Cost-Per-Impression Pricing for Display Advertising

Kristin Fridgeirsdottir, Sami Najafi-Asadolahi

Cost-Per-Impression Pricing for Display Advertising

(Authors' names blinded for peer review)

Display advertising is a $25 billion business with a promising upward revenue trend. In this paper, we consider an online display advertising setting in which a web publisher posts display ads on its website and charges based on the cost-per-impression (CPM) pricing scheme while promising to deliver a certain number of impressions to the ads posted. The publisher is faced with uncertain demand for advertising slots and uncertain supply of visits from viewers. Advertisers specify various attributes of viewers who are most related to their ads, and request their ads to be displayed only to those viewer types (targeting). We formulate the problem as a novel queuing system, where the advertising slots correspond to service channels with the service rate of each server synchronized with other active servers. We determine the publisher's optimal price to charge per impression and show that it can increase in the number of impressions made to each ad, which is in contrast to the quantity-discount commonly offered in practice. Furthermore, we show that the optimal CPM price may increase in the number of ads rotating among slots. This result is typically not expected because an increase in the number of rotating ads in the system can be interpreted as an increase in the service capacity. However, the effective service rate to each ad depends negatively on the number of rotating ads, which leads to the opposite impact.

Key words: Queueing Systems; Online Advertising; Pricing; Markov Chains; Cost-per-impression

---

1. Introduction

Display advertising is currently a $25 billion business (Anandan 2012), which is expected to reach $200 billion in "a few short years," according to Google. As a result, Google is now investing heavily in display advertising (Peterson 2011) and not only focusing on its sponsored search advertising, where textual ads are displayed along with search results. In addition, display advertising is expected to continue to grow at a faster pace and overtake sponsored search advertising by 2015 (Fredricksen 2011). This paper focuses on a common online advertising setting in which web publishers post display ads on their websites for an upon-agreed number of impressions (Guaranteed Delivery), and charge based on the cost-per-impression (CPM[1]) pricing scheme (i.e., an advertiser pays a certain price for each impression made to his ad). The publishers are often faced with uncertain demand from advertisers requesting advertising space to post their ads and uncertain supply

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

of visits from viewers. That makes – even though, the publisher guarantees to serve an ad with a certain number of impressions – the completion of the service highly uncertain as it depends on how many viewers would visit the page where the ad is posted. In such an inherently uncertain environment, pricing is one of the most challenging operational decisions that web publishers face, and mostly ad-hoc approaches are currently used. It is now generally believed that the ability to determine the CPM price of display ads optimally in this highly uncertain environment is a key to the web publishers' revenue increase. However, optimal pricing of display ads has not received much attention in the literature, in contrast with pricing of sponsored search ads, which is quite well researched (see, e.g., Edelman et al. 2007 and references within).

In view of this gap, this paper has three main objectives. First, we develop a modeling framework that captures the fundamental operational challenges faced by web publishers posting ads on their websites and charging based on the CPM pricing scheme while promising to deliver a certain number of impressions on the ads posted. The publishers are faced with uncertain demand for advertising space through an advertising network (an online intermediary matching and sending advertisers to related websites) and uncertain traffic to their websites. Advertisers specify various attributes of viewers who are most related to their ads, and request their ads to be displayed only to those viewer types (targeting). Second, we use this model to determine the publishers' optimal price to charge per impression and investigate the impact of various factors, such as the number of advertising spaces and the number of promised impressions on its behavior. And third, we would like to derive improved understanding about the publishers' operational decisions when they face uncertainties from both the demand and supply sides and other real factors such as targeting.

Advertising networks Advertising networks are online companies that connect web publishers who want to sell their impressions or clicks (i.e., online inventory), with advertisers who want to run their ads on relevant websites. Large publishers often sell around 60% of their inventory through advertising networks and smaller ones often sell their entire inventory. We focus on publishers that receive their demand through ad networks. However, our model also applies to the setting where direct sales channels are used with advertisers not willing to wait for an ad space to become available. This scenario is very common when there is intense competition of web publishers attracting advertisers. The setting where a web publisher posts ads sent through an advertising network and charges based on the CPM pricing scheme captures about 25% of the display advertising market (IAB 2011).

We consider a common type of advertising network, known as a blind network. A blind network is sone where advertisers clearly define their desired *slot categories* for their ads in advance when registering with the ad network. (For instance, they may request a right hand side slot on a sport page.) But, they do not know the exact website that their ads will be posted on. Contextweb, Valueclick, and Clicksor are examples of blind ad networks. A more recent generation of blind ad networks are known as *targeted networks* in which advertisers (in addition to specifying the slot categories) can *target* specific audience segments (viewer types) based on their demographic, geographic and (cookie-based) behavioral attributes. Chitika is an example of a targeted ad network.

Furthermore, advertising networks often work with *immediate* (unfilled) inventories. That is, an advertiser's demand is sent to a web publisher only if it has a space available to post the ad in the advertiser's requested category. Otherwise, the ad network does not offer any ad space with this publisher and automatically directs the demand to other available publishers.

Note that since advertising networks often contain thousands of websites, it is rare that an advertiser's desired slot on a page, with his targeted the viewer types, is unavailable. However, even in such unusual cases, advertising networks do not keep advertisers waiting for the next available publisher. Instead, they direct advertisers to available publishers that participate in one of their partner networks.

Transaction steps The general steps for the transactions made between advertisers and publishers through an advertising network are the following.

(1) A web publisher has slots available and approaches the advertising network. The publisher registers each group of equivalent slots in terms of size, format, and page (the typical audience segments or viewer types that visit that page, e.g., sport, travel, etc.) as a separate "subsystem" with a different tracking code and a chosen price (per impression). The price that the publisher chooses for each subsystem is often called the subsystem's ask-price. Publishers are mostly free to determine their ask-prices. Nevertheless, some networks, such as Clicksor, have a more selective process. In these networks, publishers are often segmented into two main groups of *premium* and *non-premium* publishers. Premium publishers can freely choose their ask-prices, but the slot prices for non-premium publishers are set automatically by advertising networks. Advertising networks often do not reveal the price information to non-premium publishers, but they guarantee to pay no less than a pre-agreed minimum payment to the publisher for posting ads. In this paper, we restrict our attention to the common case that publishers can freely set their own ask-prices. (2) An advertiser requests his ad to be posted on a related participating website in the network. When registering his request, the advertiser clearly defines his target slot and viewers types (e.g., a leaderboard slot on top of a travel page, displayed to viewers from California with a potential interest in photography), as well as the maximum price he is willing to pay (bid-price). In addition, most

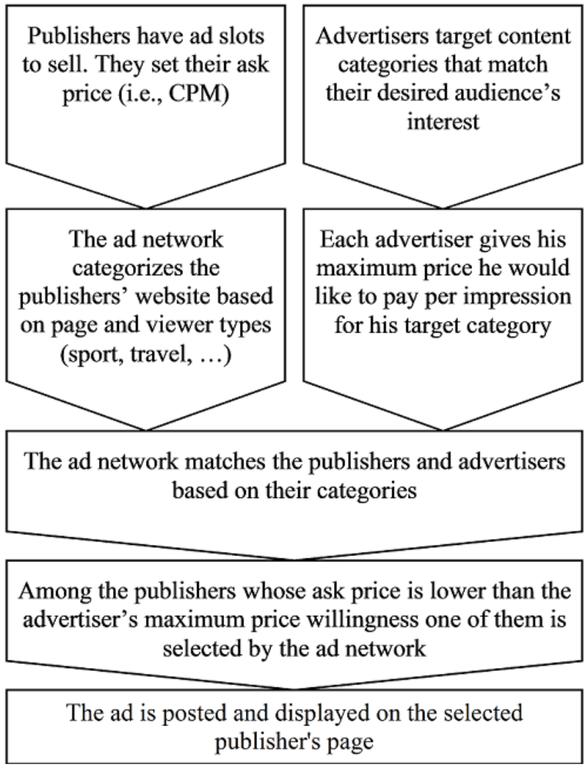

Figure 1 The general steps for transactions between advertisers and web publishers through an advertising network

advertisers choose mostly from either of the two following contracts. The first contract is known as Guaranteed Delivery (GD) in which the advertiser requests a certain number of impressions or clicks to be made on his ad. The second contract is known as Fixed-Advertising-Campaign-Length where the advertiser merely specifies a start and an end date and requests his ad to be posted for that fixed time frame (see, e.g., Akella et al. 2009). In this paper, we focus on the common case that the advertiser requests a certain number of impressions, and is charged based on the CPM pricing scheme. We refer to the information about an advertiser's desired slot category, and the viewer types he targets as his chosen "ad campaign".

