that match the characteristics of the website category. The network eliminates the publishers whose CPM (called the ask price) is higher than the advertisers' willingness-to-pay. Finally, the ad network sends the advertisers to one of these websites. Ultimately, the publisher receives a certain rate of demand that depends on the price it offers for a certain number of impressions. Figure 1.20 shows the summary of the steps for transaction between advertisers and web publishers through advertising networks.

A2. Proof of Proposition 1

Lemma 10 *Given $x \in \mathbb{N}$, $i \in \mathbb{N}$, and $\kappa \in \mathbb{R}$, the following result holds:*

$$\sum_{k_1=1}^{x} \sum_{k_2=k_1}^{x} \dots \sum_{k_i=k_{i-1}}^{x} \kappa = \binom{x+i-1}{i} \kappa. \tag{1.14}$$

Proof

Proofs of all lemmas are provided in Appendix C.

Lemma 11 *Given $k \in \mathbb{N} \cap [0, x-1]$ and $x \in \mathbb{N} \cup \{0\}$, the following result holds:*

$$\sum_{i=k}^{x+k-1} \binom{i}{k} = \binom{x+k}{k+1}. \tag{1.15}$$

Lemma 12 *Given $x \in \mathbb{N} \cap [n-1, \infty)$ with $n \in \mathbb{N}$ and $r \in \mathbb{R}_+$,*

$$\sum_{i=0}^{n-1} \binom{x+i-1}{i} r^i (1+r)^{n-i-1} = \sum_{i=0}^{n-1} \binom{x+n-1}{i} r^i. \tag{1.16}$$

Proof of Proposition 1

To prove Equations (1.4) and (1.5) we list the flow balance equations and show that the probabilities are of the form:

$$\pi_{\mathbf{k}_1} = Ar^i(1+r)^{n-i-1}, \tag{1.17}$$

$$\pi_{\mathbf{k}_2} = Ar^n, \tag{1.18}$$

where $\mathbf{k}_1 = (k_1, k_2, ..., k_i, 0, ..., 0)$ represents the state of having i slots occupied (one slot with k_1 impressions left to satisfy, another one with k_2 impressions left, etc.) and $n - i$ empty slots and $\mathbf{k}_2 = (k_1, k_2, ..., k_n)$ represents the state of having all n slots occupied. We use the following notation for compactness and define:

$$\mathbf{k}_1 \triangleq \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j \mathbf{e}_j^T, \text{ where } |\mathcal{G}_{>0}(\mathbf{k})| < n,$$
$$\text{and } \mathbf{k}_2 \triangleq \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j \mathbf{e}_j^T, \text{ where } |\mathcal{G}_{>0}(\mathbf{k})| = n,$$

in which, unless otherwise specified, $k_{j_1} \geq k_{j_2}$ for $j_1 \geq j_2$. $\mathbf{e}_j^T$ is an n-tuple row vector where its j^{th}component is 1 and the rest of its components are zero. Moreover,

$$\mathcal{G}_{>0}(\mathbf{k}) \triangleq \{j \mid k_j > 0\},\ \mathcal{G}_x(\mathbf{k}) \triangleq \{j \mid k_j = x\}, \tag{1.19}$$
$$\mathcal{G}_0(\mathbf{k}) \triangleq \{j \mid k_j = 0\},\ \mathcal{G}_h(\mathbf{k}) \triangleq \{j \mid k_j = h\}.$$

Intuitively $|\mathcal{G}_{>0}(\mathbf{k})|$ and $|\mathcal{G}_0(\mathbf{k})|$ are the number of occupied and empty slots, respectively. $|\mathcal{G}_x(\mathbf{k})|$ is the number of slots whose impressions left to satisfy are equal to x, i.e., no impressions have been delivered for these slots.

Next, we show by summing over the relevant $\pi's$ that the probabilities of finding a certain number of advertisers in the system are of the form:

$$\begin{aligned} \mathbb{P}_i &= \binom{x+i-1}{i} A r^i (1+r)^{n-i-1}, \qquad i = 0, 1, 2, ..., n-1, \\ \mathbb{P}_n &= \binom{x+n-1}{n} A r^n, \qquad i = n. \end{aligned} \tag{1.20}$$

Finally, we use the fact that $\sum_{i=0}^{n} \mathbb{P}_i = 1$ and solve for A.

We first consider a Markov chain where the state of the system is the vector $\mathbf{k} = \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j \mathbf{e}_j^T$ and k_j represents the number of impressions left to satisfy in a slot. (We do not distinguish between the slots.) After identifying the possible transitions of the system, we list the flow balance equations. The flow balance equations are equations are as follows:

i) For $\mathbf{k} = (0, ..., 0) = \mathbf{0}$ we have:

$$r\pi_{\mathbf{0}} = \sum_{i=1}^{n} \pi_{\mathbf{v}_i^T}, \ \textit{where } \mathbf{v}_i^T \triangleq \sum_{j=1}^{i} \mathbf{e}_j^T. \tag{1.21}$$

For example, if $\mathbf{k} = (0,0,0)$ then the transition equation becomes: $r\pi_{(0,0,0)} = \pi_{(1,0,0)} + \pi_{(1,1,0)} + \pi_{(1,1,1)}$.

ii) The second group of states are when $\mathbf{k} = \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j \mathbf{e}_j^T$ with $|\mathcal{G}_{>0}(\mathbf{k})| < n$ (some slots are empty) and $|\mathcal{G}_x(\mathbf{k})| = 0$ (the impressions left to satisfy in all slots are less than x). The transition balance equation for $\mathbf{k}$ becomes:

$$(1+r)\pi_{\mathbf{k}} = \pi_{\mathbf{k}+\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|}} + \sum_{q=|\mathcal{G}_{>0}(\mathbf{k})|+1}^{|\mathcal{G}_0(\mathbf{k})|} \pi_{\mathbf{k}+\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|}+\mathbf{w}^T_q}, \tag{1.22}$$

where $\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|} \triangleq \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} \mathbf{e}^T_j$ and $\mathbf{w}^T_q \triangleq \sum_{t=|\mathcal{G}_{>0}(\mathbf{k})|+1}^{q} \mathbf{e}^T_j$. For example, if we take $x = 5$ and $n = 3$ then for the state $\mathbf{k} = (4,0,0)$ we have $\mathcal{G}_{>0}(\mathbf{k}) = \{1\}$ and $\mathcal{G}_0(\mathbf{k}) = \{2,3\}$, $\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|} = (1,0,0)$, $\mathbf{w}^T_2 = (0,1,0)$, and $\mathbf{w}^T_3 = (0,1,1)$. Hence, the flow balance equation becomes: $(1+r)\pi_{(4,0,0)} = \pi_{(5,0,0)} + \pi_{(5,1,0)} + \pi_{(5,1,1)}$.

iii) If $\mathbf{k} = \sum_{j=1}^{|\mathcal{G}_x(\mathbf{k})|} x\mathbf{e}^T_j + \sum_{j=|\mathcal{G}_x(\mathbf{k})|+1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j\mathbf{e}^T_j$ where $|\mathcal{G}_{>0}(\mathbf{k})| < n$ (some slots are empty), $|\mathcal{G}_x(\mathbf{k})| > 0$ (the impressions left to satisfy in some slots are x) then the flow balance equation becomes:

$$(1+r)\pi_{\mathbf{k}} = r\pi_{(\mathbf{k}-x\mathbf{e}^T_1)^T\mathbf{D}_{n\times n}}, \textit{ where } \mathbf{D}_{n\times n} \triangleq [\mathbf{e}_2 \vdots \mathbf{e}_3 \vdots \ldots \vdots \mathbf{e}_{n-1} \vdots \mathbf{0}_{1\times n}]_{n\times n}. \tag{1.23}$$

For example, if we take $x = 5$ and $n = 4$ and $\mathbf{k} = (5,5,4,0)$ then $\mathbf{k} - x\mathbf{e}^T_1 = (0,5,4,0)$ and $\mathbf{D}_{4\times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}_{4\times 4}$. Hence, $(\mathbf{k} - x\mathbf{e}^T_1)^T\mathbf{D}_{n\times n} = (5,4,0,0)$.
Therefore, the flow balance equation becomes: $(1+r)\pi_{(5,5,4,0)} = r\pi_{(5,4,0,0)}$.

iv) If $\mathbf{k} = \sum_{j=1}^{|\mathcal{G}_x(\mathbf{k})|} x\mathbf{e}^T_j + \sum_{j=|\mathcal{G}_x(\mathbf{k})|+1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j\mathbf{e}^T_j$, where $|\mathcal{G}_{>0}(\mathbf{k})| = n$ (all slots are occupied), $|\mathcal{G}_x(\mathbf{k})| > 0$ (the impressions left to satisfy in some slots are x) then the flow balance equation becomes:

$$\pi_{\mathbf{k}} = r\pi_{(\mathbf{k}-x\mathbf{e}^T_1)^T\mathbf{D}_{n\times n}} \textit{where} \mathbf{D}_{n\times n} = [\mathbf{e}_2 \vdots \mathbf{e}_3 \vdots \ldots \vdots \mathbf{e}_{n-1} \vdots \mathbf{0}_{1\times n}]_{n\times n}. \tag{1.24}$$

For example, if $x = 5$ and $n = 4$ and $\mathbf{k} = (5,5,4,3)$ then $\mathbf{k} - x\mathbf{e}_1^T = (0,5,4,3)$ and $\mathbf{D}_{4\times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}_{4\times 4}$. Hence, $(\mathbf{k} - x\mathbf{e}_1^T)^T \mathbf{D}_{n\times n} = (5,4,3,0)$. Therefore, the flow balance equation becomes: $\pi_{(5,5,4,3)} = r\pi_{(5,4,3,0)}$.

v) If $\mathbf{k} = \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j \mathbf{e}_j^T$ with $|\mathcal{G}_{>0}(\mathbf{k})| = n$, (all slots are occupied), $|\mathcal{G}_x(\mathbf{k})| = 0$ (the impressions left to satisfy in all slots are less than x) then the flow balance equation becomes:

$$\pi_{\mathbf{k}} = \pi_{\mathbf{k} + \mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|}}, \text{ where } \mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|} = \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} \mathbf{e}_j^T. \tag{1.25}$$

For example, if $x = 5$ and $n = 4$ and $\mathbf{k} = (4,3,2,1)$ then $\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|} = (1,1,1,1)$ and the flow balance equation becomes: $\pi_{(4,3,2,1)} = \pi_{(5,4,3,2)}$.

Next we verify that the functional form stated in Equations (1.17) and (1.18) satisfies the Flow Balance Equations i) - v):

i) By inserting Equations (1.17) and (1.18) into the flow balance equation we obtain a left hand side of $Ar(1+r)^{n-1}$ and a right hand side of $A[r(1+r)^{n-2} + r^2(1+r)^{n-3} + ... + r^{n-1}(1+r)^0 + r^n]$. We use induction to show that both sides are equal, i.e., $\sum_{j=1}^{n-1} r^j(1+r)^{n-j-1} + r^n = r(1+r)^{n-1}$. We start with $n = 1$ and note that both sides are equal to r. We now assume that the equality holds for $n = k$, i.e., $\sum_{j=1}^{k-1} r^j(1+r)^{k-j-1} + r^k = r(1+r)^{k-1}$. In order to show that the equality then holds for $n = k+1$ we need to show that

$\sum_{j=1}^{k} r^j(1+r)^{k-j} + r^{k+1} = r(1+r)^k$. We have that

$$\sum_{j=1}^{k} r^j(1+r)^{k-j} + r^{k+1} = (1+r)\sum_{j=1}^{k-1} r^j(1+r)^{k-j-1} + r^k + r^{k+1}. \quad (1.26)$$

Using the induction assumption we obtain

$$(1+r)[r(1+r)^{k-1} - r^k] + r^k + r^{k+1} = r(1+r)^k - r^{k+1} + r^{k+1} = r(1+r)^k, \quad (1.27)$$

which completes the induction proof.

ii) Using a similar approach as for Case i) we need to show that

$$A(\sum_{j=0}^{n-i-1} r^{i+j}(1+r)^{n-i-j-1} + r^n) = (1+r)Ar^i(1+r)^{n-i-1}, \quad (1.28)$$

which is the same as showing

$$\sum_{j=0}^{n-i-1} r^j(1+r)^{n-i-j-1} + r^{n-i} = (1+r)^{n-i}. \quad (1.29)$$

To simplify the notation we set $m = n - i$. We then need to show that $\sum_{j=0}^{m-1} r^j(1+r)^{m-j-1} + r^m = (1+r)^m$. We prove this equality by induction. If $m = 1$ both sides of the equality are $1+r$. Let us now assume that the equality holds for $m = k$, i.e., $\sum_{j=0}^{k-1} r^j(1+r)^{k-j-1} + r^k = (1+r)^k$. We now need to show that the equality holds for $m = k+1$, i.e., that $\sum_{j=0}^{k} r^j(1+r)^{k-j} + r^{k+1} = (1+r)^{k+1}$. We have

$$\sum_{j=0}^{k} r^j(1+r)^{k-j} + r^{k+1} = (1+r)\sum_{j=0}^{k-1} r^j(1+r)^{k-j-1} + r^k + r^{k+1}. \quad (1.30)$$

Using the induction assumption this equals to $(1+r)[(1+r)^k - r^k] + r^k + r^{k+1} =$

$(1+r)^{k+1}$, which completes the induction proof.

iii) We need to verify that $(1+r)Ar^{k+j}(1+r)^{n-k-j-1} = rAr^{k-1+j}(1+r)^{n-k-j}$, which always holds.

iv) We need to verify that $Ar^n = rAr^{n-1}$, which always holds.

v) This equation always holds.

When deriving A, we first need to formulate $\mathbb{P}_i$, the probability that there are i advertisers in the system, and then use the fact that $\sum_{i=0}^{n} \mathbb{P}_i = 1$ to solve for A. First, we know that $\mathbb{P}_0 = \pi_{\mathbf{0}} = A(1+r)^{n-1}$. Let us then consider i, $0 < i < n$. The probability of having i advertisers in the system where each has k_j impressions left to satisfy with

$$\mathbf{k} = \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j \mathbf{e}_j^T, \ |\mathcal{G}_{>0}(\mathbf{k})| = i < n, \ |\mathcal{G}_x(\mathbf{k})| = 0. \tag{1.31}$$

Without loss of generality, let k_j be increasing in j, i.e., $k_1 \leq k_2 \leq ... \leq k_i$. We observe that $\mathbb{P}_i$ is the sum over all possible values that k_j can take while there are i people in the system ($|\mathcal{G}_{>0}(\mathbf{k})| = i$). Hence, by Lemma 17,

$$\mathbb{P}_i = \sum_{k_1=1}^{x} \sum_{k_2=k_1}^{x} ... \sum_{k_i=k_{i-1}}^{x} \pi_{\mathbf{k}} = \binom{x+i-1}{i} \pi_{\mathbf{k}}. \tag{1.32}$$

As a result $\mathbb{P}_i = \binom{x+i-1}{i} Ar^i (1+r)^{n-i-1}$ for $i < n$ and $\mathbb{P}_n = \binom{x+n-1}{n} Ar^n$. Moreover, since $\sum_{i=0}^{n} \mathbb{P}_i = 1$ we have that

$$\sum_{i=0}^{n-1} \binom{x+i-1}{i} Ar^i (1+r)^{n-i-1} + \binom{x+n-1}{n} Ar^n = 1, \tag{1.33}$$

which gives

$$A = \frac{1}{\sum_{j=0}^{n-1} \binom{x+j-1}{j} r^j (1+r)^{n-j-1} + \binom{x+n-1}{n} r^n}. \tag{1.34}$$

Finally, using Lemma 19 we get $A = \frac{1}{\sum_{j=0}^{n} \binom{x+n-1}{j} r^j}$, which completes the proof.

■

Appendix B

B1. Proofs of Other Propositions

Proof of Proposition 2 (i) In order to show $\mathbb{P}_n$ is increasing in r we show its derivative with respect to r is always positive. By differentiating $\mathbb{P}_n$ with respect to r and simplifying we get:

$$\frac{\partial \mathbb{P}_n}{\partial r} = \frac{\sum_{i=0}^{n} \binom{x+n-1}{n}\binom{x+n-1}{i} r^{n+i-1}(n-i)}{\sum_{i=0}^{n} \binom{x+n-1}{i} r^i} \geq 0, \tag{1.35}$$

which is always positive. Hence, $\mathbb{P}_n$ is increasing in r.

(ii) After some calculations we get

$$\mathbb{P}_n(x+1) - \mathbb{P}_n(x) = \frac{r^n \sum_{i=0}^{n} r^i \binom{x+n-1}{n}\binom{x+n}{i}\left(\frac{n-i}{x}\right)}{\sum_{i=0}^{n} \binom{x+n-1}{i} r^i \sum_{i=0}^{n} \binom{x+n}{i} r^i}, \tag{1.36}$$

which is always positive. Hence, $\mathbb{P}_n(x+1) - \mathbb{P}_n(x) \geq 0$.

(iii) For the last part we prove $\mathbb{P}_{n+1} \leq \mathbb{P}_n$ using contradiction. Let us assume $\mathbb{P}_{n+1} > \mathbb{P}_n$. This is the same as saying

$$r\frac{x+n}{n+1} \sum_{i=0}^{n} \frac{r^i (x+n-1)!}{i!(x+n-1-i)!} > \sum_{i=0}^{n+1} \frac{r^i (x+n)!}{i!(x+n-i)!}. \tag{1.37}$$

Next we reindex the sum on the right hand side of (1.37) by setting $i = j+1$ and we get

$$\sum_{i=0}^{n} \frac{r^i (x+n)!}{(n+1)i!(x+n-1-i)!} > \frac{1}{r} + \sum_{j=0}^{n} \frac{r^j (x+n)!}{(j+1)!(x+n-1-j)!}. \tag{1.38}$$

By comparing the sums term by term we see that each term on the left hand side of

(1.38) is smaller than the corresponding one on the right hand side, which contradicts the assumption of $\mathbb{P}_{n+1} > \mathbb{P}_n$. Hence, we must have $\mathbb{P}_{n+1} \leq \mathbb{P}_n$. ■

Lemma 13 *Given any natural numbers $x \in \mathbb{N}$ and $n \in \mathbb{N}$,*

$$\sum_{i=0}^{n}\sum_{j=0}^{n-1}\binom{x+n-1}{i}\binom{x+n}{j}r^{i+j} \geq \sum_{i=0}^{n}\binom{x+n-1}{n}\binom{x+n}{i}r^{n+i}(n-i). \tag{1.39}$$

Proofs of all lemmas are provided in the Appendix C.

Lemma 14 *Let $Q(x) = Q_N(x)/Q_D(x)$, where*

$$Q_N(x) = \left(\sum_{i=0}^{n}\binom{x+n-1}{i}\binom{x+n}{n}r^{n+i} + \sum_{i=0}^{n}\binom{x+n-1}{n}\binom{x+n}{i}r^{n+i}(n-i)\right) \tag{1.40}$$

and

$$Q_D(x) = \left(\sum_{i=0}^{n}\binom{x+n-1}{i}r^i\sum_{i=0}^{n}\binom{x+n}{i}r^i\right). \tag{1.41}$$

Then for any x, $n \in \mathbb{N}$, and $r \in \mathbb{R}_+$, $Q(x)$ is increasing in x.

Lemma 15 *Given any natural numbers $x \in \mathbb{N}$ and $n \in \mathbb{N}$,*

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}r^{i+j} \geq \sum_{i=0}^{n}\binom{x+n-1}{n}\binom{x+n-1}{i}r^{n+i}(n+1-i). \tag{1.42}$$

Lemma 16 *Given any $x, n \in \mathbb{N} \cup \{0\}$, and $r \in \mathbb{R}_+$*

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}r^{i+j}(n+1-i)(n+i-2j) \geq 0. \tag{1.43}$$

Proof of Proposition 3 (i)We set $L_n(x) = rx(1 - \mathbb{P}_n(x))$. We need to show that $\Delta L_n(x) = L(x+1) - L(x) \geq 0$. We have

$$\Delta L_n(x) = rx(\mathbb{P}_n(x) - \mathbb{P}_n(x+1)) + r(1 - \mathbb{P}_n(x+1)). \tag{1.44}$$

Focusing on the first term in $\Delta L_n(x)$ we get

$$x(\mathbb{P}_n(x) - \mathbb{P}_n(x+1)) = \frac{xr^n[\binom{x+n-1}{n}\sum_{i=0}^{n}\binom{x+n}{i}r^i - \binom{x+n}{n}\sum_{i=0}^{n}\binom{x+n-1}{i}r^i]}{\sum_{i=0}^{n}\binom{x+n-1}{i}r^i\sum_{i=0}^{n}\binom{x+n}{i}r^i}. \tag{1.45}$$

Knowing that $\binom{x+n}{i} = \binom{x+n-1}{i} + \binom{x+n-1}{i-1}$ and after some simplification we get

$$x(\mathbb{P}_n(x) - \mathbb{P}_n(x+1)) = -\frac{r^n\binom{x+n-1}{n}\sum_{i=0}^{n}\binom{x+n}{i}r^i(n-i)}{\sum_{i=0}^{n}\binom{x+n-1}{i}r^i\sum_{i=0}^{n}\binom{x+n}{i}r^i}. \tag{1.46}$$

The result in (1.46) also shows that the full state probability is increasing in x. In addition, we can see that

$$1 - \mathbb{P}_n(x+1) = 1 - \frac{\binom{x+n}{n}r^i}{\sum_{i=0}^{n}\binom{x+n}{i}r^i} = \frac{\sum_{i=0}^{n-1}\binom{x+n}{i}r^i}{\sum_{i=0}^{n}\binom{x+n}{i}r^i}. \tag{1.47}$$

Therefore we can simplify $\Delta L_n(x)$ as

$$\Delta L_n(x) = r\frac{-\sum_{i=0}^{n}\binom{x+n-1}{n}\binom{x+n}{i}r^{n+i}(n-i) + \sum_{i=0}^{n-1}\binom{x+n}{i}r^i\sum_{i=0}^{n}\binom{x+n-1}{i}r^i}{\sum_{i=0}^{n}\binom{x+n-1}{i}r^i\sum_{i=0}^{n}\binom{x+n}{i}r^i}. \tag{1.48}$$

Now to show $\Delta L_n(x)$ is positive we need to show its numerator is always positive. That is,

$$\sum_{i=0}^{n}\sum_{j=0}^{n-1}\binom{x+n-1}{i}\binom{x+n}{j}r^{i+j} - \sum_{i=0}^{n}\binom{x+n-1}{n}\binom{x+n}{i}r^{n+i}(n-i) \geq 0. \tag{1.49}$$

This is always true according to Lemma 20. Therefore, $L_n(x)$ is increasing in x.

(ii) In order to prove that $L_n(x)$ is concave in x we need to show that

$$\Delta L_n(x) = 1 - \frac{\sum_{i=0}^{n} \binom{x+n-1}{i}\binom{x+n}{n} r^{n+i} + \sum_{i=0}^{n} \binom{x+n-1}{n}\binom{x+n}{i} r^{n+i}(n-i)}{\sum_{i=0}^{n} \binom{x+n-1}{i} r^i \sum_{i=0}^{n} \binom{x+n}{i} r^i}, \tag{1.50}$$

is decreasing in x. But this is always true based on Lemma 21.

(iii) As $r = \lambda/\mu$ it is equivalent to showing that L is concave increasing in r given μ is fixed. We know that $L_n(x) = rx(1-\mathbb{P}_n) = rx - rx\mathbb{P}_n$. Hence, we get $\frac{\partial L_n(x)}{\partial r} = x - x\frac{\partial(r\mathbb{P}_n)}{\partial r}$ and $\frac{\partial^2 L_n(x)}{\partial r^2} = -x\frac{\partial^2(r\mathbb{P}_n)}{\partial r^2}$. We first show that $\frac{\partial L_n(x)}{\partial r} \geq 0$. We have that:

$$\frac{\partial(r\mathbb{P}_n)}{\partial r} = \frac{\binom{x+n-1}{n} r^n [\sum_{i=0}^{n} \binom{x+n-1}{i} r^i (n+1-i)]}{[\sum_{i=0}^{n} \binom{x+n-1}{i} r^i]^2}. \tag{1.51}$$

Hence, in order to ensure that $\frac{\partial L_n(x)}{\partial r} \geq 0$ we need to show that:

$$[\sum_{i=0}^{n} \binom{x+n-1}{i} r^i]^2 - [\binom{x+n-1}{n} \sum_{i=0}^{n} \binom{x+n-1}{i} r^{i+n}(n+1-i)] \geq 0, \tag{1.52}$$

which is true according to Lemma 23. Hence, $L_n(x)$ is increasing in r. Now we show that $\frac{d^2 L_n(x)}{dr^2} \leq 0$. Note that showing $\frac{d^2 L_n(x)}{dr^2} \leq 0$ is equivalent to showing $\frac{\partial^2(r\mathbb{P}_n)}{\partial r^2} \geq 0$. So we work with the latter one. From the relation (54) in the paper we have

$$\frac{\partial(r\mathbb{P}_n)}{\partial r} = \frac{\binom{x+n-1}{n} r^{n+1} [\sum_{i=0}^{n} \binom{x+n-1}{i} r^{i-1} (n+1-i)]}{[\sum_{i=0}^{n} \binom{x+n-1}{i} r^i]^2}. \tag{1.53}$$

Furthermore,

$$\frac{\partial^2(r\mathbb{P}_n)}{\partial r^2} = \frac{\binom{x+n-1}{n} r^{n-1}}{[\sum_{i=0}^{n} \binom{x+n-1}{i} r^i]^3} [\sum_{i=0}^{n}\sum_{j=0}^{n} \binom{x+n-1}{i}\binom{x+n-1}{j} r^{i+j}(n+1-i)(n+i-2j)]. \tag{1.54}$$

Based on Lemma 24 we have that $\frac{\partial^2(r\mathbb{P}_n)}{\partial r^2} \geq 0$.Hence, $L_n(x)$ is concave increasing in r.

(iv) The proof follows from the third part of Proposition 2 and Equation (6) in the paper. ■

Proof of Proposition 4 In the Erlang's loss system with n servers the probability of a full system is

$$\mathbb{P}_n^E = \frac{\frac{(xr)^n}{n!}}{\sum_{i=0}^{n} \frac{(xr)^i}{i!}} = \frac{\frac{r^n}{n!}}{\sum_{i=0}^{n} \frac{r^i}{i! x^{n-i}}}. \tag{1.55}$$

While for the web publisher's system it is

$$\mathbb{P}_n = \frac{\binom{x+n-1}{n} r^n}{\sum_{i=0}^{n} \binom{x+n-1}{i} r^i}. \tag{1.56}$$

Starting from (1.56) and simplifying yields

$$\begin{aligned} \mathbb{P}_n &= \frac{\frac{r^n}{n!}}{\sum_{i=0}^{n} \frac{r^i}{i! x(x+1)(x+2)\ldots(x+n-1-i)}} \geq \frac{\frac{r^n}{n!}}{\sum_{i=0}^{n} \frac{r^i}{i! x^{n-i}}} \\ &= \mathbb{P}_n^E. \end{aligned}$$

Therefore $\mathbb{P}_n \geq \mathbb{P}_n^E$. Moreover, from Equation (6) in the paper we have that the average number of advertisers in the web publisher's system is $L = rx(1 - \mathbb{P}_n)$. According to Harel (1990) the same formula holds for the average number of jobs in the Erlang's loss system, i.e., $L^E = rx(1 - \mathbb{P}_n^E)$. As $\mathbb{P}_n \geq \mathbb{P}_n^E$ we have that $L \leq L^E$.
■

Proof of Proposition 5 For the web publisher's system we showed in Proposition 1 that the probability distribution of the number of the advertisers in the system is:

$$\begin{aligned} \mathbb{P}_i &= \frac{\binom{x+i-1}{i} r^i (1+r)^{n-i-1}}{\sum_{i=0}^{n} \binom{x+n-1}{i} r^i}, \quad i = 0, 1, 2, ..., n-1, \\ \mathbb{P}_n &= \frac{\binom{x+i-1}{i} r^i}{\sum_{i=0}^{n} \binom{x+n-1}{i} r^i}. \end{aligned} \tag{1.57}$$

With $x = 1$ and $n \to \infty$ we will get the distribution of the bulk service system with infinite capacity. The distribution with $x = 1$ and finite n is:

$$\begin{aligned} \mathbb{P}_i &= \frac{r^i (1+r)^{n-i-1}}{\sum_{i=0}^{n} \binom{n}{i} r^i} = \frac{r^i (1+r)^{n-i-1}}{(1+r)^n}, \quad \text{for } i = 0, 1, 2, ..., n-1, \\ \mathbb{P}_n &= \frac{r^n}{\sum_{k=0}^{n} \binom{n}{k} r^k} = \frac{r^n}{(1+r)^n}. \end{aligned} \tag{1.58}$$

Using the binomial identity $\sum_{k=0}^{n} \binom{n}{k} r^k = (1+r)^n$ we get

$$\begin{aligned} \mathbb{P}_i &= \frac{r^i (1+r)^{n-i-1}}{(1+r)^n} = \frac{r^i}{(1+r)^{i+1}}, \quad i = 0, 1, 2, ..., n-1, \\ \mathbb{P}_n &= \frac{r^n}{\sum_{k=0}^{n} \binom{n}{k} r^k} = (\frac{r}{1+r})^n. \end{aligned} \tag{1.59}$$

Now if we pass $n \to \infty$ then $\mathbb{P}(n) \to 0$, i.e., the probability of a full system is zero. However, $\mathbb{P}_i = \frac{r^i}{(1+r)^{i+1}}, \ \forall i \in N$.

For the second part, when $n \to \infty$ we get the expected value of the number of people in the system as

$$L = \mathbb{E}_i = \sum_{i=0}^{\infty} \frac{i r^i}{(1+r)^{i+1}} = \frac{r}{(1+r)^2} \sum_{i=1}^{\infty} i \left(\frac{r}{1+r} \right)^{i-1}. \tag{1.60}$$

Now taking $\frac{r}{1+r} = q$ we obtain

$$\mathbb{E}_i = \frac{r}{(1+r)^2}\sum_{i=1}^{\infty} iq^{i-1} = \frac{r}{(1+r)^2}\frac{\partial}{\partial q}\left(\sum_{i=1}^{\infty} q^i\right) \tag{1.61}$$
$$= \frac{r}{(1+r)^2}\frac{\partial}{\partial q}\left(\frac{q}{1-q}\right) = \frac{r}{(1+r)^2}\left(\frac{1}{1-q}\right)^2 = r, \tag{1.62}$$

which is the desired result. To obtain the variance we first need to obtain the second moment

$$\mathbb{E}_{i^2} = \sum_{i=0}^{\infty}\frac{i^2 r^i}{(1+r)^{i+1}} = \sum_{i=0}^{\infty}\frac{(i(i-1)+i)r^i}{(1+r)^{i+1}} = \mathbb{E}_{i(i-1)} + \mathbb{E}_i. \tag{1.63}$$

But on the other side $\mathbb{E}_{i(i-1)}$ is obtained as

$$\mathbb{E}_{i(i-1)} = \sum_{i=0}^{\infty}\frac{i(i-1)r^i}{(1+r)^{i+1}} = \frac{r^2}{(1+r)^3}\sum_{i=1}^{\infty} i(i-1)q^{i-2} = \frac{r^2}{(1+r)^3}\frac{\partial^2}{\partial q^2}\left(\sum_{i=2}^{\infty} q^i\right) \tag{1.64}$$
$$= \frac{r^2}{(1+r)^3}\frac{\partial}{\partial q}\left(\frac{q(2-q)}{(1-q)^2}\right) = \frac{r^2}{(1+r)^3}\left(\frac{2}{(1-q)^3}\right) = 2r^2.$$

Hence, using (1.61), (1.63), and (1.64) we get $\mathbb{E}_{i^2} = 2r^2 + r$. Therefore $Var_i = \mathbb{E}_{i^2} - \mathbb{E}_i^2 = r^2 + r$. ■

Proof of Proposition 6 (i) Using (7) in the paper we get

$$\frac{\partial^2 R(\lambda)}{\partial \lambda^2} = x\mu(\frac{\partial^2 L(\lambda)}{\partial \lambda^2}p(\lambda) + L\frac{\partial^2 p(\lambda)}{\partial \lambda^2} + 2\frac{\partial L(\lambda)}{\partial \lambda}\frac{\partial p(\lambda)}{\partial \lambda}). \tag{1.65}$$

Knowing that $p(\lambda)$ is positive and concave decreasing and L is concave increasing we have that $\frac{\partial^2 R(\lambda)}{\partial \lambda^2} \leq 0$.

(ii) The expression for the optimal price follows from the FONC. ■

Proof of Proposition 7 (i) By abusing the notation slightly we denote the optimal revenue with $n+1$ slots as

$$R^*_{n+1} = \lambda^*(n+1)(1 - \mathbb{P}_{n+1}(\lambda^*(n+1)))p(\lambda^*(n+1))x. \tag{1.66}$$

Using optimality and part (iii) of Proposition 2 we get

$$R^*_{n+1} \geq \lambda^*(n)(1 - \mathbb{P}_{n+1}(\lambda^*(n)))p(\lambda^*(n))x \geq \lambda^*(n)(1 - \mathbb{P}_n(\lambda^*(n)))p(\lambda^*(n))x = R^*_n, \tag{1.67}$$

which completes the proof for this part.

(ii) For the second part we again abuse the notation slightly and denote the optimal revenues with $x+1$ impressions as

$$R^*_{x+1} = L_{x+1}(\lambda^*(x+1))\mu p(\lambda^*(x+1)). \tag{1.68}$$

Using optimality and parts (ii) and (iii) of Proposition 3 we get

$$R^*_{x+1} \geq L_{x+1}(\lambda^*(x))\mu p(\lambda^*(x)) \geq L_x(\lambda^*(x))\mu p(\lambda^*(x)) = R^*_x, \tag{1.69}$$

which completes the proof for this part.

(iii) For the third part we note that the busy probability $\mathbb{P}_n$ depends only on $r = \lambda/\mu$, not on λ and μ separately. By adapting our notation we denote the optimal revenues with μ as the arrival rate of the viewer as

$$R^*(\mu) = R(\lambda^*(\mu), \mu) = \lambda^*(\mu)(1 - \mathbb{P}_n(\lambda^*(\mu)/\mu))p(\lambda^*(\mu))x. \tag{1.70}$$

According to Part (i) of Proposition 2, $\mathbb{P}_n$ is increasing in λ (and r) and thus decreasing in μ. Using that fact and optimality we have for $\mu^1 \geq \mu^2$ that

$$R^*(\mu^1) \geq \lambda^*(\mu^2)(1-\mathbb{P}_n(\lambda^*(\mu^2)/\mu^1))p(\lambda^*(\mu^2))x \geq \lambda^*(\mu^2)(1-\mathbb{P}_n(\lambda^*(\mu^2)/\mu^2))p(\lambda^*(\mu^2))x = R^*(\mu^2), \tag{1.71}$$

which completes the proof. ■

Proof of Proposition 8 The proof involves using the FONC and the Implicit Function Theorem as well as comparing terms in multiple sums.

We need to show that $\frac{\partial \lambda^*}{\partial x} \leq 0$. By Implicit Function Theorem we get $\frac{\partial \lambda^*}{\partial x}$ as

$$\frac{\partial \lambda^*}{\partial x} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial \lambda^*}} = -\frac{\frac{\partial}{\partial x}(L^{'}(\lambda^*))p(\lambda^*) + \frac{\partial}{\partial x}(L(\lambda^*))p^{'}(\lambda^*)}{L^{''}(\lambda^*)p(\lambda^*) + L^{'}(\lambda^*)p^{'}(\lambda^*) + L^{'}(\lambda^*)p^{'}(\lambda^*) + L(\lambda^*)p^{''}(\lambda^*)}. \tag{1.72}$$

Note that since x is discrete we are slightly abusing the Implicit Function Theorem. However, we treat x for the remainder as discrete and, e.g., $\frac{\partial}{\partial x}(L^{'}(\lambda^*))$ corresponds to $\Delta(L^{'}(\lambda))\big|_{\lambda=\lambda^*} = L^{'}_{x+1}(\lambda^*) - L^{'}_{x}(\lambda^*)$. Since, $p(\lambda^*) > 0$, $p^{'}(\lambda^*) < 0$, $p^{''}(\lambda^*) < 0$ and $L(\lambda^*) > 0$, $L^{'}(\lambda^*) > 0$, $L^{''}(\lambda^*) < 0$, therefore the denominator is negative. Hence, we need just to show

$$\frac{\partial}{\partial x}(L^{'}(\lambda^*))p(\lambda^*) + \frac{\partial}{\partial x}(L(\lambda^*))p^{'}(\lambda^*) \leq 0. \tag{1.73}$$

Using the FONC, $L^{'}(\lambda^*)p(\lambda^*) + L(\lambda^*)p^{'}(\lambda^*) = 0$, we are are left with showing that

$$g(\lambda^*) \triangleq \frac{\partial}{\partial x}(L(\lambda^*))L^{'}(\lambda^*) - \frac{\partial}{\partial x}(L^{'}(\lambda^*))L(\lambda^*) \geq 0. \tag{1.74}$$

Without loss of generality we set $\mu = 1$ and thus $\lambda^* = r$. Now we have $L = rx(1-\mathbb{P}_x)$ and then

$$\frac{\partial L}{\partial x} = r(x+1)(1-\mathbb{P}_{x+1}) - rx(1-\mathbb{P}_x). \tag{1.75}$$

(We denote $\mathbb{P}_n$ with $\mathbb{P}_x$ to emphasize the dependence on x.) Also from the proof of Proposition 3 we have that $L' = x(1 - f_x)$ where f_x is

$$f_x \triangleq \frac{\sum_{i=0}^{n} \binom{x+n-1}{i} \binom{x+n-1}{n} r^{n+i} (n-i+1)}{\sum_{i=0}^{n} \sum_{j=0}^{n} \binom{x+n-1}{i} \binom{x+n-1}{j} r^{i+j}}. \tag{1.76}$$

Hence, we get

$$\frac{\partial}{\partial x}(L^{'}(\lambda^*)) = 1 - x(f_{x+1} - f_x) - f_{x+1.} \tag{1.77}$$

Using (1.75) and (1.77) in (2.49) and after some algebra we get

$$g(\lambda^*) = (1 - \mathbb{P}_x)(f_{x+1} - f_x) - (1 - f_x)(\mathbb{P}_{x+1} - \mathbb{P}_x). \tag{1.78}$$

Next we calculate each term in $g(\lambda^*)$ by inserting the relevant functions. Using the relation (5) in the paper as well as (1.76) in (1.78) we get

$$f_{x+1} - f_x \tag{1.79}$$

$$= \frac{\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \binom{x+n-1}{i} \binom{x+n-1}{j} \binom{x+n}{n} \binom{x+n}{k} r^{i+j+n+k} (n-k+1)}{\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{l=0}^{n} \binom{x+n}{i} \binom{x+n}{j} \binom{x+n-1}{k} \binom{x+n-1}{l} r^{i+j+k+l}} - \frac{\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \binom{x+n}{i} \binom{x+n}{j} \binom{x+n-1}{n} \binom{x+n-1}{k} r^{i+j+n+k} (n-k+1)}{\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{l=0}^{n} \binom{x+n}{i} \binom{x+n}{j} \binom{x+n-1}{k} \binom{x+n-1}{l} r^{i+j+k+l}}. \tag{1.80}$$

Multiplying the both sides of (1.79) and simplifying the right side gives

$$(1 - \mathbb{P}_x)(f_{x+1} - f_x) = \frac{\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{l=0}^{n-1} \binom{x+n-1}{i} \binom{x+n-1}{j} \binom{x+n}{n} \binom{x+n}{k} \binom{x+n-1}{l} r^{i+j+n+k+l} (n-k+1)}{\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{l=0}^{n} \sum_{h=0}^{n} \binom{x+n}{i} \binom{x+n}{j} \binom{x+n-1}{k} \binom{x+n-1}{l} \binom{x+n-1}{h} r^{i+j+k+l+h}} - \frac{\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{l=0}^{n-1} \binom{x+n-1}{i} \binom{x+n-1}{j} \binom{x+n-1}{n} \binom{x+n-1}{k} \binom{x+n-1}{l} r^{i+j+n+k+l} (n-k+1)}{\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \sum_{l=0}^{n} \sum_{h=0}^{n} \binom{x+n}{i} \binom{x+n}{j} \binom{x+n-1}{k} \binom{x+n-1}{l} \binom{x+n-1}{h} r^{i+j+k+l+h}}. \tag{1.81}$$

Using (1.76) we get

$$\mathbb{P}_{x+1} - \mathbb{P}_x = \frac{\frac{1}{x}\sum_{i=0}^{n}\binom{x+n}{i}\binom{x+n-1}{n}r^{i+n}(n-i)}{\sum_{i=0}^{n}\sum_{j=0}^{n}\binom{x+n}{i}\binom{x+n-1}{j}r^{i+j}}. \tag{1.82}$$

Now we use (1.76) and (1.82) to obtain $(1-f_x)(\mathbb{P}_{x+1}-\mathbb{P}_x)$, the second term in $g(\lambda^*)$. After some simplification we obtain

$$\begin{aligned}
&(1-f_x)(\mathbb{P}_{x+1}-\mathbb{P}_x)\\
&= \frac{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n-1}{n}\binom{x+n}{k}\binom{x+n}{l}r^{i+j+n+k+l}(\frac{n-k}{x})}{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n}\sum_{h=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n}{k}\binom{x+n-1}{l}\binom{x+n}{h}r^{i+j+k+l+h}}\\
&- \frac{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\binom{x+n-1}{n}\binom{x+n-1}{i}\binom{x+n-1}{n}\binom{x+n}{j}\binom{x+n}{k}r^{i+j+k+2n}(n-i+1)(\frac{n-j}{x})}{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n}\sum_{h=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n}{k}\binom{x+n-1}{l}\binom{x+n}{h}r^{i+j+k+l+h}}.
\end{aligned} \tag{1.83}$$

Adding Equations (1.81) and (1.83) gives (notice the denominators are in fact the same)

$$\begin{aligned}
&g(\lambda^*)\\
&= \frac{x\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n-1}\binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n}{n}\binom{x+n}{k}\binom{x+n-1}{l}r^{i+j+k+l}(n-k+1)}{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n}\sum_{h=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n}{k}\binom{x+n-1}{l}\binom{x+n}{h}r^{i+j+k+l+h}}\\
&- \frac{x\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n-1}\binom{x+n}{i}\binom{x+n}{j}\binom{x+n-1}{n}\binom{x+n-1}{k}\binom{x+n-1}{l}r^{i+j+k+l}(n-k+1)}{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n}\sum_{h=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n}{k}\binom{x+n-1}{l}\binom{x+n}{h}r^{i+j+k+l+h}}\\
&- \frac{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n-1}{n}\binom{x+n}{k}\binom{x+n}{l}r^{i+j+k+l}(n-k)}{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n}\sum_{h=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n}{k}\binom{x+n-1}{l}\binom{x+n}{h}r^{i+j+k+l+h}}\\
&+ \frac{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\binom{x+n-1}{n}\binom{x+n-1}{i}\binom{x+n-1}{n}\binom{x+n}{j}\binom{x+n}{k}r^{i+j+k+n}(n-i+1)(n-j)}{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n}\sum_{h=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n}{k}\binom{x+n-1}{l}\binom{x+n}{h}r^{i+j+k+l+h}}.
\end{aligned} \tag{1.84}$$

To show that $g(\lambda^*) \geq 0$ we are left with showing that its numerator is always positive as the denominator is clearly positive. To show this we need to systematically group the terms in the numerator in the four sums together and show that the sum of the terms in each group are positive. We do the grouping according to the power of r. Let us assume that the power of r is z where $0 \leq z \leq 4n$. If $g(\lambda^*) \geq 0$ then the coefficient of r^z for each and every z, $0 \leq z \leq 4n$, needs be positive. We divide the range into four parts: $0 \leq z < n$, $n \leq z < 2n$, $2n \leq z < 3n$, $3n \leq z \leq 4n$. Here we will illustrate the proof for $0 \leq z < n$. The other ranges are proved similarly.