We note that in practice, there are also some slot categories that no advertiser may be interested in. These categories are often referred to as *orphan categories* (Arpita et al. 2009). Since there is little demand for orphan categories (mostly due to an overly low probability that the posted ads are clicked on), some networks try to sell them using *auctions* at lower prices through *advertising exchanges* (Balseiro et al. 2014). Ad exchanges are platforms that facilitate the buying and selling of the orphan slot categories from multiple ad networks. In this paper, we focus on Guaranteed Delivery contracts only. (See, e.g., Balseiro et al. 2012, Muthukrishnan 2009, and Celis et al. 2012 for a survey of researches on auction pricing in advertising exchanges.)

(3) The advertising network, through some selection procedure, selects and sends the ad to one of the available subsystems that is matched to the advertiser's ad campaign and bid price. The selected publisher displays the ad to the targeted viewer groups specified by the advertiser's campaign.

(4) The advertiser pays the ask-price to the ad network. The ad network often takes about 25-50% of the payment as commission and transfers the rest to the publisher's account. Figure 1 summarizes the steps in an online transaction between publishers and advertisers through an ad network.

Web publishers seek the CPM prices that maximize their expected revenues based on the uncertain arrival of advertisers through the ad network and the uncertain supply of visits from viewers.

To capture the dynamics of the display advertising settings, with advertisers approaching the publisher at any time and viewers uploading the website at any time, we model the publisher's system as a queueing system in which advertisers act as customers who arrive at the system requesting to be served with certain number of impressions, viewers act as servers, and the slots act as serving channels. The resulting queueing system is new and despite complicated dynamics, a closed-form solution of the steady-state probability can be determined.

The main contributions in this paper 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), and targeted viewer types with different prices, and allow several ads to randomly share an advertising slot (random ad rotation). This model can serve as a building block for studying more complicated operational challenges faced by a web publisher such as competition. (See Sections 3 and 5.)

2. We derive a closed-form solution of the steady-state 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 3 and 4.)

3. We demonstrate that the optimal price can increase in the number of impressions that are offered. 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 has recently started to charge a higher price per impression for contracts delivering a large number of impressions[2].

[2] Confirmed by Prof. Preston McAfee, former VP and Research Fellow at Yahoo! (currently, a chief economist at Microsoft)

4. We provide further insights by showing that, in certain conditions, the optimal CPM price increases in the number of ads rotating among the slots. This result is typically not expected given our common intuition from the supply-demand relationship: an increase in the number of rotating ads in the system can be interpreted as an increase in the service capacity. However, the fact that the effective service rate to each ad depends on the number of rotating ads causes an opposite impact. (See Section 4.)

5. We provide an analysis based on a real data-set from a major Norwegian web publisher to support our assumptions and results, along with a simulation analysis. (See Section 5.)

The remainder of this paper is structured as follows. Section 2 reviews the related literature and Section 3 describes the model. Section 4 details the optimal CPM price for web publishers. Section 5 considers some extensions and Section 6 provides concluding remarks.

2. Literature Review

There are two streams of literature related to our research. The first is 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 in *operations research and 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 our paper, he does not consider the uncertainties involved with the advertisers' demands and viewers' supplies. 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. Fjell (2009) uses a deterministic economic model to analyze the choice between CPM and CPC when a web publisher is both a price taker in the market for display ads and faces a decreasing number of viewers visiting its website. His results show that if the CTR is less than the CPM to CPC ratio the publisher should choose the CPM contract otherwise CPC should be chosen. The prices are assumed to be determined exogenously by the market and the supply and demand uncertainties are not considered. McAfee et al. (2010) consider a deterministic model for a web publisher selling maximally representative allocations to advertisers based on the GD contract. Lewis and Reiley

(2011) measure the impact of advertising on sales through an experiment performed between Yahoo! and a major retailer. They find that online display advertising can have a significant impact on a retailer's sales. Najafi-Asadolahi and Fridgeirsdottir (2014) 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.

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 (2009) study the problem of dynamically determining the subscription fee and the size of advertising space on a website. Unlike our paper, all these papers are focused on capacity management problems not pricing 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. (2006) develop a model that determines how ads on a website should be scheduled in a planning horizon to maximize revenue. Their problem belongs to the class of NP-hard problems, and they develop a heuristic to solve it. They also provide a good overview of other related papers on scheduling. In a related work, Turner et al. (2011) develop a model for the dynamic in-game ad scheduling problem faced by a leading network provider of in-game ad space.

There is a growing literature on online display advertising from a revenue management perspective, which mainly focuses on the optimal display ads allocation problems. Examples of recent works include Balseiro et al. (2014), J. Yang et al. (2010) and Alaei et al. (2009) (for a reference of traditional revenue management models, see, e.g., Talluri and van Ryzin 2004). Ciocan and Farias (2012) develop an algorithm for a large class of dynamic allocation problems with unknown demand, with applications in display ad slot allocation and network revenue management. Chen (2011) considers a mechanism design approach for a monopolistic web publisher deciding whether to allocate its display ad slots to guaranteed contracts or to the spot market. Balseiro et al. (2014) consider a similar problem for a web publisher with a single slot, and use a stochastic control approach to characterize an asymptotically optimal efficient allocation policy. Balseiro et al. (2013) study auctions for online display advertising exchanges and show that ignoring advertisers' budgets in these markets can result in substantial revenue losses for publishers. Araman and Popescu

(2010) study the ad allocation problem for more 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). Araman and Fridgeirsdottir (2011) consider a similar web publisher setting to our paper and study pricing and capacity management for a CPM system where advertisers are willing to wait. Their setting does not allow for closed-form solutions and they derive asymptotically optimal solutions.

3. The Model

In this section, we formulate the problem of a web publisher facing uncertain demand from advertisers requesting space to display their ads. The web publisher's website consists of a single webpage with similar slots for ads. In Section 5.1, we will generalize this setting and consider a website with multiple pages. The web publisher uses an ad network that supplies it with the demand. Advertisers request a certain number of impressions (visits) made to their ads by their targeted viewer types (through their chosen ad campaigns). However, the supply of (targeted) viewers is uncertain.

Advertisers' Arrivals and Impression Request Let $\mathcal{K}$ be the set of ad campaigns with size $|\mathcal{K}|$. An advertiser arriving at the ad network, chooses an ad campaign $\kappa_h \in \mathcal{K}$, $h = 1, 2, ..., |\mathcal{K}|$, where κ_h is a vector of attributes of the slot and viewer types that he aims to target. These attributes may include: (i) the slot size and format with the page content, (ii) the keywords related to the page content, (iii) the demographic information of the targeted viewers (e.g., their gender, age, ethnicity, education, income, geographical location), (iv) the device and operating systems that viewers use (e.g., PCs, laptops, smartphones, etc.), and (v) other (behavioral) attributes that the ad network may learn by tracking viewers' activities through the *cookie files* posted in their devices (e.g., viewers with potential *interest in photography)* (see, e.g., Bharadwaj et al. 2010).

We assume that advertisers with the ad campaign κ_h arrive at the publisher's page according to a Poisson process with rate λ_{κ_h}. Assuming Poisson arrivals is common in service settings (see, e.g., Van Mieghem 2000, Cao et al. 2002, Savin et al. 2005). While this assumption captures the stochastic nature of advertisers' demand and may be appropriate for some ad networks, it is unlikely to be a good universal estimator of all advertisers' arrival distributions. We retain the Poisson assumption to maintain the analytical tractability.

Each arriving advertiser with the campaign κ_h requests his ad to be posted on one of the slots on the publisher's page until displayed to (impressed by) x_{κ_h} unique viewers (impression goal). In reality, the requested number of impressions, x_{κ_h}, can be random across advertisers, i.e., x_{κ_h} is a random variable. We will look at this as an extension in Section 5.2.

As explained in the introduction, advertisers' requests are sent to the web publisher's system as long as the publisher has a space available to serve. That is, if all of the publisher's ad spaces are already occupied by advertisers, the network does not send advertisers to the publisher's system. Rather, it sends advertisers to other available systems that match their request. This implies that the publisher's page is a *loss* system. Note that if advertisers approach a web publisher directly (not using any ad network), they may be willing to wait for an available space (e.g., an advertiser directly approaching CNN.com). In that case, if the waiting time is short (e.g., advertisers are served shortly after their arrivals) then since no (or little) queue is formed, the arrivals and the service mechanisms of the system would still be close to those of a loss system. Thus, the main results and managerial insights would be similar as well.

In addition, when the web publisher has a slot available, it usually does not leave it empty; rather it places a default (filler) ad in there (remnant advertising). A default ad is often the publisher's own ad (house ad), or a run-of-network ad that the ad network sends to fill the place (e.g., a public service announcement). In both cases, a default ad generates 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.

Viewers' Arrivals and Targeting Let $\mathcal{V}$ be the set of viewer types with size $|\mathcal{V}|$. Each element in $\mathcal{V}$, denoted by $v_j \in \mathcal{V}$, $j = 1, 2, ..., |\mathcal{V}|$, is a vector of the attributes for one viewer type, which we call "type v_j". Viewers type v_j arrive at the publisher's page according to a Poisson process with rate μ_{v_j}. This assumption is consistent with empirical studies, e.g., Cao et al. (2002), that show that the viewers' traffic tends locally to Poisson distribution. In addition, in Section 5.2, we show that our results provide accurate estimates for the publisher's model even when the viewers' arrivals are non-Poisson. We let the binary variable $v_j^{\kappa_h}$ be $v_j^{\kappa_h} = 1$ if the ad campaign κ_h targets viewers type v_j, and $v_j^{\kappa_h} = 0$, otherwise. We let $V(\kappa_h)$ be the set of viewer types that are targeted by ad the ad campaign κ_h, i.e., $V(\kappa_h) = \{v_j | v_j^{\kappa_h} = 1, v_j \in \mathcal{V}\}$, and $K(v_j)$ be the set of ad campaigns that target viewers type v_j, i.e., $K(v_j) = \{\kappa_h | v_j^{\kappa_h} = 1, \kappa_h \in \mathcal{K}\}$. Without loss of generality, we assume that each viewer type is targeted by at least one ad campaign and each ad campaign targets at least one viewer type, i.e., $\underset{\kappa_h \in \mathcal{K}}{\cup} V(\kappa_h) = \mathcal{V}$, $\underset{v_j \in \mathcal{V}}{\cup} K(v_j) = \mathcal{K}$. When a viewer type v_j visits the publisher's page, only ads from the campaigns set $K(v_j)$ are displayed to him, which we refer to as the *page-version* v_j. Figure 2 shows an illustrative example of three ad campaigns targeting different viewer types. For convenience, we denote the viewer type that is targeted by only the ad campaign κ_h with v_h.

Case A: In this case, the set of ad campaigns and viewer types are $\mathcal{K} = \{\kappa_1, \kappa_2, \kappa_3\}$ and $\mathcal{V} = \{v_1, v_2, ..., v_7\}$, respectively. In addition, the set of viewer types targeted by the ad campaigns

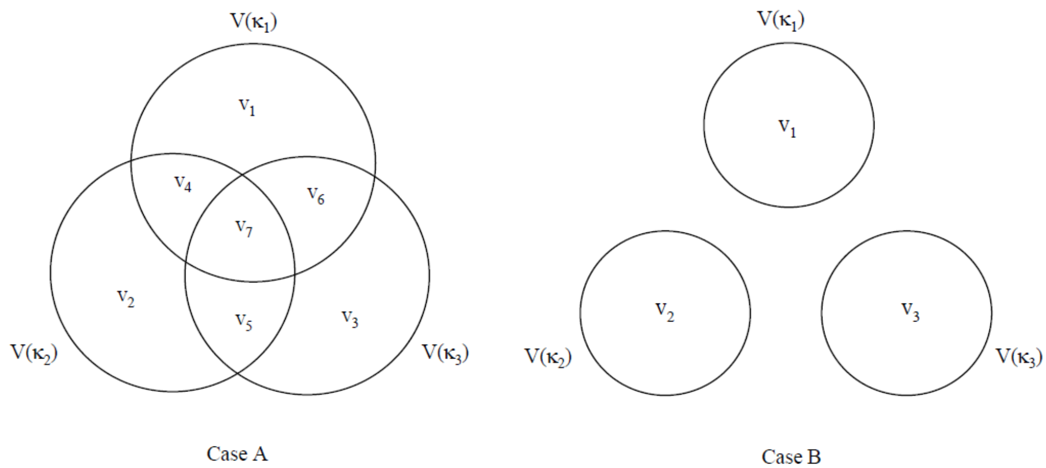

Figure 2 An example of three ad campaigns targeting various viewer types.

κ_1, κ_2, and κ_3 are $V(\kappa_1) = \{v_1, v_4, v_6, v_7\}$, $V(\kappa_2) = \{v_2, v_4, v_5, v_7\}$, and $V(\kappa_3) = \{v_3, v_5, v_6, v_7\}$, respectively. When a viewer type v_i visits the publisher's page, he sees only ads from the campaigns that belong to the set $K(v_i)$, which have targeted the viewer (i.e., page-version v_j). For example, a viewer type v_1 see ads only from the campaign set $K(v_1) = \{\kappa_1\}$, a viewer type v_4 sees ads only from the campaigns set $K(v_4) = \{\kappa_1, \kappa_2\}$, and a viewer type v_7 sees ads only from the set $K(v_7) = \{\kappa_1, \kappa_2, \kappa_3\}$.

Case B: In this case, we have $\mathcal{K} = \{\kappa_1, \kappa_2, \kappa_3\}$, $\mathcal{V} = \{v_1, v_2, v_3\}$. In addition, each ad campaign is targets exactly one viewer type (i.e., $V(\kappa_h) = \{v_h\}, h = 1, 2, 3$) while each viewer type is targeted exactly by one ad campaign (i.e., $K(v_j) = \{\kappa_j\},\ j = 1, 2, 3$). That is, there is a *one-to-one correspondence* between the elements of $\mathcal{K}$ and $\mathcal{V}$.[3]

We note that in some ad networks, advertisers may not target the viewer types specifically. Instead, they may leave the task of targeting related viewer types to the ad network. In such cases, using the ad campaign and viewer type attribute vectors κ_h and v_j, the ad network determines a relevant score sometimes referred to as *quality of service* (QoS) or *placement quality*, which we denote by $Q(\kappa_h, v_j)$. A higher value of $Q(\kappa_h, v_j)$ suggests that an ad from the campaign κ_h has a greater chance to be considered and clicked by viewers type v_j. There is no standard way to obtain $Q(\kappa_h, v_j)$, however one common measure for $Q(\kappa_h, v_j)$ is that it is viewed equivalent to the click-through rate (CTR) that an ad with the campaign κ_h is estimated to perceive if it is shown to viewers type v_j (see, e.g., Balseiro et al. 2013). Ad networks typically use empirical data to estimate $Q(\kappa_h, v_j)$ and post an ad on the page-version v_j only if a minimal QoS level criteria such

[3] Note that it is also possible that multiple viewer types are targeted by exactly one advertiser. This situation, is a special scenario of Case B because we may consider all those viewer types as a unified type.

as $Q(\kappa_h, v_j) \geq \tau$, for some given τ, is satisfied. Here, we do not consider the QoS criteria directly, but in Section 4, we will capture it indirectly by considering a generalized price-demand function for the publisher's page.

Ads Rotation The publisher can typically serve more advertisers on each page-version v_j than there are slots. For example, multiple ads could share the same slot with each ad randomly displayed to the viewers 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) or AdCycle's ad management software[4]. We denote S_{κ_h} as the maximum number of ads from the campaign κ_h accepted into the publisher's system. In addition, we denote n_{v_j} to be the number of slots on the page-version v_j. If $S_{v_j} := \sum_{\kappa_h \in K(v_j)} S_{\kappa_h} > n_{v_j}$ then, when a viewer type v_j arrives, a subset of n_{v_j} ads from the pool of S_{v_j} is selected randomly and displayed to him (random ad rotation). For example, in Figure 2, Case A, the publisher may accept up to 7 ads from each campaign into the system, i.e., $S_{\kappa_1} = S_{\kappa_2} = S_{\kappa_3} = 7$, whereas each page-version has only 4 slots, i.e., $n_{v_j} = 4$, $j = 1, 2, ..., 7$. Thus, for example, when a viewer type v_7 arrives the publisher displays 4 ads from a pool of maximum $S_{v_7} = \sum_{\kappa_h \in K(v_7)} S_{\kappa_h} = S_{\kappa_1} + S_{\kappa_2} + S_{\kappa_3} = 21$ related ads from the campaigns that have targeted the viewer, i.e., $K(v_7) = \{\kappa_1, \kappa_2, \kappa_3\}$.