Let $B(x,n,z)$ be the coefficient of r^z in the numerator for any given z. If $0 \leq z < n$ and $l = z - i - j - k$. Then after some algebra we obtain $B(x,n,z)$ as

$$B(x,n,z) = \sum_{i=0}^{z} \sum_{j=0}^{z-i} \sum_{k=0}^{z-i-j} H_{i,j,k}(x,n) C_{i,j,k}(x,n), \tag{1.85}$$

where $H_{i,j,k}(x,n)$ is

$$H_{i,j,k}(x,n) \triangleq \binom{x+n-1}{i}\binom{x+n-1}{j}\binom{x+n-1}{n}\binom{x+n-1}{k}\binom{x+n-1}{z-i-j-k}(x+n)^2 \geq 0, \tag{1.86}$$

and $C_{i,j,k}(x,n)$ is

$$C_{i,j,k}(x,n) \triangleq (\frac{n-k+1}{x+n-k}) - \frac{x(n-k+1)}{(x+n-i)(x+n-j)} - (\frac{n-k}{x+n-k})(\frac{1}{x+n-(\underbrace{z-i-j-k}_{l})}). \tag{1.87}$$

We can see that $H_{i,j,k}(x,n) \geq 0$. Hence, we only need to show $C_{i,j,k}(x,n) \geq 0$. After some simplification in (1.87) we get

$$\begin{aligned} C_{i,j,k}(x,n) \geq & \frac{ij(n-l-1) + n^2(n-i-j-l-1) + kx(n-l)}{(n+x-i)(n+x-j)(n+x-k)(n+x-k)} \\ & + \frac{nx(2n-2i-2j-l-2) + x^2(n-i-j-1)}{(n+x-i)(n+x-j)(n+x-k)(n+x-k)} \\ & \geq 0. \end{aligned} \tag{1.88}$$

Having in mind that $z = i + j + k + l$ and $0 \leq z < n$ we notice that in the right side of (1.88) each term in the numerator (and the denominator) is positive. Hence,

$C_{i,j,k}(x,n) \geq 0$. Given that other ranges for z hold we have that $g(\lambda^*) \geq 0$, which ensures $\frac{\partial \lambda^*}{\partial x} \leq 0$. As the price is decreasing in λ, we have proved that the optimal price is increasing in x. ■

Proof of Proposition 9 *i)* From the paper we know that the revenue function can be expressed as

$$R = \mu L p(\lambda). \tag{1.89}$$

Taking the first derivative with respect to λ and making it equal to zero yields

$$F = \left.\frac{\partial R}{\partial \lambda}\right|_{\lambda^*} = \frac{\partial L}{\partial \lambda} p(\lambda) + L \left.\frac{\partial p(\lambda)}{\partial \lambda}\right|_{\lambda^*} = 0, \tag{1.90}$$

which yields

$$\left.\frac{\partial L}{\partial \lambda}\right|_{\lambda^*} = -\frac{L\frac{\partial p(\lambda)}{\partial \lambda}}{p(\lambda)}, \tag{1.91}$$

which is always necessarily positive as $\frac{\partial p(\lambda)}{\partial \lambda} \leq 0$.

Using (1.90) we obtain

$$\left.\frac{\partial F(n,\lambda^*)}{\partial n}\right|_{n^*} = \frac{\partial F}{\partial \lambda^*}\frac{\partial \lambda^*}{\partial n} + \left.\frac{\partial F}{\partial n}\right|_{n^*} = 0, \tag{1.92}$$

from which, we get

$$\left.\frac{\partial \lambda^*}{\partial n}\right|_{n^*} = -\frac{\left.\frac{\partial F(n,\lambda^*)}{\partial n}\right|_{n^*}}{\left.\frac{\partial F(n,\lambda^*)}{\partial \lambda^*}\right|_{n^*}}, \tag{1.93}$$

which completes the proof of part (i).

Using (1.92), (1.93), (1.90) yields

$$\frac{\partial F}{\partial n^*} = \frac{d^2L}{d\lambda^* dn^*}p + \frac{d\lambda^*}{dL}\frac{dp}{dn^*} + \frac{dL}{dn^*}\frac{dp}{d\lambda^*} + L\frac{d^2p}{d\lambda^* dn^*} \leq 0, \tag{1.94}$$

which is necessarily negative because of the assumption (b). On the other side from (1.90) we get

$$\frac{\partial F(n, \lambda^*)}{\partial \lambda^*} = \frac{d^2L}{d\lambda^{*2}}p + 2\frac{\partial L}{\partial \lambda^*}\frac{\partial p}{\partial \lambda^*} + L\frac{d^2p}{d\lambda^{*2}}. \tag{1.95}$$

Since $\frac{\partial L}{\partial \lambda^*}\frac{\partial p}{\partial \lambda^*} \leq 0$, hence by assumption (c) it follows that

$$\frac{\partial F(n, \lambda^*)}{\partial \lambda^*} \leq 0. \tag{1.96}$$

From (1.93), (1.94), and (1.96) it follows that

$$\left.\frac{\partial \lambda^*}{\partial n}\right|_{n^*} \geq 0, \tag{1.97}$$

which completes the proof for part (ii).

For part (iii), from the chain rule we obtain

$$\frac{dp}{dn^*} = \frac{\partial p(n^*, \lambda^*)}{\partial n^*} = \frac{\partial p}{\partial \lambda^*}\frac{\partial \lambda^*}{\partial n^*} + \frac{\partial p}{\partial n^*}, \tag{1.98}$$

which is necessarily negative as $\frac{\partial p}{\partial \lambda^*} \leq 0$. ■

B2. Filler Ads

In this section, we show that considering a charge for "filler" ads does not affect the increasing property of the optimal price with respect to the requested impressions.

In order to see the reason for this issue, first consider the price function to depend only on λ (demand rate) and n (number of slots). In addition, assume that displaying each filler ad generates the revenue $e > 0$ per impression for the publisher, which can be considered as a transfer price if the ad is for a different division of the company that the publisher belongs to, or a low fee charged to a non-profit organization[6]. Our task is now to show that the charge for filler ads, e, does not play a role in the monotonicity of the optimal price.

We can modify the revenue function to include the price as follows:

$$R(\lambda, \mu, x, n) = Lp(\lambda, n)\mu + (n - L)e\mu, \tag{1.99}$$

where n is the number of slots and L is the number of advertisers in the publisher's system (in the steady state condition). Then $(n - L)$ is the average number of empty slots and $(n - L)e\mu$ is the average revenue of displaying $(n - L)\mu$ filler ads per time unit. For the next step, we apply $L = rx(1 - \mathbb{P}_n(\lambda, x, n))$, with $r = \lambda/\mu$ to (1.99). Our problem now reduces to

$$R(\lambda, \mu, x, n) = \lambda(1 - \mathbb{P}_n(\lambda, \mu, x))x(p(\lambda, n) - e) + ne\mu. \tag{1.100}$$

Note that for a high value of e, i.e., $e \geq p(\lambda, n)$, the maximum of (1.99) with respect to λ becomes $R^* = ne\mu$. This is because a large e makes the first term negative,

[6] In the context of advertising networks, e can also be considered as the flat rate charged to the ad network by the publisher for displaying low rate run-of-network ads.

which leads to $\lambda^* = 0$. That is, the publisher denies all the arriving advertisers.

Given this, it only remains to show that e does not play a role in the maximization problem above. In order to see that, we note that the two terms e, and $ne\mu$ in (1.100) are both independent of λ. Hence, it is easy to see that the maximization of (1.100) becomes equivalent to the maximization of

$$\max_{\lambda} \widehat{R}(\lambda, \mu, x, n) = \lambda(1 - \mathbb{P}_n(\lambda, \mu, x))x\widehat{p}(\lambda, n), \tag{1.101}$$

where the price function is defined as

$$\widehat{p}(\lambda, n) = p(\lambda, n) - e \geq 0. \tag{1.102}$$

We can now see that (1.101) has the same form as the basic model considered in our paper. In addition, since $p(\lambda, n)$ has the following properties $p'_\lambda(\lambda, n) \leq 0$, $p''_\lambda(\lambda, n) \leq 0$ (i.e., the necessary technical conditions for Proposition 6 to hold), so does $\widehat{p}(\lambda, n)$. As a result, the revenue function in (1.101) with the new price function $\widehat{p}(\lambda, n)$ satisfies the necessary conditions for Proposition 6. Therefore, at the optimal level, the price function $\widehat{p}^*(\lambda^*(x), n)$ would increase in x. However, it is easy to see that in (1.101), $p^*(\lambda^*(x), n)$ increases in x as well. This is because $\widehat{p}'^*_x(\lambda^*(x), n) \geq 0$ is the same as $(p^*(\lambda^*(x), n) - e)'_x \geq 0$. However, as e is a constant $(p^*(\lambda^*(x), n) - e)'_x \geq 0$ reduces to $p'^*_x(\lambda^*(x), n) \geq 0$. Therefore, the result follows and we can conclude that charging for filler ads does not change our monotonicity results.

■

Appendix C

Lemma 17 *Given $x \in \mathbb{N}$, $i \in \mathbb{N}$, and $\kappa \in \mathbb{R}$, the following result holds:*

$$\sum_{k_1=1}^{x} \sum_{k_2=k_1}^{x} \ldots \sum_{k_i=k_{i-1}}^{x} \kappa = \binom{x+i-1}{i} \kappa. \tag{1.103}$$

Proof We prove the lemma with induction. For the case $i = 1$, as mentioned earlier, $B_1 = x = \binom{x+1-1}{1}$. Now let us assume that the formula holds for B_i for $i = s$, i.e., $B_s = \binom{x+s-1}{s}$ and for any x. We then need to show that it also holds for $i = s+1$, i.e., $B_{s+1} = \binom{x+s}{s+1}$. Let us condition our counting of terms on the value of k_{s+1}. We first assume k_{s+1} takes the value of 1. The number of the terms in this case will be exactly the same as for the problem with s filled slots which is equal to $\binom{x+s-1}{s}$ according to the induction assumption. If $k_{s+1} = 2$ the other indices can vary from 2 to x. They can not take 1 anymore because all the states with 1 are already counted for in the case with $k_{s+1} = 1$. The number of terms in this case will be similar as the first case except we only have $x - 1$ values to choose from, i.e., $\binom{x+s-2}{s}$. With a similar reasoning for $k_{s+1} = 3$ we obtain $\binom{x+s-3}{s}$. Repeating the same reasoning we can see that $B_{s+1} = \binom{x+s-1}{s} + \binom{x+s-2}{s} + \binom{x+s-3}{s} + \ldots + \binom{s}{s}$. By using Lemma 18 we obtain that this summation is equal to $\binom{x+s}{s+1}$, which completes the proof. ■

Lemma 18 *Given $k \in \mathbb{N} \cap [0, x-1]$ and $x \in \mathbb{N} \cup \{0\}$, the following result holds:*

$$\sum_{i=k}^{x+k-1} \binom{i}{k} = \binom{x+k}{k+1}. \tag{1.104}$$

Proof We prove the lemma by induction. For $x = 1$ we have both sides equal to 1. Let us assume that for $x = s$ we have $\sum_{i=k}^{s+k-1} \binom{i}{k} = \binom{s+k}{k+1}$. We then need to show that for $x = s+1$ we have $\sum_{i=k}^{s+k} \binom{i}{k} = \binom{s+k+1}{k+1}$. We can see that $\sum_{i=k}^{s+k} \binom{i}{k} =$

$\sum_{i=k}^{s+k-1}\binom{i}{k}+\binom{s+k}{k}$ and by using the induction assumption we have $\sum_{i=k}^{s+k}\binom{i}{k}=\binom{s+k}{k+1}+\binom{s+k}{k}$. Using the Pascal's rule, $\binom{a-1}{b}+\binom{a-1}{b-1}=\binom{a}{b}$, we obtain $\sum_{i=k}^{s+k}\binom{i}{k}=\binom{s+k+1}{k+1}$, which completes the proof. ■

Lemma 19 *Given* $x \in \mathbb{N} \cap [n-1, \infty)$ *with* $n \in \mathbb{N}$ *and* $r \in \mathbb{R}_+$,

$$\sum_{i=0}^{n-1}\binom{x+i-1}{i}r^i(1+r)^{n-i-1}=\sum_{i=0}^{n-1}\binom{x+n-1}{i}r^i. \tag{1.105}$$

Proof We prove the lemma by induction. If $n=1$ then both sides are equal to 1. Let us assume the equality holds for $n=k$, i.e.,

$$D(k) \triangleq \sum_{i=0}^{k-1}\binom{x+i-1}{i}r^i(1+r)^{k-i-1}-\sum_{i=0}^{k-1}\binom{x+k-1}{i}r^i=0. \tag{1.106}$$

Then we need to show it also holds for $n=k+1$, i.e., that

$$D(k+1)=\sum_{i=0}^{k}\binom{x+i-1}{i}r^i(1+r)^{k-i}-\sum_{i=0}^{k}\binom{x+k}{i}r^i=0. \tag{1.107}$$

We start from $D(k+1)$ and try to reach to $D(k)$. We obtain

$$D(k+1)=(1+r)\sum_{i=0}^{k-1}\binom{x+i-1}{i}r^i(1+r)^{k-i-1}+\binom{x+k-1}{k}r^k-\sum_{i=0}^{k}\binom{x+k}{i}r^i. \tag{1.108}$$

Using the induction assumption we get

$$\begin{aligned}
&D(k+1) && (1.109)\\
&=(1+r)\sum_{i=0}^{k-1}\binom{x+k-1}{i}r^i+\binom{x+k-1}{k}r^k-\sum_{i=0}^{k-1}\binom{x+k}{i}r^i-\binom{x+k}{k}r^k\\
&=\sum_{i=0}^{k-1}\left[\binom{x+k-1}{i}-\binom{x+k}{i}\right]r^i-\sum_{i=0}^{k-1}\binom{x+k-1}{i}r^{i+1}+\binom{x+k-1}{k}r^k-\binom{x+k}{k}r^k,
\end{aligned}$$

and in the end using Pascal's rule twice and setting the index in the first sum to $i = j - 1$, we get

$$\begin{aligned} D(k+1) &= \sum_{j=0}^{k-2} \binom{x+k-1}{j} r^{j+1} + \binom{x+k-1}{k-1} r^k - \sum_{i=0}^{k-1} \binom{x+k-1}{i} r^{i+1} \\ &= \sum_{j=0}^{k-1} \binom{x+k-1}{j} r^{j+1} - \sum_{i=0}^{k-1} \binom{x+k-1}{i} r^{i+1} \\ &= D(k) = 0, \end{aligned} \tag{1.110}$$

which completes the proof. ■

Lemma 20 *Given any natural numbers $x \in \mathbb{N}$, and $n \in \mathbb{N}$,*

$$\sum_{i=0}^{n} \sum_{j=0}^{n-1} \binom{x+n-1}{i} \binom{x+n}{j} r^{i+j} \geq \sum_{i=0}^{n} \binom{x+n-1}{n} \binom{x+n}{i} r^{n+i} (n-i). \tag{1.111}$$

Proof The Lemma can be proved using the same approach as in the proof of Lemma 23. ■

Lemma 21 *Let $Q(x) = Q_N(x)/Q_D(x)$, where*

$$Q_N(x) = \left(\sum_{i=0}^{n} \binom{x+n-1}{i} \binom{x+n}{n} r^{n+i} + \sum_{i=0}^{n} \binom{x+n-1}{n} \binom{x+n}{i} r^{n+i} (n-i) \right) \tag{1.112}$$

and

$$Q_D(x) = \left(\sum_{i=0}^{n} \binom{x+n-1}{i} r^i \sum_{i=0}^{n} \binom{x+n}{i} r^i \right). \tag{1.113}$$

Then for any x, $n \in \mathbb{N}$, and $r \in \mathbb{R}_+$, $Q(x)$ is increasing in x

Proof We need to show that $Q(x+1) \geq Q(x)$. This is equivalent to showing that

$$
\begin{aligned}
A(x,n) &\triangleq \sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\binom{x+n-1}{i}\binom{x+n}{j}\binom{x+n}{k}\binom{x+n+1}{n}r^{i+j+k} \quad (1.114)\\
&+\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\binom{x+n-1}{i}\binom{x+n}{j}\binom{x+n+1}{k}\binom{x+n}{n}r^{i+j+k}(n-k)\\
&+\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\binom{x+n-1}{i}\binom{x+n}{j}\binom{x+n+1}{k}\binom{x+n}{n}r^{i+j+k}(n-k)\\
&-\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\binom{x+n}{i}\binom{x+n+1}{j}\binom{x+n-1}{k}\binom{x+n}{n}r^{i+j+k}\\
&-\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\binom{x+n}{i}\binom{x+n+1}{j}\binom{x+n}{k}\binom{x+n-1}{n}r^{i+j+k}(n-k)\\
\geq\ & 0.
\end{aligned}
$$

In order to show that the inequality above holds we need to show that for any z, $0 \leq z \leq 3n$, the coefficient of r^z is positive. We consider z in three separate regions, namely, $0 \leq z < n$, $n \leq z < 2n$, and $2n \leq z \leq 3n$. Here we prove the inequality for $0 \leq z < n$. The proof is similar for the other two regions. For any z, $0 \leq z < n$, the coefficient for r^z in $A(x,n)$ is

$$\sum_{i=0}^{z}\sum_{j=0}^{n-z} B(x,n,i,j,z),$$

where we set $k = z - i - j$ and $B(x, n, i, j, z)$ is obtained as

$$B(x,n,i,j,z) = \binom{x+n-1}{i}\binom{x+n}{j}\binom{x+n-1}{z-i-j}\binom{x+n-1}{n}(x+n)^2(x+n+1) \tag{1.115}$$

$$[\frac{1}{x(x+1)(x+n-z+i+j)} + \frac{(n-z+i+j)}{x(x+n-z+i+j)(x+n-z+i+j+1)} - \frac{1}{x(x+n-i)(x+n-j+1)} - \frac{(n-z+i+j)}{(x+n-i)(x+n+1-j)(x+n-z+i+j)}].$$

Since $z = i + j + k$ and $0 \leq z < n$ we have

$$\begin{aligned} & \frac{1}{x(x+1)(x+n-z+i+j)} - \frac{1}{x(x+n-i)(x+n-j-1)} \qquad (1.116) \\ = \; & \frac{(-i-in-jn+n^2)+(nx+zx-2ix-2jx)+(z-i-j)}{x(x+1)(x+n-z+i+j)(x+n-i)(x+n-j-1)} \\ \geq \; & 0. \end{aligned}$$

as all three terms in the numerator are positive. In a similar way we have

$$\frac{(n-z+i+j)}{x(x+n-z+i+j)(x+n-z+i+j+1)} - \frac{(n-z+i+j)}{(x+n-i)(x+n+1-j)(x+n-z+i+j)} \tag{1.117}$$

$$= \frac{(n-z+i+j)((n-i)+(n^2-in-jn)+(nx+xz-2ix-2jx)+ij}{x(x+n-z+i+j)(x+n-z+i+j+1)(x+n-i)(x+n+1-j)}$$

$$\geq 0.$$

Therefore, the coefficient of r^z is positive, which completes the proof for $0 \leq z < n$.

■

Lemma 22 *For $0 \leq j \leq i \leq n$ and $x \geq 1$ we have*

$$\binom{x+n-1}{i}\binom{x+n-1}{n+j-i} \geq \binom{x+n-1}{n}\binom{x+n-1}{j}. \tag{1.118}$$

Proof We prove the lemma by contradiction and assume $\binom{x+n-1}{i}\binom{x+n-1}{n+j-i} < \binom{x+n-1}{n}\binom{x+n-1}{j}$. After some algebra we have

$$n!(x-1)!j!(x+n-1-j)! < i!(x+n-1-i)!(n+j-i)!(x+-1-j+i)!.$$

With further simplifications we get

$$\Pi_{k=j+1}^{i}(n-i+k) \cdot \Pi_{k=j+1}^{i}(x+n-k) < \Pi_{k=j+1}^{i} k \cdot \Pi_{k=j+1}^{i}(x+i-k),$$

which is a contradiction as $n \geq i$. Hence, we conclude that $\binom{x+n-1}{i}\binom{x+n-1}{n+j-i} \geq \binom{x+n-1}{n}\binom{x+n-1}{j}$. ■

Lemma 23 *Given any natural numbers $x \in \mathbb{N}$ and $n \in \mathbb{N}$,*

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j} r^{i+j} \geq \sum_{i=0}^{n}\binom{x+n-1}{n}\binom{x+n-1}{i} r^{n+i}(n+1-i). \tag{1.119}$$

Proof We prove this lemma by selecting a few "convenient" terms from the double sum on the left hand side of (1.119) and then showing that their sum is always greater than the sum on the right hand side.

We focus on the double sum on the left hand side of (1.119) and notice since all its terms are positive this double sum is greater than a sum over a few of its terms.

We first list the terms where $i+j=2n$, then the term with $i+j=2n-1$, etc:

$$\begin{aligned}
&\sum_{i=0}^{n}\sum_{j=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}r^{i+j} \qquad (1.120)\\
&\geq \binom{x+n-1}{n}\binom{x+n-1}{n}r^{2n}+[\binom{x+n-1}{n}\binom{x+n-1}{n-1}\\
&+\binom{x+n-1}{n-1}\binom{x+n-1}{n}]r^{2n-1}+[\binom{x+n-1}{n}\binom{x+n-1}{n-2}\\
&+\binom{x+n-1}{n-1}\binom{x+n-1}{n-1}+\binom{x+n-1}{n-1}\binom{x+n-1}{n-1}\\
&+\binom{x+n-1}{n-2}\binom{x+n-1}{n}]r^{2n-2}+...+[\binom{x+n-1}{n}\binom{x+n-1}{n-3}\\
&+\binom{x+n-1}{n-1}\binom{x+n-1}{n-2}+\binom{x+n-1}{n-2}\binom{x+n-1}{n-1}\\
&+\binom{x+n-1}{n-3}\binom{x+n-1}{n}]r^{2n-3}+...+[\sum_{i=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{n-i}]r^{n}.
\end{aligned}$$

After some algebra we obtain

$$\begin{aligned}
&\sum_{i=0}^{n}\sum_{j=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}r^{i+j} \qquad (1.121)\\
&\geq [\sum_{i=n}^{n}\binom{x+n-1}{i}\binom{x+n-1}{2n-i}]r^{2n}+[\sum_{i=n-1}^{n}\binom{x+n-1}{i}\binom{x+n-1}{2n-1-i}]r^{2n-1}\\
&+[\sum_{i=n-1}^{n}\binom{x+n-1}{i}\binom{x+n-1}{2n-1-i}]r^{2n-1}+[\sum_{i=n-2}^{n}\binom{x+n-1}{i}\binom{x+n-1}{2n-2-i}]r^{2n-2}\\
&+...+[\sum_{i=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{n-i}]r^{n}\\
&=\sum_{j=0}^{n}[\sum_{i=j}^{n}\binom{x+n-1}{i}\binom{x+n-1}{n+j-i}]r^{j+n}.
\end{aligned}$$

Now we subtract the term $-\sum_{j=0}^{n}\binom{x+n-1}{n}\binom{x+n-1}{j}r^{n+i}(n+1-j)$ from both sides of

(1.121) to get

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}r^{i+j}-\sum_{j=0}^{n}\binom{x+n-1}{n}\binom{x+n-1}{j}r^{n+i}(n+1-j) \tag{1.122}$$

$$\geq \sum_{j=0}^{n}r^{n+j}[\sum_{i=j}^{n}\binom{x+n-1}{i}\binom{x+n-1}{n+j-i}-\binom{x+n-1}{n}\binom{x+n-1}{j}(n+1-j)].$$

On the other side Lemma 22 tells us that the below result is always correct:

$$\binom{x+n-1}{i}\binom{x+n-1}{n+j-i}\geq\binom{x+n-1}{n}\binom{x+n-1}{j} \text{ for } 0\leq\ j\leq i\leq n \text{ and } x\geq 1. \tag{1.123}$$

Replacing (1.123) in (1.122) we get

$$\sum_{j=0}^{n}r^{n+j}[\sum_{i=j}^{n}\binom{x+n-1}{i}\binom{x+n-1}{n+j-i}-\binom{x+n-1}{n}\binom{x+n-1}{j}(n+1-j)] \tag{1.124}$$

$$\geq \sum_{j=0}^{n}r^{n+j}[\sum_{i=j}^{n}\binom{x+n-1}{n}\binom{x+n-1}{j}-\binom{x+n-1}{n}\binom{x+n-1}{j}(n+1-j)]$$

$$=\sum_{j=0}^{n}r^{n+j}[\binom{x+n-1}{n}\binom{x+n-1}{j}(n+1-j)-\binom{x+n-1}{n}\binom{x+n-1}{j}(n+1-j)]$$

$$=0.$$

That shows the positivity of (1.119) and completes the proof. ■

Lemma 24 *Given any $x,n\in\mathbb{N}\cup\{0\}$, and $r\in\mathbb{R}_+$*

$$\sum_{i=0}^{n}\sum_{j=0}^{n}\binom{x+n-1}{i}\binom{x+n-1}{j}r^{i+j}(n+1-i)(n+i-2j)\geq 0. \tag{1.125}$$

Proof The lemma can be proved using a similar approach as in the proof of Lemma 23. ■

Chapter 2

Cost-Per-Click Pricing for Display Advertising

2.1 Introduction

Display advertising (including banner ads, video ads and all non-text-based ads) is a $25 billion business with a promising revenue rise for the coming years (McAfee et al. 2010). Web publishers that generate revenues from display advertising face several challenging decisions. They need to decide on how many advertising slots to have on their website, whether to hire a sales force to attract advertisers to post ads on their website or rely on advertising networks, how many impressions to promise to deliver, and how much to charge, etc. Pricing is one of the most challenging decisions web publishers face, with mostly ad-hoc approaches being used. It is now generally believed that the optimal pricing of display ads is the key to the web publishers' revenue success (Break Media 2009). Despite the vast amount of literature on the subject of online advertising relatively few people address the optimal pricing of display banner ads. Majority of work assume that ad prices are fixed and are not affected by other decision factors.

In this chapter, we fill this gap by providing systematic approaches for pricing of display ads when a web publisher faces major uncertainties from the demand from advertisers requesting space on one side and the supply from viewers on the other. This supply is usually in the form of the number of impressions or number of clicks. We focus on the price per click (cost-per-click or CPC). This is, in particular, one of the alternative schemes to the cost-per-impression (CPM)[1] considered in the first chapter. We show that the behavior of the two pricing schemes at the optimal level can be considerably different. As described in the first chapter, for instance, the optimal CPM prices decrease in the number of advertising slots, while in this chapter, we show that the optimal CPC prices may increase with the number of slots.

More importantly, we demonstrate that in CPC contracts[2] the common tendency among practitioners to convert the prices between the two schemes using the click-through rate[3] (CTR) can be misleading. To see the importance of this result, consider a web publisher facing the issue of determining the CPC price. In such a case, the optimal approach for the publisher is to find the best price by considering the CPC system with the uncertainties derived from the supply and demand directly. Alternatively, in order to avoid the complexities involved with the direct modeling, some web publishers tend to simply divide the CPM price by the CTR, and offer that as the CPC price. (Of course, this is provided that they know the CPM price

[1] Generally, CPM price refers to the price for every 1000 impressions. However, throughout this paper we slightly abuse the term and use the CPM price to refer to "price per every impression".

[2] CPC contracts are those where the publishers promise and charge based on the number of clicks.

[3] CTR is the probability that an ad is clicked. In practice, it is generally calculated by dividing the number of viewers that click on a certain ad by the total number of visitors to the publisher's page.

beforehand.) We show that this simple approach, though intuitive and broadly used in practice, can be misleading as it may incur a significant revenue loss for the publishers. The Appendix of this chapter demonstrates examples of two real websites that explicitly divide the CPM price by CTR to obtain the CPC price.

Currently, available models do not provide a formal method for addressing problems where web publishers determine the CPC prices for the systems affected by the two major uncertainties from the advertisers' demand and the viewers' supply. We formulate the problem as a queueing system as a suitable approach for capturing the dynamics of the advertisers' uncertain demand and how it can be matched with the dynamics of viewer' uncertain supply[4]. In our queueing system formulation, we assume that each of the slots in the publisher's system is a serving channel. The advertisers correspond to the customers of this queueing system requesting to be served, and the viewers act as the servers of the system. We employ a vector-valued state variable (with one entry per each slot), and for any given time, the pricing decision depends on the advertisers' demand, the viewers' supply, the number of slots in the website, and the number of clicks sold to each advertiser. This setup allows for a fairly general dynamics for this problem.

The primary contributions of this chapter are:

1. We construct a modeling framework capturing the main trade-offs in the operation of a web publisher dealing with an ad network that comes from matching supply with demand. We consider a general setting of multiple webpages, mul-

[4] It would be also possible to consider the mentioned uncertainties using alternative approaches, for instance, stochastic dynamic programing and in particular, controlled Markov jumped processes. However, it is easy to verify that due to the complexities involved with modeling the problem these approaches, though seemingly more idealistic, easily fall into the trap of "the curse of dimensionality", which makes them become fruitless. In addition, due to practical issues, web publishers mainly avoid dynamic pricing and strictly prefer to use static pricing, with minor time-based price updates per each day.

tiple types of ads (e.g. based on location and size) with different prices, and allow ads to share an advertising slot. This model can serve as a building block for studying more complicated operational issues of a web publisher such as competition. (See the next chapter for a discussion on competition)

2. We derive a closed-form solution of the probability distribution of the number of advertisers in the system. This enables us to determine the optimal price for the web publisher to charge advertisers and analyze the publisher's system in detail. (See Sections 2.3 and 2.4.)

3. We derive additional theoretical results through out this chapter. An appealing result, for instance, is that the steady state probability of the number of advertisers in the publisher's system coincides with a version of the $M/M/1/n$ system. This result is rather surprising since the two systems have completely distinctive characteristics. (See Proposition 26)

4. On the managerial side, we demonstrate that the general heuristic widely employed in practice, in which a publisher simply uses the CTR to convert the price of one scheme to the other can be misleading, resulting in a considerable revenue loss compared to the optimal policy. (See Section 2.5)

5. We provide further insights by showing that, unlike the CPM price considered in the first chapter, the optimal CPC price may increase with the number of advertising slots. This may go against our common intuition from the supply-demand relationship: since an increase in the number of empty slots in the system can be interpreted as an increase in the service capacity in the system. As a result, one may expect the opposite result to hold. (See Section 2.6)

6. Our model is among the first to bridge the gap between much of the academic

literature on pricing, which mainly focuses on deterministic pricing models, and the much more complex online display advertising systems encountered in practice. It also provides a significant contribution to the currently developing management science literature on online advertising, and helps to distance from the commonly made assumptions of the deterministic models in the marketing literature. The closed-form results of our model can also serve as decision tools to help the web publishers running advertising operations, for instance, by providing an extra layer of intelligence on top of their pricing engine software.

The remainder of Chapter 2 is organized as follows: The next section provides the relevant literature. Section 2.3 describes the model formulation. Section 2.4 discusses the web publisher's revenue maximization problem and Section 2.6 details the numerical study. Section 2.7 presents some extensions to the publisher's problem and Section 2.8 concludes and presents directions for future research.

2.2 Literature Review

There are two streams of literature related to our research. The first is the literature on online advertising within the *marketing* area, which is quite extensive. Novak and Hoffman (2000) provide an overview of advertising pricing schemes for the internet. However, there is limited literature on analytical models for optimal pricing and other decision making for a web publisher with an advertising operation. (For issues faced by advertisers such as predicting audience for advertising campaigns see, e.g., Danaher (2007) and papers referenced therein.)

The second stream of literature is on *management science*. The online advertising research within this area is limited and there are few works directly related to online advertising pricing.

In some of the earlier work, Mangàni (2003) compares the expected revenues from the CPC and the CPM schemes using a simple deterministic model. Unlike this dissertation, he does not consider the uncertainties involved with the advertisers' demands and viewers' supplies. At the same time, Chickering and Heckerman (2003) develop a delivery system that maximizes the CTR given inventory-management constraints in the form of advertisement quotas. Both of these papers assume the prices are fixed. In the first chapter, we determine the optimal price for the CPM system. However, our model there is not applicable to the CPC system. The main reason is that unlike the CPC system, the service rate for each advertiser remains fixed in the CPM system. Moreover, in the CPM system the advertisers receive service in a synchronized fashion, which does not occur in the CPC system. Lastly, the service-time for each advertiser in the CPM system is Erlang while in the CPC system it does not have the properties of Erlang.

There has been some recent literature on online search, the other section of the online advertising market. Johnson et al. (2004) consider an empirical study to examine the dynamics of online search behavior. In addition, Ghose and Yang (2009) provide an empirical analysis of search engine advertising for the sponsored searches on the internet. None of the results in these two papers can be extended to ours, as they do not develop analytical models for the price decisions. Moreover, the nature of search advertising is fundamentally different from display advertising, as it is mainly based on using auctions that we do not consider here.

Some researchers have focused on the relevant problem of pricing of goods and services on the internet. Brynjolfsson and Smith (2000) and Clemons et al. (2002) conduct empirical evaluations of price dispersions and price differentiations on the internet. Bakos and Brynjolfsson (1999, 2000) study the optimal strategies of the products bundling for a retailer selling products through the internet. Dewan et al. (2000) and (2003) examine the problem of optimal product customization and price strategy both in monopoly and in competition. Jain and Kannan (2002), and Sandararajan (2004) analyzed the optimal pricing of information goods from the economics and game theoretic standpoint. Although all of these papers consider a variety of online pricing problems, none are applicable to the CPC system, as the settings in these papers are for quite different problems.

Some authors have considered the problem of a web publisher who not only generates revenues from advertising but also from subscriptions. Baye and Morgan (2000) develop a simple economic model of online advertising and subscription fees. Prasad et al. (2003) model two offerings to viewers of a website: a lower fee with more ads and a higher fee with fewer ads. Kumar and Sethi (2008) study the problem of dynamically determining the subscription fee and the size of advertising space on a website. They use optimal control theory to solve the problem and obtained the optimal subscription fee and the optimal advertisement level over time. Unlike our thesis, all these papers are focused on capacity management problems not price decisions, and the price is assumed to be fixed.

Scheduling of ads on a website has also recently become a popular topic. Kumar et al. (2008) develop a model that determines how ads on a website should be scheduled in a planning horizon to maximize revenue. They consider geometry and display frequency as the two most important factors specifying the ads. Their problem belongs to the class of NP-hard problems, and they develop a heuristic to solve it. They also provided a good overview of other related papers on scheduling.

As mentioned earlier, the queuing system developed in this chapter to characterize the CPC system is new. Relatively, few papers in the queuing literature consider systems similar to the CPC system. Green (1980), Brill and Green (1984), Courcoubetis and Reiman (1987), and Hong and Ott (1989) study systems with simultaneous system requirements. These papers are all related to the CPM system developed in the first chapter. Nevertheless, none of these considers the constantly changing service rates' phenomenon as occurs in the CPC system.

Finally, we end this section by a short review of related work in revenue management. For a comprehensive reference of the traditional revenue management models, we refer the reader to the book by Talluri and van Ryzin (2004) (the book does not cover the online setting). Savin et al. (2005) consider revenue management for rental businesses with two customer classes. Although considering a different problem, they have assumed uncertainty in the customers demand to their model, which has some similarity to our model. Araman and Popescu (2009) also study revenue management for traditional media, specifically broadcasting. Their model is concerned with how to allocate limited advertising space between up-front contracts and the so-called scatter market (i.e., a spot market) in order to maximize profits and meet contractual commitments. Unlike our thesis, both of these papers are concerned with the capacity decisions and price is not an issue of focus.

In the next section, we discuss about the main model.

2.3 The Model

We consider a web publisher facing uncertain demand from advertisers requesting space to display their ads[5]. Advertisers require a certain number of viewers to click

[5] Note that the advertisers' demand has two layers in nature: First of all, each advertiser requests a space for his ad. For instance, this can be one of the empty slots in the publisher's website.

on their ads. The supply of viewers is uncertain as well as their clicking behavior. The advertisers' demands are sent through an ad network[6]. The ad network supplies the web publisher with advertisers as long as the publisher has space available. If no space is available the network does not assign ads to that publisher. This implies that the publisher's website is a *loss* system (see the Appendix for details on the matching process of ad networks). Our model also applies to the setting where direct sales channels are used with advertisers not willing to wait for space to become available, which is common in the intense competition of web publishers for advertisers.

A web publisher often charges different prices based on the size of the ad, the page on which the ad is posted, and the ad's allocated position on the page; e.g., the leaderboard (the horizontal banner at the top) on the homepage of a news site is more expensive than a small square at the bottom of the lifestyle page. Hence, when a web publisher registers with an ad network it classifies similar advertising slots that are charged *the same price* and registers each group with a separate tracking code.

We assume the web publisher's website (the system) contains J pages labeled from 1 to J. For example, for a news site these pages could correspond to the business

Secondly, after an advertiser is provided with an empty slot, he requests an uncertain number of clicks. Hence, the advertisers' real demand is the number of clicks but given the publisher's space constraint. On the other side, the viewers' supply is also the number of clicks. As a result, both supply and (real) demand are of the same nature.

[6] In many instances, the ad networks guarantee the requested number of clicks. However, some ad networks such as Yahoo! do not explicitly guarantee the number of clicks. Instead, in its contracts, a sales representative of Yahoo! simply converts the requested number of clicks into the number of impressions based on his/her own estimation (no methodology or decision making tools are applied, and this estimation is completely based on the sales person's own estimation). If after delivering the number of impressions, the requested number of clicks is not satisfied, the sale representative assigns extra impressions to be delivered for the second time based on his estimation, and completely free of charge. However, if the company sees that the number of clicks is not met after a few times of adding extra impressions (bad ads), it ends the contract one-sided. It is quite interesting to note that in Yahoo's sales approach, though the company does not guarantee the number of clicks, it delivers much more numbers of impressions and clicks than requested by advertisers completely free of charge in order to satisfy the advertisers.

page, travel page, etc. Each page can have several groups of ads where the same price is charged within each group. For instance, the top of the page can display two equally sized ads, while several small ads can be placed at the bottom (rectangles). This leads to two ad groups. More formally, for each page j we group the ads into M^j groups (the subsystems) of equivalent slots, where each subsystem m, $1 \leq m \leq M^j$, contains $n^{j,m}$ equivalent slots. We denote by $\lambda^{j,m}$ the rate with which the advertisers arrive requesting space in subsystem (j, m). An advertiser requesting a slot in group m on page j requires his ad to be posted on the website until clicked $X^{j,m}$ times by the viewers. $X^{j,m}$ is a random variable, which varies from one advertiser to the next. We denote the expected value of $X^{j,m}$ with $\mathbb{E}(X^{j,m}) = x^{j,m}$.

Publishers do not usually leave a slot empty; rather they place a "default" ad in there. A default ad (or a filler ad) can often be the publisher's own ad or a run-of-network ad that the ad network sends to fill the place. In both cases, a default ad generates a minimal revenue. Hence, when a revenue generating ad is sent to the publisher the filler ad would be replaced by a proper revenue generating ad.

Let $p^{j,m}$ denote the price per click for a banner posted in subsystem (j, m). The subsystems on page j are differentiated by a set of attributes, e.g., the size, the location, and the relevance of the content offered. The viewers' preference for these attributes could have them consider and click on a banner from one of the subsystems.

We denote the traffic rate of viewers to a page j by μ^j. Next, we propose a choice model capturing the viewers "click behavior".

Viewers Choice Model When arriving at page j, a viewer considers ads in each subsystem based on the attributes of the ads in the subsystem and his preference for these attributes. We model the viewers' preference by the coefficient vector

$V^T = (1/v_1, ..., 1/v_l) \geq 0$ (see Anderson et al. 1992), where each component indicates the preference weight that viewers give to each attribute (such as size and location). Then the viewer's choice to consider the ads in subsystem (j, m) (see, e.g., Danaher and Mullarkey 2003; Lohtia et al. 2003 for similar applications) can be expressed by the following Multinomial-Logit (MLN) function:

$$\varpi_{jm} = \frac{\exp(-A_{jm}^T V)}{\sum_{j=1}^{J} \sum_{m=1}^{M^j} \exp(-A_{jm}^T V)}, \tag{2.1}$$

where ϖ_{jm} is the probability that a viewer selects subsystem (j, m). In this formula $A_{jm}^T = (a_{jm}^1, ..., a_{jm}^l)$ is the attributes' vector with a_{jm}^i, $i = 1, ..., l$, referring to the magnitude of each attribute of ads in subsystem (j, m), such as size and location (see, e.g., Talluri and van Ryzin (2004a, 2004b) and Vulcano and van Ryzin (2010) for similar applications of MLN choice models to describe individuals' choice behavior).

We note that when v_i tends to zero for a certain attribute i the choice probability in Equation (2.1) depends only on attribute i. Alternatively, when v_i is very high, the viewers become insensitive to attribute i. In the same way, if for *all* attributes, v_i tends to infinity, the viewers become indifferent towards the attributes and consider ads in all subsystems with an equal probability.

As previously mentioned, viewers arrive at the publisher's website with rate μ per time unit and their attentions are captured by subsystem (j, m) based on the choice model described by Equation (2.1). We naturally assume that viewers always prefer to consider and click on real ads compared to default (filler) ones in a subsystem. For example, CNN.com or Financial Times frequently display their own default ads. These ads are often not designed to be clicked on as a publishers' major aim from displaying default ads is often to strengthen its own brand recognition. In addition, if viewers choose subsystem (j, m) but it only has filler ads, we assume that they also consider the ads in another subsystem (g, h), (the ads in subsystem g on page h) with probability $\alpha_{j,m}^{g,h}$, or leave the website without considering any ads with probability $1 - \alpha_{j,m}^{g,h}$.

Rotation of ads The publisher can often serve more advertisers than there are slots. For example, two ads could share the same slot with each ad displayed to the viewers randomly based on pre-assigned display weights. Random weight-based ad rotations are commonly used by ad management software such as Double-Click for Publishers (DCP) by Google (DCP 2010). We denote $s^{j,m}$ as the number of sets of ads being served in subsystem (j, m). In other words, each slot in subsystem (j, m) is randomly rotated among $s^{j,m}$ ads.

The Optimization Problem The publisher's goal is to maximize its total revenue rate by determining the right prices to charge. The revenue rate for each subsystem consists of the payments made by the advertisers multiplied by the "effective" demand rate for that subsystem. Each payment consists of the price per click, denoted by $p^{j,m}$, multiplied by the number of clicks requested, $X^{j,m}$. We capture the price-sensitivity of the advertisers with the price-demand function, $p^{j,m}(\lambda^{j,m})$, which is assumed to be continuous and decreasing in the arrival rate of the advertisers. (In Sections 2.7.2 and 2.7.1 we consider the price also to depend on the number of clicks.) Even though it might not be trivial for the publisher to determine this function, we assume it can do so with trial and error. For instance, ad networks often encourage publishers to start by offering low prices and then gradually increase them to the appropriate value. The process of advertisers being matched to web publishers based on type preference and willingness-to-pay can be modeled specifically. However, ultimately it will lead to a price-demand relationship. We will not model the process in detail here but in the Appendix we provide a description, from one of the ad networks, of the matching process.

Note that an advertiser chooses his desired subsystem in advance when registering with the ad network. For instance, he may request a right hand side banner on the sport page. If that subsystem is fully occupied at the publisher's site then the network does not offer slots in this subsystem. Given that the publisher registers each subsystem separately with the ad network, we consider the advertisers' demand for each subsystem to be independent.

In addition, it is common that only a part of the advertisers' demand per time unit can usually be met by the publisher. This means that, the real demand rate for each subsystem is scaled down by the probability that there are advertising slots available. We denote the probability of having i advertisers in subsystem (j, m) by $\mathbb{P}_i^{j,m}$, $i \in \{0, ..., s^{j,m}n^{j,m}\}$. Note that a total of $s^{j,m}n^{j,m}$ advertisers can be served with $s^{j,m}$ advertisers sharing the same slot on a random display basis.

As we have a one-to-one relationship between the prices and the arrival rates of the advertisers, we will optimize the revenue rate with respect to the arrival rates and then determine the prices from the price-demand functions, $p^{j,m}(\lambda^{j,m})$. The optimization problem of the publisher of maximizing its expected revenue rate can be formulated as follows:

$$\max_{\mathbf{\Lambda}_1,...,\mathbf{\Lambda}_J} R(\mathbf{\Lambda}_1, ..., \mathbf{\Lambda}_J) = \sum_{j=1}^{J} \sum_{m=1}^{M^j} \lambda^{j,m}(1 - \mathbb{P}^{j,m}_{s^{j,m}n^{j,m}}(\lambda^{j,m}; X^{j,m}, n^{j,m}, s^{j,m}, \mu^{j,m}_{eff}))p^{j,m}(\lambda^{j,m})\mathbb{E}(X^{j,m})$$

$$\mathbf{\Lambda}_j = \left(\lambda^{j,1}, ..., \lambda^{j,M^j}\right)^t \in [0, +\infty)^{M^j}, \quad j = 1, ..., J. \tag{2.2}$$

In this formula $\mathbb{P}^{j,m}_{s^{j,m}n^{j,m}}$ is the probability that the subsystem (j,m) is fully occupied. Therefore, $\lambda^{j,m}(1-\mathbb{P}^{j,m}_{s^{j,m}n^{j,m}}(\lambda^{j,m};X^{j,m},n^{j,m},s^{j,m},\mu^{j,m}_{eff}))$ is the effective advertisers' arrival rate at subsystem (j,m). In addition, $\mu^{j,m}_{eff}$ is the viewers' effective arrival rate at subsystem (j,m), which is obtained through the following proposition.

Proposition 25 *The effective viewers' arrival rate at subsystem (j,m), $\mu^{j,m}_{eff}$, is given by $\mu^{j,m}_{eff}=\mu(1-\mathbb{P}^{j,m}_0)\mathbb{P}^{j,m}_A$, where $\mathbb{P}^{j,m}_A$, the probability that subsystem (j,m) is considered, is given by*

$$\mathbb{P}^{j,m}_A=\varpi_{j,m}+\sum_{\substack{g=1\\(g,h)\neq(j,m)}}^{J}\sum_{h=1}^{M^j}\left(\varpi_{g,h}\mathbb{P}^{g,h}_0\alpha^{j,m}_{g,h}+\frac{\alpha^{g,h}_{j,m}\alpha^{j,m}_{g,h}\mathbb{P}^{j,m}_0\mathbb{P}^{g,h}_0(\varpi_{j,m}+\varpi_{g,h})}{1-\alpha^{g,h}_{j,m}\alpha^{j,m}_{g,h}\mathbb{P}^{j,m}_0\mathbb{P}^{g,h}_0}\right), \tag{2.3}$$

where μ is the rate at which advertisers consider all subsystems and $\alpha^{g,h}_{j,m}$ ($\alpha^{j,m}_{g,h}$) is the fraction of the advertisers that faced filler ads in subsystem (j,m) (subsystem (g,h)) who approach subsystem (g,h) (subsystem (j,m)).

The proof of this proposition and other results are provided in the Appendix.

In order to obtain the optimal CPC price, we first need to characterize $\mathbb{P}^{j,m}_{n^{j,m}s^{j,m}}$ for each subsystem (j,m) [7].

Probability Distribution Recall that advertisers' and viewers' arrivals to subsystem (j,m) are Poisson [8]. This property is important as it helps us to characterize $\mathbb{P}^{j,m}_{n^{j,m}s^{j,m}}$ by modeling subsystem (j,m)'s dynamics using Markov transitional balance equations. In our analysis, we look at each subsystem (j,m) as a queuing system

[7] Note that we analyze the operation of the web publisher from a steady state point of view. Dynamic pricing would also be possible to consider. Nevertheless, according to our discussions with practitioners, the publishers price decisions are not frequently reset, so tends to have more a static nature, rather than a dynamic one.

[8] We will relax this assumtion later.

where the slots correspond to serving channels. In this queuing system, the arriving advertisers request to be served. Each viewer triggers part of the service to one of the advertisers by clicking on his ad. Each advertiser completes his service when $X^{j,m}$ viewers click on his ad. For convenience, during this section we refer to the arbitrary subsystem (j, m) as a *system* and drop all indices j and m. Without loss of generality, we also set $s = 1$. To be able to fully characterize each of these queuing models, we make the following assumptions.

Assumption 1 *The advertisers' demands follow a Poisson process with a stationary rate* λ. In practice the Poisson process might be inhomogeneous, i.e., λ can be a deterministic function of time, or it can even be a doubly stochastic process, where λ itself is a random variable or constitutes a stochastic process. However, it is easy to verify that the corresponding state-space and the transition equations become too complicated even for very simple special cases. As a result, we restrict our focus only to the homogenous Poisson process (but in Section 2.7.2 we conduct a simulation analysis for the *stochastic* rates and in Section 2.7.1 for *non-Poisson* arrivals on both advertisers' and viewers' sides).

Assumption 2 *The viewers visit the publisher's system based on a Poisson process with stationary rate* μ. This assumption has been criticized in the literature as some research supports that web traffic shows self similarity, long range dependence and heavy tailed distribution, which are not properties of the Poisson process (see, e.g., Gong et al. 2005). Nevertheless, there are several studies that recognize that the Poisson distribution is an appropriate assumption (see, e.g., Cao et al. 2002). We construct our main model assuming Poisson arrivals. However, we show in Section 2.7.1 that our results provide accurate estimates for the publisher's model even when the viewers' arrivals are non-Poisson.

Assumption 3 *Advertisers request the same number of clicks, i.e.,* $\mathbb{E}(X) = x$. We make this assumption for tractability reasons. However, in Section 2.7.2, we consider different generalizations of this assumption and show through simulations that even if the advertisers choose different numbers of clicks according to the random variable X and are charged a price depending on X, the problem can be well approximated by assuming that all advertisers request $x = \mathbb{E}(X)$ with a single price charged.

Having Markovian arrival and service processes we can now model the system using Markov chains. Note that even though we are ultimately interested in keeping track of the number of advertisers in the system, in order to model the system's dynamics we need to keep track of the system at a more detailed level; the number of clicks left to be delivered for each slot within the subsystem. When an advertiser arrives, he is randomly assigned to one of the available slots with equal probability as they are equivalent. This random ad-to-slot allocation means that we can keep track of the dynamics of the system without distinguishing between the slots. Let us define the state of the system and its transitions.

We formulate the system as a queuing model with the state vector

$$\mathbf{k} =(k_1, k_2, ..., k_n), \quad 0 \leq k_h \leq x, \tag{2.4}$$

in which each component represents the number of clicks left to satisfy in one of the slots without distinguishing among the slots. For instance, k_h indicates that there is an ad in the system, which needs to be clicked k_h times more to leave the system. If $k_h = 0$, it indicates that the corresponding slot is empty. Alternatively, if $k_h = x$ it indicates that an ad of a new advertiser has just been placed in the slot. Note that as we do not distinguish between the slots (all slots in the subsystem are equivalent) any combination of the same components does not lead to a new state. For example, $(3,4,2)$, $(4,3,2)$, and $(2,3,4)$ all refer to the same state. In order to illustrate how the state transitions work, we take the following examples:

Suppose that the system is in the state $(k_1, k_2, ..., k_i, 0, ..., 0)$, where the first i components are positive and the rest is zero (empty slots). The viewers consider the system with the effective rate μ_{eff} (see Proposition 1). We then assure that each viewer clicks on one of the ads in the system with the probability β, or leaves the subsystem with $1-\beta$. Given that a viewer clicks on one of the ads, each of the i equivalent ads has an equal chance, $1/i$, to be clicked. As a result, the state of the system makes a transition to the new state

$$\mathbf{k}' = (k_1 - 1, k_2, ..., k_i, 0, ..., 0)$$

with rate $\widehat{\mu}/i$, where $\widehat{\mu} = \mu_{eff}\beta$. Next, consider the state of the system to be $\mathbf{k} = (\underbrace{k_1, k_1, ..., k_1}_{i}, k_{i+1}, ..., k_h, 0, ..., 0)$. We observed that in state $\mathbf{k}$, i ads have the same number of clicks left. Since we do not distinguish among the ads, the viewer can click on one of the ads in this group with an i/h chance, while the other ads have a $1/h$ chance each to be clicked. As a result, the state of the system makes a transition to the new state

$$\mathbf{k}' = (k_1 - 1, \underbrace{k_1, ..., k_1}_{i-1}, k_{i+1}, ..., k_h, 0, ..., 0)$$

with rate $i\widehat{\mu}/h$. Lastly, consider the state of the system to be $\mathbf{k} = (k_1, k_2, ..., k_i, \underbrace{0, ..., 0}_{n-i})$. Now, if an advertiser arrives at the subsystem, the publisher assigns one of the empty slots to his ad, and the state will make a transition to $(k_1, k_2, ..., k_i, x, \underbrace{0, ..., 0}_{n-i-1})$ with rate λ.

In order to find $\pi_{\mathbf{k}}$, the probability of finding the system in state $\mathbf{k}$, we characterize all possible states and transitions of the system and solve the flow balance equations.

Note that the publisher's system is significantly different from other systems studied so far in the queuing literature. The reason for this difference is that each advertiser receives his service not with a fixed rate, but with a constantly changing rate, as the probability of viewers considering an ad depends on the number of displayed ads in the system at any point in time. For example, if there are three ads displayed, since the ads are equivalent, each ad has a one-third chance of absorbing a viewer's attention; whereas if there are five, the chance reduces to one-fifth. The constantly changing service rate, which depends on the state of the system, complicates the analysis of this system.

For the purpose of the next proposition, we index the state $\mathbf{k}$'s components from $h = 1$ to $h = n$. For example, if $\mathbf{k}$ =$(2, 3, 5)$ then $h(2) = 1$, $h(3) = 2$, and $h(5) = 3$. Furthermore, we define $\mathcal{G}_c(\mathbf{k}) = \{h \mid k_h = c\}$ to be the set of slots in the subsystem whose number of remaining clicks are c, $c = 0, 1, ..., x$, where h refers to the index of c. $|\mathcal{G}_c(\mathbf{k})|$ refers to the size of $\mathcal{G}_c(\mathbf{k})$ indicating how many slots in the system have remaining clicks equal to c. For instance if $\mathbf{k}$ =$(2, 2, 3, 5)$ then $|\mathcal{G}_2(\mathbf{k})| = 2$, $|\mathcal{G}_3(\mathbf{k})| = 1$, and $|\mathcal{G}_5(\mathbf{k})| = 1$. The next proposition gives the closed form solution of the steady-state probability of the number of advertisers in the system.

Proposition 26 *Let k_s $(0 \leq k_s \leq x)$ be the number of clicks left in slot s $(0 \leq s \leq n)$ in the system. Define the state of the system as $\mathbf{k} = \sum_{s=1}^{S} c_s \mathbf{v}_{c_s}$, with $\mathbf{v}_{c_s} \triangleq \sum_{h \epsilon \mathcal{G}_{c_s}(\mathbf{k})} \mathbf{e}_h^T$, and $\mathcal{G}_{c_s}(\mathbf{k}) = \{h \mid k_h = c_s\}$ (i.e., the set of components in $\mathbf{k}$ with value*

c_s). Furthermore, define $|\mathcal{G}_{c_s}(\mathbf{k})|$ as the size of $\mathcal{G}_{c_s}(\mathbf{k})$ $(0 \leq c_s \leq S)$ (i.e., the number of components in $\mathbf{k}$ whose values are c_s), where S is the number the of the groups of slots whose remaining clicks are the same. Then the steady-state probability of the system for state $\mathbf{k}$ exists, and is expressed as:

$$\pi_{\mathbf{k}} = \frac{(\sum_{s=1}^{S} |\mathcal{G}_{c_s}(\mathbf{k})|)!}{\Pi_{s=1}^{S} |\mathcal{G}_{c_s}(\mathbf{k})|!} r^{\left(\sum_{s=1}^{S} |\mathcal{G}_{c_s}(\mathbf{k})|\right)}, \quad r = \frac{\lambda}{\widehat{\mu}}. \tag{2.5}$$

Furthermore, the steady-state probability of having i advertisers in the system is:

$$\mathbb{P}(i) = \frac{(rx)^i}{\sum_{j=0}^{n} (rx)^j}, \quad i = 0, 1, 2, ...n. \tag{2.6}$$

It is surprising to see that Equation (2.6) in Proposition 26 (the probability of the number of jobs) coincides with that of an $M/M/1/n$ system with $\rho = rx$, where ρ is often regarded as the traffic intensity. This coincidence is interesting since the two systems have considerably different characteristics. Now, let us explore the reason for this coincidence. we consider a similar system where the advertisers' arrival rate and the requested number of clicks take the values $\lambda_1 = \lambda x$, and $x_1 = 1$ respectively. It is easy to verify that Equation (2.6) remains unchanged. However, having $\lambda_1 = \lambda x$ and $x_1 = 1$ suggests that the advertisers arrive at the system with rate λ_1 and request only one click. In Figure 2.1 we setup a transition diagram for this system where the state vector is an n-tuple vector with i ones (which indicate there are i advertisers in the system) and $n - i$ zeros. That is, the state of the subsystem is expressed as $\mathbf{k} = (\overbrace{1, 1, 1..., 1}^{i}, 0, ..., 0)$.

The state of this new system can be collapsed into a one dimensional state space based on the number of advertisers in the system (rather than the vector $\mathbf{k}$). This

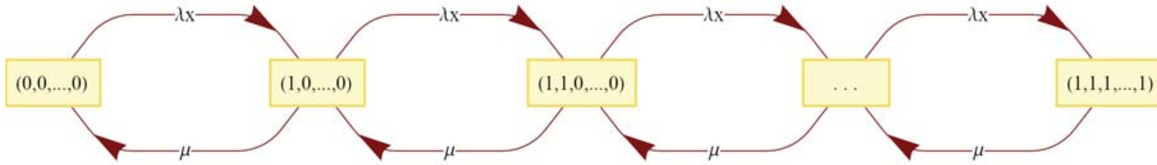


Figure 2.1: An illustration of the CPC system transition diagram while the advertisers arrival rate and the requested numver of clicks are $\widehat{\lambda} = \lambda x$, and $\widehat{x} = 1$ respectively. It is easy to verify that the probability distribution of the number of the advertisers in the system does not change in comparison to the CPC system in which the advertisers' arrival rate is λ and the requested number of clicks is x.

system is indeed the $M/M/1/n$ queuing system. It is quite interesting that two systems with very different dynamics have the same steady state probability of the number of jobs in the system.

The other interesting observation regarding Proposition 26 is that $\pi_{\mathbf{k}}$ does not depend on the actual clicks remaining in each slot. This result is consistent with the first chapter, where we show that in the CPM setting the probability distribution of the number of ads is independent of the number of impressions remaining for each ad.

In the next two propositions, we show some structural properties of the average number of advertisers in the system and the busy probability through the next two propositions. They will be useful when considering the pricing problem of the web publisher in the next section.

Proposition 27 *$\forall x^{j,m}, n^{j,m}$ in the subsystem (j,m) the full state probability, $\mathbb{P}_n^{j,m}$ defined by (2.6) satisfies:*

(i) $\frac{\partial \mathbb{P}^{j,m}_{n^{j,m}}}{\partial r^{j,m}} \geq 0$,

(ii) $\mathbb{P}^{j,m}_{n^{j,m}}(x^{j,m}+1) - \mathbb{P}^{j,m}_{n^{j,m}}(x^{j,m}) \geq 0$,

(iii) $\mathbb{P}^{j,m}_{n^{j,m}+1}(x^{j,m}) \leq \mathbb{P}^{j,m}_{n^{j,m}}(x^{j,m})$.

This proposition is quite intuitive as one would expect the value of full state probability at any point in time to be decreasing in the number of slots, and at any given number of slots to increase in both the intensity rate and number of clicks. Nevertheless, we shall show in Section 2.6 that $\mathbb{P}^{j,m}_{n^{j,m}}$ is not necessarily concave in the number of clicks.

Proposition 28 *Using Proposition (27) the average number of advertisers in the subsystem* (j, m)*,* $L^{j,m}_{n^{j,m}}(x^{j,m})$*, and the increment* $\Delta L^{j,m}_{n^{j,m}}(x^{j,m}) = L^{j,m}(x^{j,m} + 1) - L^{j,m}(x^{j,m})$ *satisfies* $\forall x^{j,m}, n^{j,m}$:

(i) $\Delta L^{j,m}_{n^{j,m}}(x^{j,m}) \geq 0,$

(ii) $\Delta L^{j,m}_{n^{j,m}}(x^{j,m} + 1) \leq \Delta L^{j,m}_{n^{j,m}}(x^{j,m}), \quad r^{j,m}x^{j,m} > 1,$

(iii) $\frac{\partial L^{j,m}}{\partial r^{j,m}} \geq 0, \frac{\partial^2 L^{j,m}}{\partial r^{j,m^2}} \leq 0, \quad r^{j,m}x^{j,m} > 1,$

(iv) $L^{j,m}_{n^{j,m}}(x^{j,m}) \leq L^{j,m}_{n^{j,m}+1}(x^{j,m}).$

Proposition (28) (i) and (ii) imply that the average number of advertisers in the web publisher's system is increasing concave in the number of impressions. Furthermore, (iii) implies that the average number of advertisers in the system is increasing concave in the intensity rate $r^{j,m}$. Part (iv) also mentions that the average number of advertisers in the web publisher's system increases in the number of slots.

2.4 The Optimal Price

Having fully characterized the probabilistic properties of the web publisher's operation in one of the subsystems, we now turn to the question of finding the optimal pricing policy for the web publisher. Recall that $\alpha^{g,h}_{j,m}$ is the probability that a viewer who considers subsystem (j, m) and is faced with default ads, could check subsystem (g, h) as well. For the purpose of tractability, at this stage we consider that the value of $\alpha_{j,m}^{g,h}$ is sufficiently small to be ignored. This means that we focus on the case, where there are no inter-flow rates between the subsystems, and hence $\mu_{eff} = \mu$. Note that it is common in practice that there are no inter-flow between two subsystems, which are located on two different pages. Generally, an average of 30% to 80% of websites' viewers consist of those who only visit a single page and then leave the system (Alexa 2010). This percentage is often regarded as *Bounce Rate*. In addition, a large publisher such as Yahoo! provides only one subsystem in most of its pages, for instance games pages (McAfee 2010). Obviously, the viewers visiting the games' page and facing default ads usually do not consider them. They rather tend to stay in the same page to continue playing their games. As a result the interflow rate becomes ignorable.

Recall that the advertisers' arrivals at the subsystems are independent. The reason is that advertisers choose their targeted subsystem in advance when they register their ad with the ad network. Note that as the advertisers' and viewers' arrivals at each subsystem are independent from other subsystems, the revenue maximization problem indicated in (2.2) becomes separable in each subsystem. Therefore, we conveniently restrict our focus to one subsystem only.

The web publisher's objective is to determine the optimal price to charge per click that maximizes the revenue rate. (Without loss of generality we will ignore any cost components.) The revenue objective function in (2.2) can be characterized by the following formulation:

$$\begin{aligned} \underset{\lambda}{\text{Max}}\ R(\lambda) &= \lambda(1-\mathbb{P}_n(\lambda;\widehat{\mu},n,x))p(\lambda)x, \\ &s.t. \\ \lambda &\in [0,+\infty). \end{aligned} \tag{2.7}$$

The following proposition ensures the existence of the optimal solution and gives the optimal price.

Proposition 29 *Let $p(\lambda)$ be a nonnegative concave decreasing function defined on the support $[0,+\infty)$. Then the revenue rate $R(\lambda)$ define by (2.7) is a nonnegative concave function of λ. Furthermore, the optimal advertisers' arrival rate λ^* satisfies:*

$$\left.\frac{\partial L(\lambda)}{\partial \lambda}\right|_{\lambda^*} p(\lambda^*) + \left.\frac{\partial p(\lambda)}{\partial \lambda}\right|_{\lambda^*} L(\lambda^*) = 0. \tag{2.8}$$

The first part of Proposition (29) guarantees the concavity of the objective function defined by (2.7). Note that in order to ensure concavity we need $p(\lambda)$ to be concave. Even though this might seem a restrictive assumption it includes a *linear price* which is widely applied in Economics and Management Science literature. In addition, the numerical analysis indicates that many convex price functions give a unimodal revenue function as well. (Other weaker conditions such as assuming concave payment rate $\lambda p(\lambda)$ or monotonicity of the price elasticity $-\frac{\partial\lambda}{dp}\frac{p}{\lambda}$ do not seem sufficient.) Furthermore, the second part implies that at the optimal arrival rate, the relative change in the average number of advertisers in the system is the opposite of the relative change in the optimal price.

The next proposition gives the insightful result that the web publisher is better off with having more slots, offering higher number of clicks and having more traffic on his website. In the following proposition, we denote $R(\lambda^*)$ by $R_{n,x}(\lambda^*(n,x);\widehat{\mu})$ to emphasize n, x, and $\widehat{\mu}$.

Proposition 30 *The optimal revenue rate $R_{n,x}(\lambda^*(n,x);\widehat{\mu})$ defined by 2.7 satisfies*

(i) $R_{n,x}(\lambda^*(n,x);\widehat{\mu}) \leq R_{n+1,x}(\lambda^*(n+1,x);\widehat{\mu})$,

(ii) $R_{n,x}(\lambda^*(n,x);\widehat{\mu}) \leq R_{n,x+1}(\lambda^*(n,x+1);\widehat{\mu})$,

(iii) $R_{n,x}(\lambda^*(n,x);\widehat{\mu}_1) \leq R_{n,x}(\lambda^*(n,x);\mu_2), \quad \widehat{\mu}_1 \leq \widehat{\mu}_2$.

The following proposition states the counter-intuitive result that the optimal price does not follow the economies-of-scale property with respect to x.

Proposition 31 *If the price-demand function, $p(\lambda) \in \mathcal{C}^2[0,+\infty)$, is concave decreasing in the advertisers' arrival rate, λ then:*

(i) λ^* is decreasing in x,

(ii) $p(\lambda^*)$ is increasing in x.

The proposition above is quite interesting as one could expect the opposite, i.e., the price to be lower when more clicks are offered. In order to understand what drives these results, we observe that there are two competing forces. First, the higher the number of clicks the longer it takes to serve each advertiser, which means that the web publisher does not need as many advertisers as before. Second, a higher number of clicks makes fewer advertisers interested, i.e., λ^* is an implicit monotone decreasing function of x. Therefore, the web publisher is more likely to face more empty spaces in the long-run. However, the first effect seems to always dominate, which results in a higher price with lower demand. Practically speaking, the web publisher should not offer quantity discounts from an operational point of view within the advertising network. All the same, there could be marketing reasons for offering a quantity discount. We will explore these in Section (2.6).

2.5 The Simple Heuristic Pricing

Many web publishers that promise based on numbers of clicks tend to use a simple heuristic approach for obtaining the CPC prices. This heuristic is based on dividing the optimal CPM prices by the CTR[9] to calculate the optimal CPC prices[10]. In this section, we study the shortcomings of this heuristic by comparing its revenue with the optimal revenue obtained by using the correct model.

Let us look at the simple heuristic in more detail. Consider a publisher's system that has advertisers arriving with rate λ, viewers arriving with rate μ, and each advertiser requesting x clicks. In addition, let β be the fraction of the viewers clicking on one of the ads in the subsystem considered before leaving. Since the publisher does not know how to obtain the optimal CPC price (p^*_{cpc}) directly, it charges the scaled CPM price $p^*_{cpc} = p^*_{cpm}/CTR$ as the publisher considers selling x clicks to be on average equivalent to selling $N = x/CTR$ impressions. We give the publisher the benefit of the doubt and assume that the publisher knows to how to obtain the optimal CPM price correctly[11].

In order to examine this popular approach, we consider a CPC system at which

[9] The CTR is the probability that an ad is clicked. In practice, it is generally calculated by dividing the number of viewers that have clicked on a certain ad by the total number of visitors to the publisher's system.

[10] Examples of real publishers implementing this campaign include Clickz.com. Clickz.com also provides a special page to help users calculate the CPC price using the CPM price and the CTR. The page is available at: http://www.clickz.com/cpa-calculator

[11] Obviously, using a wrong CPM price will only lead to an increased error.

the advertisers arrive with rate λ, and the viewers arrive with rate $\mu = 100$. The numbers are illustrative and only for the purpose of example. We set the price function per impression to be $p_{cpm}(\lambda) = 0.005 - 0.01\lambda^c$. In order to determine the CPC prices, a publisher using the CPM prices, considers a CPM system at which the advertisers arrive with rate λ, the viewers arrive with rate $\mu = 100$, and the number of impressions sold to each advertiser are $N = x/CTR$. Then the CPC price it offers would be $p_{cpc}(\lambda) = p^*_{cpm}(\lambda)/CTR$. The CTR used in practice is the observed value for the ratio of the number of people clicking on a certain ad to the total number of people visiting the publisher's system over a certain period, for instance one week. In other words, the CTR value that practitioners observe is the average chance that an ad would have in order to be clicked in a stable condition. The following proposition gives the CTR value that a publisher observes in the log-run.

Proposition 32 *The observed CTR value in the long-run converges to*

$$CTR(\lambda, \mu, x, n, \beta) = \beta \sum_{i=1}^{n} \frac{1}{i} \frac{\mathbb{P}_i^{cpc}(\lambda, \widehat{\mu}, x, n)}{1 - \mathbb{P}_0^{cpc}(\lambda, \widehat{\mu}, x, n)}, \tag{2.9}$$

where $r = \lambda/\widehat{\mu}$, $\widehat{\mu} = \mu\beta$, and $\mathbb{P}_i^{cpc}$ is the probability of having i ads in the publisher's system.

As can be seen in the above proposition, the value of the observed CTR depends on advertisers' and viewers' effective arrival rates, the number of requested clicks, and the number of slots in its system. Note that in this proposition, β/i refers to the expected probability of an ad to be clicked (the actual CTR) on each visit of viewers considering a particular subsystem when there are i ads displayed. Moreover, $\mathbb{P}_i^{cpc}/1 - \mathbb{P}_0^{cpc}$ refers to the proportion of the time that there are i ads in the publisher's system given that the system has at least one ad. Obviously, the reason for considering the conditional probability is that the actual CTR is non-existent for the periods that the publisher's system is empty.

Furthermore, Proposition 32 is quite interesting as it implies that the CTR's observed value, in fact, changes with the price. By changing the price, the publisher is in fact affecting the advertisers' arrival rate λ. This is due to the one-to-one relationship between price and the arrival rate. Naturally, the change in the value of λ leads to the change in the value of the observed CTR. For the purpose example, we set $\beta = 0.5$.[12]

For the rest of our analysis, we take the following *steps*:

Step 1. First, we obtain the optimal price per impression (p_{cpm}), and accordingly the optimal advertisers' arrival rate of the equivalent CPM system (λ_{cpm}). In order to find these two values, we apply the closed form results from the first chapter for CPM systems. In Chapter 1, we showed that the closed-form steady-state probability of the number of advertisers in a CPM system is expressed as

$$\mathbb{P}_i^{cpm}(\lambda, \mu, N, s) = \frac{\binom{N+i-1}{i} r^i (1+r)^{s-i-1}}{\sum_{j=0}^{s} \binom{N+s-1}{j} r^j}, \; i < s,$$
$$\mathbb{P}_s^{cpm}(\lambda, \mu, N, s) = \frac{\binom{N+s-1}{s} r^s}{\sum_{j=0}^{s} \binom{N+s-1}{j} r^j},$$

where N is the number of impressions being sold, s is the number of slots in the publisher's system, and $r = \lambda/\mu$. We represent the optimal advertisers' arrival rate for the considered corresponding CPM system with λ^*_{cpm}. λ^*_{cpm} is obtained

[12] Note that we have made an extensive search for a possible appropriate candid function for β in the marketing literature. Nevertheless, our attempt has been unsuccessful. As a result, various potential functions for β (as functions of the number of slots, etc) were examined. However, it was observed that regardless of the chosen function for β the obtained insights tended to be the almost the same. Therefore, for convenience of presentation we select a fixed value for β.

from the following maximization problem:

$$\max_{\lambda_{cpm}} R_{cpm}(\lambda_{cpm}) = \lambda_{cpm}(1 - \mathbb{P}_n^{cpm}(\lambda_{cpm}, \mu, N, n))p_{cpm}(\lambda_{cpm})N \quad (2.11)$$

$$s.t.$$

$$N = \left\lceil \frac{x}{CTR(\lambda_{cpm}, \mu, x, n, \beta)} \right\rceil, \quad (2.12)$$

$$CTR(\lambda_{cpm}, \mu, x, n, \beta) = \beta \sum_{i=1}^{n} \frac{1}{i} \frac{\mathbb{P}_i^{cpc}(\lambda_{cpm}, \widehat{\mu}, x, n)}{1 - \mathbb{P}_0^{cpc}(\lambda_{cpm}, \widehat{\mu}, x, n)}, \quad (2.13)$$

$$\lambda_{cpm} \geq 0, \quad (2.14)$$

where [.] in (2.12) refers to the integer sign.

Step 2. Next, we compute the obtained revenue of the CPC system using the corresponding CPM system's solution as follows (note that here $\mathbb{P}_n^{cpc}$ is the full-state probability for the CPC system):

$$R_{cpc}(\lambda_{cpm}^*) = \lambda_{cpm}^*(1 - \mathbb{P}_n^{cpc}(\lambda_{cpm}^*, \widehat{\mu}, x, n)) \left(\frac{p_{cpm}(\lambda_{cpm}^*)}{CTR(\lambda_{cpm}^*, \mu, x, n, \beta)} \right) x, \quad (2.15)$$

where CTR is obtained using Equation (2.13). Equation (2.15) implies that the publisher commits himself to delivering x clicks, each with the price $p_{cpc}(\lambda_{cpm}^*) = p_{cpm}(\lambda_{cpm}^*)/CTR$, while the advertisers' effective arrival rate is

$$\lambda_{cpm}^*(1 - \mathbb{P}_n^{cpc}(\lambda_{cpm}^*, \widehat{\mu}, x, n)).$$

Step 3. Then, we find the optimal advertisers' arrival rate of the CPC system, λ_{cpc}^*, and compute the optimal revenue using the "correct" model as follows:

$$
\begin{aligned}
R^*_{cpc}(\lambda^*_{cpc}) &= \max_{\lambda_{cpc}} \lambda_{cpc}(1 - \mathbb{P}^{cpc}_n(\lambda_{cpc}, \mu, x, n))p_{cpc}(\lambda_{cpc})x \quad (2.16)\\
& \quad s.t.\\
p_{cpc}(\lambda_{cpc}) &= \frac{p_{cpm}(\lambda_{cpc})}{CTR(\lambda_{cpc}, \mu, x, n, \beta)}, \quad (2.17)\\
CTR(\lambda_{cpc}, \mu, x, n, \beta) &= \beta \sum_{i=1}^{n} \frac{1}{i} \frac{\mathbb{P}^{cpc}_i(\lambda, \widehat{\mu}, x, n)}{1 - \mathbb{P}^{cpc}_0(\lambda, \widehat{\mu}, x, n)}, \quad (2.18)\\
\lambda_{cpc} &\geq 0. \quad (2.19)
\end{aligned}
$$

Sterp 4. Finally, we obtain the relative revenue gap using the following formula:

$$
Gap = \frac{R^*_{cpc}(\lambda^*_{cpc}) - R_{cpc}(\lambda^*_{cpm})}{R^*_{cpc}(\lambda^*_{cpc})} \times 100 \quad (2.20)
$$

Table 2.1 shows the relative revenue gap between the optimal revenue obtained by the correct pricing, and the revenue obtained by the heuristic conversion of the CPM prices for different numbers of slots and different numbers of clicks.

c = 0.5	*Number of slots (n)*						
Number of Clicks (x)	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$
$x = 3,000$	26%	35%	39%	41%	41%	41%	41%
$x = 10,000$	20%	25%	25%	25%	23%	22%	21%
$x = 20,000$	17%	20%	20%	18%	17%	16%	15%
$x = 50,000$	14%	15%	14%	13%	12%	11%	10%
$x = 70,000$	13%	13%	12%	11%	10%	10%	9%
$x = 100,000$	12%	12%	11%	10%	9%	9%	8%

Table 2.1: The relative performance gap $\frac{R^*_{cpc}(\lambda^*_{cpc}) - R_{cpc}(\lambda^*_{cpm})}{R^*_{cpc}(\lambda^*_{cpc})} \times 100(\%)$

As can be seen from the table, the relative revenue gap between the optimal and the heuristic policies ranges between 8% at $n = 8$ and $x = 100,000$, and 41% for $5 \leq n \leq 8$ and $x = 3000$. For example, at $x = 10,000$ the relative revenue gap is on average around 23%, while at $x = 100,000$ the gap is approximately 10%. Furthermore, from Table 2.1, it can be observed that the relative gap decreases in the number of clicks. This observation is quite intuitive since with the increased number of clicks the advertisers stay longer in the publisher's system. As a result, the system converges to a deterministic system in which all the ads are occupied and the CTR converges to the fixed value $CTR = \beta/n$. Therefore, the relative gap caused by the conversion tends to diminish as the number of clicks increases substantially. From the table, we also observe that the relative gap is concave in the number of slots. Obviously, as n increases the gap caused by conversion increases. Nevertheless, when n grows too large the publisher's system behavior converges to that of a deterministic always empty system in which the value of CTR becomes zero.

Note that the number of clicks sold in practice often does not grow so large. For instance, our observation from the real data of a leading Scandinavian publisher for the past six months suggests that the number of clicks sold to advertisers are often around $x = 20,000$ or even less. The reason for selling relatively lower numbers of clicks is mainly due to the low CTR value (often ranging between 0.1% and 0.13% on average), which makes the advertisers stay for a very long time to complete their requested service.

2.6 Numerical Analysis

In the previous sections, we obtained the closed-form solution for the steady state probability distribution of the number of advertisers in the web publisher's system. Furthermore, we determined the conditions that ensure a unique optimal price. In this section we derive important insights about the optimal price behavior with respect to factors such as the number of slots, the requested clicks for different pricing functions.

Advertising Slots

First, we consider the sensitivity with respect to the number of advertising slots and show that the behavior of the optimal price is non-obvious with respect to the number of advertising slots.

In our analysis, we let the viewers' arrival rate be $\mu = 1,000$. In addition, we let the price-demand function be $p(\lambda) = 0.5 - \lambda^c$, $c > 0$. Figures 2.2 and 2.3 show the relationship between the optimal price and the number of slots when the requested numbers of clicks are $x = 3,000$ and $x = 10,000$ respectively. As can be seen from the figures for low numbers of requested clicks the optimal price decreases in the number of slots. Nevertheless, for high numbers of requested clicks, the optimal price increases in the number of slots. The reason of this behavior is due to the trade-off between the publisher's serving capacity on one side and the increased service time on the other: Adding an extra slot to the publisher's system increases the system's capacity to serve more advertisers. However, as more advertisers are being served, each advertiser has less chance to be recognized by a viewer and be clicked. As a result, the advertisers' average service time is increased as they need to stay longer in the publisher's system to complete their service. With the increased service time the publisher can now serve fewer advertisers per time unit. As a result, the increased service time reduces the publisher's overall serving capacity, which may dominate the extra service capacity added as a result of an additional slot. As can be seen in Figure 2.2, for small numbers of requested clicks the capacity increase as a result of an added slot is more than the capacity, which is lost as a result of the increased

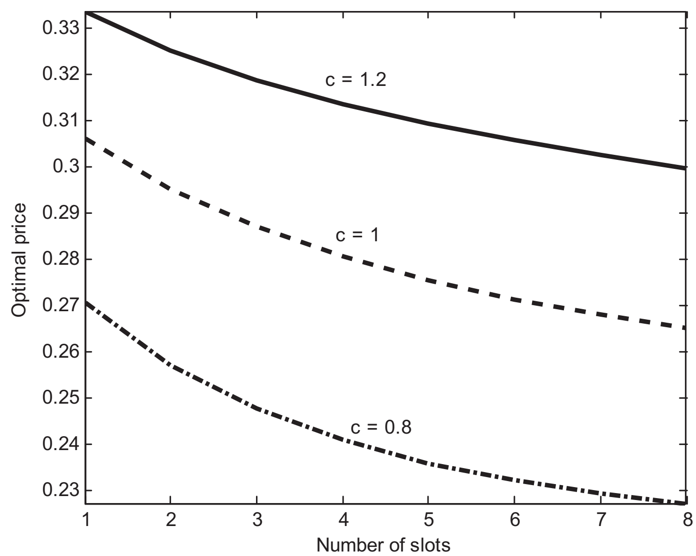


Figure 2.2: Optimal price vs slots for low number of clicks

service time. Hence, the overall serving capacity increases and the publisher can serve more advertisers. In order to absorb more advertisers, the publisher reduces its offered price. However, for large numbers of clicks the capacity loss due to the increased service time dominates the direct capacity increase as a result of adding an extra slot. Hence, the overall serving capacity decreases and the publisher can serve fewer advertisers in the long-run. In order to respond to this change, the publisher increases its offered price.