The Optimization Problem The publisher's goal is to maximize its total revenue rate by determining the right prices to charge advertisers each time their ads are displayed to one of their targeted viewers. The revenue rate gained from advertisers served on the page-version $v_j \in \mathcal{V}$ consists of the payments made by advertisers multiplied by their "actual" demand rate for the page-version v_j (defined later). The payment of an advertiser on the page-version v_j consists of the price per each impression, denoted by p_{v_j} (specified below), multiplied by the number of impressions, denoted by x_{v_j}. Thus, the total payment made by an advertiser with the ad campaign $\kappa_h \in \mathcal{K}$ is $\sum_{v_j \in V(\kappa_h)} x_{v_j} p_{v_j}$, where $x_{\kappa_h} = \sum_{v_j \in V(\kappa_h)} x_{v_j}$. The following proposition characterizes the advertisers' demand process for any page-version v_j.[5]

Proposition 1 *Let μ_{κ_h} be the arrival rate of viewers whose types are targeted by the ad campaign $\kappa_h \in \mathcal{K}$, i.e., $\mu_{\kappa_h} = \sum_{v_{j'} \in V(\kappa_h)} \mu_{v_{j'}}$. Then the total demand for a page version $v_j \in \mathcal{V}$, from advertisers with the campaigns $\kappa_h \in K(v_j)$, is Poisson with rate*

$$\lambda_{v_j} = \sum_{\kappa_h \in K(v_j)} \frac{\mu_{v_j}}{\mu_{\kappa_h}} \lambda_{\kappa_h}, \forall v_j \in \mathcal{V}. \tag{1}$$

[4] http://www.adcycle.com/

[5] Note that each impression made an ad is now valued and priced based on a viewer's type who has visited the page. Naturally some viewers might be more valuable to an advertiser than others. For example, viewers with an income level greater than \$100k may be more valuable to an advertiser than college students. Thus, when the ad is impressed by these more valuable viewers, the advertiser is charged more.

Proof: Advertisers with the campaign κ_h arrive at the publisher's system according to a Poisson process with rate λ_{κ_h}, and request their ads to be displayed to the set of viewer types $V(\kappa_h)$. As the arrival of viewers type $v_j \in V(\kappa_h)$ is Poisson, the probability that the ad is displayed to the viewer-type $v_j \in V(\kappa_h)$ is μ_{v_j}/μ_{κ_h}, where $\mu_{\kappa_h} = \sum_{v_{j'} \in V(\kappa_h)} \mu_{v_{j'}}$. Thus, the demand for the page-version v_j is Poisson with rate $\frac{\mu_{v_j}}{\mu_{\kappa_h}}\lambda_{\kappa_h}$, any ad campaign $\kappa_h \in K(v_j)$. Hence, the total demand for the page-version $v_j \in \mathcal{V}$ is Poisson with rate $\lambda_{v_j} = \sum_{\kappa_h \in K(v_j)} \frac{\mu_{v_j}}{\mu_{\kappa_h}}\lambda_{\kappa_h}$. ■

We capture the price-sensitivity of the advertisers with a price-demand function, $p_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})$, which is assumed to be continuous and (weakly) decreasing in the advertisers' arrival rate, the number of impressions, and the number of ad spaces. The decreasing relation between the price and the number of impressions captures the fact that, *given all other parameters are fixed*, advertisers receive lower prices (better deals) if they are willing to purchase larger numbers of impressions. In addition, advertisers often do not want to see their ads posted on pages on which the ads are hardly recognized, due to many ads posted on the page (*slot congestion*). For this reason, advertisers request pages with a lower ad congestion as they provide a higher chance for the ads to be seen and clicked on. As explained before, advertisers often perceive such a higher chance as a higher QoS or the placement quality. Publishers are aware that to deliver a desired quality of service, there is a trade-off between the website's profitability and the slot congestion, which prevents the publisher from becoming overly greedy and posting too many ads on its page (see, e.g., Mookerjee et al. 2012). To capture this trade-off, we assume that the publisher incurs a congestion penalty, in the form of a price discount, for adding an extra space on each page version (i.e., $\partial p_{v_j}/\partial S_{v_j} \leq 0$). The decreasing relation between p_{v_j} and S_{v_j} prevents the publisher from posting too many ads (on each page-version), so the click chance of the posted ads does not reduce significantly.

Even though it might not be trivial for the publisher to determine the price 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 values. Furthermore, publishers such as Yahoo! have started looking into estimating the price-demand relationship. The process of advertisers being matched to web publishers based on type preference and willingness-to-pay can be modelled specifically. However, ultimately it will lead to a price-demand relationship. We will not model the process in detail here but in Section 1, we have provided a description of the matching process common in ad networks.

For popular websites often only a part of the advertisers' demand can be met by the publisher. This means that the actual demand rate for each subsystem is scaled down by the probability that there are advertising spaces available at the arrival time of an advertiser. However, as arrivals are

Poisson, we can invoke the *PASTA* property that Poisson arrivals always see time averages. Thus, the arrival-time probability of having i advertisers served in the page-version v_j is identical to its steady-state probability (Gross and Harris 1998), which we denote by $\mathbb{P}_i^{v_j}$, $i \in \{0, ..., S_{v_j}\}, v_j \in \mathcal{V}$.

We note that as the price for each page-version, $p_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})$, is a one-to-one function of λ_{v_j}, and λ_{v_j} is a function of λ_{κ_h} (Proposition 1), we can optimize the revenue rate with respect to λ_{κ_h} directly and then determine the optimal λ_{v_j} and $p_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})$, accordingly. The optimization problem of the publisher can be formulated as:

$$\max_{\forall \lambda_{\kappa_h} \geq 0, \ \kappa_h \in \mathcal{K}} R = \sum_{v_j \in \mathcal{V}} \lambda_{v_j}(1 - \mathbb{P}_{S_{v_j}}^{v_j}(\lambda_{v_j}; x_{v_j}, S_{v_j}, \mu_{v_j}))p_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})x_{v_j}, \tag{2}$$

Subject to:

$$\lambda_{v_j} = \sum_{\kappa_h \in K(v_j)} \frac{\mu_{v_j}}{\mu_{\kappa_h}} \lambda_{\kappa_h}, \ v_j \in \mathcal{V}.$$

In the above formula, $\mathbb{P}_{S_{v_j}}^{v_j}$ is the steady-state probability that the page-version v_j page is fully occupied. Hence, $\lambda_{v_j}(1 - \mathbb{P}_{S_{v_j}}^{v_j})$ is the effective arrival rate of advertisers with the ad campaigns $\kappa_h \in K(v_j)$ who are accepted by the page-version v_j[6]. Note that if there is a one-to-one correspondence between the elements of $\mathcal{K}$ and $\mathcal{V}$ (e.g., as in Figure 2, Case B), the demand rate from the advertisers with a particular ad campaign is the same as the demand rate for the page-version that they have targeted, i.e., $\lambda_{v_h} = \lambda_{\kappa_h}$ $(h = 1, 2, ..., |\mathcal{K}| = |\mathcal{V}|)$. Thus, the optimization problem (2) can be re-expressed based on the ad-campaigns instead of the viewer-types:

$$\max_{\forall \lambda_{\kappa_h} \geq 0, \ \kappa_h \in \mathcal{K}} R = \sum_{\kappa_h \in \mathcal{K}} \lambda_{\kappa_h}(1 - \mathbb{P}_{S_{\kappa_h}}^{\kappa_h}(\lambda_{\kappa_h}; x_{\kappa_h}, S_{\kappa_h}, \mu_{\kappa_h}))p_{\kappa_h}(\lambda_{\kappa_h}, x_{\kappa_h}, S_{\kappa_h})x_{\kappa_h}, \tag{3}$$

where since the right side is separable over $\kappa_h \in \mathcal{K}$, we can focus on optimizing the revenue generated by one campaign.

In order to optimize (2), we first need to characterize the steady-state probability $\mathbb{P}_{S_{v_j}}^{v_j}(\lambda_{v_j}; x_{v_j}, S_{v_j}, \mu_{v_j})$. To obtain $\mathbb{P}_{S_{v_j}}^{v_j}$, we focus on only one page-version (*a single page with only one version*), drop all indices and refer to the advertisers targeting version v_j merely as "advertisers", and viewers type v_j merely as "viewers".

[6] Note that if an advertiser is not accepted into the page-version v_j, his request is immediately forwarded to another relevant publisher with available space on its page-version v_j. For example, an advertiser may request his ad to be viewed by three types of viewers a, b and c. The publisher can only serve the ad when viewers a or b arrive, but not when viewers type c arrive (because there are already enough ads posted when viewers type c arrive). In that situation, the publisher served only part of the contract, and the rest is forwarded to and immediately served by another relevant publisher who can expose the ad to viewers type c.

The Steady-State Probability Distribution In this section, we first set $S = n$, i.e., there is no rotation of ads. However, we will consider ads rotation later in the section. Having Markovian arrival and service processes, we can now model the system using continuous-time 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 impressions left to be delivered for each slot.