Numbers of Clicks

Next, we illustrate additional interesting insights about the optimal price by focusing on its relation with the number of clicks. For this purpose, we set the number of slots to be $n = 4$, and consider the rest of the setup to be the same as before. Figure 2.4 gives an idea of the relationship between the optimal price and the number of clicks for alternative values of the number of slots, n.

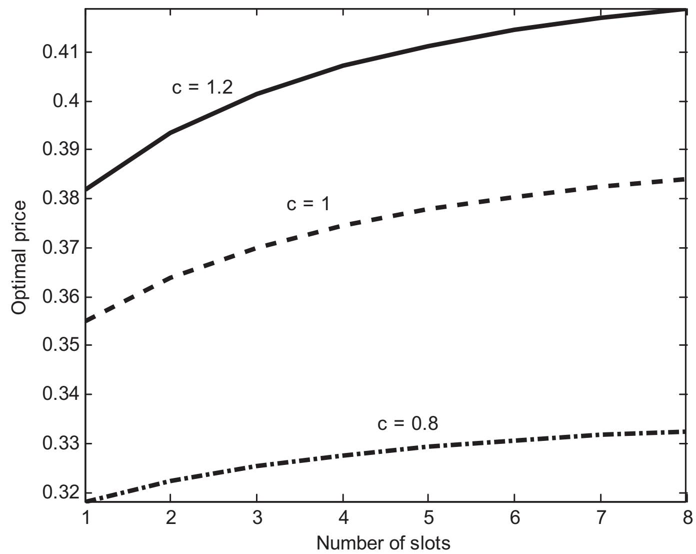


Figure 2.3: Optimal price vs slots for high number of clicks

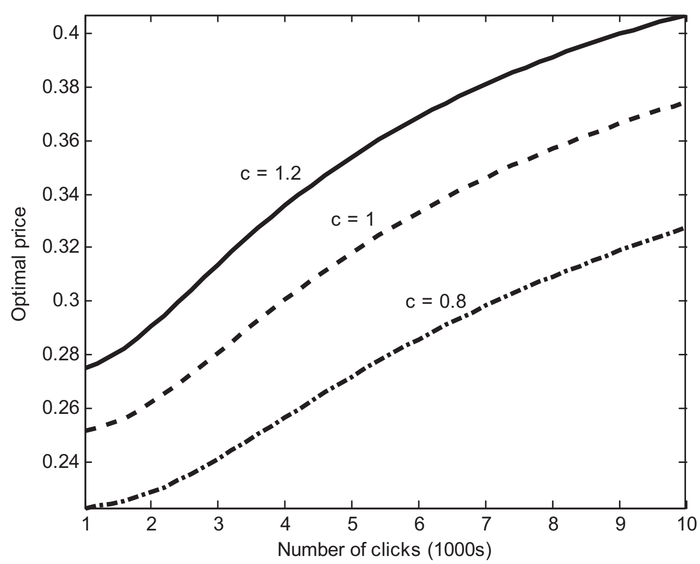


Figure 2.4: Optimal price vs number of clicks

As observed in Figure 2.4, the optimal price increases in the number of clicks. The reason for this behavior is that an increased number of clicks increases the service time. That is, each advertiser stays longer in the system to be fully served. Therefore, the publisher can serve fewer advertisers per time unit and as a result, the publisher increases its offered price.

We note that in general, advertisers are often attracted by quantity discounts. Therefore, it would make sense to offer a price per click that decreases with the number of clicks explicitly. In order to consider this issue, we set the price to depend not only on the arrival rate of advertisers λ, but also on the number of clicks x. We consider the following price function:

$$p(\lambda) = 0.50 - \lambda - 10^{-5}x. \tag{2.21}$$

We continue to explore the sensitivity with respect to the number of advertising slots. Figure 2.5 illustrates the relationship between the optimal revenue and the number of slots. We find that the optimal revenue is not increasing anymore, but becomes a concave function with a global maximum. Obviously, the reason for such a behavior is because of the trade-off made between the price increase as a result of the service capacity loss (due to the increased service time), and price decrease as a result of the promised quantity discount.

2.7 Extensions

There are many directions the CPC model can be extended to. In this section, we discuss two of them, leaving the rest for future research.

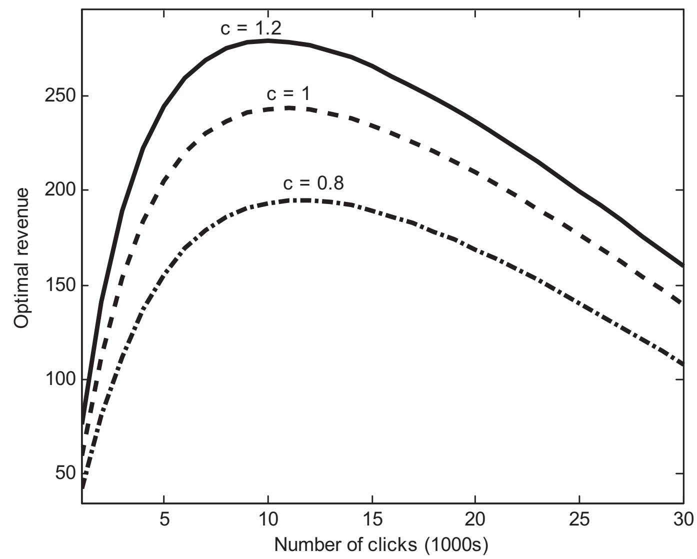


Figure 2.5: Optimal revenue vs number of clicks

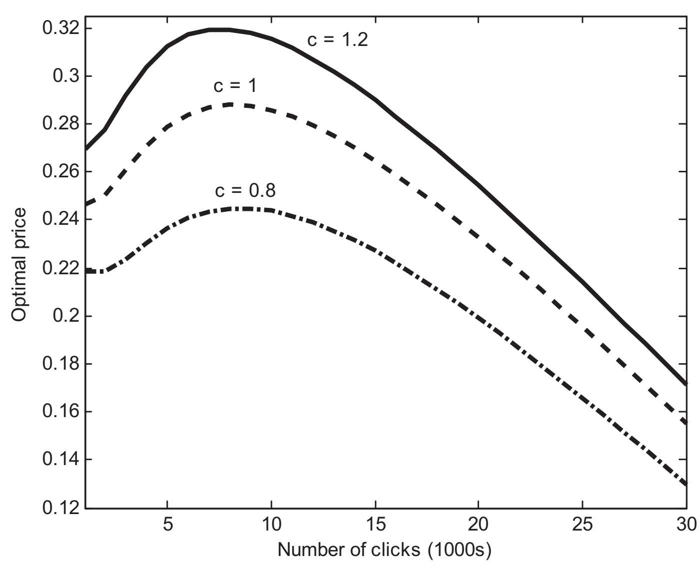


Figure 2.6: Optimal price vs number of clicks

2.7.1 Non-Poisson Arrivals

In Section 2.3 we assumed that the advertisers' arrivals at the web publisher from the ad network follow a Poisson process (Assumption 2), which might not be the case in reality. In addition, the viewers' arrival process might not be Poisson either (Assumption 3). In this section, we explore other distributions for both the demand and supply sides.

In our simulation study, we specifically examine the amount of revenue a publisher can lose by using the base model's solution obtained in Section 2.4 (based on Poisson arrivals, a single number of clicks offered, and a single price charged) to determine the price, while the clicks requested are random and both the advertisers' and the viewers' arrival processes are non-Poisson.

We let the viewers' arrival rate be $\widehat{\mu} = 1$. For the advertisers' interarrival time distributions, we consider the following distributions: Normal with mean $1/\lambda$ and standard deviation $1/\lambda$, Erlang-2 with mean $1/\lambda$ and standard deviation $1/\sqrt{2}\lambda$, Erlang-4 with mean $1/\lambda$ and standard deviation $1/2\lambda$, uniform with the two parameters 0 and $2/\lambda$, exponential with rate λ, and finally deterministic arrivals. For the viewers' inter arrival time distributions, we consider the same distributions with λ replaced with $\widehat{\mu} = 1$. The number of slots is set to be $n = 4$. We choose the pricing function to be $p(\lambda, X) = 0.5 - \lambda^{0.8} - 10^{-6}X$ where the random number of requested clicks, X, follows a Normal distribution with mean $x = \mathbb{E}(X) = 1000$ and standard deviation 500. The steps of each simulation process are as follows:

First, we obtain the advertisers' optimal arrival rate, $\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2}$, when the advertisers' interarrival times follow the generic distribution $\mathbf{D}_1$, the viewers' interarrival times follow $\mathbf{D}_2$, and each advertiser requests a different number of clicks according to a random variable X. This includes simulating the publisher's system for a range of values of λ and then selecting $\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2}$, the rate that gives the highest simulated revenue. We represent the revenue related to $\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2}$ with $R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2})$.

Second, we compute the optimal value for λ using the closed-form solution provided in Equation (2.7) by assuming the price function to be $p(\lambda, x) = 0.5 - \lambda^{0.8} - 10^{-6}X$, where $x = \mathbb{E}(X) = 1000$. We represent this optimal value with $\lambda^*_{x,Exp}$. If the web publisher uses our analytical solution with the average demand x, for a system that does not have Poisson arrivals of advertisers and viewers, and each advertiser requests X clicks its "real" revenue would become $R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{x,Exp})$. See the Appendix for a detailed schematic graph illustrating the explained steps.

Finally, we compute the revenue gap using the following formula

$$Gap = \frac{R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2}) - R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{x,Exp})}{R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2})} \times 100(\%).$$

Table 2.2 shows the relative revenue performance gaps for the different interarrival time distributions considered for advertisers' and viewers' arrivals as well as the random the number of requested clicks, X, that results in generating adjusted price for each click request. We observe that the computed revenue gaps are between $0.77\% - 4.25\%$. This suggests that the Poisson policy while considering the average of the requested clicks tends to be an accurate estimate for the publisher's model even when both the viewers' and the advertisers' arrivals are non-Poisson and the price is adjusted based on each advertiser's requested clicks.

Interarrival dist.	*Viewers*					
Advertisers	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Det.*	*Exp.*
Erlang-2	2.31%	2.30%	2.15%	2.26%	2.03%	2.19%
Erlang-4	1.57%	1.80%	1.74%	1.25%	1.42%	1.53%
Normal	1.99%	3.1%	3.08%	3.18%	2.37%	2.92%
Uniform	1.39%	1.94%	1.74%	1.22%	1.55%	2.06%
Det.	0.77%	1.41%	1.16%	1.01%	0.83%	0.97%
Exp.	4.25%	3.82%	2.36%	4.03%	3.57%	–

Table 2.2: The relative performance gap $\frac{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})-R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{x,Exp})}{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})} \times 100(\%)$

2.7.2 Model's Reliability Under More General Conditions

Our purpose from this section is to show that closed-form results obtained in the Model's section are relatively accurate estimates when the publisher's system operates in more general conditions than assumed in this chapter. We explore this observation by simulating the publisher's operation while relaxing some of the restrictive assumptions, which we had made primarily for tractability. We show that the gaps between the closed-form values of L (the average number of advertisers in the system) and $\mathbb{P}_n$ (the probability that the system is full) obtained from the stylized model, and their corresponding simulated values for more general model are often less than 1%. As a result, the solution of our stylized model is a relatively accurate estimate for more general models. We specifically investigate the model's performance by relaxing the following assumptions from our stylized model:

1. **Uncertain Click Requests** In Section 2.3 we assumed that all advertisers request x clicks. This assumption was principally made for tractability. In this section, we consider the number of clicks requested by each advertiser to be a random variable X following a certain probability distribution. For example, in many ad networks, the advertisers can choose the number of clicks from a limited option list, i.e., x_1, ...,x_k. As a result, it can be thought that in the

long-run any value x_i is selected by a proportion of advertisers, namely p_i, where $0 \leq p_i \leq 1$. Similarly, in some ad networks, although there is no limit on selecting the number of clicks, it is natural to consider that advertisers select the number of clicks based on a continuous probability distribution.

2. **Asymmetric Click Through Rates** The second issue that we are concerned with is each ad having the same chance to be clicked by the viewers entering the publisher's system. That is the CTR is the same for all ads. Although the ads belong to the same subsystem, and thus it is natural to assume the ads to have similar click through rates, it might be more realistic to assume that due to some external factors such as the designs of ads, they may demonstrate different levels of attractiveness. In this section we consider this issue as well.

3. **Non-stationary Arrival Rates** In Section 2.3 we assumed that advertisers' Poisson arrival process is stationary. Nevertheless, in reality, the arrival rates might change over time. In this section, we consider this issue as well.

In order to explore items $(1)-(3)$, we let the advertisers' arrivals follow a Poisson process with rate Λ, where Λ itself follows the truncated Normal distribution with mean 1 and standard deviation 0.2. That is, $\Lambda \triangleq \psi \mathbf{1}_{\{\psi>0\}}$, and $\psi \sim \mathcal{N}(1, 0.2)$. Furthermore, we assume that the viewers' arrivals at the publisher's system, follow a Poisson distribution with rate $\mu = 10$ per time unit. Note that as the viewers' arrivals remain unaffected by the ad network filter, we consider the viewers' rate constant over time. The numbers are just illustrative and only for the purpose of our example. We let the number of slots be $n = 4$. Each arriving advertiser requests X clicks, where X follows a truncated normal distribution with mean m (the average of the requested clicks) and standard deviation $0.5m$. That is, $X \triangleq Y\mathbf{1}_{\{Y>0\}}$ clicks, where $Y \sim N(m, 0.5m)$. The high standard deviation of Y makes the number of clicks requested by one advertiser be considerably different from the next.

On the viewers' side, we index the slots from 1 to n. Upon arrival at the publisher's website, a viewer observes the ads in the system, and decides to click on one of them based on his preference. The viewer's preference in clicking on the slots is asymmetric. In order to consider this asymmetry in our analysis, we assume that the viewer selects the ads with the index value $I \triangleq [B \times N_t + 1]$, with $B \sim Beta(0.1, 0.1)$, where $[.]$ is the integer sign, and N_t is the number of ads displayed on the page upon the viewer's arrival. The selection of Beta distribution is arbitrary and for the purpose of example[13]. It can be observed that the majority of time the viewer clicks on the ads whose indices are 1, and 4, while the other two ads, if available, attract less of his attention to click[14].

For convenience, we refer to the abovementioned system as the *Extended System* and denote its average and full-state probability by L^{ES} and $\mathbb{P}_n^{ES}$ respectively. In addition, we refer the stylized system developed in the Model's section as the *Base System* and denote its average and full-state probability by L and $\mathbb{P}_n$ as before. We vary the average of the requested clicks from $m = 0$ to $m = 1000$ with steps of 50 clicks. For each value of m, we conduct a discrete event simulation with a time horizon of $T = 50,000$ time units. Note that in every time unit, on average, one advertiser and 10 viewers arrive at the system. Therefore, each of the simulations consists of an average $550,000$ "events" where each event is either arrival of an advertiser, or a viewer. Furthermore, note that the reason for considering L^{ES} and $\mathbb{P}_n^{ES}$ only is that we need these two parameters only for the web publisher's optimization

[13] Note that we have considered a variety of different distributions for W. However, all of them lead to similar result. Therefore, we decided to mention the results based on Beta distribution, which appeared to us as a good representative example for modeling the asymmetric clicking process.

[14] This can be, for instance, because the slots 1 and 4 are located in a more visible positions.

problem in (2.7) above.

We observe, the ratio $\left(L - L^{ES}\right)/L$ is less than roughly 0.03% for all values of m. In addition, the ratio $\left(\mathbb{P}_n - \mathbb{P}_n^{ES}\right)/\mathbb{P}_n$ is less than about 0.2% throughout[15]. As a result, the closed-form results obtained from our stylized system are almost identical to the simulated results obtained for the extended system. Hence the base system is a relatively accurate estimate for more complex circumstances. We also note that as m becomes very large, the ratios turn to zero very quickly, confirming that the two systems perform in a very similar way.

2.8 Conclusion

In this chapter, we have presented a revenue optimization model for a web publisher selling his advertising space through an advertising network. The web publisher generates revenue by displaying ads on its website and charges according to the CPC pricing scheme. The web publisher operation is modeled with a queuing system, where the arrival process corresponds to the advertisers sent by the ad network for posting their ads, the service process corresponds to the viewers visiting the website, with the advertising slots playing the role of servers.

A primary feature of most advertising networks is that they only deal with *immediate inventories*. This means that when a publisher's system is full, the network ignores it. Instead, the network directs the advertisers to other available systems within the advertisers' selected category. In queuing terms this corresponds to a system with no waiting spaces.

The queuing model developed is different from models existing in current lit-

[15]To conserve the space, we have moved the simulation's results graph to the Appendix.

erature. Despite the complexity of the model, we are able to derive a closed-form solution of the probability distribution of the number of advertisers in the publisher's system for any number of banners and any number of requested clicks. This enabled us to set up the revenue-maximizing problem of the web publisher and derive the optimal price to charge per click, which was one of the purposes of this chapter.

We derive additional theoretical results through this chapter. An interesting result, for instance, is that the steady state probability of the number of advertisers in the publisher's system coincides with that of the $M/M/1/n$ system. This result is rather surprising since the two systems have different dynamics.

On the managerial side, we demonstrate that the general heuristic widely employed in the CPC contracts where a publisher simply uses the CTR to convert the price of one scheme to the other can be misleading, resulting in a considerable revenue loss compared to the optimal policy. In addition, we provided further insights by showing that, unlike the CPM price considered in the first chapter, the optimal CPC price may increase with the number of slots. This may not seem intuitive in comparison to our common understanding from the supply-demand relationship, since an increase in the number of empty slots in the system can be interpreted as an increase in the serving capacity in the system. As a result, one may expect the opposite result to hold.

We considered the model's robustness by considering random click requests, nonstationary arrival rates, and asymmetry in click-through rates through an extensive simulation study and concluded that the solutions of the simulated systems are only minimally different from our basic model.

We believe that, in view of the model's high flexibility in response to numerous uncertain circumstances, the results we obtain can be naturally integrated into automated pricing software to generate relatively optimal prices, and enhance revenue performance of countless numbers of websites selling their slots through advertising networks.

Our model is among the first to bridge the gap between much of the academic literature on pricing, which mainly focuses on deterministic pricing models, and the much more complex online display advertising systems encountered in practice. It also provides a significant contribution to the currently developing management science literature on online advertising, and help to distance from the commonly made assumptions of the deterministic models in the marketing literature. The closed-form results of our model can also serve as decision tools to help the web publishers running advertising operations, for instance, by providing an extra layer of intelligence on top of their pricing engine software.

In conclusion, we do not claim that our model solves all significant issues regarding the CPC optimal pricing. However, we think that the modeling framework developed in this chapter can provide a good basis for multiple research directions, which would explore analytically many relevant issues in online advertising.

Appendix A

A1. Overview of the publisher's problem

In this section, we explain in detail how the publishers sell their slots to the advertisers through ad networks[16]. In most ad networks the advertisers do not approach the web publisher directly, but arrive at the system through the ad network. An ad network is a company that connects the web publishers who want to sell their slots (also called online inventory), with the advertisers who want to run their ads in the relevant websites. Large publishers often sell around 60% of their inventory through ad networks. However smaller publishers often sell their entire inventory through ad networks. In our dissertation, we consider a common type of ad networks, known as blind networks. A blind network is such that advertisers place their ads, and clearly define the category of target websites for their ads (e.g., based on the size and format of their ads (i.e., leaderboard, or small rectangle), the target websites' contents, the average number of viewers per week visiting the target websites, and viewers geographical locations), but do not know the exact places where their ads are being placed. They will only know their ads will be placed in one of the target websites within the category of their request. Contextweb, Valueclick, and Clicksor are examples of these large blind networks.

Note that most ad networks work with *immediate* inventories. That is, when a publisher's website falls into the advertiser's target category, and all slots are already occupied the ad network never wastes time waiting for that publisher to become available. Instead, it directs the ads to one of the other websites within

[16] The current market size of display advertising within ad networks is significant with the 2009 predicted revenue of \$5.2 billion *only* in the United States. It is currently considered the fastest growing sector of online advertising market, and is anticipated to reach \$7.6 billion by 2012 (Think Equity 2007).

that category, and the selection of the target website is usually made randomly. In several networks, if the number of the available websites in the requested category is low or there are no available websites at the time of the advertiser's request, the ad network again does not wait for one of the sites to become available. Instead it uses the available websites in the requested category from other ad networks in partnership programs in order to increase the number of available websites. Usually the partnerships of such ad networks are managed by other companies called *ad exchanges* (i.e., ADSDAQ.com). Ad exchanges can be ad networks too, their main difference being that in addition they have direct access to resources of several other ad networks as well. Ad exchanges' performances are sometimes considered similar to *stock exchanges* like NASDAQ, since they provide a *single virtual market* for multiple networks to sell their inventories. An example of a large ad exchange is ADSDAQ (we can easily notice the similarity of its name to NASDAQ).

The important question here is, if the advertisers and the publishers do not directly interact with each other, is the price still set by the web publishers themselves or do the networks automatically choose this for them? The answer is for many of the ad exchanges like ADSDAQ it is still the publishers who determine their prices. Nevertheless, some other networks like Clicksor have different policies. Clicksor divides its publishers into the two main categories of *premium* and *non-premium*. The premium publishers are still free to choose their own prices. However, the slot prices for the non-premium publishers are automatically determined by the network. These networks also do not reveal the price information to their non-premium publishers, though they guarantee to pay at least a minimum amount of payment to these publishers. In our dissertation, we restrict our attention to the networks in which the publishers are completely free to select their prices.

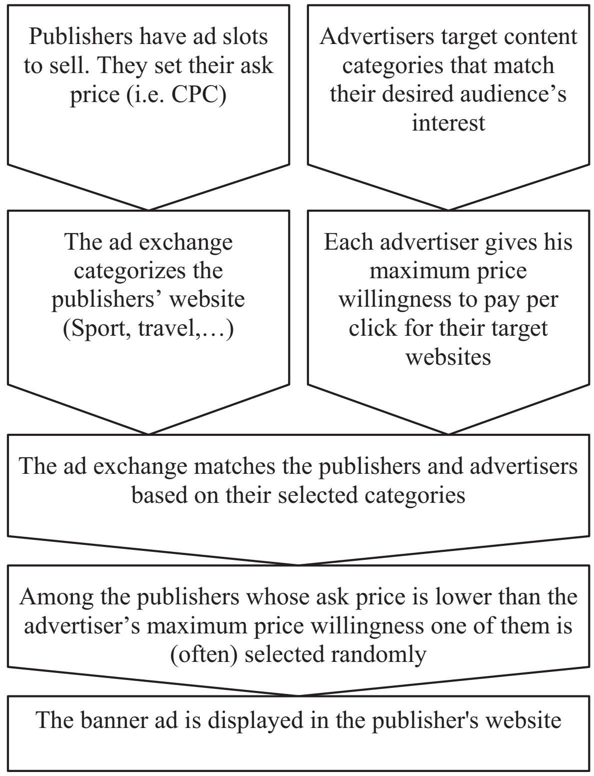


Figure 2.7: The general steps for transaction between advertisers and web publishers through advertising exchages

The price (per click) determined by the publishers is called the publisher's (or the slot's) *ask-price*. Note that, when registering the slots in the ad network, a web publisher registers each group of equivalent slots in his website (based on such factors as size, format, and location), with a separate, unique code and then sets a different ask-price for each group (the size and format of each group follow the standards of Internet Advertising Bureau (IAB)). Therefore, each publisher can register several subsystems. However, since the advertisers clearly determine their target group when registering in the network, each single group of slots can be viewed on its own as an independent, separate CPC subsystem having *equivalent slots* that are *priced the same.*

In order to match the slots' ask price with the advertisers' willingness to pay, the advertisers are always asked to bid the maximum price they are willing to pay while registering with the ad network. The ad network then uses these maximum bid prices to screen the websites in the advertisers target category with ask prices lower than the advertiser's maximum. After this step, the network sends the advertiser's ad to one of these websites. The advertisers then pay the ask price to the ad network and the network, taking a certain percentage of the publisher's revenue as commission, ranging from 25% to 50%, transfers the rest to the publisher's account, usually on a monthly basis. Figure 2.7 summarizes the necessary steps to take for online transactions between publishers and the advertisers through the ad networks.

In addition to the pricing mechanism explained above, there are two further pricing schemes that are employed; however, we do not consider these in this dissertation. In the first scheme, there are some networks like Adtoll (Adtoll.com), in which the publishers' steps are almost the same, but the steps for advertisers are a little different. Adtoll Company allows the advertisers to see the full list of websites on hand in their selected target category, and also allows them to choose the list's most favoured websites. This process especially coincides with the abovementioned steps if we assume the websites in the target category are equivalent in the advertiser's point of view, and as a result, have the same chance of being selected. The second pricing scheme is the pure auction model, which is (partly) practiced by such networks as Right Media, acquired by Yahoo! in 2003. In these networks, first the publisher determines an ask-price for his slot in a quite similar way to the one described above. However, if not sold with the determined price, in order to avoid unsold inventory, the slot will be auctioned, in the same way as products are auctioned in the eBay website. Finally, the network sells the slot to the advertiser with the highest bid. Pure auctioning of the slots is an interesting field. Nevertheless, we leave it for future works.

A2. Proof of Proposition 26

We consider a Markov chain in which the state of the system is defined to be the vector $\mathbf{k}_{n\times 1} \triangleq (i_1, i_2, ..., i_n) = \sum_{s=1}^{r} c_s \mathbf{v}_{c_s} \in (\mathbb{N} \cup \{0\})^n$. $c_s \in \mathbb{N} \cup \{0\}$ with $s \in \mathbb{N} \cap [1, r]$ represents the identical number of clicks that is left in a group of slots. Further, the vector $\mathbf{v}_{c_s} \in (\mathbb{N} \cup \{0\})^n$ is defined as $\mathbf{v}_{c_s} \triangleq \sum_{j \epsilon \mathcal{G}_{c_s}(\mathbf{k})} \mathbf{e}_j^T$ wherein the set $\mathcal{G}_{c_s}(\mathbf{k})$ is characterized as $\mathcal{G}_{c_s}(\mathbf{k}) \triangleq \{j \,|\langle \mathbf{k}, \mathbf{e}_j \rangle = c_s\}$ in which $\langle \mathbf{k}, \mathbf{e}_j \rangle$ is defined as the inner product of the two vectors $\mathbf{k}$ and $\mathbf{e}_j$, the unit jth vector. We need to identify all the possible states of the system and obtain the transition balance equations for every state. In view of the complexity of the transition equations, there is not any standard technique to solve them in a single system. Thus we illustrate the results hold by verification. The symmetric CPC system has in general 10 distinct transition equations as follows:

i) For $\mathbf{k} = (0, ..., 0) = \mathbf{0}_{n\times 1}$ the flow balance is straightforward to obtain. $\mathbf{k}$ can either go to $\left(\mathbf{k} + \mathbf{x}\mathbf{e}_1^T\right)$ with rate λ or come from the sate $\left(\mathbf{k} + \mathbf{e}_1^T\right)$ with rate μ. As a result the flow balance equation becomes:

$$r\pi_{\mathbf{k}} = \pi_{\mathbf{k}+\mathbf{e}_1^T}. \tag{2.22}$$

ii) If $\mathbf{k} = i\mathbf{e}_1^T$ with $i \in \mathbb{N} \cap [1, x-1]$ then state $\mathbf{k}$ can either go to sate $\left(\mathbf{k} - \mathbf{e}_1^T\right)$ with rate μ or to state $\left(\mathbf{k} + \mathbf{x}\mathbf{e}_2^T\right)$ with rate λ. It also can either come from the state $\left(\mathbf{k} + \mathbf{e}_1^T\right)$ with rate μ or the state $\left(\mathbf{k} + \mathbf{e}_2^T\right)$ with the rate $\lambda/2$. Hence the balance equation becomes:

$$(1 + r)\pi_{\mathbf{k}} = \pi_{\mathbf{k}+\mathbf{e}_1^T} + \frac{1}{2}\pi_{\mathbf{k}+\mathbf{e}_2^T}.$$

iii) If $\mathbf{k} = i\mathbf{v}_i$, with $\mathbf{v}_i = \sum_{j=1}^{k} \mathbf{e}_j^T$ in which $k \in \mathbb{N} \cap [1, n-1]$ and $i \in \mathbb{N} \cap [1, x-1]$ then the state $\mathbf{k}$ transits to state $\left(\mathbf{k} + \mathbf{x}\mathbf{e}_{k+1}^T\right)$ with rate λ. It also transits to state $\left(\mathbf{k} - \mathbf{e}_1^T\right)$ with rate μ. Further, the two states $\left(\mathbf{k} + \mathbf{e}_1^T\right)$ and $\left(\mathbf{k} + \mathbf{e}_{k+1}^T\right)$ transit to the state $\mathbf{k}$ with rates μ and $\left(\frac{\mathbf{1}_{\{i \neq 1\}} + (k+1)\mathbf{1}_{\{i=1\}}}{k+1}\right)\mu$ respectively. As a result the flow balance equation becomes:

$$(1+r)\pi_{\mathbf{k}} = \frac{1}{k}\pi_{\mathbf{k}+\mathbf{e}_1^T} + \left(\frac{\mathbf{1}_{\{i \neq 1\}} + (k+1)\mathbf{1}_{\{i=1\}}}{k+1}\right)\pi_{\mathbf{k}+\mathbf{e}_{k+1}^T}. \tag{2.23}$$

iv) If $\mathbf{k} = i\mathbf{v}_i$ with $\mathbf{v}_i = \sum_{j=1}^{n} \mathbf{e}_j^T$ and $i \in \mathbb{N} \cap [1, x-1]$ then $\mathbf{k}$ can either come from $\left(\mathbf{k} + \mathbf{e}_1^T\right)$ with the rate μ/n or go to the state $\left(\mathbf{k} - \mathbf{e}_1^T\right)$ with the rate μ. Therefore, the flow balance equation becomes:

$$\pi_{\mathbf{k}} = \frac{1}{n}\pi_{\mathbf{k}+\mathbf{e}_1^T}. \tag{2.24}$$

v) Define $\vartheta(\mathbf{k}, z) \triangleq \sum_{s=1}^{z} |\mathcal{G}_{c_s}(\mathbf{k})|$ for any $z \in \mathbb{N} \cap [1, r]$. The the state $\mathbf{k} = \sum_{s=1}^{r} c_s \mathbf{v}_{c_s}$ with $0 < c_1 < c_2 < ... < c_r$, $c_s \in \mathbb{N} \cap [1, x-1]$ in which $\mathbf{v}_{c_s} = \sum_{u \epsilon \mathcal{G}_{c_s}(\mathbf{k})} \mathbf{e}_u^T$ and $\sum_{s=1}^{r} |\mathcal{G}_{c_s}(\mathbf{k})| \in \mathbb{N} \cap [1, n-1]$ goes to the state $\mathbf{k} + \mathbf{x}\mathbf{e}_{(\vartheta(\mathbf{k},r)+1)}^T$ with the transition rate λ, and to one of the sates $\mathbf{k} - \mathbf{e}_{(\vartheta(\mathbf{k},z-1)+1)}^T$,for all $z \in \mathbb{N} \cap [1, r]$, with the transition rate $\frac{|\mathcal{G}_{c_z}(\mathbf{k})|}{\vartheta(\mathbf{k},z)}\mu$. $\mathbf{k}$ also comes from either the state $\mathbf{k} + \mathbf{e}_{(\vartheta(\mathbf{k},r)+1)}^T$ with rate $\left(\frac{\mathbf{1}_{\{c_1 > 1\}} + (|\mathcal{G}_{c_1}(\mathbf{k})|+1)\mathbf{1}_{\{c_1 = 1\}}}{\vartheta(\mathbf{k},r)+1}\right)\mu$ or from one of the states $\mathbf{k} + \mathbf{e}_{(\vartheta(\mathbf{k},z-1)+1)}^T$ with all $z \in \mathbb{N} \cap [1, r]$, with rate

$$\left(\frac{\mathbf{1}_{\{c_{z+1} > c_z + 1\}} + \mathbf{1}_{\{z=r\}} + (\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_{z+1} = c_z + 1\}}}{\vartheta(\mathbf{k}, r)}\right)\mu.$$

As a result the flow balance equation becomes:

$$(1+r)\pi_{\mathbf{k}} = \left(\frac{\mathbf{1}_{\{c_1>1\}} + (\left|\mathcal{G}_{c_1}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_1=1\}}}{\vartheta(\mathbf{k},r)+1} \right) \pi_{\mathbf{k}+\mathbf{e}^T_{(\vartheta(\mathbf{k},r)+1)}} \tag{2.25}$$
$$+ \sum_{z=1}^{r} \left(\frac{\mathbf{1}_{\{c_{z+1}>c_z+1\}} + \mathbf{1}_{\{z=r\}} + (\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_{z+1}=c_z+1\}}}{\vartheta(\mathbf{k},r)} \right) \pi_{\mathbf{k}+\mathbf{e}^T_{(\vartheta(\mathbf{k},z-1)+1)}}.$$

vi) If $\mathbf{k} = \sum_{s=1}^{r} c_s \mathbf{v}_{c_s}$ with $0 < c_1 < c_2 < ... < c_r$, $c_s \in \mathbb{N} \cap [1, x-1]$ and $\mathbf{v}_{c_s} = \sum_{u \epsilon \mathcal{G}_{c_s}(\mathbf{k})} \mathbf{e}^T_u$ and $|\mathcal{G}_0(\mathbf{k})| = 0$ the system goes to one the possible states $\mathbf{k} - \mathbf{e}^T_{(\vartheta(\mathbf{k},z-1)+1)}$ with all $z \in \mathbb{N} \cap [1, r]$ with the transition rate $\frac{|\mathcal{G}_{c_z}(\mathbf{k})|}{n}\mu$, and comes from one the possible states $\mathbf{k} + \mathbf{e}^T_{(\vartheta(\mathbf{k},z-1)+1)}$ with all $z \in \mathbb{N} \cap [1, r]$ with the transition rate

$$\left(\frac{\mathbf{1}_{\{c_{j+1}>c_j+1\}} + \mathbf{1}_{\{z=r\}} + (\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_{z+1}=c_z+1\}}}{n} \right) \mu.$$

As a result the flow balance equation becomes:

$$\pi_{\mathbf{k}} = \sum_{z=1}^{r} \left(\frac{\mathbf{1}_{\{c_{j+1}>c_j+1\}} + \mathbf{1}_{\{z=r\}} + (\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_{z+1}=c_z+1\}}}{n} \right) \pi_{\mathbf{k}+\mathbf{e}^T_{(\vartheta(\mathbf{k},z-1)+1)}}. \tag{2.26}$$

For example, take $\mathbf{k} = (1,1,2,2,3)$ then $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, $|\mathcal{G}_{c_1}(\mathbf{k})| = 2$, $|\mathcal{G}_{c_2}(\mathbf{k})| = 2$, $|\mathcal{G}_{c_3}(\mathbf{k})| = 1$. Moreover both $c_2 = c_1 + 1$, and $c_3 = c_2 + 1$ hold. Also $\mathbf{k} + \mathbf{e}^T_{(\vartheta(\mathbf{k},0)+1)} = (2,1,2,2,3)$, $\mathbf{k} + \mathbf{e}^T_{(\vartheta(\mathbf{k},1)+1)} = (1,1,3,2,3)$, and $\mathbf{k} + \mathbf{e}^T_{(\vartheta(\mathbf{k},2)+1)} = (1,1,2,2,4)$. Therefore, the balance equation becomes: $\pi_{(1,1,2,2,3)} = \frac{3}{5}\pi_{(2,1,2,2,3)} + \frac{2}{5}\pi_{(1,1,3,2,3)} + \frac{1}{5}\pi_{(1,1,2,2,4)}$.

vii) For $\mathbf{k} = x\mathbf{v}_x$, where $\mathbf{v}_x = \sum_{u=1}^{|\mathcal{G}_x(\mathbf{k})|} \mathbf{e}^T_u$ and $|\mathcal{G}_x(\mathbf{k})| < n$ with an analogous argument the balance equation becomes:

$$(1+r)\pi_{\mathbf{k}} = r\pi_{\mathbf{k}-\mathbf{x}\mathbf{e}^T_{|\mathcal{G}_x(\mathbf{k})|}} + \left(\frac{\mathbf{1}_{\{x>1\}} + (|\mathcal{G}_x(\mathbf{k})|+1)\,\mathbf{1}_{\{x=1\}}}{|\mathcal{G}_x(\mathbf{k})|+1}\right)\pi_{\mathbf{k}+\mathbf{e}^T_{(|\mathcal{G}_x(\mathbf{k})|+1)}}. \tag{2.27}$$

viii) For $\mathbf{k} = x\mathbf{v}_x$, with $\mathbf{v}_x = \sum_{u=1}^{n} \mathbf{e}_u^T$ the balance equation turns out to be:

$$\pi_{\mathbf{k}} = r\pi_{\mathbf{k}-\mathbf{e}_j^T}. \tag{2.28}$$

ix) If $\mathbf{k} = x\mathbf{v}_x + \sum_{s=1}^{r} c_s\mathbf{v}_{c_s}$, where $\mathbf{v}_x = \sum_{u=1}^{|\mathcal{G}_x(\mathbf{k})|} \mathbf{e}_u^T$, and $\mathbf{v}_{c_s} = \sum_{u=|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},s-1)}^{|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},s)} \mathbf{e}_u^T$, $|\mathcal{G}_0(\mathbf{k})| > 0$ the system goes to the state $\mathbf{k}+\mathbf{x}\mathbf{e}^T_{(|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},r)+1)}$ with rate λ or to $\mathbf{k}-\mathbf{e}_1^T$ with rate $\left(\frac{|\mathcal{G}_x(\mathbf{k})|}{|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},r)}\right)\mu$ or to $\mathbf{k}-\mathbf{e}^T_{(|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},z-1)+1)}$ for $z \in \mathbb{N}\cap[1,r]$ with the transition rate $\left(\frac{\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right|}{|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},r)}\right)\mu$. Moreover, the state $\mathbf{k}$ comes from either the state $\mathbf{k}+\mathbf{e}^T_{(|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},r)+1)}$ with rate $\left(\frac{1+\left|\mathcal{G}_{c_1}(\mathbf{k})\right|\mathbf{1}_{\{c_1=1\}}}{|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},r)+1}\right)\mu$ or from $\mathbf{k}+\mathbf{e}^T_{(|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},z-1)+1)}$, $z \in \mathbb{N}\cap[1,r]$ with the transition rate

$$\left(\frac{\mathbf{1}_{\{c_{z+1}>c_z+1\}}+\mathbf{1}_{\{z=r\}}+(\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right|+1)\mathbf{1}_{\{c_{z+1}=c_z+1\}}}{|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},r)}\right)\mu.$$

Hence the flow balance equation becomes:

$$\begin{aligned}(1+r)\pi_{\mathbf{k}} = r\pi_{\mathbf{k}-\mathbf{e}_1^T} &+ \sum_{z=1}^{r}\left(\frac{\mathbf{1}_{\{c_{z+1}>c_z+1\}}+\mathbf{1}_{\{z=r\}}+(\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right|+1)\mathbf{1}_{\{c_{z+1}=c_z+1\}}}{|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},r)}\right)\pi_{\mathbf{k}+\mathbf{e}^T_{(|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},z-1)+1)}} \\ &+ \left(\frac{1+\left|\mathcal{G}_{c_1}(\mathbf{k})\right|\mathbf{1}_{\{c_1=1\}}}{|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},r)+1}\right)\pi_{\mathbf{k}+\mathbf{e}^T_{(|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},r)+1)}}.\end{aligned}$$

x) If $\mathbf{k} = x\mathbf{v}_x + \sum_{s=1}^{r} c_s\mathbf{v}_{c_s}$, with $\mathbf{v}_x = \sum_{u=1}^{|\mathcal{G}_x(\mathbf{k})|} \mathbf{e}_u^T$, $\mathbf{v}_{c_s} = \sum_{u=|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},s-1)}^{|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},s)} \mathbf{e}_u^T$, $|\mathcal{G}_0(\mathbf{k})| = 0$

then we obtain:

$$\pi_{\mathbf{k}} = r\pi_{\mathbf{k}-\mathbf{e}_1^T} + \sum_{z=1}^{r}\left(\frac{\mathbf{1}_{\{c_{z+1}>c_z+1\}} + \mathbf{1}_{\{z=r\}} + (\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_{z+1}=c_z+1\}}}{n}\right)\pi_{\mathbf{k}+\mathbf{e}^T_{(|\mathcal{G}_x(\mathbf{k})|+\vartheta(\mathbf{k},z-1)+1)}}. \tag{2.29}$$

Verifying the items 1 to 4, as well as 7 and 8 are immediate. Therefore, we need just to verify the solution for items $5, 6, 7, 9$, and 10.We only verify the solution for number 5. The rest are verified the same way.

For number 5 we need to show

$$(1+r)\pi_{\mathbf{k}} = \left(\frac{\mathbf{1}_{\{c_1>1\}} + (\left|\mathcal{G}_{c_1}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_1=1\}}}{\vartheta(\mathbf{k}, r) + 1}\right)\pi_{\mathbf{k}+\mathbf{e}^T_{(\vartheta(\mathbf{k},r)+1)}} + \sum_{z=1}^{r}\left(\frac{\mathbf{1}_{\{c_{z+1}>c_z+1\}} + \mathbf{1}_{\{z=r\}} + (\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_{z+1}=c_z+1\}}}{\vartheta(\mathbf{k}, r)}\right)\pi_{\mathbf{k}+\mathbf{e}^T_{(\vartheta(\mathbf{k},z-1)+1)}}. \tag{2.30}$$

We show this in multiple stages:

I. For the first case assume that $c_1 > 1$ and $c_{j+1} > c_j + 1$ for all $1 \leq j \leq r$. For the left side we get $(1+r)\pi_{\mathbf{k}} = \binom{\vartheta(\mathbf{k},r)}{\left|\mathcal{G}_{c_1}(\mathbf{k})\right| \; \left|\mathcal{G}_{c_2}(\mathbf{k})\right| \ldots \left|\mathcal{G}_{c_r}(\mathbf{k})\right|}(1+r)r^{\vartheta(\mathbf{k},r)}$. For convenience define $R_1(\mathbf{k}) \triangleq \left(\frac{\mathbf{1}_{[c_1>1]} + (\left|\mathcal{G}_{c_1}(\mathbf{k})\right|+1)\mathbf{1}_{\{c_1=1\}}}{\vartheta(\mathbf{k},r)+1}\right)\pi_{\mathbf{k}+\mathbf{e}^T_{(\vartheta(\mathbf{k},r)+1)}}$. Then for the first term in the right side we get $R_1(\mathbf{k}) = \frac{1}{\vartheta(\mathbf{k},r)+1}\binom{\vartheta(\mathbf{k},r)+1}{\left|\mathcal{G}_{c_1}(\mathbf{k})\right| \; \left|\mathcal{G}_{c_2}(\mathbf{k})\right| \ldots \left|\mathcal{G}_{c_r}(\mathbf{k})\right|}r^{k+1}$. After some simplifications we obtain

$$R_1(\mathbf{k}) = \binom{\vartheta(\mathbf{k}, r)}{\left|\mathcal{G}_{c_1}(\mathbf{k})\right| \; \left|\mathcal{G}_{c_2}(\mathbf{k})\right| \ldots \left|\mathcal{G}_{c_r}(\mathbf{k})\right|}r^{k+1} = \frac{\vartheta(\mathbf{k}, r)!}{\prod_{s=1}^{r}\left|\mathcal{G}_{c_s}(\mathbf{k})\right|!}r^{k+1}. \tag{2.31}$$

For the second term in the right side define

$$R_2(\mathbf{k},\mathbf{z}) \triangleq \left(\frac{\mathbf{1}_{\{c_{z+1}>c_z+1\}} + \mathbf{1}_{\{z=r\}} + (\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_{z+1}=c_z+1\}}}{\vartheta(\mathbf{k},r)} \right) \pi_{\mathbf{k}+\mathbf{e}^T_{(\vartheta(\mathbf{k},z-1)+1)}}. \tag{2.32}$$

Considering the assumptions $R_2(\mathbf{k},\mathbf{z})$ becomes as

$$R_2(\mathbf{k},\mathbf{z}) = \frac{1}{\vartheta(\mathbf{k},r)} \binom{\vartheta(\mathbf{k},r)}{\left|\mathcal{G}_{c_1}(\mathbf{k})\right| \ \left|\mathcal{G}_{c_2}(\mathbf{k})\right| \ldots \left(\left|\mathcal{G}_{c_z}(\mathbf{k})\right| - 1\right) \ 1 \ldots \left|\mathcal{G}_{c_r}(\mathbf{k})\right|} r^k, \tag{2.33}$$

that after some algebraic manipulation it is reduced to

$$R_2(\mathbf{k},\mathbf{z}) = \left|\mathcal{G}_{c_z}(\mathbf{k})\right| \frac{(\vartheta(\mathbf{k},r)-1)!}{\prod_{s=1}^{r} \left|\mathcal{G}_{c_s}(\mathbf{k})\right|!} r^k. \tag{2.34}$$

Hence the whole right hand side $R.H.S = R_1(\mathbf{k}) + \sum_{z=1}^{r} R_2(\mathbf{k},\mathbf{z})$, becomes

$$R.H.S = \frac{\vartheta(\mathbf{k},r)!}{\prod_{s=1}^{r} \left|\mathcal{G}_{c_s}(\mathbf{k})\right|!} r^{k+1} + \sum_{z=1}^{r} \left|\mathcal{G}_{c_z}(\mathbf{k})\right| \frac{(\vartheta(\mathbf{k},r)-1)!}{\prod_{s=1}^{r} \left|\mathcal{G}_{c_s}(\mathbf{k})\right|!} r^k. \tag{2.35}$$

After some simplifications, knowing that $\sum_{z=1}^{r} \left|\mathcal{G}_{c_z}(\mathbf{k})\right| = \vartheta(\mathbf{k},r)$ we obtain

$$R.H.S = \frac{\vartheta(\mathbf{k},r)!}{\prod_{s=1}^{r} \left|\mathcal{G}_{c_s}(\mathbf{k})\right|!} r^{k+1} + \frac{\vartheta(\mathbf{k},r)!}{\prod_{s=1}^{r} \left|\mathcal{G}_{c_s}(\mathbf{k})\right|!} r^k, \tag{2.36}$$

which completes the proof.

II. If $c_1 = 1$ then we get $\left(\frac{\mathbf{1}_{\{c_1>1\}} + (\left|\mathcal{G}_{c_1}(\mathbf{k})\right|+1)\mathbf{1}_{\{c_1=1\}}}{\vartheta(\mathbf{k},r)+1} \right) = \left(\frac{1+\left|\mathcal{G}_{c_1}(\mathbf{k})\right|}{\vartheta(\mathbf{k},r)+1} \right)$. Hence for $R_1(\mathbf{k})$ we get

$$R_1(\mathbf{k}) = \left(\frac{1+\left|\mathcal{G}_{c_1}(\mathbf{k})\right|}{\vartheta(\mathbf{k},r)+1} \right) \binom{\vartheta(\mathbf{k},r)+1}{\left(1+\left|\mathcal{G}_{c_1}(\mathbf{k})\right|\right) \ \left|\mathcal{G}_{c_2}(\mathbf{k})\right| \ldots \left|\mathcal{G}_{c_r}(\mathbf{k})\right|} r^{k+1}, \tag{2.37}$$

where after some simplification of (2.37) we get

$$R_1(\mathbf{k}) = \binom{\vartheta(\mathbf{k},r)}{\left|\mathcal{G}_{c_1}(\mathbf{k})\right| \ \left|\mathcal{G}_{c_2}(\mathbf{k})\right| \ldots \left|\mathcal{G}_{c_r}(\mathbf{k})\right|} r^{k+1} = \frac{\vartheta(\mathbf{k},r)!}{\prod_{s=1}^{r} \left|\mathcal{G}_{c_s}(\mathbf{k})\right|!} r^{k+1}, \tag{2.38}$$

which is the same as what we obtained previously for (2.31). From this point on the rest of steps are the same.

III. Finally, if for any arbitrary $z \in \mathbb{N} \cap [1, r-1]$, $c_{z+1} = c_z + 1$, then

$$\left(\frac{\mathbf{1}_{\{c_{z+1} > c_z + 1\}} + \mathbf{1}_{\{z=r\}} + (\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right| + 1)\mathbf{1}_{\{c_{z+1} = c_z + 1\}}}{\vartheta(\mathbf{k},r)} \right) = \left(\frac{1 + \left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right|}{\vartheta(\mathbf{k},r)} \right). \tag{2.39}$$

Thus $R_2(\mathbf{k}, \mathbf{z})$ turns out to become

$$R_2(\mathbf{k},\mathbf{z}) = \left(\frac{1 + \left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right|}{\vartheta(\mathbf{k},r)} \right) \binom{\vartheta(\mathbf{k},r)}{\left|\mathcal{G}_{c_1}(\mathbf{k})\right| \ \left|\mathcal{G}_{c_2}(\mathbf{k})\right| \ldots \left(\left|\mathcal{G}_{c_z}(\mathbf{k})\right| - 1\right) \ \left(\left|\mathcal{G}_{c_{z+1}}(\mathbf{k})\right| + 1\right) \ldots \left|\mathcal{G}_{c_r}(\mathbf{k})\right|} r^k. \tag{2.40}$$

After some manipulation (2.40) is simplified to $R_2(\mathbf{k}, \mathbf{z}) = \left|\mathcal{G}_{c_z}(\mathbf{k})\right| \frac{(\vartheta(\mathbf{k},r)-1)!}{\prod_{s=1}^{r} \left|\mathcal{G}_{c_s}(\mathbf{k})\right|!} r^k$, which is the same result as in (2.34). As a result (2.36) always remains unchanged, and all the steps for verification will be identical afterwards. Hence the proof becomes complete. ■

Appendix B

B1. Comparison of The CPC System with Erlang's Loss System

As previously mentioned, the queuing model developed in this chapter is new. One of the queuing models in the literature related to ours is $M/E_x/n/n$, the so-called *Erlang's loss system*. As in our system, this system does not have any waiting space and the only jobs in the system are the ones being served by one of the n servers. The difference comes from the operation of the servers.

In Erlang loss system, the servers operate independently, while in the CPC system the service rate of each server depends on the number of active servers in the system at any point in time. So the servers are not independent anymore. The Erlang loss formula, which represents the probability distribution of the number of jobs in the system is the following:

$$\mathbb{P}_i^E = \frac{\frac{(rx)^i}{i!}}{\sum_{j=0}^{n} \frac{(rx)^j}{j!}}, \quad 0 \leq i \leq n,$$

which we can compare to the distribution for the CPC system:

$$\mathbb{P}_i = \frac{(rx)^i}{\sum_{j=0}^{n} (rx)^j}, \quad 0 \leq i \leq n.$$

If $n = 1$ the two formulas yield the same results as expected. Nevertheless, for $n \geq 2$, the inter-dependencies of slots in the CPC system start playing a role. The following proposition compares the probability of the system being full for Erlang's loss system and the CPC system.

Proposition 33 *The probability of a fully occupied system is higher for the cost-per-click system than for Erlang's loss system, i.e.,* $\mathbb{P}_n \geq \mathbb{P}_n^E$*. In addition, the average number of jobs in the cost-per-click system is more than the average number of jobs in the Erlang's loss system, i.e.,* $L \geq L_E$*.*

This proposition is quite intuitive. Since the service rate of each server in Erlang's loss system is more than the effective service rate in the CPC system, $\mathbb{P}_n^E$ and L_E become smaller than $\mathbb{P}_n$ and L respectively.

Moreover, note that one important difference between the CPC system and Erlang loss system is that we do not need to define the *n-tuple* vector $\mathbf{k}$ to characterize Erlang system as we had to for the CPC, since its characterization is much simpler. This observation confirms that the CPC system indeed contributes to the queuing literature as it cannot be characterized in a similar way to the traditional systems.

B2. Proofs of Other Propositions

Lemma 34 *For any* $n \in \mathbb{N}$*, and* $\rho \in \mathbb{R}^+$*,* $n - (n+1)\rho + \rho^{n+1} \geq 0$.

Proof The proof is with induction. For $n = 1$ the verification is immediate. Assuming that for $n = k$, $k - (k+1)\rho + \rho^{k+1} \geq 0$. For $n = k+1$ we get

$$k + 1 - (k+2)\rho + \rho^{k+2} = \left(k - (k+1)\rho + \rho^{k+1}\right) + (\rho - 1)\left(\rho^{k+1} - 1\right). \tag{2.1}$$

Due to the induction assumption, the first term is always positive. For the second term if $\rho \geq 1$ then $\rho^{k+1} \geq 1$. Hence $(\rho - 1)\left(\rho^{k+1} - 1\right) \geq 0$. Also if $\rho \leq 1$ then $\rho^{k+1} \leq 1$.As a result $(\rho - 1)\left(\rho^{k+1} - 1\right) \geq 0$ and this completes the proof. ■

Proof of Proposition 25 In order to derive $\overline{\mu}^{j,m}$, we note that there are two streams of viewers that consider subsystem (j,m). The first stream consists of viewers who have initially considered subsystem (j,m) (for instance, ads on the top of page j), which we denote by $W^1_{j,m} := \mu\varpi_{j,m}$. Out of those viewers, $S^1_{j,m} := W^1_{j,m}(1-\mathbb{P}_0^{j,m})$ can see real ads while $B^1_{j,m} := W^1_{j,m}\mathbb{P}_0^{j,m}$ only see filler ads displayed on subsystem (j,m). Thus, $W^{1,(g,h)}_{(j,m)} := \alpha^{g,h}_{j,m}B^1_{j,m}$ of the viewers consider ads in the subsystem (g,h) (the subsystem h located in page g) and the rest leave the system. From those viewers who consider the subsystem (g,h), $B^{1,(g,h)}_{(j,m)} = W^{1,(g,h)}_{(j,m)}\mathbb{P}_0^{g,h}$ see only filler ads in the subsystem (g,h). Therefore, $W^2_{j,m} = \alpha^{j,m}_{g,h}B^{1,(g,h)}_{(j,m)}$ of the viewers check back subsystem (j,m) for real ads, while the rest leave the website. In short, $W^2_{j,m}$ is the fraction of the $W^1_{j,m}$ viewers who had initially considered ads in subsystem (j,m), but after experiencing a loop have come back to recheck subsystem (j,m) for the second time[1]. Note that theoretically the same loop of procedures can be repeated infinitely. However, in practice $\alpha^{g,h}_{j,m}$ or $\alpha^{j,m}_{g,h}$ might be near zero meaning that viewers may leave the website quickly just after once or twice check of empty (of real ads) subsystems. Given this, in loop κ, we find

$$\begin{aligned} S^\kappa_{j,m} &= W^\kappa_{j,m}(1-\mathbb{P}_0^{j,m}), \\ W^\kappa_{j,m} &= \mu\varpi_{j,m}\left(\alpha^{g,h}_{j,m}\alpha^{j,m}_{g,h}\mathbb{P}_0^{j,m}\mathbb{P}_0^{j,h}\right)^{\kappa-1}, \ \kappa = 1,2,... \end{aligned} \tag{2.2}$$

where $S^\kappa_{j,m}$ is the fraction of the $\mu\varpi_{j,m}$ viewers who had initially approached the subsystem (j,m) in the first loop and after a few checks eventually consider ads that are posted in subsystem (j,m) in loop κ. As a result, the overall number of viewers

[1]Since the subsystems can belong to different pages, the publisher might post new ads in a subsystem before a viewer re-checks it. In addition, some ads may leave the subsystems and give their place to other ads or filler ads.

in the first stream (in interaction with subsystem (g,h)) can be obtained as

$$S_{j,m} = \sum_{\kappa=1}^{\infty} S_{j,m}^{\kappa} = \mu\varpi_{j,m}(1-\mathbb{P}_0^{j,m}) \sum_{\kappa=1}^{\infty} \left(\alpha_{j,m}^{g,h}\alpha_{g,h}^{j,m}\mathbb{P}_0^{j,m}\mathbb{P}_0^{g,h}\right)^{\kappa-1}$$
$$= \frac{\mu\varpi_{j,m}(1-\mathbb{P}_0^{j,m})}{1-\alpha_{j,m}^{g,h}\alpha_{g,h}^{j,m}\mathbb{P}_0^{j,m}\mathbb{P}_0^{j,h}}. \tag{2.3}$$

The second stream of viewers includes those who had initially chosen subsystem (g,h) but finally had to approach subsystem (j,m). Based on a similar argument it can be shown that

$$S_{(g,h)}^{\kappa,(j,m)} = W_{(g,h)}^{\kappa,(j,m)}(1-\mathbb{P}_0^{j,m}), \tag{2.4}$$
$$W_{(g,h)}^{\kappa,(j,m)} = \mu\varpi_{g,h}\mathbb{P}_0^{g,h}\left(\alpha_{mh}\alpha_{hm}\mathbb{P}_0^{j,m}\mathbb{P}_0^{j,h}\right)^{\kappa-1}, \ \kappa = 1,2,...$$

where $S_{(g,h)}^{\kappa,(j,m)}$ is the fraction of the $\mu\varpi_{g,h}$ viewers who had first selected to consider ads in subsystem (g,h), but shifted to consider ads in subsystem (j,m) instead, in loop κ. Thus, the total number of viewers in the second stream is

$$S_{(g,h)}^{(j,m)} = \sum_{\kappa=1}^{\infty} S_{(g,h)}^{\kappa,(j,m)} = \mu\varpi_{g,h}\mathbb{P}_0^{g,h}(1-\mathbb{P}_0^{j,m}) \sum_{\kappa=1}^{\infty} \left(\alpha_{mh}\alpha_{hm}\mathbb{P}_0^{j,m}\mathbb{P}_0^{j,h}\right)^{\kappa-1}$$
$$= \frac{\mu\varpi_{g,h}\mathbb{P}_0^{g,h}(1-\mathbb{P}_0^{j,m})}{1-\alpha_{mh}\alpha_{hm}\mathbb{P}_0^{j,m}\mathbb{P}_0^{j,h}}. \tag{2.5}$$

Proof of Proposition 27

i) Showing $\frac{\partial \mathbb{P}_n}{\partial r} \geq 0$ is the same as showing $\frac{\partial \mathbb{P}_n}{\partial \rho} \geq 0$ in which $\rho = rx$. From Proposition (26) we have $\mathbb{P}_n = \frac{\rho^n(1-\rho)}{1-\rho^{n+1}}$. Taking the derivative of $\mathbb{P}_n$ with respect to ρ, we get

$$\frac{\partial \mathbb{P}_n}{\partial \rho} = \frac{\rho^{n-1}\left(n-(n+1)\rho+\rho^{n+1}\right)}{\left(1-\rho^{n+1}\right)^2}. \tag{2.6}$$

By Lemma 34 the numerator is always positive and the proof is complete.

ii) Since x is discrete, showing $\mathbb{P}_n(x+1)-\mathbb{P}_n(x) \geq 0$ directly is a challenging task. Instead, knowing that

$$\mathbb{P}_n(x) = \frac{(rx)^n\,(1-rx)}{1-(rx)^{n+1}},\ x \in \mathbb{N}. \tag{2.7}$$

We consider the alternative real-valued continuous function $\mathbb{Q}_n(y)$ defined by

$$\mathbb{Q}_n(y) = \frac{(ry)^n\,(1-ry)}{1-(ry)^{n+1}},\ y \in \mathbb{R}^+. \tag{2.8}$$

It is evident that for the points in which $y = x \in \mathbb{N}$ the two functions are the same; namely, $\mathbb{P}_n(x) = \mathbb{Q}_n(x)$. In order to show $\mathbb{P}_n(x)$ is increasing in x, we are enough to show $\mathbb{Q}_n(y)$ is increasing in y. This is true because $\mathbb{Q}_n(y)$ being increasing in y also implies $\mathbb{Q}_n(y+1) - \mathbb{Q}_n(y) \geq 0$ for any $y \in \mathbb{R}^+$ including any natural number x. Hence $\mathbb{Q}_n(x+1) - \mathbb{Q}_n(x) \geq 0$ for any $y \in \mathbb{N}$, which implies $\mathbb{P}_n(x+1)-\mathbb{P}_n(x) \geq 0$. But showing $\mathbb{Q}_n(y)$ is increasing in y is the same as showing $\frac{\partial \mathbb{P}_n}{\partial \rho} \geq 0$, which was proven in part (i). This completes the proof for part (ii).

In order to show $\mathbb{P}_n(x+1) - \mathbb{P}_n(x) \geq 0$ directly we have

$$\mathbb{P}_n(x+1) - \mathbb{P}_n(x) = \frac{(r(x+1))^n \sum_{i=0}^n (rx)^i - (rx)^n \sum_{i=0}^n (r(x+1))^i}{[\sum_{i=0}^n (r(x+1))^i][\sum_{i=0}^n (rx)^i]}.$$

After some manipulation, we get

$$\mathbb{P}_n(x+1) - \mathbb{P}_n(x) = \frac{\sum_{i=0}^n r^{n+i}(x^i(x+1)^n - x^n(x+1)^i)}{[\sum_{i=0}^n (r(x+1))^i][\sum_{i=0}^n (rx)^i]}.$$

But this is clear to see that $x^i(x+1)^n - x^n(x+1)^i \geq 0$. This is true because after some manipulation we get $\frac{x}{x+1} \leq 1$ that always holds. As a result $P_n(x+1) - P_n(x) \geq 0$ and this completes the second proof for part (ii).

iii) As in part (ii) showing directly that $\mathbb{P}_{n+1}(x) - \mathbb{P}_n(x) \geq 0$ is not easy. Hence we consider the alternative continuous function as $\mathbb{T}_x(m)$ defined by

$$\mathbb{T}_x(m) = \frac{(rx)^m (1-ry)}{1-(ry)^{m+1}};\ m \in \mathbb{R}. \tag{2.9}$$

We can say $\mathbb{T}_x(m)$ is a continuous version of $\mathbb{P}_n$ as when m gets natural values the two functions become the same. As a result to show $\mathbb{P}_n$ is decreasing in n we are enough to show $\mathbb{T}_x(m)$ is decreasing in m, namely to show $\frac{\partial \mathbb{T}_x(m)}{\partial m} \leq 0$. This is true because $\mathbb{T}_x(m)$ decreasing over any real number m automatically implies its decreasing over any natural number, namely $\mathbb{T}_x(m) - \mathbb{T}_x(m+1) \geq 0$ for any $m \in \mathbb{N}$. But this is the same as saying $\mathbb{P}_{n+1}(x) - \mathbb{P}_n(x) \geq 0$, $x \in \mathbb{N}$ that we are looking for. Thus we calculate $\frac{\partial \mathbb{T}_x(m)}{\partial m}$ to get

$$\frac{\partial \mathbb{T}_x(m)}{\partial m} = -\frac{(x-1)x^n Log(x)}{\left(1-x^{n+1}\right)^2}, \tag{2.10}$$

which is clearly negative, and this completes the proof of part (iii). ■

Proof of Proposition 28 (i) To show L is increasing in x we have $L(x+1) - L(x) = \sum_{i=0}^{n} i\left(\mathbb{P}_n(x+1) - \mathbb{P}_n(x)\right)$. After some simplifications we obtain

$$L(x+1) - L(x) = \frac{\left(\sum_{i=0}^{n}\sum_{j=0}^{n} i r^{i+j}\left[(x+1)^i x^j - x^i (x+1)^j\right]\right)}{\left(\sum_{j=0}^{n} r^j (x+1)^j\right)\left(\sum_{j=0}^{n} r^j x^j\right)}. \tag{2.11}$$

Hence we need to show $Z(x)$ defined as below is positive. That is,

$$Z(x) \triangleq \sum_{i=0}^{n}\sum_{j=0}^{n}(x+1)^{i+j} i r^{i+j}\left(\left(\frac{x}{x+1}\right)^{j} - \left(\frac{x}{x+1}\right)^{i}\right) \geq 0. \tag{2.12}$$

After some algebraic operations $Z(x)$ can be represented as the sum of polynomials increasing in orders of r. That is,

$$Z(x) = \sum_{s=0}^{n} G_1(x,s)r^s + \sum_{s=n+1}^{2n} G_2(x,s)r^s,$$

where,

$$G_1(x,s) \triangleq \sum_{i=0}^{s}(x+1)^s i\left(\left(\frac{x}{x+1}\right)^{s-i} - \left(\frac{x}{x+1}\right)^{i}\right), \ (x,s) \in \mathbb{N}\times\mathbb{Z}\cap[0,n], \tag{2.13}$$

and

$$G_2(x,s) \triangleq \sum_{i=s-n}^{n}(x+1)^s i\left(\left(\frac{x}{x+1}\right)^{s-i} - \left(\frac{x}{x+1}\right)^{i}\right), \ (x,s) \in \mathbb{N}\times\mathbb{Z}\cap[n+1,2n]. \tag{2.14}$$

We separately show that $G_1(x,s)$ and $G_2(x,s)$ are always positive for $s = 2k$ and for $s = 2k+1$.

i-1) For $s = 2k$, $G_1(x,s)$ is simplified to:

$$G_1(x,2k) = (x+1)^{2k} r^{2k}\sum_{i=0}^{k} i(2k-2i)\left(\left(\frac{x}{x+1}\right)^{i} - \left(\frac{x}{x+1}\right)^{2k-i}\right),$$

in which $\left(\frac{x}{x+1}\right)^{i} - \left(\frac{x}{x+1}\right)^{2k-i} \geq 0$ since $2k-i \geq i$. So the result follows. With a similar argument, for $s = 2k$,

$$G_2(x,2k) = (x+1)^{2k} r^{2k} \sum_{i=2k-n}^{k} (2k-2i) \left(\left(\frac{x}{x+1} \right)^i - \left(\frac{x}{x+1} \right)^n \right).$$

However, it can be easily verified that $\left(\frac{x}{x+1}\right)^i - \left(\frac{x}{x+1}\right)^n \geq 0$, which completes the proof.

i-2) For $s = 2k+1$ then simplify $G_1(x,s)$ as follows:

$$\begin{aligned} G_1(x,2k+1) &= (x+1)^{2k+1} r^{2k+1} \sum_{i=0}^{k-1} (2k+1-2i) \left(\left(\frac{x}{x+1} \right)^i - \left(\frac{x}{x+1} \right)^{2k+1-i} \right) \\ &\quad + k \left(\left(\frac{x}{x+1} \right)^k - \left(\frac{x}{x+1} \right)^{k+1} \right). \end{aligned} \tag{2.15}$$

We can easily see that $\left(\frac{x}{x+1}\right)^i - \left(\frac{x}{x+1}\right)^{2k+1-i} \geq 0$, since $2k+1-i \geq 0$. Also this is clear to see that $\left(\frac{x}{x+1}\right)^k - \left(\frac{x}{x+1}\right)^{k+1} \geq 0$ always holds. Therefore, $G_1(x,2k+1) \geq 0$. In a similar way,

$$\begin{aligned} G_2(x,2k+1) &= r^{2k+1}(x+1)^{2k+1} \sum_{i=2k-n}^{k-1} i(2k+1-2i) \left(\left(\frac{x}{x+1} \right)^i - \left(\frac{x}{x+1} \right)^n \right) \\ &\quad + k \left(\left(\frac{x}{x+1} \right)^k - \left(\frac{x}{x+1} \right)^n \right), \end{aligned} \tag{2.16}$$

which is always positive, and the proof is complete.

In order to prove L's convexity, we note that convexity only holds for $rx > 1$. Convexity of L on a real continuum of x implies its convexity over discrete values as well. Hence, assuming x a real variable and taking twice differentiations of L and making simplifications, we get

$$\frac{\partial^2 L}{\partial x^2} = \frac{\left(\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} r^{i+j+k} x^{i+j+k-2} (i^2 - ij)\,(i+j-1-2k) \right)}{\left(\sum_{j=0}^{n} (rx)^j \right)^3}. \tag{2.17}$$

However, it can be checked that for $rx > 1$ we have $\sum\limits_{i=0}^{n} \sum\limits_{j=0}^{n} \sum\limits_{k=0}^{n} r^{i+j+k} x^{i+j+k-2} (i^2 - ij)\,(i+j-1-2k) \leq 0$. Hence the result follows.

(ii) To show L is increasing in r, we have $L = \sum\limits_{i=0}^{n} i(rx)^i / \sum\limits_{j=0}^{n} (rx)^j$. Taking the derivative of L we get $\frac{\partial L}{\partial r} = \left(\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} (rx)^{i+j-1} (i^2 - ij) \right) / \left(\sum\limits_{j=0}^{n} (rx)^j \right)^2$. We observe that $\frac{\partial L}{\partial r} \geq 0$ if and only if $\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} (rx)^{i+j-1} (i^2 - ij) \geq 0$. Using the new indexing $s = i + j$ and simplifying, we get

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (rx)^{i+j-1} (i^2 - ij) = \sum_{s=2}^{n} \left(\sum_{i=1}^{s} i(2i-s) \right) (rx)^{s-1} + \sum_{s=n+1}^{2n} \left(\sum_{i=s-n}^{n} i(2i-s) \right) (rx)^{s-1}. \tag{2.18}$$

After some simplification, we obtain

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (rx)^{i+j-1} (i^2 - ij) = \sum_{s=2}^{n} \frac{1}{6} s(s+1)(s+2)(rx)^{s-1} + \sum_{s=n+1}^{2n} \frac{1}{6} (2n-s)(2n+1-s)(2n+2-s)(rx)^{s-1}. \tag{2.19}$$

Therefore, $\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} (rx)^{i+j-1} (i^2 - ij) \geq 0$, and the proof is complete. The proof to show the convexity of L in r has a similar procedure to part (i) and is omitted. ■

Lemma 35 *For any $n \in \mathbb{N}$ and $r \in \mathbb{R}$, $\sum_{i=0}^{n} \sum_{j=0}^{n} (rx)^i (n-i+1)(n+i-2j) \geq 0$.*

Proof Define the function $F(x) \triangleq \sum_{i=0}^{n} \sum_{j=0}^{n} (rx)^i (n-i+1)(n+i-2j)$. After some manipulation $F(x)$ is simplified to $F(x) = \sum_{i=0}^{n} (rx)^i [\sum_{j=0}^{n} (n-i+1)(n+i-2j)]$. On the other side the sum $\sum_{j=0}^{n} (n-i+1)(n+i-2j)$ is reduced to: $\sum_{j=0}^{n} (n-i+1)(n+i-2j) = i(n+1)(n+1-i)$.The result follows by replacing the reduced sum in $F(x)$. ■

Proof of Proposition 29 i) The first derivative of $R(\lambda)$ with respect to λ is

$$\frac{\partial R(\lambda)}{\partial \lambda} = \frac{\partial \Gamma(\lambda)}{\partial \lambda} p(\lambda) + \frac{\partial p(\lambda)}{\partial \lambda} \Gamma(\lambda). \tag{2.20}$$

Thus the second derivative becomes

$$\frac{\partial^2 R(\lambda)}{\partial \lambda^2} = \frac{\partial^2 \Gamma(\lambda)}{\partial \lambda^2} p(\lambda) + 2 \frac{\partial \Gamma(\lambda)}{\partial \lambda} \frac{\partial p(\lambda)}{\partial \lambda} + \frac{\partial^2 p(\lambda)}{\partial \lambda^2} \Gamma(\lambda). \tag{2.21}$$

To show $\frac{\partial^2 R(\lambda)}{\partial \lambda^2} \leq 0$, Assuming $\frac{\partial p(\lambda)}{\partial \lambda} \leq 0$ and $\frac{\partial^2 p(\lambda)}{\partial \lambda^2} \leq 0$, we need to show $\Gamma(\lambda)$ is increasing concave, namely $\frac{\partial \Gamma(\lambda)}{\partial \lambda} \geq 0$ and $\frac{\partial^2 \Gamma(\lambda)}{\partial \lambda^2} \leq 0$. By definition,

$$\Gamma(\lambda) = \lambda x (1 - \mathbb{P}_n(\lambda)) = \mu r x (1 - \mathbb{P}_n(\lambda)). \tag{2.22}$$

In order to show the result, without loss of generality, we take the derivative of $\Gamma(\lambda)$ with respect to r for a given μ. For the first derivative, we get

$$\frac{\partial \Gamma(\lambda)}{\partial r} = \mu x - \mu \frac{\partial \left(r \mathbb{P}_n(\lambda) \right)}{\partial r}. \tag{2.23}$$

Thus in order to show Γ is increasing, we need just to show $r\mathbb{P}_n(\lambda)$ is increasing in r (for a given μ). We observe that

$$r\mathbb{P}_n(\lambda) = \frac{r^{i+1}x^i}{\sum_{i=0}^{n}(rx)^i}. \tag{2.24}$$

Taking the derivative of (2.24) and simplifying gives

$$\frac{\partial\left(r\mathbb{P}_n(\lambda)\right)}{\partial r} = \left(\sum_{i=0}^{n}(n+1-i)(rx)^{i+n}\right) / \left(\sum_{i=0}^{n}(rx)^i\right)^2 \geq 0. \tag{2.25}$$

Thus $\Gamma(\lambda)$ is increasing in r. Now taking the second derivative of $\Gamma(\lambda)$ with respect to r gives

$$\frac{\partial^2\Gamma(\lambda)}{\partial r^2} = -\mu\frac{\partial^2\left(r\mathbb{P}_n(\lambda)\right)}{\partial r^2}. \tag{2.26}$$

Therefore, $\Gamma(\lambda)$ is concave if and only if $\frac{\partial^2(r\mathbb{P}_n(\lambda))}{\partial r^2} \geq 0$. To show this we take the derivative of (2.25) to get

$$\frac{\partial^2\left(r\mathbb{P}_n(\lambda)\right)}{\partial r^2} = \left(\sum_{i=0}^{n}\sum_{j=0}^{n}(rx)^{i+n-1}(n-i+1)(n+i-2j)\right) / \left(\sum_{i=0}^{n}(rx)^i\right)^3. \tag{2.27}$$

However, based on the lemma 35, $\sum_{i=0}^{n}\sum_{j=0}^{n}(rx)^{i+n-1}(n-i+1)(n+i-2j) \geq 0$. Therefore, $\frac{\partial^2(r\mathbb{P}_n(\lambda))}{\partial r^2} \geq 0$ and the result follows.

(ii) The proof of part (ii) is immediate from (2.20). ■

Lemma 36 *Let $\Gamma(\lambda) = \lambda x(1-\mathbb{P}_n(\lambda))$, $x \in \mathbb{N}$, $n \in \mathbb{N}$, $\lambda \in \mathbb{R}_+$. Then we have $\Gamma_x(\lambda) \leq \Gamma_{x+1}(\lambda)$.*

Proof In order to show $\mathit{\Gamma}_x(\lambda) \leq \mathit{\Gamma}_{x+1}(\lambda)$ we write $(\mathit{\Gamma}_{x+1}(\lambda) - \mathit{\Gamma}_x(\lambda))$ as

$$\left(\frac{\mathit{\Gamma}_{x+1}(\lambda) - \mathit{\Gamma}_x(\lambda)}{\mu}\right) = rx[\mathbb{P}_n(x) - \mathbb{P}_n(x+1)] + r[1 - \mathbb{P}_n(x+1)] \tag{2.28}$$

Furthermore, from (2.7) , for $(\mathbb{P}_n(x) - \mathbb{P}_n(x+1))$ we get

$$rx[\mathbb{P}_n(x) - \mathbb{P}_n(x+1)] = \frac{-rx\left(\sum_{i=0}^{n} r^{n+i}(x^i(x+1)^n - x^n(x+1)^i)\right)}{\left(\left(\sum_{i=0}^{n}(r(x+1))^i\right)\left(\sum_{i=0}^{n}(rx)^i\right)\right)}. \tag{2.29}$$

Likewise, for $(1 - \mathbb{P}_n(x+1))$ we obtain

$$r[1 - \mathbb{P}_n(x+1)] = \left(r\sum_{i=0}^{n-1}(r(x+1))^i\right) \Big/ \left(\sum_{i=0}^{n}(r(x+1))^i\right). \tag{2.30}$$

Replacing (2.29) and (2.30) in (2.28) and simplifying gives

$$\begin{aligned}&\left(\frac{\mathit{\Gamma}_{x+1}(\lambda) - \mathit{\Gamma}_x(\lambda)}{\mu}\right)\\ &= \frac{-rx\sum_{i=0}^{n} r^{n+i}(x^i(x+1)^n - x^n(x+1)^i) + r[\sum_{i=0}^{n-1}(r(x+1))^i][\sum_{i=0}^{n}(r(x+1))^i]}{[\sum_{i=0}^{n}(r(x+1))^i][\sum_{i=0}^{n}(rx)^i]}.\end{aligned} \tag{2.31}$$

The first sum in the numerator of (2.31) can be simplified as

$$-\sum_{i=0}^{n} r^{n+i+1}(x^{i+1}(x+1)^n - x^{n+1}(x+1)^i) = -\sum_{i=0}^{n-1} r^{n+i+1}(x^{i+1}(x+1)^n - x^{n+1}(x+1)^i). \tag{2.32}$$

Likewise, after some algebraic operations, the second sum in the numerator of (2.31) becomes

$$r[\sum_{i=0}^{n-1}(r(x+1))^i][\sum_{i=0}^{n}(r(x+1))^i] = \sum_{i=1}^{n-1} r^i \sum_{j=0}^{i} x^j(1+x)^{i-j-1} + \sum_{i=n+1}^{2n} r^i \sum_{j=i-1-n}^{n-1} x^{i-j-1}(1+x)^j, \tag{2.33}$$

where some manipulations on (2.33) yields

$$\sum_{i=1}^{n-1} r^i \sum_{j=0}^{i} x^j (1+x)^{i-j-1} + \sum_{i=0}^{n-1} r^{i+n+1} \sum_{j=i}^{n-1} x^{i+n-j}(1+x)^j \geq \sum_{i=0}^{n-1} r^{i+n+1} \sum_{j=i}^{n-1} x^{i+n-j}(1+x)^j. \tag{2.34}$$

However, it is straightforward to see that $\left(\sum_{j=i}^{n-1} x^{i+n-j}(1+x)^j\right)$ is simplified to

$$\sum_{j=i}^{n-1} x^{i+n-j}(1+x)^j = x^{i+1}(x+1)^n - x^{n+1}(x+1)^i. \tag{2.35}$$

Replacing (2.32), (2.33), (2.34), and (2.35) in (2.31) we obtain

$$\left(\frac{\Gamma_{x+1}(\lambda) - \Gamma_x(\lambda)}{\mu}\right) \tag{2.36}$$
$$\geq \frac{-\sum_{i=0}^{n-1} r^{n+i+1}(x^{i+1}(x+1)^n - x^{n+1}(x+1)^i) + \sum_{i=0}^{n-1} r^{i+n+1}(x^{i+1}(x+1)^n - x^{n+1}(x+1)^i)}{[\sum_{i=0}^{n}(r(x+1))^i][\sum_{i=0}^{n}(rx)^i]} = 0.$$

Therefore, $\Gamma_{x+1}(\lambda) \geq \Gamma_x(\lambda)$ and the proof is complete. ■

Proof of Proposition 30 (i) Adapting our notation, we denote the optimal revenue with $n+1$ slots as

$$R^*_{n+1} = R_{n+1}(\lambda^*(n+1)) = \lambda^*(n+1)(1 - \mathbb{P}_{n+1}(\lambda^*(n+1)))p(\lambda^*(n+1))x. \tag{2.37}$$

Using optimality and part (iii) of Proposition 27 we get

$$R^*_{n+1} \geq \lambda^*(n)(1 - P_{n+1}(\lambda^*(n)))p(\lambda^*(n))x \geq \lambda^*(n)(1 - P_n(\lambda^*(n)))p(\lambda^*(n))x = R^*_n, \tag{2.38}$$

which completes the proof for this part.

(ii) For the second part we again adapt our notation and denote the optimal revenue with $x+1$ clicks as

$$R^*_{x+1} = R_{x+1}(\lambda^*(x+1)) = \mathit{\Gamma}_{x+1}(\lambda^*(x+1))\mu p(\lambda^*(x+1)). \tag{2.39}$$

Using optimality and Lemma 36 we have

$$R^*_{x+1} \geq \mathit{\Gamma}_{x+1}(\lambda^*(x))\mu p(\lambda^*(x)) \geq \mathit{\Gamma}_{x}(\lambda^*(x))\mu p(\lambda^*(x)) = R^*_x, \tag{2.40}$$

which completes the second part of the proof.