When an advertiser arrives, he is randomly assigned to one of the available slots with an equal probability as the slots 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 now define the state of the system and its transitions. We formulate the system as a queueing model with the state vector

$$\mathbf{k} = (k_1, k_2, ..., k_n), \quad 0 \le k_h \le x, \quad h = 1, 2, ..., n, \tag{4}$$

in which each component represents the number of impressions left to be satisfied 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 displayed 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 system 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. For convenience, we consider that $\mathbf{k}$'s positive components are always arranged in an increasing order followed by components whose values are zero. We illustrate how the state transitions work through the following examples:

i) 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 $n - i$ are zero. This means that there are i ads in the system with the remaining impressions $k_1, ..., k_i$ and the rest $n - i$ slots are empty. The viewers consider the system with rate μ. When a viewer arrives at the system, since all the ads are displayed, the state of the system makes a transition to the new state

$$\mathbf{k}' = (k_1 - 1, k_2 - 1, ..., k_i - 1, 0, ..., 0) = \mathbf{k} - \sum_{h=1}^{i} \mathbf{e}_h^T,$$

with rate μ where $\mathbf{e}_h$ is the h^{th} unit vector. That is, all the positive components' values reduce by one at the same time (the synchronized service), while the zero components do not change. For example, if the state of the system is $(2, 3, 4, 0, 0)$ then it makes a transition to the state $(1, 2, 3, 0, 0)$ with rate μ. We note that (unlike many typical multi-channel systems) the service channels in the

CPM system are inter-dependent through their synchronized services, which are triggered by the arrival of each viewer (synchronization), which makes the analysis of the system complicated.

ii) Next, 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 system, the publisher assigns one of the empty slots to his ad, and the state will make a transition to the state

$$\mathbf{k}'' = (k_1, k_2, ..., k_i, x, \underbrace{0, ..., 0}_{n-i-1}) = \mathbf{k} + \mathbf{x}\mathbf{e}_{i+1}^T$$

with rate λ. Once again, we note that the vectors $\mathbf{k} + \mathbf{x}\mathbf{e}_{i+1}^T, \mathbf{k} + \mathbf{x}\mathbf{e}_{i+2}^T, ...,$ and $\mathbf{k} + \mathbf{x}\mathbf{e}_n^T$ all refer to the same system state, where for convenience we represent them all with $\mathbf{k}'' = \mathbf{k} + \mathbf{x}\mathbf{e}_{i+1}^T$.

In order to find $\pi_{\mathbf{k}}$, the steady-state probability that the system is in state $\mathbf{k}$, we characterize all possible states and transitions, and solve the flow balance equations. Given the complex transition dynamics one may wonder if a nondegenerate solution exists for $\pi_{\mathbf{k}}$. Proposition 2 ensures the existence of the steady state for the publisher's Markov chain.

Proposition 2 *In the publisher's Markov chain, a unique nondegenerate solution to the stationary flow-balance equations always exists.*

The proof of Proposition 2 mainly relies on two well-known theorems for stochastic processes. The first says that a continuous-time Markov chain (CTMC) has identical and unique limiting and stationary distributions if it is irreducible and positive recurrent, while the second says that an irreducible CTMC with finite number of states is positive recurrent (ergodic). The publisher's CTMC is clearly irreducible since any state can be reached from any other. It can also be seen directly from (4) that the number of states is finite. Hence, the chain is positive recurrent. Thus, a unique steady-state distribution exists. The following proposition gives the closed-form solution of the steady-state probability distribution of the number of advertisers in the system.

Proposition 3 *The steady-state probability of a web publisher's system with n slots in the state $\mathbf{k} = \sum_{\mathbf{j=1}}^{\mathbf{i}} k_j \mathbf{e}_j^T$, where k_j $(0 \le k_j \le x)$ is the number of impressions left in slot j $(0 \le j \le i)$ and the rest $n-i$ slots are empty is:*

$$\pi_{\mathbf{k}}(r, x, n) = \frac{\left(\frac{r}{1+r}\right)^i \left(\frac{1}{1+r}\right)^x \mathbf{1}_{i<n} + \left(\frac{r}{1+r}\right)^n \left(\frac{1}{1+r}\right)^{x-1} \mathbf{1}_{i=n}}{\sum_{j=0}^{n} \binom{x+n-1}{j} \left(\frac{r}{1+r}\right)^j \left(\frac{1}{1+r}\right)^{x-1+n-j}}, \quad r = \frac{\lambda}{\mu}, \tag{5}$$

Furthermore, the steady-state probability of having i advertisers ($0 \le i \le n$) in the system is:

$$\mathbb{P}_i(r,x,n) = \frac{\binom{x+i-1}{i}\left(\frac{r}{1+r}\right)^i \left(\frac{1}{1+r}\right)^x \mathbf{1}_{i<n} + \binom{x+n-1}{n}\left(\frac{r}{1+r}\right)^n \left(\frac{1}{1+r}\right)^{x-1} \mathbf{1}_{i=n}}{\sum_{j=0}^{n} \binom{x+n-1}{j}\left(\frac{r}{1+r}\right)^j \left(\frac{1}{1+r}\right)^{x-1+n-j}}. \tag{6}$$

Surprisingly, despite the complicated structure of (6), we may interpret it intuitively using an $M/M/1$ system. This interpretation is interesting because the two systems have considerably different mechanisms. We note that the ratio $r/(1+r)$ can be viewed as the (jump-process) probability (in an $M/M/1$ system) that the number of advertisers arriving at the system increases by one at any arbitrary point in time. When $i<n$, Equation (6) suggests that from a steady-state standpoint the system has i advertisers if from any idle period, the number of advertisers arriving at the system jumps from 0 to i before the number of viewers jumps from 0 to x. Technically, this is all the possible rearrangements between the i terms $r/(1+r)$ and the $x-1$ terms $1/(1+r)$, i.e., $\binom{x+i-1}{i}\left(\frac{r}{1+r}\right)^i\left(\frac{1}{1+r}\right)^{x-1}$, times the probability that the x^{th} viewer arrives last, i.e., $1/(1+r)$. In a similar way, the denominator of (6) can be explained as the (Binomial) probability that the publisher accepts up to and including n advertisers into the system. When $i=n$, Equation (6) suggests that the steady-state probability that the publisher's system has n advertisers is the probability that from any idle period, there are at least n jumps in advertisers' arrivals before the number of jumps in the viewers' arrivals reaches to x, i.e., $\sum_{k=n}^{\infty}\binom{x+n-1}{n}\left(\frac{r}{1+r}\right)^k\left(\frac{1}{1+r}\right)^{x-1} = \binom{x+n-1}{n}\left(\frac{1}{1+r}\right)^{x-2}\left(\frac{r}{1+r}\right)^n$, times the probability that the x^{th} viewer arrives last, i.e., $1/(1+r)$. However, this probability is scaled by the denominator, the probability that at most n advertisers can be served. Another interesting observation from Proposition 3 is that $\pi_{\mathbf{k}}$ does not depend on the actual number of impressions left in each slot, but it depends only on the number of filled slots.

Rotation of Ads Next, we consider the case where the slots can be shared by up to $S>n$ (real) ads. Let us consider there are m ads present in the system, which are rotated across the n slots in the system. If $m<S$ then $(S-m)$ ads would be filler ads. When a new advertiser arrives the publisher immediately replaces this real ad with one of the filler (default) ads. When a viewer arrives at the system a subset of n ads are randomly selected from the pool of S real and filler ads and displayed to the viewer[7]. We know that the number of possible subsets to select n ads out of S is $\binom{S}{n}$. To obtain the number of subsets that include a particular ad, we select that particular ad

[7] It would be possible to consider that when the number of ads in the system $m<n$ then the publisher only fills $(n-m)$ slots with filler ads. Hence, when a viewer arrives all ads are displayed. This scenario might be more realistic, however adds layers of complexity to the analytical model. To maintain analytical tractability, we assume that if $m<n$, the publisher posts $(S-m)$ filler ads in the remaining empty slots and rotates all randomly.

and then choose the remaining $n-1$ ads in the subset from the $S-1$ remaining total ads. Hence, the probability that a particular ad is displayed is

$$\mathbb{P}_{Disp} = \frac{\binom{S-1}{n-1}}{\binom{S}{n}} = \frac{n}{S}.$$

Therefore, the Markovian transitions of the system would be identical to those of a system with S slots *without* rotation while the viewers' arrival (service) rate is $\mu \times \mathbb{P}_{Disp} = \mu n/S$[8]. That is, the probability of having i advertisers in the system with random ad rotation, $\mathbb{P}_i(r,x,n,S)$, is

$$\begin{aligned} \mathbb{P}_i(r,x,n;S) &= \mathbb{P}_i(\widehat{r},x,S), \quad i=0,1,2,...,S, \\ \widehat{r} &= rS/n, \end{aligned} \tag{7}$$

where $\mathbb{P}_m(\widehat{r},x,S)$ is defined by (6). In the next two propositions, we show some structural properties of the average number of advertisers in the system and the busy probability. These will be useful when considering the pricing problem of the web publisher in the next section.