(iii) For the third part we note that the busy probability $\mathbb{P}_n$ depends only on $r = \lambda/\mu$, not on λ and μ separately. Adapting our notation we denote the optimal revenues with μ as the arrival rate of the viewer as

$$R^*(\mu) = R(\lambda^*(\mu), \mu) = \lambda^*(\mu)(1 - \mathbb{P}_n(\lambda^*(\mu)/\mu))p(\lambda^*(\mu))x. \tag{2.41}$$

According to Part (i) of Proposition 27 $\mathbb{P}_n$ is increasing in r and thus for a given λ decreasing in μ. Using that fact and optimality we have for $\mu_1 \geq \mu_2$ that

$$R^*(\mu_1) \geq \lambda^*(\mu_2)(1-\mathbb{P}_n(\lambda^*(\mu_2)/\mu_1))p(\lambda^*(\mu_2)) \geq \lambda^*(\mu_2)(1-\mathbb{P}_n(\lambda^*(\mu_2)/\mu_2))p(\lambda^*(\mu_2)) = R^*(\mu_2), \tag{2.42}$$

Which completes the proof. ■

Lemma 37 *Let $\mathit{\Gamma}(r) = rx(1-\mathbb{P}_n(x,r))$ in which $\mathbb{P}_n(x,r)$ is the full state probability. Then $\mathit{\Gamma}(r)$ is increasing concave in r (and hence in λ).*

Proof Taking the first derivative of $\Gamma(r) = \frac{\sum_{i=1}^{n}(rx)^i}{\sum_{j=0}^{n}(rx)^j}$ and simplifying gives

$$\frac{\partial \Gamma(r)}{\partial r} = \frac{\sum_{i=1}^{n}\sum_{j=0}^{n}(rx)^{i+j}(i-j)}{r\left(\sum_{j=0}^{n}(rx)^j\right)^2} = \frac{\sum_{i=1}^{n} i(rx)^i}{r\left(\sum_{j=0}^{n}(rx)^j\right)^2} \geq 0, \tag{2.43}$$

which is indicating that $\Gamma(r)$ is increasing in r. In order to get the convexity, we take the second derivative of $\Gamma(r)$. After some simplifications, we get

$$\frac{\partial^2 \Gamma(r)}{\partial r^2} = \frac{\sum_{i=1}^{n}\sum_{j=0}^{n}(rx)^{i+j}i(i-1-2j)}{r^2\left(\sum_{j=0}^{n}(rx)^j\right)^3}, \tag{2.44}$$

where after the reindexing $s = i + j$, becomes

$$\frac{\partial^2 \Gamma(r)}{\partial r^2} = \frac{\sum_{s=1}^{n}(rx)^s \sum_{i=1}^{s} i(3i-2s-1) + \sum_{s=n+1}^{2n}(rx)^s \sum_{i=s-n}^{n} i(3i-2s-1)}{r^2\left(\sum_{j=0}^{n}(rx)^j\right)^3} \leq 0, \tag{2.45}$$

which proves the convexity. To see why (2.45) holds it is effortless to verify that $\sum_{i=1}^{s} i(3i-2s-1) = 0$; $1 \leq s \leq n$, and $\sum_{i=s-n}^{n} i(3i-2s-1) = (n+1)(2n+1-s)(n-s) \leq 0$; $n+1 \leq s \leq 2n$, hence the result follows. ■

Proof of Proposition 31 We need to show that $\frac{\partial \lambda^*}{\partial x} \leq 0$. By Implicit Function Theorem, we get $\frac{\partial \lambda^*}{\partial x}$ as

$$\frac{\partial \lambda^*}{\partial x} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial \lambda^*}} = -\frac{\frac{\partial}{\partial x}(\Gamma'(\lambda^*))p(\lambda^*) + \frac{\partial}{\partial x}(\Gamma(\lambda^*))p'(\lambda^*)}{\Gamma''(\lambda^*)p(\lambda^*) + \Gamma'(\lambda^*)p'(\lambda^*) + \Gamma'(\lambda^*)p'(\lambda^*) + \Gamma(\lambda^*)p''(\lambda^*)}, \tag{2.46}$$

in which by First Order Necessary Condition we have

$$F = \Gamma'(\lambda^*)p(\lambda^*) + \Gamma(\lambda^*)p'(\lambda^*) = 0. \tag{2.47}$$

Note that since x is discrete we are slightly abusing the Implicit Function Theorem. Consider λ^* to be a continuous function of x rather than the discrete one. If we show λ^* is increasing in real x we have indeed shown it increases in any integer x. Similarly, if we show the functions $\Gamma(\lambda^*)$ and $\Gamma'(\lambda^*)$ are increasing (/ decreasing) in any increasing sequence of real values x then the monotonicity is automatically transferred to any increasing sequence of integer values x. Since, $p(\lambda^*) > 0$, $p'(\lambda^*) < 0$, $p''(\lambda^*) < 0$ and $\Gamma(\lambda^*) > 0$, $\Gamma'(\lambda^*) > 0$, $\Gamma''(\lambda^*) < 0$ (See Lemma (37)), the denominator is negative. Hence, we are left with showing

$$\frac{\partial}{\partial x}(\Gamma'(\lambda^*))p(\lambda^*) + \frac{\partial}{\partial x}(\Gamma(\lambda^*))p'(\lambda^*) \leq 0. \tag{2.48}$$

Using the *FONC*, $\Gamma'(\lambda^*)p(\lambda^*) + \Gamma(\lambda^*)p'(\lambda^*) = 0$, we need to show

$$g(\lambda^*) = \frac{\partial}{\partial x}(\Gamma(\lambda^*))\Gamma'(\lambda^*) - \frac{\partial}{\partial x}(\Gamma'(\lambda^*))\Gamma(\lambda^*) \geq 0. \tag{2.49}$$

Without loss of generality, we set $\mu = 1$ and thus $\lambda^* = r$. Now we have $\Gamma(r) = \frac{\sum_{i=1}^n (rx)^i}{\sum_{j=0}^n (rx)^j}$. Hence we get

$$\frac{\partial \Gamma(r)}{\partial x} = \frac{\left(\sum_{i=1}^n ix^{i-1}r^i\right)\left(\sum_{j=0}^n (rx)^j\right) - \left(\sum_{i=1}^n (rx)^i\right)\left(\sum_{j=0}^n jr^{j-1}r^j\right)}{\left(\sum_{j=0}^n (rx)^j\right)^2}.$$

After some simplifications, we get

$$\frac{\partial \Gamma(r)}{\partial x} = \frac{\sum_{i=1}^n \sum_{j=0}^n (rx)^{i+j}(i-j)}{x\left(\sum_{j=0}^n (rx)^j\right)^2} = \frac{\sum_{i=1}^n i(rx)^i}{x\left(\sum_{j=0}^n (rx)^j\right)^2}.$$

Hence

$$\frac{\partial \Gamma(r)}{\partial x}\frac{\partial \Gamma(r)}{\partial r} = \frac{\sum_{i=1}^n \sum_{j=1}^n (rx)^{i+j-1}ij}{\left(\sum_{j=0}^n (rx)^j\right)^4}. \tag{2.50}$$

Moreover $\frac{\partial}{\partial x}(\Gamma'(\lambda^*))$ is obtained as

$$\frac{\partial}{\partial x}(\Gamma'(\lambda^*)) = \frac{\sum_{i=1}^{n}\sum_{j=0}^{n}(rx)^{i+j-1}(i^2-2ij)}{\left(\sum_{j=0}^{n}(rx)^j\right)^3}.$$

Hence $\frac{\partial}{\partial x}(\Gamma'(\lambda^*))\Gamma(\lambda^*)$ becomes

$$\frac{\partial}{\partial x}(\Gamma'(\lambda^*))\Gamma(\lambda^*) = \frac{\sum_{i=1}^{n}\sum_{j=0}^{n}\sum_{k=1}^{n}(rx)^{i+j+k-1}(i^2-2ij)}{\left(\sum_{j=0}^{n}(rx)^j\right)^4}. \tag{2.51}$$

In order to show $g(\lambda^*) \geq 0$ we are enough to show

$$P(x,r,n) = \sum_{i=1}^{n}\sum_{j=0}^{n}(rx)^{i+j-1}ij - \sum_{i=1}^{n}\sum_{j=0}^{n}\sum_{k=1}^{n}(rx)^{i+j+k-1}(i^2-2ij) \geq 0.$$

Note that we can represent $P(x,r,n)$ as

$$P(x,r,n) = \sum_{s=1}^{3n} c_s (rx)^{s-1},$$

in which c_s is the appropriate coefficient of the term $(rx)^{s-1}$. We show that $P(x,r,n) \geq 0$ by showing $c_s \geq 0$, $2 \leq s \leq 3n$. Consider the three intervals $2 \leq s \leq n$, $n+1 \leq s \leq 2n$, and $2n+1 \leq s \leq 3n$. We need to show $c_s \geq 0$ holds in each of these intervals separately. Since the procedures of proofs for the three intervals are similar we only show for $2 \leq s \leq n$. For every $s \in \{2, ..., n\}$ the coefficient of the term $(rx)^{s-1}$ in $\sum_{i=1}^{n}\sum_{j=0}^{n}(rx)^{i+j-1}ij$ becomes $\sum_{i=1}^{s} i(s-i)$, while the coefficient of $(rx)^{s-1}$ in the second sum, $\sum_{i=1}^{n}\sum_{j=0}^{n}\sum_{k=1}^{n}(rx)^{i+j+k-1}(i^2-2ij)$ becomes $\sum_{i=1}^{s}\sum_{k=1}^{s-i}(i^2-2i(s-i-k))$. Hence we get

$$c_s = \sum_{i=1}^{s} i(s-i) - \sum_{i=1}^{s}\sum_{k=1}^{s-i}(i^2-2i(s-i-k)) = 0. \tag{2.52}$$

This is true because $\sum_{i=1}^{s} i(s-i) = \sum_{i=1}^{s} \sum_{k=1}^{s-i} (i^2 - 2i(s-i-k)) = \frac{1}{6}s(s+1)(s+2)$. Using the same procedure it can be checked that $c_s \geq 0$ for other intervals as well.

■

Appendix C

C1. Graphical Presentation of the Revenue Gap in Section 5.3

The following figure illustrates a schematic presentation of how the revenue gap is computed through the steps 1-3 in Section 5.3 (Non-Poisson Arrivals).

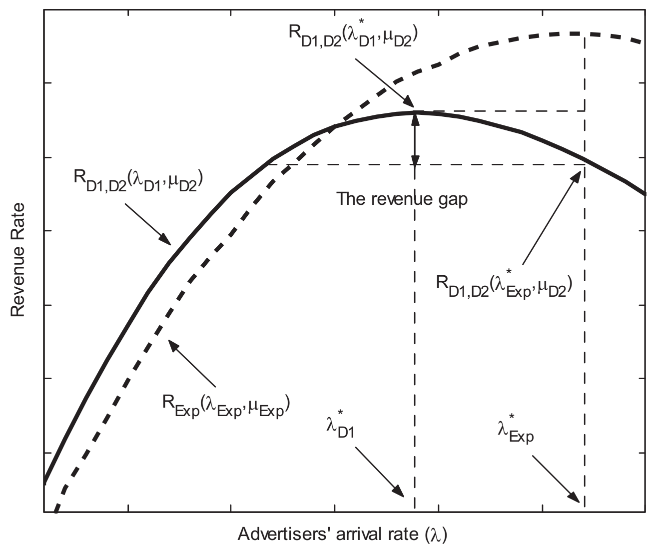
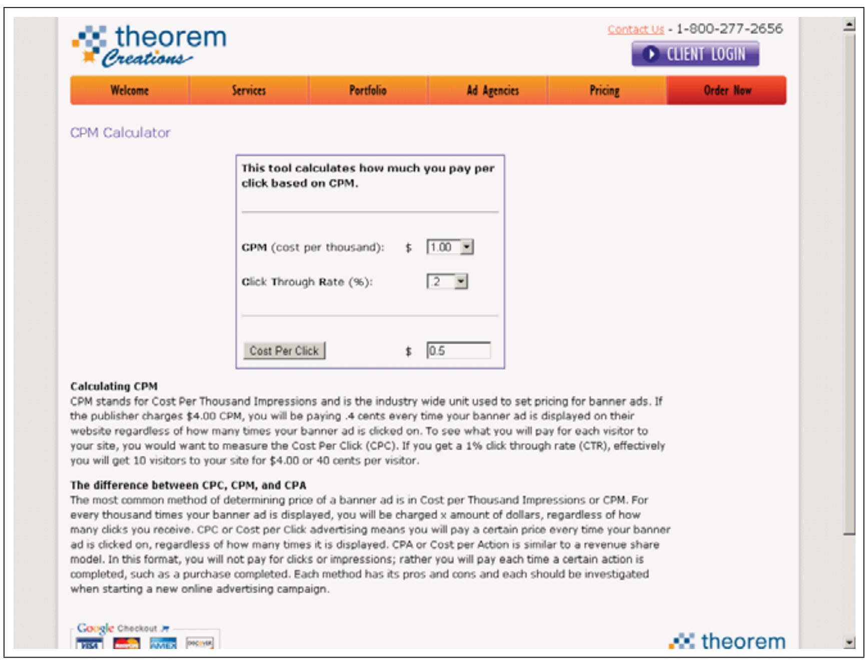


Figure 2.1: A schematic presentation of calculating the revenue gap when the advertisers' and the viewers' interarrival processes are non-Poisson, and the price function depends on the number of clicks X in which X is a random variable.

C2. Examples of Real Publishers Obtaining CPC By Dividing CPM by CTR

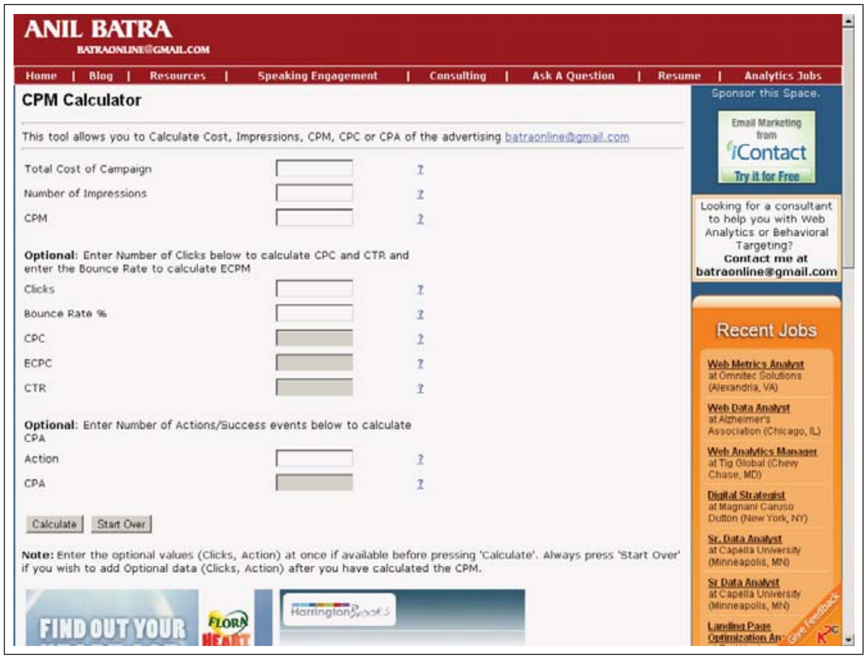
In this section, we give two examples of real publishers that obtain the CPC price by dividing the CPM price by the CTR.

Theorem Creations The first example is from the leading web banner ad design firm *Theorem Creations*. This well-known company specializes in rich media online advertising across all media platforms including Flash, Animated GIF, DART Motif, PointRoll, Eyeblaster, Yahoo Rich Media, Atlas, and other banner ad technologies. As depicted in the slide below, this company provides a relatively full explanation of how to convert the CPM prices to the CPC prices by dividing the CPM prices by the CTR. To see the company's page visit: *http://www.theoremcreations.com/ppc/cpm_calculator.php* .



Anil Batra.com The other example (below slide) belongs to the website of a leading Search and Analytics Practitioner, which recommends the same heuristic for converting the prices.

The page can be reached at: *http://www.anilbatra.com/digitalmarketing/web-analytics-jobs.asp*.


ANIL BATRA
BATRAONLINE@GMAIL.COM
Home | Blog | Resources | Speaking Engagement | Consulting | Ask A Question | Resume | Analytics Jobs
CPM Calculator
This tool allows you to Calculate Cost, Impressions, CPM, CPC or CPA of the advertising batraonline@gmail.com
Total Cost of Campaign
Number of Impressions
CPM
Optional: Enter Number of Clicks below to calculate CPC and CTR and enter the Bounce Rate to calculate ECPM
Clicks
Bounce Rate %
CPC
ECPC
CTR
Optional: Enter Number of Actions/Success events below to calculate CPA
Action
CPA
Calculate
Start Over
Note: Enter the optional values (Clicks, Action) at once if available before pressing 'Calculate'. Always press 'Start Over' if you wish to add Optional data (Clicks, Action) after you have calculated the CPM.
FIND OUT YOUR
Sponsor this Space.
Email Marketing from iContact
Try it for Free
Looking for a consultant to help you with Web Analytics or Behavioral Targeting? Contact me at batraonline@gmail.com
Recent Jobs
Web Metrics Analyst at Omnitec Solutions (Alexandria, VA)
Web Data Analyst at Alzheimer's Association (Chicago, IL)
Web Analytics Manager at Tig Global (Chew Chase, MD)
Digital Strategist at Magnani Caruso Dutton (New York, NY)
Sr. Data Analyst at Capella University (Minneapolis, MN)
Sr Data Analyst at Capella University (Minneapolis, MN)
Landing Page

Appendix D

D1. Direct Sales Channels

In Chapter 2, we focused on the CPC pricing scheme through advertising networks. Cost per click pricing through ad networks captures around 25% of the whole online advertising market (IAB 2009; Business Week 2009; Media Banker 2009). One of the related problems, which can be explored is how CPC pricing policies can be affected if a publisher sells the slots to the advertisers using its own direct sale's channel rather than using ad networks [2].

In this section, we do not intend to model the CPC pricing of the publishers using direct sales channel in detail as it focuses on a different section of the market and adds several layers of complexities, which make its analysis beyond the scope of this chapter. In this section, we only consider a special case, where the publisher deals with impatient advertisers only, and leave the more extensive analysis of this extension for the future research.

In order to consider this extension, we assume the web publisher has only a single page. We label the slots from 1 to n. The slots can be different from each other. We define the type-i advertisers to be the advertisers interested in occupying the slot i ($0 \leq i \leq n$). Similarly, we consider the type-i viewers to be the visitors whose first clicking-decision is to click is on the ad i. We let the type-i advertisers' and viewers' arrival rates be λ_i, and μ_i respectively. We let the number of slots be equal to n. Upon arrival at the publisher's system if a type-i advertisers realize that the slot i is unavailable, they consider slot j ($j \neq i$), instead, with the probability $p_{ij} = \alpha_k \, \lambda_j / \lambda_i$,

[2] Note that the share of the CPC pricing though publishers' direct sales' channels is also about 25% of the whole market. That said, the two different section are approximately equally divided. (IAB 2010; Business Week 2009; Media Banker 2009)

with $\alpha_k \leq \min\{\lambda_j/\lambda_i\}$ for all $0 \leq i, j \leq n$. k is the number of ads in the publisher's system. The structure of p_{ij} has been made for tractability. p_{ij}'s structure implies that if slot i is unavailable the advertisers are more likely to check more popular slots first.

Similarly, upon arrival at the publisher's system if a viewer faces with an empty slot or a filler ad he considers the ad positioned in slot j with the probability $p_{ij} = \gamma_{n-k}\, \mu_j/\mu_i$, where γ_{n-k} is a coefficient, which is related to the number of empty slots with $\gamma_{n-k} \leq \min\{\mu_j/\mu_i\}$ for all $0 \leq i, j \leq n$. That is, if a viewer decides to consider other ads in the system, he is more likely to consider and click on the more popular ones. The following proposition gives the *closed-form* solution of the steady state probability of the number of clicks left for each ad in the system.

Proposition 38 *Let λ_k and μ_k be the type-k advertisers' and viewers' arrival rates respectively. Then the general flow balance equation of the publisher's system is:*

$$\left((1+\gamma_{|\mathcal{G}_0(\mathbf{k})|}) \sum_{j\in\mathcal{G}_{>0}(\mathbf{k})} \mu_j + (1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|}) \sum_{j\in\mathcal{G}_0(\mathbf{k})} \lambda_j \right) \pi_{(\mathbf{k})} \tag{2.53}$$

$$= \sum_{j\in(\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k}))\cup\mathcal{G}_0(\mathbf{k})} \mu_j (1+\gamma_{|\mathcal{G}_0(\mathbf{k})\setminus\{j\}|}) \pi_{(\mathbf{k}+\mathbf{e}_j^T)} + \sum_{j\in\mathcal{G}_x(\mathbf{k})} \lambda_j (1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1}) \pi_{(\mathbf{k}-x_j\mathbf{e}_j^T)},$$

where x_k is the number of clicks requested by the type-k advertiser. Furthermore, the steady state probability of the number of clicks left to satisfy for each ad is expressed as

$$\pi_{(\mathbf{k})} = \eta \left(\Pi_{i=1}^{|\mathcal{G}_{>0}(\mathbf{k})|-1}(1+\alpha_i)\right) \left(\Pi_{i\in\mathcal{G}_{>0}(\mathbf{k})}\lambda_i\right) \left(\Pi_{j=1}^{|\mathcal{G}_0(\mathbf{k})|-1}(1+\gamma_j)\right) \left(\Pi_{j\in\mathcal{G}_0(\mathbf{k})}\mu_j\right), \tag{2.54}$$

where η is a positive coefficient, which is obtained as

$$\eta = \left(\sum_{u_1=0}^{x_1} \dots \sum_{u_n=0}^{x_n} \left(\Pi_{i=1}^{|\mathcal{G}_{>0}(\mathbf{U})|-1} (1+\alpha_i) \right) \left(\Pi_{i \in \mathcal{G}_{>0}(\mathbf{U})} \lambda_i \right) \left(\Pi_{j=1}^{|\mathcal{G}_{0}(\mathbf{U})|-1} (1+\gamma_j) \right) \left(\Pi_{j \in \mathcal{G}_{0}(\mathbf{U})} \mu_j \right) \right)^{-1} . \tag{2.55}$$

D2. Random Display Ad Rotations

Another extension to the CPC model is its natural connection to the CPM system in which multiple ads are displayed, one at a time, in the location of a single slot. In order to consider this relation, we let the number of slots be merely equal to 1. (An example of a system with a single slot is the free E-mail environment websites, which often provide a single skyscraper ad slot on the right side or top of the page.) This single slot is used to display up to n individual banner ads, one at a time. We let advertisers arrive at the system with the stationary rate λ, and request their ads to be shown to x unique viewers appearing to the system. We also let the viewers arrive at the system with the stationary rate μ. Upon arrival at the system, the viewer sees only one of the ads on hand with identical probability.

It is easy to notice that the abovementioned system, in essence, corresponds to the CPC model that we considered in Chapter 2. Hence, all the results discussed in Section 2.6 of Chapter 2 are applied here as well. For example, if the ads are given different display weights, which is in fact more common in practice, the system's performance becomes almost the same as the CPC. By the same token, the results do not change if the advertisers request randomly different clicks, or, even if their arrival rates constantly change instead of being stationary. Nevertheless, with more slots (unless there are no inflows among them), investigating the system's characteristics turns out to be excessively complex. Exploring this system is beyond the scope of this manuscript, and we leave it for future research.

D3. Slot Assignment Decisions Among Subsystems

D3.1. Slot Assignment Decisions Between Two CPC Subsystems

In Chapters 1 and 2, we focused on pricing of the display as one of the significant problems that web publishers face. In those chapters, the number of slots in each subsystem was considered fixed. One of the related problems is the capacity decisions. That is, the publisher may know the prices of each subsystem and wishes to know how to split the total capacity of each page (in terms of the possible number of slots) into different subsystems. In this section, we consider briefly this issue using a stylized model, and leave a deeper analysis of this issue for future research.

In order to start, we assume the publisher has only a single webpage. The page contains S distinct subsystems, namely $s = 1, 2, ..., S$. There are n slots in the page on the whole. We assume that the slots can be assigned to any of the subsystems with only little change in the size and shape. Each subsystem s contains n_s slots, which is the decision variable. Let λ_s be the effective arrival rate of the advertisers interested in posting their ads in subsystem s. Likewise, let μ_s be the viewers' arrival rate whose first decision is to click on one of the slots in subsystem s. Each advertiser interested in subsystem s requests x_s clicks on average. The price per click for each ad in the subsystem s is assumed to be p_s. The web publisher wishes to maximize the page's overall expected revenue rate given the best slot assignments the decision variables n_s, $s = 1, 2, ..., S$. Thus, the web publisher's maximization problem can be expressed as

$$
\begin{aligned}
\underset{\{n_1,\ldots,n_S\}}{\text{Max}} \ R(n_1,\ldots,n_S) &= \sum_{s=1}^{S} \lambda_s(1-\mathbb{P}^s_{n_s}(\lambda_s,\mu_s,x_s,n_s))p_s x_s \qquad (2.56)\\
s.t. \ \sum_{s=1}^{S} n_s &= n,
\end{aligned}
$$

where $\mathbb{P}^s_i(\lambda_s,\mu_s,x_s,n_s)$ is the probability of having i advertisers in the subsystem s. For convenience, we consider $S = 2$. That is, we assume that the page has only two subsystems, namely, high price and low price. Then, the publisher's objective function can be expressed as

$$
\underset{n_1}{\text{Max}} \ R(n_1;n) = \frac{\lambda_1 x_1 p_1(1-(r_1x_1)^{n_1})}{(1-(r_1x_1)^{n_1+1})} + \frac{\lambda_2 x_2 p_2(1-(r_2x_2)^{n-n_1})}{(1-(r_2x_2)^{n-n_1+1})}. \qquad (2.57)
$$

The subsequent proposition gives the conditions that determine the optimal policy for the slot assignments between two subsystems.

Proposition 39 *In an n-slot system with $S = 2$ independent subsystems the optimal policy for slot assignments satisfies*

$$
\begin{aligned}
\frac{\lambda_1 x_1 p_1 \mathbb{P}^1_{n_1^*-1}(x_1, n_1^*-1)}{\lambda_2 x_2 p_2 \mathbb{P}^2_{n_2^*-1}(x_2, n-n_1^*-1)} &\geq \frac{\mathbb{P}^2_0(x_2, n-n_1^*)}{\mathbb{P}^1_0(x_1, n_1^*)}, \qquad (2.58)\\
\frac{\lambda_1 x_1 p_1 \mathbb{P}^1_{n_1^*}(x_1, n_1^*)}{\lambda_2 x_2 p_2 \mathbb{P}^2_{n_2^*-2}(x_2, n-n_1^*-2)} &\leq \frac{\mathbb{P}^2_0(x_2, n-n_1^*-1)}{\mathbb{P}^1_0(x_1, n_1^*+1)}. \qquad (2.59)
\end{aligned}
$$

(2.58) implies that at the optimal allocation level, the ratio of the first subsystem's lost sales with capacity (n_1^*-1) to the second subsystem's lost sales with capacity $(n-n_1^*-1)$ is more than the ratio of the empty-state probability of the second subsystem to the first, when one unit of capacity is added to each of the subsystems. Likewise, inequality (2.59) implies that the ratio of the first subsystem's lost sales to

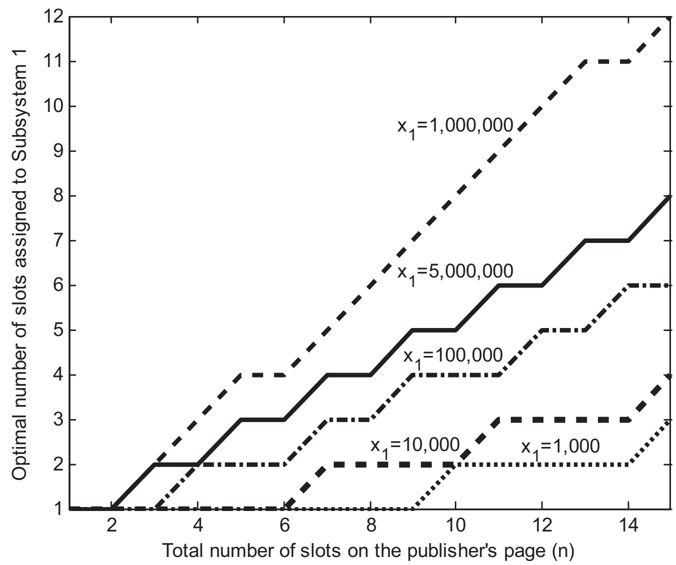


Figure 2.2: The optimal number of slots assigned to Subsystem 1. It can be observed that the optimal number of slots allocated to a subsystem is non-monotonic in the requested number of clicks.

the second subsystem's lost sales with the respective capacities n_1^*, and $n - n_1^* - 2$ is less than the ratio of the empty-state probability of the second subsystem to the first, when one unit of capacity is added to each of the subsystems.

As an illustrative example, we set the advertisers' arrival rates for both subsystems equal to 1. We also set the viewers arrival rates for the subsystems to $\mu_1 = 1,000,000$ and $\mu_2 = 500,000$ per time unit. Further, we consider the prices p_1 and p_2 to be equal to \$0.75 and \$2 per every click respectively. The advertisers for the subsystem 2 request $x_2 = 1,000,000$ clicks at all times. We vary the number of slots in the system from 1 to 15. Figure (2.2) shows the optimal number of slots assigned to subsystem 1 for different values of x_1. From the figure, it is clear that the optimal number of slots assigned to each subsystem is not monotone with respect to the requested numbers of clicks for that subsystem.

D3.2. Slot Assignments Between A CPC and A CPM Subsystems

Next, we consider the related problem of the optimal slot assignments between a CPC and a CPM subsystems. Clearly, using (2.10), the publisher's maximization problem can be expressed as

$$\operatorname*{Max}_{n^{cpm}} R(n^{cpm}) = \lambda^{cpm} x^{cpm} p^{cpm} \frac{\sum_{j=0}^{n^{cpm}-1} \binom{x^{cpm}+n^{cpm}-1}{j} r_{cpm}^{j}}{\sum_{j=0}^{n^{cpm}} \binom{x^{cpm}+n^{cpm}-1}{j} r_{cpm}^{j}} + \frac{\lambda^{cpc} x^{cpc} p^{cpc} (1 - (r_{cpc} x^{cpc})^{n-n^{cpm}})}{(1 - (r_{cpc} x^{cpc})^{n-n^{cpm}+1})}, \tag{2.60}$$

where n^{cpm} is the number of slots assigned to the CPM subsystem. Although (2.60) considers only one subsystem from each type, it is still very difficult to solve analytically. Hence, we restrict our focus on only an illustrative numerical analysis.

As in the previous section, we let the advertisers' arrival rates at both subsystems be equal to 1. We also let the viewers' arrival rates for the two subsystems be $\mu_1 = 1,000,000$, and $\mu_2 = 50,000$ per time unit. Furthermore, we consider the CPM price $p^{cpm} = 0.05$, and the CPC price $p^{cpc} = 5$. The number of impressions requested for the CPM subsystem is on average $x^{cpm} = 1,000,000$. Figure (2.3) shows the optimal number of slots assigned to the CPC subsystem for different values of requested clicks x^{cpc}.

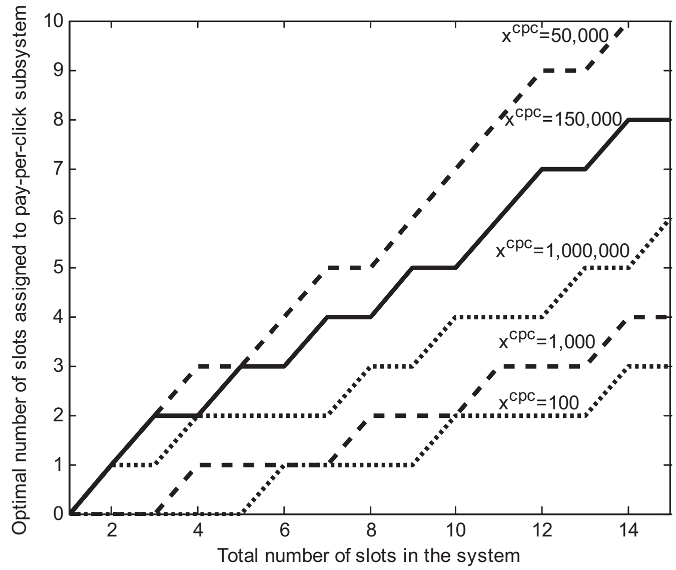


Figure 2.3: Optimal number of slots assigned to the CPC subsystem. It can be observed that the optimal number of slots allocated to the CPC subsystem is non-monotonic in the requested number of clicks.

Appendix E

Proof of Propositions in Appendix D

Proof of Proposition 38 Let $\mathcal{G}_{>0}(\mathbf{k}) = \{i \,|k_i > 0\}$, $\mathcal{G}_0(\mathbf{k}) = \{i\,|k_i = 0\}$,and $\mathcal{G}_x(\mathbf{k}) = \{i\,|k_i = x_i\}$, where x_i is the number of clicks assigned to the banner located in slot i. Let

$\mathbf{e}_i^T$ be an n tuple row vector with 1 in the ith position and zero elsewhere. Then the state of the system can be written as

$$\mathbf{k} = (k_1, k_2, ..., k_n) = \sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T + \sum_{i \in \mathcal{G}_x(\mathbf{k})} x_i \mathbf{e}_i^T . \tag{2.61}$$

Hence all the flow balance equations of the system can be represented as the following single general transition equation. That is,

$$\left((1 + \gamma_{|\mathcal{G}_0(\mathbf{k})|}) \sum_{j \in \mathcal{G}_{>0}(\mathbf{k})} \mu_j + (1 + \alpha_{|\mathcal{G}_{>0}(\mathbf{k})|}) \sum_{j \in \mathcal{G}_0(\mathbf{k})} \lambda_j \right) \pi_{(K)} \tag{2.62}$$

$$= \sum_{j \in (\mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})) \cup \mathcal{G}_0(\mathbf{k})} \mu_j (1 + \gamma_{|\mathcal{G}_0(\mathbf{k}) \backslash \{j\}|}) \pi_{(\mathbf{k} + \mathbf{e}_j^T)} + \sum_{j \in \mathcal{G}_x(\mathbf{k})} \lambda_j (1 + \alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1}) \pi_{(\mathbf{k} - x_j \mathbf{e}_j^T)} .$$

In order to illustrate better, before the rest of the proof let us see some examples for this.

Example 1 If $\mathbf{k} = (k_1, x_2, 0, k_4)$ then $n = 4, \mathcal{G}_{>0}(\mathbf{k}) = \{1, 2, 4\}, \mathcal{G}_0(\mathbf{k}) = \{3\}, \mathcal{G}_x(\mathbf{k}) = \{2\}, \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k}) = \{1, 4\}, (\mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})) \cup \mathcal{G}_0(\mathbf{k}) = \{1, 2, 4\}, |\mathcal{G}_{>0}(\mathbf{k})| = 3, |\mathcal{G}_0(\mathbf{k})| =$

1. Hence we get

$$[(1+\gamma_1)\sum_{j\in\mathcal{G}_{>0}(\mathbf{k})}\mu_j+(1+\alpha_2)\lambda_3]\pi_{(\mathbf{k})}$$
$$=\sum_{j\in(\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k}))\cup\mathcal{G}_0(\mathbf{k})}\mu_j(1+\gamma_{|\mathcal{G}_0(\mathbf{k})\setminus\{j\}|})\pi_{(\mathbf{k}+\mathbf{e}_j^T)}+\lambda_3(1+\alpha_2)\pi_{(\mathbf{k}-x_2\mathbf{e}_2^T)}. \tag{2.63}$$

By expanding, we obtain

$$\begin{aligned}[(1+\gamma_1)\sum_{j\in\{1,2,4\}}\mu_j+(1+\alpha_2)\lambda_3]\pi_{(k_1,x_2,0,k_4)} &= \sum_{j\in\{1,4\}}\mu_j(1+\gamma_1)\pi_{((k_1,x_2,0,k_4)+\mathbf{e}_j^T)}\\ &+\sum_{j\in\{3\}}\mu_j\underbrace{(1+\gamma_0)}_{1}\pi_{((k_1,x_2,0,k_4)+\mathbf{e}_j^T)}\\ &+\lambda_3(1+\alpha_2)\pi_{((k_1,x_2,0,k_4)-x_2(0,1,0,0))},\end{aligned} \tag{2.64}$$

which is simplified to

$$\begin{aligned}[(1+\gamma_1)\sum_{j\in\{1,2,4\}}\mu_j+(1+\alpha_2)\lambda_3]\pi_{(k_1,x_2,0,k_4)} &= \mu_1(1+\gamma_1)\pi_{(k_1+1,x_2,0,k_4)}\\ &+\mu_4(1+\gamma_1)\pi_{(k_1,x_2,0,k_4+1)}+\\ &\mu_3\pi_{(k_1,x_2,1,k_4)}+\lambda_3(1+\alpha_2)\pi_{(k_1,0,0,k_4)}.\end{aligned} \tag{2.65}$$

By expanding, we obtain

$$\begin{aligned}[(1+\gamma_1)\sum_{j\in\{1,2,4\}}\mu_j+(1+\alpha_2)\lambda_3]\pi_{(k_1,x_2,0,k_4)} &= \sum_{j\in\{1,4\}}\mu_j(1+\gamma_1)\pi_{((k_1,x_2,0,k_4)+\mathbf{e}_j^T)}\\ &+\sum_{j\in\{3\}}\mu_j\underbrace{(1+\gamma_0)}_{1}\pi_{((k_1,x_2,0,k_4)+\mathbf{e}_j^T)}\\ &+\lambda_3(1+\alpha_2)\pi_{((k_1,x_2,0,k_4)-x_2(0,1,0,0))},\end{aligned} \tag{2.66}$$

which is simplified to

$$[(1+\gamma_1)\sum_{j\in\{1,2,4\}}\mu_j+(1+\alpha_2)\lambda_3]\pi_{(k_1,x_2,0,k_4)} = \mu_1(1+\gamma_1)\pi_{(k_1+1,x_2,0,k_4)} \tag{2.67}$$

$$+\mu_4(1+\gamma_1)\pi_{(k_1,x_2,0,k_4+1)}$$

$$+\mu_3\pi_{(k_1,x_2,1,k_4)}+\lambda_3(1+\alpha_2)\pi_{(k_1,0,0,k_4)}.$$

Example 2 If in (2.62) we have $\mathcal{G}_{>0}(\mathbf{k})=\{1,2,...,n\}$, and $\mathcal{G}_0(\mathbf{k})=\mathcal{G}_x(\mathbf{k})=\{\}$ then $|\mathcal{G}_{>0}(\mathbf{k})|=n$, $|\mathcal{G}_0(\mathbf{k})|=0$,and $\mathbf{k}=\sum_{i=1}^{n}k_i\mathbf{e}_i^T$ $(k_i<x_i)$. Therefore, we get

$$\left(\sum_{j=1}^{n}\mu_j\right)\pi_{(\sum_{i=1}^{n}k_i\mathbf{e}_i^T)}=\sum_{j=1}^{n}\mu_j\pi_{(\sum_{i=1}^{n}k_i\mathbf{e}_i^T+\mathbf{e}_j^T)}.$$

Example 3 If $\mathcal{G}_0(\mathbf{k})=\{1,2,...,n\}$, and $\mathcal{G}_{>0}(\mathbf{k})=\mathcal{G}_x(\mathbf{k})=\{\}, \mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})=\{\}$then $|\mathcal{G}_{>0}(\mathbf{k})|=0$ and $|\mathcal{G}_0(\mathbf{k})|=n$. Thus (2.62) is simplified to

$$\left(\sum_{j=1}^{n}\lambda_j\right)\pi_{(\sum_{i=1}^{n}k_i\mathbf{e}_i^T)}=(1+\gamma_{n-1})\sum_{j=1}^{n}\mu_j\pi_{(\mathbf{e}_j^T)}.$$

Going back to the proof, consider the equation (2.62).In order to find a solution for it, we present it the in following way:

$$\sum_{j\in\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})\mu_j[\pi_{(\mathbf{k})}-\pi_{(\mathbf{k}+\mathbf{e}_j^T)}] \tag{2.68}$$

$$+\sum_{j\in\mathcal{G}_x(\mathbf{k})}[(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})\mu_j\pi_{(\mathbf{k})}-\lambda_j(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1})\pi_{(\mathbf{k}-x_j\mathbf{e}_j^T)}] \tag{2.69}$$

$$+\sum_{j\in\mathcal{G}_0(\mathbf{k})}[\lambda_j(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|})\pi_{(\mathbf{k})}-\mu_j(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|-1})\pi_{(\mathbf{k}+\mathbf{e}_j^T)}]=0. \tag{2.70}$$

This is a vector difference equation, where does not have any standard way to solve. In order to obtain the solution of this equation, we treat it like an identity and make each of the disjoint sums (2.68), 2.69), and (2.70) equal to zero, and obtain some results. Later we show that the obtained results, indeed, construct the equation's unique solution. From (2.68)-(2.70), we obtain respectively

$$\pi_{(\mathbf{k})} = \pi_{(\mathbf{k}+\mathbf{e}_j^T)}; \quad \forall j \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k}), \tag{2.71}$$

$$(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})\mu_j \pi_{(\mathbf{k})} = \lambda_j(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1})\pi_{(\mathbf{k}-x_j\mathbf{e}_j^T)}; \quad \forall j \in \mathcal{G}_x(\mathbf{k}), \tag{2.72}$$

$$\lambda_j(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|})\pi_{(\mathbf{k})} = \mu_j(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|-1})\pi_{(\mathbf{k}+\mathbf{e}_j^T)}; \quad \forall j \in \mathcal{G}_0(\mathbf{k}). \tag{2.73}$$

Now using the following two lemmas, we verify the solution.

Lemma 40 *Let* $\mathbf{k} = \sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T + \sum_{i \in \mathcal{G}_x(\mathbf{k})} x_i \mathbf{e}_i^T$ *be the vector of the number of clicks left in each slot. If the identities (2.68), (2.69), and (2.70) hold then the following relation always holds.*

$$\pi_{(\mathbf{k})} = \pi_{\left(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T + \sum_{i \in \mathcal{G}_x(\mathbf{k})} x_i \mathbf{e}_i^T\right)} = \frac{\prod_{i \in \mathcal{G}_x(\mathbf{k})} \lambda_i \prod_{i=|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|}^{|\mathcal{G}_{>0}(\mathbf{k})|-1} (1+\alpha_i)}{\prod_{i \in \mathcal{G}_x(\mathbf{k})} \mu_i \prod_{i=|\mathcal{G}_0(\mathbf{k})|}^{|\mathcal{G}_x(\mathbf{k})|+|\mathcal{G}_0(\mathbf{k})|-1} (1+\gamma_i)} \pi_{\left(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T\right)}. \tag{2.74}$$

Proof By (2.82) we have $(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})\mu_j \pi_{(\mathbf{k})} = \lambda_j(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1})\pi_{(\mathbf{k}-x_j\mathbf{e}_j^T)}$ for $\forall j \in \mathcal{G}_x(\mathbf{k})$. Hence we get

$$\pi_{\left(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T + \sum_{i \in \mathcal{G}_x(\mathbf{k})} x_i \mathbf{e}_i^T\right)} = \frac{\lambda_{j_1}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1})}{\mu_{j_1}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})} \pi_{\left(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T + \sum_{i \in \mathcal{G}_x(\mathbf{k}) \backslash \{j_1\}} x_i \mathbf{e}_i^T\right)}, \quad j_1 \in \mathcal{G}_x(\mathbf{k}).$$

Similarly, for any $j_2 \in \mathcal{G}_x(\mathbf{k})/\{j_1\} \subset \mathcal{G}_x(\mathbf{k})$ we have

$$\pi_{\left(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})} k_i\mathbf{e}_i^T + \sum_{i\in\mathcal{G}_x(\mathbf{k})\setminus\{j_1\}} x_i\mathbf{e}_i^T\right)} = \frac{\lambda_{j_2}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-2})}{\mu_{j_2}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|+1})}\pi_{\left(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})} k_i\mathbf{e}_i^T + \sum_{i\in\mathcal{G}_x(\mathbf{k})\setminus\{j_1,j_2\}} x_i\mathbf{e}_i^T\right)}.$$

Continuing with a similar way get

$$\begin{aligned}
&\pi_{\left(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})} k_i\mathbf{e}_i^T + \sum_{i\in\mathcal{G}_x(\mathbf{k})} x_i\mathbf{e}_i^T\right)} \\
&= \frac{\lambda_{j_1}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1})}{\mu_{j_1}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})}\frac{\lambda_{j_2}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1})}{\mu_{j_2}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})}\cdots\frac{\lambda_{j_{|\mathcal{G}_x(\mathbf{k})|}}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|})}{\mu_{j_{|\mathcal{G}_x(\mathbf{k})|}}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|+|\mathcal{G}_x(\mathbf{k})|-1})}\pi_{\left(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})} k_i\mathbf{e}_i^T\right)} \\
&= \frac{\prod_{j\in\mathcal{G}_x(\mathbf{k})}\lambda_j \prod_{j=|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|}^{|\mathcal{G}_{>0}(\mathbf{k})|-1}(1+\alpha_j)}{\prod_{j\in\mathcal{G}_x(\mathbf{k})}\mu_j \prod_{j=|\mathcal{G}_0(\mathbf{k})|}^{|\mathcal{G}_x(\mathbf{k})|+|\mathcal{G}_0(\mathbf{k})|-1}(1+\gamma_j)}\pi_{\left(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})} k_i\mathbf{e}_i^T\right)},
\end{aligned}$$

which completes the proof. ■

Lemma 41 *Let* $\mathbf{k} = \sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})} k_i\mathbf{e}_i^T$ *be the vector of the number of clicks left in each slot. If all the identities (2.68), (2.69), and (2.70) hold then:*

$$\pi_{\left(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})} \mathbf{e}_i^T\right)} = \frac{\prod_{i\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})}\lambda_i \prod_{i=1}^{|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|-1}(1+\alpha_i)}{\prod_{i\in\mathcal{G}_x(\mathbf{k})}\mu_i \prod_{i=|\mathcal{G}_0(\mathbf{k})|+|\mathcal{G}_x(\mathbf{k})|}^{n-1}(1+\gamma_i)}\pi_{(\mathbf{0})}. \tag{2.75}$$

Proof By (2.83) we have

$$\pi_{(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})}\mathbf{e}_i^T)} = \pi_{(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\}}\mathbf{e}_i^T+\mathbf{e}_{j_1}^T)} \tag{2.76}$$
$$= \frac{\lambda_{j_1}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\}|})}{\mu_{j_1}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})\cup\mathcal{G}_x(\mathbf{k})|})}\pi_{(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\}}\mathbf{e}_i^T)}, \quad j_1 \in \mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k}).$$

We note that $|\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\}| = |\mathcal{G}_{>0}(\mathbf{k})| - |\mathcal{G}_x(\mathbf{k})| - 1$, and $|\mathcal{G}_0(\mathbf{k})\cup\mathcal{G}_x(\mathbf{k})| = |\mathcal{G}_0(\mathbf{k})| + |\mathcal{G}_x(\mathbf{k})|$.

In addition, with a an analogous argument we get

$$\pi_{(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\}}\mathbf{e}_i^T)} = \pi_{(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\}\backslash\{j_2\}}\mathbf{e}_i^T+\mathbf{e}_{j_2}^T)} \tag{2.77}$$
$$= \frac{\lambda_{j_2}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\}\backslash\{j_2\}|})}{\mu_{j_2}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})\cup\mathcal{G}_x(\mathbf{k})|+1})}\pi_{(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\}\backslash\{j_2\}}\mathbf{e}_i^T)},$$
$$j_2 \in \mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\},$$

where $|\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1\}\backslash\{j_2\}| = |\mathcal{G}_{>0}(\mathbf{k})| - |\mathcal{G}_x(\mathbf{k})| - 2$. Combining the two results above gives

$$\pi_{(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})}\mathbf{e}_i^T)} =$$
$$\frac{\lambda_{j_1}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|-1})}{\mu_{j_1}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|+|\mathcal{G}_x(\mathbf{k})|})}\frac{\lambda_{j_2}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|-2})}{\mu_{j_2}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|+|\mathcal{G}_x(\mathbf{k})|+1})}\pi_{(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k})\backslash\{j_1,j_2\}}\mathbf{e}_i^T)}, \quad j_1, j_2 \in \mathcal{G}_{>0}(\mathbf{k})\backslash\mathcal{G}_x(\mathbf{k}).$$

Keeping on this way we get

$$\pi_{(\sum_{i\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})}\mathbf{e}_i^T)} = \frac{\lambda_{j_1}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|-1})}{\mu_{j_1}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|+|\mathcal{G}_x(\mathbf{k})|})}\frac{\lambda_{j_2}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|-2})}{\mu_{j_2}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|+|\mathcal{G}_x(\mathbf{k})|+1})}\cdots \frac{\lambda_{j_{(|\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})|)}}(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|-|\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})|})}{\mu_{j_{(|\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})|)}}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|+|\mathcal{G}_x(\mathbf{k})|+|\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})|-1})}\pi_{(\mathbf{0})}.$$

Furthermore, it is easily observed that

$$1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|+|\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})|} = 1+\alpha_0 = 1,$$

and

$$1+\gamma_{|\mathcal{G}_0(\mathbf{k})|+|\mathcal{G}_x(\mathbf{k})|+|\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})|-1} = 1+\gamma_{n-1}.$$

Hence the result follows and the lemma is proven. ∎

Going back to the proof, consider the equation (2.62).In order to find a solution for it, we present it in the following way:

$$\sum_{j\in\mathcal{G}_{>0}(\mathbf{k})\setminus\mathcal{G}_x(\mathbf{k})}(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})\mu_j[\pi_{(\mathbf{k})}-\pi_{(\mathbf{k}+\mathbf{e}_j^T)}] \tag{2.78}$$

$$+\sum_{j\in\mathcal{G}_x(\mathbf{k})}[(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})\mu_j\pi_{(\mathbf{k})}-\lambda_j(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1})\pi_{(\mathbf{k}-x_j\mathbf{e}_j^T)}] \tag{2.79}$$

$$+\sum_{j\in\mathcal{G}_0(\mathbf{k})}[\lambda_j(1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|})\pi_{(\mathbf{k})}-\mu_j(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|-1})\pi_{(\mathbf{k}+\mathbf{e}_j^T)}]=0. \tag{2.80}$$

This is a vector difference equation, where does not have any standard way to solve. In order to obtain the solution of this equation, we treat it like an identity and make each of the disjoint sums (2.78), 2.79), and (2.80) equal to zero, and obtain some results. Later we show that the obtained results, indeed, construct the equation's unique solution. From (2.78)-(2.80) we obtain respectively

$$\pi_{(\mathbf{k})} = \pi_{(\mathbf{k}+\mathbf{e}_j^T)}; \quad \forall j \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k}), \tag{2.81}$$

$$(1+\gamma_{|\mathcal{G}_0(\mathbf{k})|})\mu_j \pi_{(\mathbf{k})} = \lambda_j (1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|-1})\pi_{(\mathbf{k}-x_j\mathbf{e}_j^T)}; \quad \forall j \in \mathcal{G}_x(\mathbf{k}), \tag{2.82}$$

$$\lambda_j (1+\alpha_{|\mathcal{G}_{>0}(\mathbf{k})|})\pi_{(\mathbf{k})} = \mu_j (1+\gamma_{|\mathcal{G}_0(\mathbf{k})|-1})\pi_{(\mathbf{k}+\mathbf{e}_j^T)}; \quad \forall j \in \mathcal{G}_0(\mathbf{k}). \tag{2.83}$$

By Lemma 40, we showed

$$\pi_{\mathbf{k}} = \pi_{(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T + \sum_{i \in \mathcal{G}_x(\mathbf{k})} x_i \mathbf{e}_i^T)} = \frac{\prod_{i \in \mathcal{G}_x(\mathbf{k})} \lambda_i \prod_{i=|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|}^{|\mathcal{G}_{>0}(\mathbf{k})|-1} (1+\alpha_i)}{\prod_{i \in \mathcal{G}_x(\mathbf{k})} \mu_i \prod_{i=|\mathcal{G}_0(\mathbf{k})|}^{|\mathcal{G}_x(\mathbf{k})|+|\mathcal{G}_0(\mathbf{k})|-1} (1+\gamma_i)} \pi_{(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T)}.$$

However, from (2.81), we have

$$\pi_{(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T)} = \pi_{(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} \mathbf{e}_i^T)}. \tag{2.84}$$

Hence we get $\pi_{\mathbf{k}}$ as

$$\pi_{\mathbf{k}} = \pi_{(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} k_i \mathbf{e}_i^T + \sum_{i \in \mathcal{G}_x(\mathbf{k})} x_i \mathbf{e}_i^T)} = \frac{\prod_{i \in \mathcal{G}_x(\mathbf{k})} \lambda_i \prod_{i=|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|}^{|\mathcal{G}_{>0}(\mathbf{k})|-1} (1+\alpha_i)}{\prod_{i \in \mathcal{G}_x(\mathbf{k})} \mu_i \prod_{i=|\mathcal{G}_0(\mathbf{k})|}^{|\mathcal{G}_x(\mathbf{k})|+|\mathcal{G}_0(\mathbf{k})|-1} (1+\gamma_i)} \pi_{(\sum_{i \in \mathcal{G}_{>0}(\mathbf{k}) \backslash \mathcal{G}_x(\mathbf{k})} \mathbf{e}_i^T)}. \tag{2.85}$$

Using Lemma 41, we get

$$\pi_{\mathbf{k}} = \left(\frac{\prod\limits_{i \in \mathcal{G}_x(\mathbf{k})} \lambda_i \prod\limits_{i=|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|}^{|\mathcal{G}_{>0}(\mathbf{k})|-1} (1+\alpha_i)}{\prod\limits_{i \in \mathcal{G}_x(\mathbf{k})} \mu_i \prod\limits_{i=|\mathcal{G}_0(\mathbf{k})|}^{|\mathcal{G}_x(\mathbf{k})|+|\mathcal{G}_0(\mathbf{k})|-1} (1+\gamma_i)} \right) \left(\frac{\prod\limits_{i \in \mathcal{G}_{>0}(\mathbf{k}) \setminus \mathcal{G}_x(\mathbf{k})} \lambda_i \prod\limits_{i=1}^{|\mathcal{G}_{>0}(\mathbf{k})|-|\mathcal{G}_x(\mathbf{k})|-1} (1+\alpha_i)}{\prod\limits_{i \in \mathcal{G}_x(\mathbf{k})} \mu_i \prod\limits_{i=|\mathcal{G}_0(\mathbf{k})|+|\mathcal{G}_x(\mathbf{k})|}^{n-1} (1+\gamma_i)} \right) \pi_{(\mathbf{0})}, \tag{2.86}$$

where after some manipulation becomes

$$\pi_{\mathbf{k}} = \frac{\prod\limits_{i \in \mathcal{G}_{>0}(\mathbf{k})} \lambda_i \prod\limits_{i=1}^{|\mathcal{G}_{>0}(\mathbf{k})|-1} (1+\alpha_i)}{\prod\limits_{i \in \mathcal{G}_{>0}(\mathbf{k})} \mu_i \prod\limits_{i=|\mathcal{G}_0(\mathbf{k})|}^{n-1} (1+\gamma_i)} \pi_{(\mathbf{0})}. \tag{2.87}$$

Hence $\pi_{\mathbf{k}}$ with the above relation is the solution of the system. Now if we take

$$\pi_{\mathbf{0}} = \eta \prod_{i=1}^{n} \mu_i \prod_{i=1}^{n-1} (1+\gamma_i), \tag{2.88}$$

then we get $\pi_{\mathbf{k}}$ as

$$\pi_{\mathbf{k}} = \eta \prod_{i \in \mathcal{G}_{>0}(\mathbf{k})} \lambda_i \prod_{i=1}^{|\mathcal{G}_{>0}(\mathbf{k})|-1} (1+\alpha_i) \prod_{i \in \mathcal{G}_0(\mathbf{k})} \mu_i \prod_{i=1}^{|\mathcal{G}_0(\mathbf{k})|-1} (1+\gamma_i). \tag{2.89}$$

In order to obtain η, knowing that $\sum\limits_{\mathbf{k}} \pi_{(\mathbf{k})} = 1$, we obtain

$$\eta \sum_{\mathbf{k}} \prod_{i \in \mathcal{G}_{>0}(\mathbf{k})} \lambda_i \prod_{i=1}^{|\mathcal{G}_{>0}(\mathbf{k})|-1} (1+\alpha_i) \prod_{i \in \mathcal{G}_0(\mathbf{k})} \mu_i \prod_{i=1}^{|\mathcal{G}_0(\mathbf{k})|-1} (1+\gamma_i) = 1. \tag{2.90}$$

Hence

$$\eta = \left(\sum_{k_1=0}^{x_1} \cdots \sum_{k_n=0}^{x_n} \prod_{i \in \mathcal{G}_{>0}(\mathbf{k})} \lambda_i \prod_{i=1}^{|\mathcal{G}_{>0}(\mathbf{k})|-1} (1+\alpha_i) \prod_{i \in \mathcal{G}_0(\mathbf{k})} \mu_i \prod_{i=1}^{|\mathcal{G}_0(\mathbf{k})|-1} (1+\gamma_i) \right)^{-1} . \tag{2.91}$$

Finally, the probability that there are t advertisers in the system becomes

$$\Pr(t) = \sum_{\substack{\mathbf{k} \\ |\mathcal{G}_{>0}(\mathbf{k})|=t}} \pi_{(\mathbf{k})} . \tag{2.92}$$

■

Proof of Proposition 39 Using the definition of $R(n_1; n)$ and after some simplification we reach to

$$R(n_1;n) - R(n_1-1;n) = \frac{p_1(r_1x_1)^{n_1}\mu_1(1-r_1x_1)^2}{(1-(r_1x_1)^{n_1})(1-(r_1x_1)^{n_1})} - \frac{p_2(r_2x_2)^{n-n_1}\mu_2(1-r_2x_2)^2}{(1-(r_2x_2)^{n-n_1})(1-(r_2x_2)^{n-n_1+1})} . \tag{2.93}$$

However, it is easy to see that

$$\frac{p_1(r_1x_1)^{n_1}\mu_1(1-r_1x_1)^2}{(1-(r_1x_1)^{n_1})(1-(r_1x_1)^{n_1})} = \lambda_1 x_1 p_1 \mathbb{P}^1_{n_1-1}(x_1, n_1-1)\mathbb{P}^1_0(x_1, n_1), \tag{2.94}$$

and

$$\frac{p_2(r_2x_2)^{n-n_1}\mu_2(1-r_2x_2)^2}{(1-(r_2x_2)^{n-n_1})(1-(r_2x_2)^{n-n_1+1})} = \lambda_2 x_2 p_2 \mathbb{P}^2_{n_2^*-1}(x_2, n-n_1-1)\mathbb{P}^2_0(x_2, n-n_1). \tag{2.95}$$

Hence from $R(n_1; n) - R(n_1 - 1; n) \geq 0$ and the first condition follows. The second condition is obtained with a similar procedure and the result follows. ■

Chapter 3

Competition between Publishers

3.1 Introduction

The Internet has revolutionized the way companies do business. It has created opportunities for new businesses and made business processes more efficient. Many companies are taking advantage of the Internet to reach out to more customers and are allocating increasing portions of their marketing budgets towards online advertising. Display advertising is a growing business with about $25 billion revenue in 2010. This revenue is expected to rise with a promising rate over the coming years (McAfee et al. 2010).

Web publishers are often affected by intense competition that they need to consider when planning their advertising operation. For instance, just recently Google, the online search ad giant, has announced that it has decided to enter the online display advertising business competition. As a result, the company has started a vigorous advertising campaign, called Watch This Space, for its display advertising platform (New York Times 2010).

In this competitive environment one of the most challenging issues faced by web publishers is how the publishers' (display) pricing policies are affected when the publishers are competing with other firms. Despite the vast amount of literature on the subject of online advertising and in particular display advertising little work has addressed competition among web publishers and the strategic pricing implications for them. In this chapter, we fill this gap by providing a stylized extension of our basic model in the first chapter incorporating competitive settings.

Currently, available models do not provide a formal method for addressing problems where web publishers determine the CPM prices for the websites affected by the intense competition. In this chapter, we explore the interactions of two web publishers in a competitive setting and provide various interesting insights about their strategic behavior at equilibrium. First, we focus on steady-state equilibriums (SSE), which tend to be significantly more general than equilibriums obtained merely in a one-stage game. The reason for this is that by considering SSE, we study the strategic behavior of the publishers in the limit when the game is played many times. As a result, the players learn from the past and become more sophisticated decision makers.

In addition, as the SSE game is very difficult for analytical tractability, we consider an alternative stylized model similar to SSE game. This is called infinitely repeated competition game of incomplete information on side between two publishers. The incomplete information feature, where one publisher enjoys private information, is not discussed in the SSE competition as it is intractable analytically. By private information, we mean that the first publisher knows about the real advertisers' arrival rates into the competition setting. However, this information is private only to the first publisher. The second publisher is unaware of this critical information and decides on its prices based on its observations from the two publishers' past price decisions. Note that the repeated incomplete information competition still tends to be too challenging to analyze analytically. As a result, we limit our focus only on zero-sum competitions, where the publishers' payoffs always sum to zero at each interaction stage. That is, a revenue gain for a publisher at one stage is the other one's lost opportunity cost.

The primary contributions of this chapter are:

1. We construct a modeling framework capturing the main trade-offs in the operations of two web publishers interacting with each other in a competitive environment. We consider the steady-state equilibrium competition as well as the repeated competition game of incomplete information on one side. Both of the competition models discussed in this chapter appear to be new compared to the work in the literature.

2. In the steady-state equilibrium game between two publishers we demonstrate the following observations:

   - We observe that for two publishers competing in the same market, the optimal managerial policy is to choose a mixture of cooperation and competition rather than a pure competition.

   - We observe that making larger contracts (with more impressions) with advertisers, not only benefits the publisher but also its competitor. That is, if a publisher offers larger contracts the revenue of both publishers increase at equilibrium.

   - We observe that competing with a publisher that has more slots on its website, may be less profitable for the competitor. That is, an increase in the number of slots in a publisher's system will lead its competitor's revenue to decrease. However, we note that the increase in the number of slots has a non-obvious impact on the publisher's own revenue, which depends on the current number of slots on the website and the number of slots added. For instance, if many slots are added both publishers' revenues can decrease. We observe a similar behavior with respect to the web traffic.

- An increase in a publisher's marginal cost will cause both publishers' prices to increase at equilibrium. However, the publisher whose cost has increased loses revenues, while the other achieves more.

3. In the repeated competition game of incomplete information on one side with show that the publisher possessing private information can always guarantee a higher payoff for itself by misleading the other publisher through strategic price manipulations during the history of the game. In addition, we show that the competition always has an equilibrium. That is, the game always has a *value*.

The remainder of this chapter is organized as follows: The next section provides the relevant literature. Section 3.3 describes the model formulation and the results for the SSE competition games. Section 3.4 discusses the model formulation and the results for zero-sum repeated competition games of incomplete information on one side and Section 3.5 concludes.

3.2 Literature Review

There are two streams of literature related to our research. The first is the literature on online advertising within the *marketing* area, which is quite extensive. Novak and Hoffman (2000) provide an overview of advertising pricing schemes for the internet. However, there is limited literature on analytical models for optimal pricing and other decision making for a web publisher with an advertising operation. (For issues faced by advertisers such as predicting audience for advertising campaigns see, e.g., Danaher (2007) and papers referenced therein.)

The second stream of literature is on *management science*. The online advertising research within this area is limited and there are few works directly related to online advertising pricing.

In some of the earlier work, Mangàni (2003) compares the expected revenues from the CPC and the CPM schemes using a simple deterministic model. At the same time, Chickering and Heckerman (2003) develop a delivery system that maximizes the CTR given inventory-management constraints in the form of advertisement quotas. Both of these papers assume the prices are fixed. Unlike our work, none of these works consider competition among web publishers.

There has been some recent literature on online search, the other section of the online advertising market. Johnson et al. (2004) consider an empirical study to examine the dynamics of online search behavior. In addition, Ghose and Yang (2009) provide an empirical analysis of search engine advertising for the sponsored searches on the internet. None of the results in these two papers can be extended to ours, as they do not develop analytical models for the price decisions to be applicable in a competition setting. Moreover, the nature of search advertising is fundamentally different from display advertising, as it is mainly based on using auctions that we do not consider here.

Some researchers have focused on the relevant problem of pricing of goods and services on the internet. Brynjolfsson and Smith (2000) and Clemons et al. (2002) conduct empirical evaluations of price dispersions and price differentiations on the internet. Bakos and Brynjolfsson (1999, 2000) study the optimal strategies of the products bundling for a retailer selling products through the internet. Dewan et al. (2000) and (2003) examine the problem of optimal product customization and price strategy both in monopoly and in competition. Jain and Kannan (2002), and Sandararajan (2004) analyzed the optimal pricing of information goods from economics and a game theoretic standpoint. Although all of these papers consider a variety of online pricing problems, none are applicable to our work, as the settings in these papers are for quite different problems than websites competing together.

Some authors have considered the problem of a web publisher who not only generates revenues from advertising but also from subscriptions. Baye and Morgan (2000) develop a simple economic model of online advertising and subscription fees. Prasad et al. (2003) model two offerings to viewers of a website: a lower fee with more ads and a higher fee with fewer ads. Kumar and Sethi (2008) study the problem of dynamically determining the subscription fee and the size of advertising space on a website. They use optimal control theory to solve the problem and obtained the optimal subscription fee and the optimal advertisement level over time. Unlike our thesis, all these papers are focused on capacity management problems not price decisions, and the price is assumed to be fixed. In addition, competition is not considered in any of these papers.

Scheduling of ads on a website has also recently become a popular topic. Kumar et al. (2008) develop a model that determines how ads on a website should be scheduled in a planning horizon to maximize revenue. They consider geometry and display frequency as the two most important factors specifying the ads. Their problem belongs to the class of NP-hard problems, and they develop a heuristic to solve it. They also provided a good overview of other related papers on scheduling.

The game setting developed in this chapter to characterize the competing of web publishers is new. There are few papers in the literature that consider competition in online advertising. In fact, we could not find any paper that explicitly studies this problem. However, we find more papers focusing on the issue of traditional advertising. Erickson (1985) uses a dynamic model of advertising rivalry between competitors in a duopoly and obtains analytical results for the case of pure market share rivalry in a mature market. In addition, Erickson (1995) uses a dynamic model of oligopolistic advertising competition, in which competitors are assumed to make a series of single-period advertising decisions with salvage values attached to achieved sales in each period. Espinosa and Mariel (1998) develop a dynamic model of oligopolistic advertising competition. Their model examines predatory advertising and informative advertising as particular cases. Using a differential game framework and comparing the open-loop and feedback equilibria to the efficient outcome, they find that for the informative advertising competition game, advertising levels are closer to the collusive outcomes in a feedback equilibrium. In the case of predatory advertising, expenditures are inefficiently high in a feedback equilibrium and the open-loop solution is more efficient. None of these papers considers the impact of competition on prices as decision variables.

Finally, as this chapter is still related to pricing and revenue management, we end this section by a short review of related work in revenue management. For a comprehensive reference of the traditional revenue management models, we refer the reader to the book by Talluri and van Ryzin (2004) (the book does not cover the online setting). Savin et al. (2005) consider revenue management for rental businesses with two customer classes. Although considering a different problem, they have assumed uncertainty in the customers demand to their model, which has some similarity to our model. Araman and Popescu (2009) also study revenue management for traditional media, specifically broadcasting. Their model is concerned with how to allocate limited advertising space between up-front contracts and the so-called scatter market (i.e., a spot market) in order to maximize profits and meet contractual commitments. Unlike our thesis, both of these papers are concerned with the capacity decisions and price is not an issue of focus.

In the next section, we discuss about the main model.

3.3 The Steady-State Competition

Web publishers are often affected by intense competition that they need to consider when planning their advertising operation. In this section, we extend our basic model to incorporate competitive settings. For illustration purposes, we focus on the case with two publishers (a duopoly) where each has one type of slots, i.e., each website can be considered as a single webpage, which contains a single subsystem as considered in Chapter 1. Advertisers wanting to post their ads arrive with rate λ and consider both publishers. We model their choice of a publisher based on a Binomial-logit (BNL) model, which is widely used in the management science literature (see, e.g., Talluri and van Ryzin (2004a, 2004b) and Vulcano and van Ryzin (2010)).[1]

[1] Another possibility would be to consider the inverse demand functions, instead of choice models, to characterize the competitive setting (see, e.g., De Miguel and Adida 2010 or Goyal and Netessine

Let p_1 and p_2 denote the price per impression announced by Publishers 1 and 2, respectively. The two websites are differentiated by a set of attributes, e.g., the price, the number of slots, and the number of impressions offered. The advertisers' preference for these attributes could have them select a slot from the more expensive website[2].

The advertisers' preference for the attributes is modeled by the coefficient vector $M^T = (1/m_1, 1/m_2, 1/m_3) > 0$ (see Anderson et al. (1992)), where each component indicates the preference weight that advertisers give to each attribute. The probability that an advertiser will choose Publisher i can be expressed as

$$\varpi_i(A_1^T, A_2^T) = \frac{\exp(-A_i^T M)}{\exp(-A_1^T M) + \exp(-A_2^T M)}, \tag{3.1}$$

where $A_i^T = (p_i, n_i, x_i)$ is the attributes' vector with $A_i^T(s),\ s = 1, 2, 3$, referring to the price, number of slots, and number of impressions.

When m_s tends to zero for a certain attribute $A_i^T(s),\ s \in \{1, 2, 3\}$, the choice probability in Equation (3.1) depends only on attribute $A_i^T(s)$. Alternatively, when m_s is very high, the advertisers become quite insensitive to attribute $A_i^T(s)$. In the same way, if for *all* attributes, m_s tends to infinity, the advertisers become indifferent towards the websites' features and choose either website with an equal probability.

Advertisers arrive with rate λ to consider both publishers. As soon as advertisers have determined which publisher to approach based on the choice model described in Equation (3.1), they check that publisher's availability. If Publisher i's website is

2007). Nevertheless, our initial exploration indicates that both approaches tend to yield equivalent insights. In this section, we restrict our focus on the choice models, as used by Anderson (1992), and leave the other approach for future research.

[2] Note that this setting is better suited for publishers serving advertisers that approach them directly and are not willing to wait for display.

fully occupied the arriving advertisers are rejected. We assume that each rejected advertiser then chooses to approach Publisher j with probability α_{ij} or leaves with probability $1-\alpha_{ij}$. Let c_i be the marginal cost and Θ_i be the fixed cost for Publisher i. Then Publisher i's equilibrium profit rate function is given by

$$\max_{p_i} \Pi_i = \overline{\lambda}^i (1 - \mathbb{P}^i_{n_i}(\overline{\lambda}^i))(p_i - c_i) - \Theta_i, \tag{3.2}$$

where $\overline{\lambda}^i$ is the advertisers' effective rate at Publisher i's website.

When deriving $\overline{\lambda}^i$, we note that there are two streams of advertisers that approach Publisher i. The first stream consists of advertisers who have initially selected Publisher i, which we denote by $W^1_{ii} := \lambda\varpi_i(A^T_1, A^T_2)$. Out of those $S^1_{ii} := W^1_{ii}(1-\mathbb{P}^i_{n_i}(\overline{\lambda}^i))$ have their ads displayed on Publisher i's website, while the rest $B^1_{ii} := W^1_{ii}\mathbb{P}^i_{n_i}(\overline{\lambda}^i)$ are rejected. Then, $W^1_{ij} := \alpha_{ij}B^1_{ii}$ of the rejected advertisers decide to approach Publisher j, and $B^1_{ij} = W^1_{ij}\mathbb{P}^j_{n_j}(\overline{\lambda}^j)$ of those are again rejected by Publisher j. Therefore, $W^2_{ii} = \alpha_{ji}B^1_{ij}$ of them reconsider Publisher i, while the rest leave. In short, W^2_{ii} is the fraction of the W^1_{ii} advertisers who had initially selected Publisher i, but have undergone a complete rejection loop and have arrived at Publisher i's website for the second time. Note that theoretically the same loop of procedures can be repeated infinitely. Generally, in loop κ, we have

$$\begin{aligned} S^\kappa_{ii} &= W^\kappa_{ii}(1 - \mathbb{P}^i_{n_i}(\overline{\lambda}^i)), \\ W^\kappa_{ii} &= \lambda\varpi_i(A^T_1, A^T_2)\left(\alpha_{ij}\alpha_{ji}\mathbb{P}^i_{n_i}(\overline{\lambda}^i)\mathbb{P}^j_{n_j}(\overline{\lambda}^j)\right)^{\kappa-1}, \ \kappa = 1, 2, ... \end{aligned} \tag{3.3}$$

where S^κ_{ii} is the fraction of the $\lambda\varpi_i(A^T_1, A^T_2)$ advertisers who had selected Publisher i in the first loop and are eventually admitted into Publisher i's system in loop κ.

As a result, the overall number of advertisers in the first stream can be obtained as

$$\begin{aligned} S_{ii} &= \sum_{\kappa=1}^{\infty} S_{ii}^{\kappa} = \lambda\varpi_i(A_1^T, A_2^T)(1-\mathbb{P}_{n_i}^i(\overline{\lambda}^i)) \sum_{\kappa=1}^{\infty} \left(\alpha_{ij}\alpha_{ji}\mathbb{P}_{n_i}^i(\overline{\lambda}^i)\mathbb{P}_{n_j}^j(\overline{\lambda}^j)\right)^{\kappa-1} \\ &= \frac{\lambda\varpi_i(A_1^T, A_2^T)(1-\mathbb{P}_{n_i}^i(\overline{\lambda}^i))}{1-\alpha_{ij}\alpha_{ji}\mathbb{P}_{n_i}^i(\overline{\lambda}^i)\mathbb{P}_{n_j}^j(\overline{\lambda}^j)}. \end{aligned} \tag{3.4}$$

The second stream of advertisers includes those who had initially chosen Publisher j but finally had to place their ads on Publisher i's website. Based on a similar argument it can be shown that

$$\begin{aligned} S_{ji}^{\kappa} &= W_{ji}^{\kappa}(1-\mathbb{P}_{n_i}^i(\overline{\lambda}^i)), \\ W_{ji}^{\kappa} &= \lambda\varpi_j(A_1^T, A_2^T)\mathbb{P}_{n_j}^j(\overline{\lambda}^j)\left(\alpha_{ij}\alpha_{ji}\mathbb{P}_{n_i}^i(\overline{\lambda}^i)\mathbb{P}_{n_j}^j(\overline{\lambda}^j)\right)^{\kappa-1}, \ \kappa = 1, 2, ... \end{aligned} \tag{3.5}$$

where S_{ji}^{κ} is the fraction of the $\lambda\varpi_j(A_1^T, A_2^T)$ advertisers who had first selected Publisher j, but made a contract with Publisher i in loop κ. Thus, the total number of advertisers in the second stream is

$$\begin{aligned} S_{ji} &= \sum_{\kappa=1}^{\infty} S_{ji}^{\kappa} = \lambda\varpi_j(A_1^T, A_2^T)\mathbb{P}_{n_j}^j(\overline{\lambda}^j)(1-\mathbb{P}_{n_j}^j(\overline{\lambda}^j)) \sum_{\kappa=1}^{\infty} \left(\alpha_{ij}\alpha_{ji}\mathbb{P}_{n_i}^i(\overline{\lambda}^i)\mathbb{P}_{n_j}^j(\overline{\lambda}^j)\right)^{\kappa-1} \\ &= \frac{\lambda\varpi_j(A_1^T, A_2^T)\mathbb{P}_{n_j}^j(\overline{\lambda}^j)(1-\mathbb{P}_{n_j}^j(\overline{\lambda}^j))}{1-\alpha_{ij}\alpha_{ji}\mathbb{P}_{n_i}^i(\overline{\lambda}^i)\mathbb{P}_{n_j}^j(\overline{\lambda}^j)}. \end{aligned} \tag{3.6}$$

Figure 3.1 illustrates the interaction of the publishers in loop κ. The following proposition summarizes this result.[3]

Proposition 42 *In a two publisher competitive setting, the effective advertisers'*

[3] Note that here we consider the *steady-state equilibrium* (SSE), which tends to be different from the equilibrium obtained through a one-stage game. Clearly, in the one shot game, the rejected advertisers cannot approach the alternative publisher as this would require the game to be repeated. Nevertheless, the SSE considers the behavior of the publishers in the limit when the game is played infinite times.

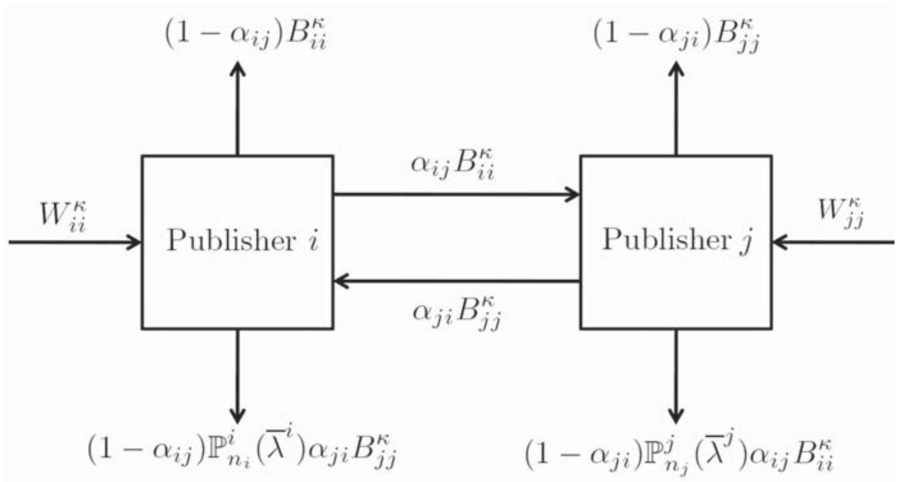


Figure 3.1: An illustrative presentation of the Publishers' interaction in loop κ. From the $W_{ij}^{\kappa}=\alpha_{ij}B_{ii}^{\kappa}$ advertisers approaching Publisher j, $(1-\alpha_{ji})\mathbb{P}_{n_j}^{j}(\overline{\lambda}^{j})\alpha_{ij}B_{ii}^{\kappa}$ leave the game while the rest $\alpha_{ji}\mathbb{P}_{n_j}^{j}(\overline{\lambda}^{j})\alpha_{ij}B_{ii}^{\kappa}$ consider approaching Publisher i as $W_{ii}^{\kappa+1}$.

arrival rate at Publisher i's website in equilibrium is $\overline{\lambda}^{i}(1-\mathbb{P}_{n_i}^{i}(\overline{\lambda}^{i}))$ *with* $\overline{\lambda}^{i}$ *given by*

$$\overline{\lambda}^{i}=S_{ii}+S_{ji}=\frac{\lambda(\varpi_i(A_1^T,A_2^T)+\varpi_j(A_1^T,A_2^T)\alpha_{ji}\mathbb{P}_{n_j}^{j}(\overline{\lambda}^{j}))}{1-\alpha_{ij}\alpha_{ji}\mathbb{P}_{n_i}^{i}(\overline{\lambda}^{i})\mathbb{P}_{n_j}^{j}(\overline{\lambda}^{j})}, \tag{3.7}$$

where λ *is the rate at which advertisers consider both publishers and* α_{ij} *(*α_{ji}*) is the fraction of the rejected advertisers by Publisher i (Publisher j) who approach Publisher j (Publisher i).*

We note that Publisher i's profit is a function of $\mathbb{P}_{n_i}^{i}(\overline{\lambda}^{i})$ and p_i. However, $\mathbb{P}_{n_i}^{i}(\overline{\lambda}^{i})$ is a function of $\overline{\lambda}^{i}$, which is again a function of $\mathbb{P}_{n_i}^{i}(\overline{\lambda}^{i})$ and p_i. Providing analytical results for these complex relationships is beyond the scope of this chapter. Instead, we provide a numerical analysis that reveals interesting insights about the strategic behavior of the publishers. Table 1 shows the results of the impact of several parameters on the optimal prices, the optimal profit rates, and other factors in the competition at equilibrium. We found the results to be consistent throughout various sets of parameter values that we considered. For illustration purposes, the choice

	Dependent variable									
Indep. var.	p_i	p_j	Π_i	Π_j	$\overline{\lambda}^i$	$\overline{\lambda}^j$	$\mathbb{P}^i_{n_i}$	$\mathbb{P}^j_{n_j}$	L_i	L_j
α_{ij}	$\nearrow$	$\nearrow$	$\nearrow$	$\nearrow$	$\nearrow$	$\searrow$	$\nearrow$	$\searrow$	$\nearrow$	$\searrow$
x_i	$\nearrow$	$\nearrow$	$\nearrow$	$\nearrow$	$\searrow$	$\nearrow$	$\nearrow$	$\nearrow$	$\nearrow$	$\nearrow$
n_i	$\searrow$	$\searrow$	$\curvearrowright$	$\searrow$	$\nearrow$	$\searrow$	$\searrow$	$\searrow$	$\searrow$	$\searrow$
μ^i	$\searrow$	$\searrow$	$\curvearrowright$	$\searrow$	$\searrow$	$\searrow$	$\searrow$	$\searrow$	$\searrow$	$\searrow$
c_i	$\nearrow$	$\nearrow$	$\searrow$	$\nearrow$	$\searrow$	$\nearrow$	$\searrow$	$\nearrow$	$\searrow$	$\nearrow$

Table 3.1: The effect of an increase in each of the indep. variables on the dep. variables

model considered has price as a single attribute.

From Table 3.1 it can be seen that an increase in the fraction α_{ij}, the proportion of advertisers rejected by Publisher i that approach Publisher j, leads to an increase in both publishers' profits. The reason for this behavior is that an increase in α_{ij} increases the advertisers' arrival rate at Publisher j's website. Publisher j responds to this demand change by increasing its price. The increase in Publisher j's price will affect the advertisers' choice decisions described by Equation (3.1) in favor of Publisher i. In order to respond to the consequent demand increase, Publisher i increases its price, which leads to more advertisers choosing Publisher j. This feedback cycle is repeated until the steady-state equilibrium is reached. It can be seen that at equilibrium[4], the prices as well as profits of both publishers increase. Furthermore, a counter-intuitive observation is that although Publisher i sends more advertisers to Publisher j's website, Publisher j's website becomes emptier at the limit. In addition, the profit increase for both publishers suggests that the cooperation of Publisher i with its competitor (e.g., by recommending Publisher j to its rejected advertisers) not only benefits Publisher j but also Publisher i itself.

In addition, Table 3.1 indicates that an increase in the number of impressions for Publisher i leads to an increase in both publishers' prices. This is because the increase

[4] Given that the equlibrium exists for the game.

makes Publisher i's system more congested. Naturally, Publisher i responds to this issue by increasing its website's price as observed in the monopolistic case. However, Publisher i's price increase will affect the advertisers' choice decisions in favor of Publisher j, which leads to an increase in its website's demand. As a consequence, Publisher j's response would be to increase its price. The eventual outcome of these interactions is that both the equilibrium prices and the profits increase. From this observation, it can be inferred that larger contracts (in terms of number of impressions) with advertisers, not only benefit the publisher offering them but also its competitor.

Furthermore, we observe that increasing the number of slots on Publisher i's website has an interesting impact on its profit. When Publisher i increases the number of slots, it decreases its price to attract more advertisers, which leads to fewer advertisers approaching Publisher j. To respond to its demand reduction, Publisher j lowers its price, which results in profit loss. However, we note that the increase in the number of slots has a non-obvious impact on Publisher i's own profit, which depends on the current number of slots on Publisher i's website, and the number of slots added. For instance, if many slots are added both publishers' profits can decrease. Generally, this observation suggests that competing with publishers having many ad slots can be less profitable for the competitors as more slots may not mean additional profit due to the subsequent price war between the publishers.

Similarly, we observe that an increase in Publisher i's number of viewers (i.e., web traffic) has a non-obvious impact on its profit. This is because an increase in the number of viewers for Publisher i, enables it to serve more advertisers. As a result, the number of rejected advertisers who consider approaching Publisher j would decrease. In response to the decreased demand, Publisher j reduces its price leading to more advertisers choosing its website. In addition, the decrease in price causes Publisher j's profits to decline. To respond to the decrease in its demand, Publisher i reduces its price. This interaction continues until the game reaches an equilibrium, where the publishers do not change prices anymore. We note that the impact of the increased traffic on Publisher i's own profit is non-obvious. Generally, the profit function seems to be concave with respect to the viewers' arrival rate. That is, depending on the value of other decision factors, a slight increase in Publisher i's website traffic may increase its profit. However, a dramatic increase can lead to a loss for both publishers. This observation suggests that in the competition setting more web traffic may not mean more profit for a publisher. In addition, a substantial traffic increase for one of the publishers may lead both publishers to lose profit as a consequence of a price war.

Finally, as can be seen from the table, an increase in Publisher i's marginal cost will cause an increase in the equilibrium prices for both publishers with the difference that Publisher i loses profit, while at the same time Publisher j gains more. The reason for this behavior is that in response to the increase in its cost, Publisher i raises its price to be able to maintain its profitability. However, this price increase causes more advertisers to choose Publisher j, leading it to increase its price and improve its overall profit.