Proposition 4 $\forall x,S,n$ $(n \le S)$ *the full state probability,* $\mathbb{P}_S$*, defined by (6) satisfies:*
(i) $\frac{\partial \mathbb{P}_S}{\partial r} \ge 0,$ *(ii)* $\mathbb{P}_S(x+1) - \mathbb{P}_S(x) \ge 0,$ *(iii)* $\mathbb{P}_{S+1}(x) \le \mathbb{P}_S(x).$

This proposition confirms the intuition that the web publisher is busier if there is more demand and less traffic, more impressions, and fewer advertising spaces. Nevertheless, our numerical analysis indicates that $\mathbb{P}_S$ is not necessarily concave in the number of impressions.

Proposition 5 $\forall x,S$ *the average number of advertisers,* $L_S(x)$*, and the increment* $\Delta L_S(x) = L(x+1) - L(x)$ *satisfy:*
(i) $\Delta L_S(x) \ge 0,$ *(ii)* $\Delta L_S(x+1) \le \Delta L_S(x),$ *(iii)* $\frac{\partial L}{\partial r} \ge 0, \frac{\partial^2 L}{\partial r^2} \le 0,$ *(iv)* $L_S(x) \le L_{S+1}(x).$

Parts (i) and (ii) in Proposition 5 imply that if the web publisher's system contains only one subsystem (i.e., a page containing similar ads with one version) then 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 if the number of impressions it offers is larger. However, as x increases, this impact levels off. In addition, Part (iii) implies that the average number of advertisers in the system is increasing and concave in the intensity rate r. That is, the publisher is busier with more demand, less traffic, or a higher demand-to-traffic ratio. However, as r increases, the impact levels off. Part (iv) states that the average number of advertisers in the web publisher's system increases in the number of advertising spaces. Proposition 4 and 5 are crucial when solving the optimal pricing problem in the next section.

[8] That is, in the proof of Proposition 2, all the transition flow balance equations would be the same with the only difference that μ is replace by $\mu n/S$. Other than that, all the steps of the proof for the system with random ad rotation would be identical to the proof of Proposition 2.

4. The Optimal Price

Having fully characterized the probabilistic properties of the web publisher's operation, we now turn to the task of finding the optimal pricing policy. The web publisher's objective is to determine the optimal price to charge per click that maximizes the revenue rate defined in (10). The next proposition ensures the existence of the optimal solution and gives the optimal price.

Proposition 6 *Let the price function, $p_{v_j}(\lambda_{v_j}; x_{v_j}, S_{v_j})$, $v_j \in \mathcal{V}$, be concave and (weakly) decreasing in λ_{v_j}. Then the publisher's revenue, R, is concave in $\Lambda = (\lambda_{\kappa_1}, ..., \lambda_{\kappa_{|\mathcal{K}|}})$. In addition, the advertisers' optimal arrival rate for each ad campaign, $\lambda^*_{\kappa_h}$, $\kappa_h \in \mathcal{K}$, is the unique solution to:*

$$\sum_{v_j \in V(\kappa_h)} \left(\frac{\partial L_{v_j}(\lambda_{v_j}; x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}} p_{v_j}(\lambda_{v_j}; x_{v_j}, S_{v_j}) + \frac{\partial p_{v_j}(\lambda_{v_j}; x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}} L_{v_j}(\lambda_{v_j}; x_{v_j}, S_{v_j}) \right) \mu_{v_j} = 0, \quad (8)$$

where $\lambda_{v_j} = \sum_{\kappa_h \in K(v_j)} \frac{\mu_{v_j}}{\mu_{\kappa_h}} \lambda_{\kappa_h}$, $v_j \in \mathcal{V}$ and $L_{v_j}(\lambda_{v_j}; x_{v_j}, S_{v_j}) = r_{v_j} \frac{S_{v_j}}{n_{v_j}} x_{v_j} (1 - \mathbb{P}_{S_{v_j}}(\lambda_{v_j}; x_{v_j}, S_{v_j}, \mu_{v_j}))$ is the average number of advertisers served on page version v_j.

In order to ensure concavity of the objective function, we need $p_{v_j}(\lambda_{v_j}; x_{v_j}, S_{v_j})$ to be weakly concave in λ_{v_j}. Even though this might seem a restrictive assumption it includes a *linear price,* which is widely applied in economics and management science literature (see, e.g., Adida and DeMiguel 2011). In addition, our numerical analysis indicates that many convex price functions give a unimodal revenue function as well.

The next proposition confirms the somewhat nonintuitive results that the web publisher may be worse off by having more spaces on its page-version v_j, and offering more impressions; but is always better off by having more viewer traffic to its page-version v_j. We denote the revenue rate generated from each page version v_j by $R_{S_{v_j}, x_{v_j}}(\lambda_{v_j}(S_{v_j}, x_{v_j}, \mu_{v_j}); \mu_{v_j})$ to emphasize the dependence on S_{v_j}, x_{v_j}, and μ_{v_j}.

Proposition 7 *The overall optimal revenue of the publisher, $\sum_{v_j \in \mathcal{V}} R_{S_{v_j}, x_{v_j}}(\lambda^*_{v_j}(S_{v_j}, x_{v_j}, \mu_{v_j}); \mu_{v_j})$, is: (i) concave in S_{v_j} with a global maximum at some $S^*_{v_j}$, (ii) concave in x_{v_j} with a global maximum at some $x^*_{v_j}$, (iii) increasing in μ_{v_j}.*

This proposition implies that it is possible that the publisher loses revenue by serving more advertisers on a page-version. The reason for this behavior is due to the trade-off between the two opposing forces: (i) the number of ad spaces and (ii) price. Adding an extra space to a page-version enables the publisher to serve more advertisers at the same time, which increases the revenue. However, with an additional space, the publisher reduces the price (as a congestion penalty) to compensate offering a more congested page (e.g., lower click-through rate), which drives down the revenue. If the revenue gained from adding an extra space is less than the revenue lost due to the price discount, the optimal revenue decreases.

The following proposition states the counter-intuitive result from a marketing point of view that, when there is a one-to-one correspondence between the elements of $\mathcal{K}$ and $\mathcal{V}$, then the publisher may optimally increase the price per impression when offering more impressions.

Proposition 8 *Let* $p_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})$ *be weakly concave in* λ_{v_j} *with* $\frac{\partial p_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j} \partial x_{v_j}} \leq 0$, *and* $K(v_j) = \{\kappa_j\}, V(\kappa_j) = \{v_j\}$, *for all* $j = 1, 2, ..., |\mathcal{V}| = |\mathcal{K}|$. *Then:*
(i) $\lambda^*_{v_j}$ *is decreasing in* x_{v_j}. *That is,* $\frac{\partial \lambda^*_{v_j}}{\partial x_{v_j}} \leq 0$,
(ii) $\frac{dp_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{dx_{v_j}} \geq 0$, *if and only if* $\frac{\partial \lambda^*_{v_j}}{\partial x_{v_j}} \leq -\frac{\partial p(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial x_{v_j}} / \frac{\partial p(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}}$.

This proposition is interesting as one typically expects the opposite, i.e., the optimal price to be lower when more impressions are offered. In order to understand what drives these results, we note that the higher the number of impressions is, the longer it takes to serve each advertiser. Since the service time for each ad is longer, the web publisher does not need as many advertisers (per time unit) to fill the ad spaces (i.e., *empty buffers*) and keep the system busy (i.e., eliminate the downtime). This results in a lower *optimal* arrival rate. In addition, since the price has a decreasing relationship with x_{v_j} and $\lambda^*_{v_j}$, an increase in x_{v_j} lowers the price while a decrease in $\lambda^*_{v_j}$ raises it. Part (ii) suggests that if the price increase due to the lower $\lambda^*_{v_j}$ is greater than the price decrease caused by a higher x_{v_j}, then the publisher finds it optimal to increase the price rather than giving a quantity discount.