3.4 Repeated Competition of Incomplete Information on One Side

Finding analytical results for the SSE game is quite difficult. As a result, in this section, we consider an alternative game setting that, while similar to the SSE competition, is stylized enough to provide us with analytical results. In our analysis, we limit our focus on two identical publishers and zero-sum competitions. In a zero-sum competition if one publisher makes a profit by selling impressions to an advertiser, the other publisher loses exactly the same amount of revenue as the lost opportunity cost. As in the previous section there are two publishers, Publisher 1 and Publisher 2. Nature chooses a state $\lambda \in \Lambda$ according to a commonly known probability on Λ, and the first publisher, but not the second publisher, is informed about the nature's choice. After each stage of the game, both publishers are informed of each others' actions. The game is repeated infinitely and the chosen state of λ remains constant throughout play. Although the chosen state λ, along with actions of players, determines the stage payoffs, during the play the second publisher learns nothing about its correct payoff and how far its pricing decision is from being optimal.

This can be represented by an infinite game tree where nature takes the first action of choosing λ and afterward never takes a second action. The information set of the second publisher is determined by the past behavior of both publishers and the information set of the first publisher is determined by the past behavior of both publishers and by choice of nature.

The Publishers Behavior Strategies In all states of nature, the first and second publisher have the same finite countable sets of CPM prices to choose from as their actions. The two price sets are denoted by $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively[5], and we assume that $|\mathcal{P}_1|$ and $|\mathcal{P}_2|$ are both at least two. A behavior strategy of Publisher 1 is an infinite sequence of $\mathbf{P}_1 = (\alpha^1, \alpha^2, ...)$ such that for each l, α^l is a mapping from $\Lambda \times (\mathcal{P}_1 \times \mathcal{P}_2)^{l-1}$ to $\Delta(\mathcal{P}_1)$, the publisher's mixed strategy space (i.e., the set

[5]Note that unlike the previous section the prices are not continuous but the price sets are assumed to be *countable* and *finite*. This assumption is necessary for tractability.

of all probability distributions on the set $\mathcal{P}_1$). A behavior strategy for Publisher 2 is an infinite sequence $\mathbf{P}_2 = (\beta^1, \beta^2, ...)$ such that for each l, β^l is a mapping from $(\mathcal{P}_1 \times \mathcal{P}_2)^{l-1}$ to $\Delta(\mathcal{P}_2)$, the set of all probability distributions on the set $\mathcal{P}_2$.

The Publishers Payoff Functions Define the set of finite play-histories of length l to be $\mathcal{H}_l := \Lambda \times (\mathcal{P}_1 \times \mathcal{P}_2)^l$, and define $\mathcal{H}_l^\lambda$ to be the subset $\{\lambda\} \times (\mathcal{P}_1 \times \mathcal{P}_2)^l$. For every $h \in \mathcal{H}_l$ with $h = (\lambda, p_1^1, p_2^1, ..., p_1^l, p_2^l)$ define

$$g_1(h) = \frac{1}{n}\sum_{k=1}^{n} \pi_1(p_1^k, p_2^k) \text{ and } g_2(h) = \frac{1}{n}\sum_{k=1}^{n} \pi_2(p_1^k, p_2^k),$$

where $\pi_i(p_1^k, p_2^k)$, $i = 1, 2$, is the profit (i.e., payoff) of Publisher i at stage k when the state of the nature is $\lambda \in \Lambda$ and Publisher 1 has selected $p_1^k \in \mathcal{P}_1$ and Publisher 2 has chosen $p_2^k \in \mathcal{P}_2$. As the game is zero-sum, Publisher i's profit is incurred on Publisher j as a penalty or lost opportunity cost. That is at each stage, $\pi_1(p_1^k, p_2^k) = \pi(p_1^k, p_2^k) = -\pi_2(p_1^k, p_2^k)$. Let $\mathbf{g}_n(h) = (g_1(h), g_2(h))$ be the vector of the average payoffs after n stages of interactions. That is,

$$\mathbf{g}_n = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{\pi}^{(k)}(p_1^k, p_2^k),$$

where $\boldsymbol{\pi}^{(k)} = (\pi_1, \pi_2)$. Next, we define the value of the game.

The Min-Max Function For every probability distribution $\mathbf{Q} = (q^1, ..., q^{|\Lambda|}) \in \Delta(\Lambda)$ we define the matrix $A(\mathbf{Q}) = \sum_{\lambda \in *} q^\lambda A^\lambda$, where A^λ is a $|\mathcal{P}_1| \times |\mathcal{P}_2|$ *payoff matrix* of Publisher 1 when the state of the nature is $\lambda \in *$. We define the function

$a^* : \Delta(\Lambda) \to \mathbf{R}$ by

$$a^*(\mathbf{Q}) := \max_{\sigma \in \Delta(\mathcal{P}_1)} \min_{\tau \in \Delta(\mathcal{P}_2)} \sigma A(\mathbf{Q})\tau = \min_{\tau \in \Delta(\mathcal{P}_2)} \max_{\sigma \in \Delta(\mathcal{P}_1)} \sigma A(\mathbf{Q})\tau$$

The function values $a^*(\mathbf{Q})$ represent the value of the game to Publisher 1 when both publishers believe that $\mathbf{Q}$ is the probability distribution on the state of advertisers' arrival rate into the competition setting. For the rest of analysis, we need to define cav and vex of a function.

Concave and Convex For any real valued function f on a convex space C let $cav(f)$ (respectively $vex(f)$) be the smallest concave function (largest convex function) larger or equal to (smaller or equal to) the function f. Given a real valued function f on C let G_f be the graph of f, namely $G_f := \{(x, y)|\, y = f(x)\}$. $cav(f)$ and $vex(f)$ can be obtained in the following way: Let $U_f := \{(x, y)|\, y \geq f(x)\}$ and $L_f := \{(x, y)|\, y \leq f(x)\}$. For any set S in the convex space let $Co(S)$ be the convex hull of S. $cav(f(x))$ can be defined with $Co(L_f)$ and $vex(f(x))$ by $Co(U_f)$, with $cav(f(x)) = \sup\{y|\, (x, y) \in Co(L_f)\}$ and $vex(f(x)) = \inf\{y|\, (x, y) \in Co(U_f)\}$. Furthermore, if C is a subset of a finite dimensional Euclidean space then for any $x \in C$ there will be a finite subset $S \subseteq C$ with some $\alpha \in \Delta(S)$ such that $x = \sum_{s \in S} \alpha_s s$ and $cav(f(x)) = \sum_{s \in S} \alpha_s f(s)$.

The next proposition ensures that in the two publishers game, for any probability distribution $\mathbf{Q} \in \Delta(\Lambda)$ Publisher 1 can guarantee gaining a payoff of $cav(a^*(\mathbf{Q}))$.

Proposition 43 *For any probability $\mathbf{Q} \in \Delta(\Lambda)$ in the two publishers competition Publisher 1 has a behavior strategy that guarantees a payoff of $cav(a^*(\mathbf{Q}))$ in the competition where $\mathbf{Q}$ is the probability on the total advertisers' arrival rates states* *determining nature's choice.*

Proof. As explained above for a point $x = \mathbf{Q}$ belonging to the set $\Delta(\Lambda)$ there exists a finite subset $S \subseteq \Delta(\Lambda)$ with some $\alpha \in \Delta(S)$ such that $\mathbf{Q} = \sum_{s\in S} \alpha_s s$ and $cav(a^*(\mathbf{Q})) = \sum_{s\in S} \alpha_s a^*(s)$. Note that each point s refers to a value that Publisher 1 reports to Publisher 2 in place of the true value of λ. At the start of each stage of the interactions Publisher 1 chooses a value $s \in S$ according to the probability distribution $\boldsymbol{\alpha} = (\alpha_s; s \in S)$. Once an s is chosen Publisher 1 reports that value (falsely) as the value of λ and plays according to strategy optimal in the game $A(s)$, which makes it gain $a^*(s)$ for that stage. In this way, using the probability distribution $\boldsymbol{\alpha}$ at the beginning of each stage, Publisher 1 guarantees itself $cav(a^*(\mathbf{Q})) = \sum_{s\in S} \alpha_s a^*(s)$. ■

For the rest of the analysis, we argue that $cav(a^*(\mathbf{Q}))$ is in fact the value of the repeated competition between the two publishers. In order to consider this issue, first we need to define the concept of *Approachability.*

Definition 44 *A set C is approachable for Publisher 1 with a behavior strategy σ if for all $\epsilon > 0$ there exists an N such that for all behavior strategies τ for Publisher 2 and $n \geq N$ it follows that*

$$\mathbb{E}_{\sigma,\tau}(d(C, g_n)) < \epsilon,$$

where d is the Euclidean distance.

A set C is excludable by Publisher 2 with a behavior strategy τ if there exists some $\delta > 0$ and an N such that for all choices σ of behavior strategies for Publisher 1 and $n \geq N$

$$\mathbb{E}_{\sigma,\tau}(d(C, g_n)) > \delta.$$

A set C is approachable for Publisher 1 if there exists a behavior strategy σ for Publisher 1 such that C is approachable for Publisher 1 with σ. For any probability distributions $\sigma \in \Delta(\mathcal{P}_1)$ and $\tau \in \Delta(\mathcal{P}_2)$ define

$$\begin{aligned} R_2(p_1) &: = Co(\{\sum_{p_1 \in \mathcal{P}_1} \sigma_{p_1} \pi(p_1, p_2) \mid p_2 \in \mathcal{P}_2\}),\ \forall p_2 \in \mathcal{P}_2, \\ R_1(p_2) &: = Co(\{\sum_{p_2 \in \mathcal{P}_2} \tau_{p_2} \pi(p_1, p_2) \mid p_1 \in \mathcal{P}_1\}),\ \forall p_1 \in \mathcal{P}_1. \end{aligned}$$

The next proposition states an important result about approachability, which is needed for the Publishers' competition.

Theorem 45 (Blackwell 1956) *Let C be a closed and convex subset of $\mathbf{R}^n$. C is approachable if and only if for all $\tau \in \Delta(\mathcal{P}_2)$ it holds that*

$$C \cap R_1(p_2) \neq \emptyset.$$

If C is not approachable then letting τ be any choice $\Delta(\mathcal{P}_2)$ with $C \cap R_1(p_2) = \emptyset$ the set C is excludable with a behavior strategy τ for all histories of the Publishers competition game.

Blackwell's theorem is quite useful in its application in the following proposition.

Proposition 46 *Let $\mathbf{y} \in \mathbf{R}^{|\Lambda|}$ be a (payoff) vector for Publisher 2 such that $\mathbf{y}$. $\mathbf{Q} \geq a^*(\mathbf{Q})$ for all $\mathbf{Q} \in \Delta(\Lambda)$. Then the subset $C_{\mathbf{y}} = \{\mathbf{v} \mid v^\lambda \leq y^\lambda, \forall \lambda \in \Lambda\}$ is approachable by Publisher 2.*

Proof. According to Blackwell's Theorem we need to show for any fixed mixed strategy $\sigma \in \Delta(\mathcal{P}_1)$ of the one-stage game for Publisher 1 that $C_{\mathbf{y}} \cap R_2(\sigma) \neq \emptyset$. This is equivalent to showing that there is a mixed strategy $\tau \in \Delta(\mathcal{P}_2)$ such that for all $\sigma A^{\lambda} \tau^t \leq y^{\lambda}$ for all $\lambda \in \Lambda$. ■

With $\sigma \in \Delta(\mathcal{P}_1)$ fixed, we can construct a new zero-sum *game* played between Publisher 2 and Nature. Nature chooses a probability $\mathbf{Q} \in \Delta(\Lambda)$ and Publisher 2 chooses a $\tau \in \Delta(\mathcal{P}_2)$. If Publisher 2 chooses a price $p_2 \in \mathcal{P}_2$ and Nature chooses the advertisers' arrival rate $\lambda \in \Lambda$ then the direct payoff for Publisher 2 is y^{λ}. However, as Publisher 2 should pay $\sigma A^{\lambda} \mathbf{e}_{p_2}^t$ to Nature, his overall payoff would be $y^{\lambda} - \sigma A^{\lambda} \mathbf{e}_{p_2}^t$,where the unit vector $\mathbf{e}_{p_2}^t$ implies that Publisher 2 selects the price p_2 with probability one, while the rest are selected with zero probability. By $\mathbf{y}.\mathbf{Q} \geq a^*(\mathbf{Q})$ for all $\mathbf{Q} \in \Delta(\Lambda)$ we know that for any choice $\mathbf{Q} \in \Delta(\Lambda)$ by Nature there is a price $p_2 \in \mathcal{P}_2$ such that $\mathbf{y}.\mathbf{Q} \geq \sigma A(\mathbf{Q}) \mathbf{e}_{p_2}^t$ which means that the min-max value of the game for Publisher 2 (between Publisher 2 and Nature) is at least zero. By the min-max theorem there must be a $\tau \in \Delta(\mathcal{P}_2)$ such that for all $\lambda \in \Lambda$, $y^{\lambda} - \sigma A^{\lambda} \tau^t \geq 0$, meaning that $\sigma A^{\lambda} \tau^t \leq y^{\lambda}$.

The next proposition summarizes the results obtained in this section.

Proposition 47 *A zero-sum infinitely-repeated and undiscounted competition of incomplete information on one side between two web publishers has a value and this value is* $cav(a^*(\mathbf{Q}))$*, where* $\mathbf{Q} \in \Delta(\Lambda)$ *is the distribution on the advertisers' arrival rate,* λ*, into the competition setting governing Nature's choice.*

Proof. By Proposition 43 the min-max value is at least $cav(a^*(\mathbf{Q}))$. By Proposition 46 the min-max value is no more than $cav(a^*(\mathbf{Q}))$. ■

3.5 Conclusion

In this chapter, we considered the interactions of two web publishers in a competitive setting and provided various interesting insights about their strategic behavior at equilibrium. We first studied the steady-state equilibrium (SSE) game. As mentioned earlier, the equilibriums obtained in steady-state games tend to be significantly more general than the equilibriums obtained merely in single-stage games. The reason for this is that by considering steady-state equilibriums, we study the strategic behavior of the publishers in the limit when the game is repeated several times. Therefore, the publishers learn from the past and become more sophisticated decision makers. One of the insights derived from the SSE game is that when two publishers compete in the same market, the optimal managerial policy for each of them is to choose a mixture of cooperation and competition rather than a pure competition. Furthermore, our observations indicated that making larger contracts with advertisers (with more impressions), not only benefits the publisher but also its competitor. That is, if a publisher offers larger contracts the revenue of both publishers increase at equilibrium.

Our examination of the SSE game also suggested that competing with a publisher that has more slots on its website, may be less profitable for its competitor. That is, an increase in the number of slots in a publisher's system will lead its competitor's revenue to decrease. However, we note that the increase in the number of slots has a non-obvious impact on the publisher's own revenue, which depends on the current number of slots on the website, and the number of slots added. For instance, if many slots are added both publishers' revenues can decrease. We observe a similar behavior with respect to the web traffic.

Although the SSE game is quite attractive to analyze, obtaining tractable results for it appears to be overly difficult. As a result, we considered an alternative similar game, namely, the zero-sum repeated competition of incomplete information on one side between the two publishers. This game tends to be more stylized than the SSE in some aspects. However, it is more general in some other. For instance, it deals with quite general payoff functions. In addition, it provides interesting tractable analytical results. We showed analytically that the publisher having private information about the market can always guarantee a higher payoff for itself by adopting a partially revealing strategy compared to be fully revealing or non-revealing. As a future research, the model introduced for the publishers' competition can be extended to zero-sum repeated competition of incomplete information on both sides, where the market is characterized by two pieces of information, each of which is private only to one of the publishers. One can also examine the none-zero-sum repeated competition of incomplete information on one (both) side(s).

Chapter 4

The Optimality Conditions for Continuous Demand Distributions With Independent Increments

4.1 Introduction

This section does not discuss about online advertising but is still related. The central focus of this chapter is finding the optimality conditions for the customers (in online advertising advertisers) demand process when the demand is not Poisson; rather follows any continuous distribution. Modeling the correct demand distribution of customers is of significant importance as the type of the demand process can considerably impact the suggested policies. A check of real demand processes shows that the customers' demand process can be frequently non-Poisson for a variety of reasons. For instance, one of the needed assumptions made in the Poisson process is that the average of the demand is equal to its variance. Nevertheless, in reality this assumption does rarely hold.

In this chapter, we consider the optimality condition applied by Gallego and van Ryzin (1994) to characterize Poisson demands with finite time horizon as well as the optimality condition introduced by Araman and Caldentey (2009) for Poisson demands with a stopping (or infinite) time horizon. We extend these two demand optimality conditions from Poisson to any arbitrary continuous distribution with mean λt and variance $\sigma^2 t$ at time t. We consider both finite and stopping time horizons for the demand. We show that in the both extensions an extra second order term appears in the optimality condition. This extra term explains the adjustment, which is needed when the demand is continuous.

4.2 Optimality Condition for a Finite Deterministic Time Horizon

Consider that the stochastic demand process X_t follows an arbitrary continuous distribution C with mean $\lambda_t(X_t, u_t, t)t$, and variance $\sigma_t^2(X_t, u_t, t)$, i.e.,

$$X_t \sim C(\lambda_t(X_t, u_t)t, \sigma_t(X_t, u_t)\sqrt{t}),$$

where u_t is the control variable[1]. Then, the total revenue gained through the demand from the initial time $t = 0$ to the terminal time $t = T$ is

$$J^*(x, T) = \sup_{u_t} \mathbb{E}_X[\int_0^T f_0(X_t, u_t)dX_t], \tag{4.1}$$

[1] For instance, at the simplest case where $\lambda_t(u_t) = u_t$ we aim to control the average of demand.

where $f_0(X_t, u_t)$ is the seller's revenue associated with the stochastic demand X_t, $0 \le t \le T$. Clearly, X_t can be re-expressed as $X_t = \lambda_t(X_t, u_t)t + \sigma_t(X_t, u_t)\widehat{X}_t$, where $\widehat{X}_t$ is a new process with the mean scaled to zero and variance t. Thus, based on Doob–Meyer's decomposition theorem (e.g., see Protter 2005), the process dX_t can be uniquely expressed as

$$dX_t = \lambda_t(X_t, u_t)dt + \sigma_t(X_t, u_t)d\widehat{X}_t, \tag{4.2}$$

where the increment process $d\widehat{X}_t$ has the mean 0 and variance dt. In order to show the reason that the variance of the increment process $d\widehat{X}_t$ is equal to dt, i.e., $var[d\widehat{X}_t] = dt$, we assume that the process $\widehat{X}_t$ has independent increments, i.e., $cov(d\widehat{X}_t, d\widehat{X}_s) = 0$ for any $t \neq s$. Next, we fix an arbitrarily fine partition $\{t_0, t_1, ..., t_n\}$ of the interval $[0, T]$, say into n subintervals all of equal length Δt with $\Delta t \to 0$, so that $t_i = t_0 + i\Delta t$. We shall call Δt the mesh of the subdivision. Furthermore, we note that $d\widehat{X}_t = \Delta\widehat{X}_t = \widehat{X}_{t+\Delta t} - \widehat{X}_t$ with $\Delta t \to 0$. Let us write ν the common value of the variance of the increment $\Delta\widehat{X}_t = \widehat{X}_{t+\Delta t} - \widehat{X}_t$. Thus, $\nu = var[\Delta\widehat{X}_t] = \mathbb{E}[(\Delta\widehat{X}_t)^2] - (\mathbb{E}[\Delta\widehat{X}_t])^2$ and hence since $\mathbb{E}[\Delta\widehat{X}_t] = 0$, we have for all t

$$\nu = \mathbb{E}[(\Delta\widehat{X}_t)^2].$$

In addition, we recall that that the formula for the variance of the sum of two random variables A and B when they are independent, namely

$$\begin{aligned} var(A+B) &= var(A) + var(B) + 2cov(A, B) \\ &= var(A) + var(B), \end{aligned}$$

which means that the variance is additive. We thus have by the additivity of variance over the independent increments that

$$\begin{aligned} var[(\widehat{X}_T - \widehat{X}_0)] &= \sum_{i=1}^{n} var[(\widehat{X}_{T-i\Delta t} - \widehat{X}_{T-(i+1)\Delta t})] \\ &= \sum_{i=0}^{n} var[\Delta \widehat{X}_{t_i}] = \sum_{i=0}^{n} \mathbb{E}[(\Delta \widehat{X}_{t_i})^2] \\ &= n\nu = vT/\Delta t. \end{aligned}$$

If in the limit as $\Delta t \to 0$, the limiting random variable $\widehat{X}_t$ is to have a finite variance at $t = T$ and $t = 0$, the limit of $vT/\Delta t$ must be also finite. This conclusion, which results from the independence of increments $\Delta \widehat{X}_{t_i}$ and the requirement of finite variance at each time, leads to the natural following standardization of the limiting process. A straightforward verification shows that if

$$\lim \frac{var(\Delta \widehat{X}_t)}{\Delta t} \to 1, \text{ as } \Delta t \to 0$$

then the process $\widehat{X}_t$ has finite variance under C and also $var(\widehat{X}_t) = t$ for any arbitrary value of t. As a result $var(d\widehat{X}_t) = dt$.

Next, we consider that a straightforward application of (4.2) to the optimal revenue function (4.1) gives

$$J^*(x,T) = \sup_{u_t} \mathbb{E}_X \begin{pmatrix} \int_0^T f_0(X_t, u_t)\lambda_t(X_t, u_t)dt \\ + \int_0^T f_0(X_t, u_t)\sigma_t(X_t, u_t)d\widehat{X}_t) \end{pmatrix}. \quad (4.3)$$

Note that in the above formula, since $d\widehat{X}_t$ is a martingale with $\mathbb{E}[d\widehat{X}_t] = 0$ and $f_0(X_t, u_t)\sigma_t(X_t, u_t)$ is bounded, the term $\int_0^T f_0(X_t, u_t)\sigma_t(X_t, u_t)d\widehat{X}_t)$ is a martingale transform of the process $\widehat{X}_t$ and hence is itself a martingale. Therefore its expected value is zero. That is,

$$\mathbb{E}_X[\int_0^T f_0(X_t,u_t)\sigma_t(X_t,u_t)d\widehat{X}_t)]=0. \tag{4.4}$$

Hence, the revenue function can be re-expressed as

$$J^*(x,T)=\sup_{u_t}\mathbb{E}_X[\int_0^T f_0(X_t,u_t)\lambda_t(X_t,u_t)dt], \tag{4.5}$$

where $x=X_0$. Our goal is now to derive the optimality conditions from (4.5). First of all, from (4.5), we establish that

$$J^*(x,T)=\sup_{u_t}\mathbb{E}_X\begin{pmatrix} \int_0^{\Delta t} f_0(X_t,u_t)\lambda_t(X_t,u_t)dt \\ +\int_{\Delta t}^T f_0(X_t,u_t)\lambda_t(X_t,u_t)dt \end{pmatrix}. \tag{4.6}$$

The essential observation is that, using the Mean Value Theorem, the expected value $\mathbb{E}_X[\int_0^{\Delta t} f_0(X_t,u_t,t)\lambda_t(X_t,u_t,t)dt]$ can be expressed as

$$\mathbb{E}_X[\int_0^{\Delta t} f_0(X_t,u_t)\lambda_t(X_t,u_t)dt]=f_0(x,u)\lambda(x,u)\Delta t+o_1(\Delta t), \tag{4.7}$$

where $u=u_s$ is a control function defined for $0\leq s\leq \Delta t$, and the last term $o_1(\Delta t)$ is a function of Δt that has the property that $o_1(\Delta t)/\Delta t\rightarrow 0$ as $\Delta t\rightarrow 0$. Now, the crucial use of the law of repeated expectations (Tower property) gives

$$J^*(x,T)=\sup_{u_t}\begin{pmatrix} f_0(x,u)\lambda(x,u)\Delta t+ \\ \mathbb{E}_X\mathbb{E}_{X,\Delta t}[\int_{\Delta t}^T f_0(X_t,u_t)\lambda_t(X_t,u_t)dt]+o_1(\Delta t) \end{pmatrix}. \tag{4.8}$$

In addition, an easy application of (4.7), and the change of variable $\theta_t=t-\Delta t$ to the second term $\mathbb{E}_{X,\Delta t}[\int_{\Delta t}^{T} f_0(X_t, u_t)\lambda_t(X_t, u_t)dt]$ gives

$$\begin{aligned}\mathbb{E}_{X,\Delta t}[\int_0^{T-\Delta t} & f_0(X_{\theta+\Delta t}, u_{\theta+\Delta t})\lambda_{\theta+\Delta t}(x, u_{\theta+\Delta t})d\theta_t] \\ &= J(X_{\Delta t}, T-\Delta t) = J(x-\Delta X, T-\Delta t). \qquad (4.9)\end{aligned}$$

The key observation in obtaining the second equality is noting that $X_{\Delta t} = X_0 - \Delta X = x - \Delta X$. Thus, the optimality condition reduces to

$$J^*(x,T) = \mathbb{E}_X[f_0(x,u^*)\lambda(x,u^*)\Delta t + J^*(x-\Delta X, T-\Delta t) + o_1(\Delta t)], \qquad (4.10)$$

where u^* belongs to the optimal control trajectory. Now, by applying the two dimensional Taylor's expansion to $J^*(x-\Delta X, T-\Delta t)$ and replacing in (4.10), we find

$$J^*(x,T) = \mathbb{E}_X \begin{pmatrix} f_0(x,u^*)\lambda(x,u^*)\Delta t + o_1(\Delta t) + J^*(x,T) - \\ J_T^{*\prime}\Delta t - J_x^{*'}\Delta X - \frac{1}{2}J_x^{*''}(\Delta X)^2 + o_2(\Delta t) \end{pmatrix}. \qquad (4.11)$$

Denoting $o_1(\Delta t) + o_2(\Delta t)$ by $o(\Delta t)$, we are now ready to invoke (4.2) in order to simplify (4.11) as

$$0 = \mathbb{E}_X \begin{pmatrix} f_0(x,u^*)\Delta t - J_T^{*\prime}\Delta t - J_x^{*'}\lambda(x,u^*)\Delta t - J_x^{*'}\Delta\widehat{X} \\ -\frac{1}{2}J_x^{*''}(\lambda(x,u^*)\Delta t + \sigma(x,u^*)\Delta\widehat{X})^2 + o(\Delta t) \end{pmatrix}. \qquad (4.12)$$

Note that an essential observation is that $\mathbb{E}[(\Delta\widehat{X})^2] = var[\Delta\widehat{X}] = \Delta t$, which reduces (4.12) to

$$0 = f_0(x,u^*)\Delta t - J_T^{*\prime}\Delta t - J_x^{*'}\lambda(x,u^*)\Delta t - \frac{1}{2}J_x^{*''}\sigma^2(x,u^*)\Delta t + o(\Delta t). \qquad (4.13)$$

Finally, dividing both sides of (4.13) by Δt and passing to the limit as $\Delta t \to 0^+$ gives optimality condition as follows:

$$J_T^{*\prime} = f_0(x, u^*) - J_x^{*'} \lambda(x, u^*) - \frac{1}{2} J_x^{*''} \sigma^2(x, u^*).$$

4.3 Optimality Condition for a Stopping Time

In this section, we examine an extended version of the optimality condition considered by Gallego and van Ryzin (1994), which was applied by Araman and Caldentey (2009) for characterizing Poisson demands when the time horizon is a stopping time[2]. In order to start, we consider the modified revenue function introduced by Araman and Caldentey (2009) as follows

$$\begin{aligned} J^*(x) &= \sup_{u_t, \theta} \mathbb{E}_{X,\tau}[\int_0^\tau e^{-rt} f_0(X_t, u_t) dt + e^{-r\tau} R]. \qquad (4.14) \\ dX_t &= \lambda_t(X_t, u_t) dt + \sigma_t(X_t, u_t) d\widehat{X}_t \\ \tau_\theta(x) &= \inf\{t \geq 0 : X_t = \theta\} \end{aligned}$$

In the above formula $x = X_0$ is the realization of the initial demand's value at time $t = 0$. In addition, $\tau = \tau_\theta(X_0) = \tau_\theta(x)$ is the stopping time by reference to the underlying stochastic demand process X_t reaching a prescribed level θ, which is to be chosen optimally. We suppose that $X_0 = x \neq \theta$. r is the discount factor, and R is the salvage value received by the seller at the stopping time $\tau_\theta(x)$. With an

[2]Note that considering an infinite time horizon follows the same lines of proof with the stopping time horizon and leads to the same optimality condition.

argument similar to the one used in the previous section, we find that

$$J^*(x) = \sup_{u_t,\theta} \mathbb{E}_{X,\tau} \begin{pmatrix} f_0(x,u)\Delta t + o_1(\Delta t) \\ + \int_{\Delta t}^{\tau} e^{-rt} f_0(X_t, u_t) dt + e^{-r\tau} R \end{pmatrix}. \tag{4.15}$$

To obtain (4.15), we used the Mean Value Theorem as stated in (4.7). Now, an easy change of variable $h = t - \Delta t$ gives

$$J^*(x) = \sup_{u_t,\theta} \begin{pmatrix} f_0(x,u)\Delta t + o_1(\Delta t) \\ +\mathbb{E}_{X,\tau}[\int_0^{\tau-\Delta t} e^{-(h+\Delta t)r} f_0(X_{h+\Delta t}, u_{h+\Delta t}) dh + e^{-r\tau} R] \end{pmatrix}. \tag{4.16}$$

Setting $\overline{\tau} \triangleq \tau - \Delta t$, $\overline{X}_h \triangleq X_{h+\Delta t}$, $\overline{u}_h \triangleq u_{h+\Delta t}$, and using the law of repeated expectations (Tower property), we find that

$$J^*(x) = \begin{pmatrix} f_0(x,u^*)\Delta t + o_1(\Delta t) \\ +e^{-r\Delta t} \mathbb{E}_{X,\tau} \mathbb{E}_{\overline{X},\overline{\tau}}[\int_0^{\overline{\tau}^*} e^{-hr} f_0(\overline{X}_h, \overline{u}_h) dh + e^{-r\tau^*} R] \end{pmatrix}, \tag{4.17}$$

where u^* is the optimal control trajectory and $\tau^* = \inf\{t \geq 0 : X_t = \theta^*\}$ is the stopping time when the stochastic demand process X_t reaches the prescribed optimal level θ^*. Furthermore, it is easy to observe that

$$J^*(X_{\Delta t}) = \mathbb{E}_{\overline{X},\overline{\tau}}[\int_0^{\overline{\tau}^*} e^{-hr} f_0(\overline{X}_h, \overline{u}_h) dh + e^{-r\tau^*} R].$$

Thus, replacing in (4.17) gives

$$J^*(x) = f_0(x,u^*)\Delta t + o_1(\Delta t) + e^{-r\Delta t} \mathbb{E}_{X,\tau}[J^*(X_{\Delta t})] \tag{4.18}$$

The essential observation in (4.18) is that

$$\begin{aligned} X_{\Delta t} &= X_0 - \Delta X = x - \Delta X, && (4.19) \\ \tau(X_{\Delta t}) &= \tau(x) - \Delta t = \overline{\tau}, && (4.20) \\ e^{-r\Delta t} &= 1 - r\Delta t + o_2(\Delta t). && (4.21) \end{aligned}$$

Applying (4.19)-(4.21) in (4.18), we find

$$J^*(x) = f_0(x, u^*)\Delta t + o_1(\Delta t) + (1 - r\Delta t + o_2(\Delta t))\mathbb{E}_{X,\tau}[J^*(x - \Delta X)].$$

A straightforward application of one dimensional Taylor's expansion and summarizing the sum of all error terms as the single term $o(\Delta t)$ gives

$$J^*(x) = \begin{pmatrix} f_0(x, u^*)\Delta t + o(\Delta t) \\ (1 - r\Delta t)\mathbb{E}_{X,\tau}[J^*(x) - J^{*\prime}(x)\Delta X + \frac{1}{2}J^{*''}(x)\left(\Delta X\right)^2] \end{pmatrix}.$$

Noticing that $\mathbb{E}[\Delta X] = \lambda(x, u)\Delta t$ and $var[\Delta X] = \sigma^2(x, u)\Delta t$, $J^*(x)$ reduces to

$$J^*(x) = \begin{pmatrix} J^*(x) + f_0(x, u^*)\Delta t - J^*(x)r\Delta t - \\ J^{*\prime}(x)\lambda(x, u)\Delta t + \frac{1}{2}J^{*''}(x)\sigma^2(x, u)\Delta t + o(\Delta t) \end{pmatrix}.$$

Finally, dividing both sides by Δt and passing to the limit as $\Delta t \to 0^+$ we obtain the optimality condition as follows

$$0 = f_0(x, u^*) - J^*(x)r - J^{*\prime}(x)\lambda(x, u) + \frac{1}{2}J^{*''}(x)\sigma^2(x, u).$$

4.4 Conclusion

In this chapter, we extended the optimality condition used by Gallego and van Ryzin (1994) to characterize Poisson demands with finite time horizon and also the optimality condition used by Araman and Caldentey (2009) for Poisson demands with a stopping (or infinite) time horizon to any arbitrary continuous distribution with mean λt and variance $\sigma^2 t$ at time t. As observed, in the both extensions, an extra second order term appears in the optimality condition, which is a function of the demand's variance. This extra term explains the "adjustment" needed when the demand process becomes continuous.

References

Adida E., V. DeMiguel. 2010. Supply Chain Competition with Multiple Manufacturers and Retailers. *Operations Res. Forthcoming.*

Anderson, S. P., A. de Palma, and J. F. Thisse. 1992. Discrete Choice Theory of Product Differentiation. MIT press, Cambridge MA.

Araman, V. F., R. Caldentey. 2009. Dynamic Pricing for Nonperishable Products with Demand Learning. *Operations Res.* **57** 1169-1188.

Araman, V. F., K. Fridgeirsdottir. 2010. Online Advertising: Revenue Management Approach. Working paper.

Araman, V. F., I. Popescu. 2009. Media Revenue Management with Audience Uncertainty: Balancing Upfront and Spot Market Sales. Forthcoming, M&SOM.

Bakos, Y., E. Brynjolfsson. 1999. Bundling Information Goods: Pricing, Profits, and Efficiency. *Management Sci.* **46** 563-585.

Bakos, Y., E. Brynjolfsson. 2000. Bundling and Competition on the Internet. *Marketing Sci.* **19** 63-82.

Brynjolfsson, E., M. D. Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Sci.* **46** 563-585.

Baye, M. R., J. Morgan. 2000. A simple model of advertising and subscription fees. *Economics Letters* **69** 345-351.

Break Media 2009. Learmonth M. 2009. Advertising Age – Kill or Cure: How Reckitt's Big Buy Rocked Online Ads,

http://breakmedia.break.com/file_download/220/AdAge.Reckitt.AndrewBudkofsky.9.7.09.pdf

Brill, P. H., L. Green 1984. Queues in Which Customers Receive Simultaneous Service from a Random Number of Servers: A System Point Approach. *Management Sci.* **30** 51-68.

Brynjolfsson, E., M. D. Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Sci.* **46** 563-585.

Business Week. 2009. Ad Networks Are Transforming Online Advertising. Available at http://www.businessweek.com/magazine/content/09_09/b4121048726676.htm.

Cao, J., W. S. Cleveland, D. Lin, D. X. Sun. 2002. Internet Traffic Tends Toward Poisson and Independent as the Load Increases. *Nonlinear Estimation and Classification*, eds. C. Holmes, D. Denison, M. Hansen, B. Yu, B. Mallick, Springer, New York.

Chao, X., M. Myuazawa, M. Pinedo. 1999. *Queueing Networks, Customers, Signals and Product Form Solutions*, *Wiley*.

Chickering, D. M., D. Heckerman. 2003. Targeted advertising on the web with inventory management interfaces. *Interfaces* **33** 71-77.

Goyal, M., S. Netessine. 2007. Strategic Technology Choice and Capacity Investment Under Demand Uncertainty. Management Sci. **53** 192-207.

Chickering, D. M., D. Heckerman. 2003. Targeted advertising on the web with inventory management interfaces. *Interfaces* **33** 71-77.

Clemons, E. K., I. Hann, L. M. Hitt 2002. Price Dispersion and Differentiation in Online Travel. *Management Sci.* **48** 534-549.

Courcoubetis, C., M. Reiman. 1987. Optimal control of a queueing system with simultaneous service requirements. *IEEE Transactions* **32** 717-727.

Danaher, P. J. 2007. Modeling Page Views Across Multiple Websites with an Application to Internet: Reach and Frequency Prediction. *Marketing Science* **26** 422-437.

Danaher, P. J., G. W. Mullarkey. 2003. Factors Affecting Online Advertising Recall: A Study of Students. Journal of Advertising Research. **43** 252-267.

Dewan, R., B. Jing, A. Seidmann. 2000. Adoption of Internet-Based Product Customization and Pricing Strategies. *Journal of Management Information Systems* **17.**

Dewan, R., M. Freimer, A. Seidmann. 2003. Organizing Distribution Channels for Information Goods on the Internet. *Management Sci.* **46** 483-495.

DCP (Double-Click for Publishers). 2010. Ad serving and management software devolped by Google. http://www.google.com/support/dfp_sb/bin/answer.py?answer=99601.

Erickson, G. M., 1985. A Model of Advertising Competition. *Journal of Marketing Research.* **22** 3

Erickson, G. M., 1995. Advertising Strategies in a Dynamic Oligopoly. *Journal of Marketing Research.* **32** 2

Espinosa M. P., P. Mariel. 2001. A model of optimal advertising expenditures in a dynamic duopoly. *Atlantic Economic Journal.* **29** 2

Evans D. S. 2008. The Economics of The Online Advertising Industry. *University of Chicago Working Paper.*

Feng, J., H. Bhargava, D. Pennock, David. 2008. Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms. *Informs Journal on Computing*, forthcoming.

Gallego, G., G. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management. Sci.* **40** 999-1020.

Gong, W., Y. Liu, V. Misra, D. Towsley. 2005. Self-similarity and long range dependence on the internet: a second look at the evidence, origins and implications. *Computer Networks* **48** 377-399.

Ghose, A., S. Yang. 2009. An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets, *Management Sci.* **55** 1605-1622.

Green, L. 1980. A Queueing System in Which Customers Require a Random Number of Servers. *Operations Res.* **28** 1335-1346.

Gross, D., C. M. Harris. 1998. *Fundamentals of Queueing Theory.* Wiley.

Ha, L. 2008. Online Advertising Research in Advertising Journals: A Review. *Journal of current issues and research in advertising.*

Harel, A. 1990. Convexity properties of the erlang loss formula. *Operations Res.* **38** 499-505.

Hong, M., T. Ott. 1989. On Waiting for Simultaneous Access to Two Resources. *Mathematics of Operations Res.* **14** 664-687.

IAB Internet Advertising Revenue Report. 2010. Interactive Advertising Bureau. Available at *www.iab.net.*

Jain, S., P. K. Kannan 2002. Pricing of Information Products on Online Servers. *Management Sci.* **48** 1123-1142.

Johnson, E. J., W. W. Moe, P. S. Fader, S. Bellman, G. L. Lohse. 2004. On the Depth and Dynamics of Online Search Behavior. *Management Sci.* **50** 299-308.

Kumar, S., V. S. Jacob, C. Sriskandarajah. 2008. Scheduling advertisements on a web page to maximize revenue. *European Journal of Operational Res.* **173** 1067-1089.

Kumar, S., S. P. Sethi. 2009. Dynamic pricing and advertising for web content providers. *European Journal of Operational Res.* ***197*** *924-944.*

Lohtia, R., N. Donthu, E. K. Hershberger. 2003. The Impact of Content and Design Elements on Banner Advertising Click-through Rates. Journal of Advertising Research. **43** 410-418.

McAfee, P. K. Papineni, S. Vassilvitskii. 2010. Maximally Representative Allocations for Guaranteed Delivery Advertising Campaigns. Working paper, Yahoo! Research.

Mangàni, A. 2003. Online advertising: Pay-per-view versus pay-per-click. *Journal of Revenue & Pricing Management* **2** 295-302.

Media Bankers. 2009. Company Report. Online Ad Networks: Monetizing the Long Tail. Available at http://www.mediabankers.com/.

Novak, T.P., D.L. Hoffman. 2000. Advertising and pricing models for the web. *In Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property*, D. Hurley, B. Kahin and H. Varian, eds. MIT Press, Cambridge.

New York Times. 2010. Google Campaign to Build Up Its Display Ads, September 21, 2010. Available at http://www.nytimes.com/2010/09/22/business/media/22adco.html?_r=1.

Prasad, A., V. Mahajan, B. Bronnenbert. 2003. Advertising versus pay-per-view in electronic media. *International Journal of Research in Marketing* **20** 13-30.

Protter, P (2005). Stochastic Integration and Differential Equations. *Springer-Verlag*. pp. 107–113.

A. Sundararajan. 2004. Nonlinear Pricing of Information Goods. *Management Sci.* **50** 1660-1673.

Savin, S. V., M. A. Cohen, N. Gans, Z. Katalan. 2005. Capacity Management in

Rental Businesses with Two Customer Bases. *Operations Res.* **53** 617-631.

Talluri, K., G. van Ryzin. 2004a. The Theory and Practice of Revenue Management. Kluwer Academic Press.

Talluri, K., G. van Ryzin. 2004b. Revenue Management Under a General Discrete Choice Model of Behavior. *Management Sci.* **50** 15-33.

Vulcano, G., G. van Ryzin. 2010. Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization. *M&SOM* **12** 371-392.