Practical Evidence i) We also checked this result with some real publishers. All publishers we spoke to offer quantity discounts except Yahoo!. According to Prof. Preston McAfee, the former vice president and senior research fellow at Yahoo!, (currently a chief scientist at Microsoft) Yahoo! now increases the CPM price for large contracts instead of giving a discount. However, they did not have any theoretical underpinnings why this policy yields a higher expected revenue; Rather they had come to this pricing approach through a series of trials and errors over time. Our result provided them with a theoretical explanation.

ii) In addition to approaching publishers, we obtained and studied real data from Aller Internett, a major Norwegian web publisher, which may confirm the counter-intuitive quantity-discount result. The data belongs to a period from 1^{st} October 2009 to 24^{th} February 2010 (when we conducted this empirical study). After clearing the data and considering a particular ad size and location, we did not have many data points remaining. Figure 3 illustrates the empirical relationship between the number of impressions and the CPM price. As can be seen, the CPM price

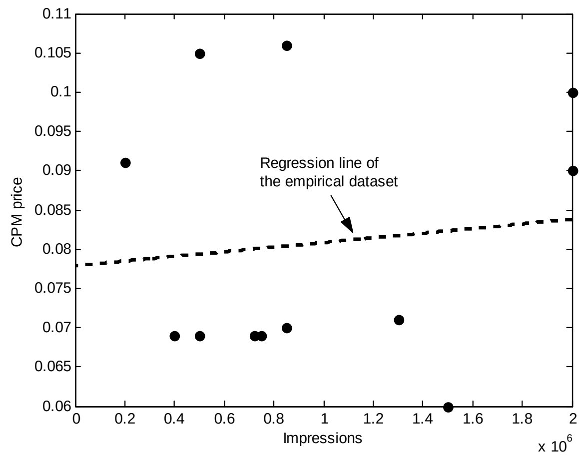

Figure 3 The relationship between the delivered number of impressions and the CPM price.

increases in the number of impressions being offered. Nevertheless, we note that despite the visible upward trend the t-statistic value for the slope of the regression line is close to zero, which implies that the price might be independent of the number of impressions for this publisher. However, the increasing relation may weakly confirm the pricing insight.[9]

Next, we consider the sensitivity of the optimal price with respect to the number of advertising spaces. When the web publisher increases the space S_{v_j}, it is typically expected to reduce the price to attract more advertisers and fill the extra space as well as due to a higher slot congestion. Nevertheless, the next proposition shows that when the publisher increases S_{v_j}, it may *increase* the price.

Proposition 9 *Let $|\mathcal{V}| = |\mathcal{K}|$ and $K(v_j) = \{\kappa_j\}, V(\kappa_j) = \{v_j\}, j = 1, 2, ..., |\mathcal{V}|$. In addition, let $p_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})$, be decreasing and weakly concave in λ_{v_j} and $S_{v_j} << x_{v_j}$. Then, there exists some $w_{v_j} \in [\frac{\mu_{v_j} n_{v_j}}{x_{v_j} S_{v_j}}, \frac{\exp(1)\mu_{v_j} n_{v_j}}{x_{v_j} S_{v_j}}]$ such that if $\lambda^*_{v_j} \geq w_{v_j}$ then:*

(i) $\lambda^*_{v_j}$ is decreasing in S_{v_j}. That is, $\frac{\partial \lambda^*_{v_j}}{\partial S_{v_j}} \leq 0$,

(ii) $\frac{dp_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})}{dS_{v_j}} \geq 0$, if and only if $\frac{\partial \lambda^*_{v_j}}{\partial S_{v_j}} \leq -\frac{\partial p_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})}{\partial S_{v_j}} / \frac{\partial p_{v_j}(\lambda_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}}$.

Part (i) states that if there is a one-to-one correspondence between the elements of $\mathcal{K}$ and $\mathcal{V}$, unlike the typical expectation, when the publisher increases the number of ad spaces it may lose some advertisers at the optimal level. The main reason for this behavior is that increasing the number of ad spaces (system buffers) enables the publisher to serve more advertisers simultaneously.

[9] Note that the reason that this publisher has offered various prices for similar impressions is that the impressions were sold by real sales agents who negotiated with advertisers. The agents were not using any pricing software to determine the prices systematically; rather the prices were set through negotiations.

However, as more ads are being served at the same time, each ad has less chance to be displayed to arriving viewers. Thus, advertisers spend more time in the system, which increases the system's non-zero buffers (occupied spaces) and reduces its downtime. Hence, the publisher does not need to speed up the arrivals. Furthermore, we note that increasing S_{v_j} and a consequent decrease in $\lambda^*_{v_j}$ have opposite impacts on the optimal price. Part (ii) mentions that if the price increase as a result of a lower $\lambda^*_{v_j}$ (publisher needing fewer arrivals) is greater than the price decrease as a result of a higher S_{v_j} (publisher discounting as a congestion penalty), then the optimal price increases.

5. Extensions

There are several directions that the CPM model can be extended to. In this section, we discuss three important extensions, leaving the rest for future research.

5.1. Multiple Pages and Types of Ads

Until now, we have considered that the publisher's system consists of a single page. In this section, we extend the base model to examine multiple pages. In order to start, we assume the publisher's website contains L pages labeled from 1 to L. 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 where the same price is charged within each group. For example, the top of the page may display two equally sized ads, while several small ads may be placed at the bottom (rectangles). This would lead to two ad-groups. More formally, for each page l, we group the ads into M^l groups of equivalent slots. Each ad-group m^l, $1 \leq m^l \leq M^l$, has $|\mathcal{V}|$ different versions. We denote the version-v_j of each ad-group m^l (i.e., group m on page l) by (m^l, v_j) and refer to it as a *subsystem.* We assume that each subsystem (m^l, v_j) contains $n^{m^l}_{v_j}$ equivalent slots. The advertisers with the ad campaign $\kappa_h \in \mathcal{K}$ who wish to post an ad in the group m^l arrive at the publisher's system according to a Poisson process with rate $\lambda^{m^l}_{\kappa_h}$. They request that their ads are displayed to $x^{m^l}_{\kappa_h}$ unique viewers of type $v_j \in V(\kappa_h)$, and pay $p^{m^l}_{v_j}$ dollars each time their ads are viewed. Viewers type $v_j \in \mathcal{V}$ visit page l according to a Poisson process with rate $\mu^l_{v_j}$ and use the page content based on an exponential distribution. When a viewer type v_j visits page l, only ads from the subsystems (m^l, v_j) are displayed to him. Thus, the arrival rate for each subsystem (m^l, v_j) is the same as the one for the version v_j of page l, i.e., $\forall (m^l, v_j)$, $\mu^{m^l}_{v_j} = \mu^l_{v_j}$. After visiting page l, viewers type v_j, may decide to visit a different page $h(\neq l)$ with probability $\alpha^{lh}_{v_j}$, or leave the publisher's system without visiting any other pages with probability $\alpha^{l0}_{v_j} = 1 - \sum_{h \neq l} \alpha^{lh}_{v_j}$. The probability $\alpha^{lh}_{v_j}$ may be viewed as the *traffic shaping* probability as referred to by some recent literature, e.g., Chakrabarti and Vee 2012. Chakrabarti and Vee 2012 suggest that targeted ad networks (e.g., Chitika) may wish to influence these traffic shaping probabilities (e.g., $\alpha_{v_j}^{lh}$) by including relevant links on page l in order to encourage viewers to visit page h as well, and by so, increase the viewers' traffic to each page as well as the ads' click chance. For tractability, we assume that viewers do not return to a page that they have already visited once (no feedback). The next proposition presents the overall arrival process of viewers type v_j at page l (from outside and from other pages).

Proposition 10 *In the publisher's system with no viewers' feedback, the arrivals of viewers type v_j at page l, from outside and from other pages, follow a Poisson process with overall rate*

$$\widehat{\mu}_{v_j}^{l} = \mu_{v_j}^{l} + \sum_{h \neq l} \mu_{v_j}^{h} \alpha_{v_j}^{hl}. \tag{9}$$

The proof follows from the fact that viewers' arrivals and the time they spend on each page are both exponential. Thus, the publisher's website, from arriving viewers' arrival and service perspective, is an Open Jackson Network with no feedback in which the indicated properties hold (see, e.g., Gross and Harris 1998).

We denote the price-demand function for subsystem (m^l, v_j) by $p_{v_j}^{m^l}(\lambda_{v_j}^{m^l}, x_{v_j}^{m^l}, S_{v_j}^{m^l})$, which is decreasing with respect to $\lambda_{v_j}^{m^l}$, $x_{v_j}^{m^l}$, and $S_{v_j}^{m^l}$. Let $\mathbb{P}_i^{m^l, v_j}$, $i \in \{0, ..., S_{v_j}^{m^l}\}$, $v_j \in \mathcal{V}$, be the probability of having i advertisers in subsystem (m^l, v_j). The optimization problem for L pages and the different types of ads on each page is:

$$\begin{aligned} \max_{\forall \lambda_{\kappa_h}^{m^l} \geq 0,\ \kappa_h \in \mathcal{K}} R &= \sum_{l=1}^{L} \sum_{m=1}^{M^l} \sum_{j=1}^{|\mathcal{V}|} \lambda_{v_j}^{m^l} (1 - \mathbb{P}_{S_{v_j}^{m^l}}^{m^l, v_j} (\lambda_{v_j}^{m^l}; x_{v_j}^{m^l}, S_{v_j}^{m^l}, \widehat{\mu}_{v_j}^{l})) p_{v_j}^{m^l} (\lambda_{v_j}^{m^l}, x_{v_j}^{m^l}, S_{v_j}^{m^l}) x_{v_j}^{m^l} \\ \lambda_{v_j}^{m^l} &= \sum_{\kappa_h \in K(v_j)} \frac{\widehat{\mu}_{v_j}^{l}}{\widehat{\mu}_{\kappa_h}^{l}} \lambda_{\kappa_h}^{m^l},\ v_j \in \mathcal{V}, \end{aligned} \tag{10}$$

in which $\widehat{\mu}_{\kappa_h}^{l} = \sum_{v_{j'} \in V(\kappa_h)} \widehat{\mu}_{v_{j'}}^{l}$ and $\mathbb{P}_{S_{v_j}^{m^l}}^{m^l, v_j}(\lambda_{v_j}^{m^l}; x_{v_j}^{m^l}, S_{v_j}^{m^l}, \widehat{\mu}_{v_j}^{l})$ is the steady-state full-state probability of the subsystem (m^l, v_j), which is given by

$$\mathbb{P}_{S_{v_j}^{m^l}}^{m^l, v_j}(\lambda_{v_j}^{m^l}; x_{v_j}^{m^l}, S_{v_j}^{m^l}, \widehat{\mu}_{v_j}^{l}) = \frac{\binom{x_{v_j}^{m^l} + S_{v_j}^{m^l} - 1}{S_{v_j}^{m^l}} \left(\frac{\widehat{r}_{v_j}^{m^l}}{1 + \widehat{r}_{v_j}^{m^l}} \right)^{S_{v_j}^{m^l}} \left(\frac{1}{1 + \widehat{r}_{v_j}^{m^l}} \right)^{x_{v_j}^{m^l} - 1}}{\sum_{j=0}^{S_{v_j}^{m^l}} \binom{x_{v_j}^{m^l} + S_{v_j}^{m^l} - 1}{j} \left(\frac{\widehat{r}_{v_j}^{m^l}}{1 + \widehat{r}_{v_j}^{m^l}} \right)^{j} \left(\frac{1}{1 + \widehat{r}_{v_j}^{m^l}} \right)^{x_{v_j}^{m^l} - 1 + S_{v_j}^{m^l} - j}}, \quad \widehat{r}_{v_j}^{m^l} = \frac{\lambda_{v_j}^{m^l}}{\widehat{\mu}_{v_j}^{l}} \frac{S_{v_j}^{m^l}}{n_{v_j}^{m^l}}.$$

5.2. Non-Poisson Arrivals

In Section 3 we assumed that the advertisers' arrivals at the web publisher from the ad network follow a Poisson process, which might not be the case in reality. In addition, the viewers' arrival process might not be Poisson either while the number of requested impressions is not necessarily

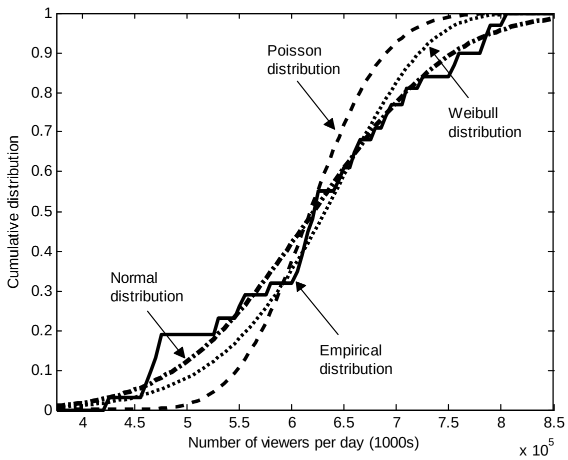

Figure 4 The empirical distribution of the viewers' arrivals, and other fitted distributions.

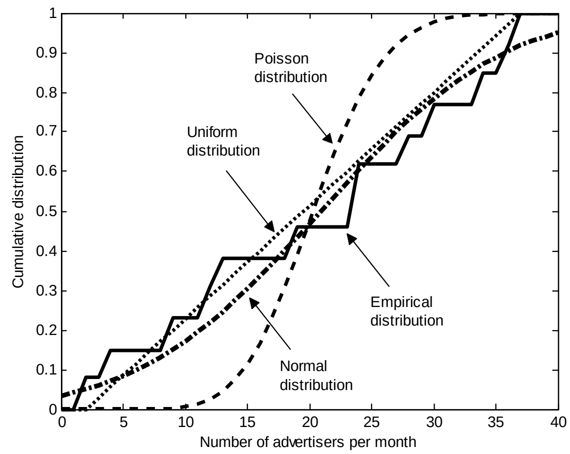

Figure 5 The empirical distribution of the advertisers' arrivals, and other fitted distributions.

the same for all advertisers. Figures 4 and 5 show an example of the empirical distributions, based on real data from Aller Internett for the advertisers' and viewers' arrivals as well as other fitted distributions. Our analysis indicates that for the arrival distribution of the viewers, the Poisson, Weibull, and Normal distributions pass the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests at the 5% significance level. However, for the advertisers' arrival distribution only the Uniform and Normal distributions pass the tests, not the Poisson. In this section, we explore other distributions for both the demand and supply sides while considering the advertisers requesting random numbers of impressions.

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 4 (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 across advertisers and both advertisers' and viewers' arrival processes are non-Poisson. For illustrative purposes, we focus only on one subsystem and drop the indices.

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 $1/(\sqrt{2}\lambda)$, uniform with the two parameters 0 and $2/\lambda$, and finally exponential with rate λ. For the viewers' inter-arrival time distributions, we consider the same distributions with λ replaced with μ. The selection of the distributions is mainly inspired by the data belonging to Aller Internett as in Figures 4 and 5.

The number of slots is set to be $n = 4$, and there is no ad rotation. We choose the pricing function to be $p(\lambda) = 0.02 - 0.2\lambda^{0.8} - 10^{-7}X$ where X is a random variable following the normal

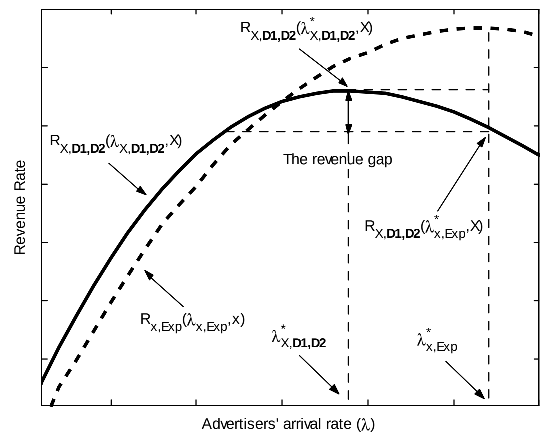

Figure 6 A schematic illustration of the revenue gap with non-Poisson arrivals and random impressions request.

distribution with mean $\mathbb{E}(X) = 1000$ and standard deviation 500. The steps of each simulation process are as follows. Using simulation, 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. We represent the revenue related to $\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2}$ with $R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2})$. Then, we obtain the optimal value for λ using the closed-form solution provided in (3) with Poisson arrivals of advertisers and viewers, where the number of impressions is the same for all advertisers, i.e., $x = \mathbb{E}(X) = 1000$. 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 impressions its "real" revenue would become $R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{x,Exp})$ (see Figure 6). Finally, we obtain 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 1 shows the relative revenue performance gaps for the different interarrival time distributions considered for advertisers' and viewers' arrivals as well as the random number of requested impressions X that results in generating instantaneously adjusted price for each impression request. We observe that the computed revenue gaps are between $0.06\% - 0.95\%$. This suggests that the Poisson policy, while considering the expected value of the number of requested impressions for all advertisers' requests, tends to be an accurate estimate for the publisher's model when both the viewers' and the advertisers' arrivals are non-Poisson and the price is adjusted based on each advertiser's requested impressions[10].

[10] In addition, in the Electronic Companion, we examine the model's performance considering other distributions for

Interarrival dist.	*Advertisers* ($\mathbf{D}_1$)			
Viewers($\mathbf{D}_2$)	*Erlang-2*	*Normal*	*Uniform*	*Exponential*
Erlang-2	0.22%	0.40%	0.45%	0.26%
Normal	0.67%	0.13%	0.95%	0.15%
Uniform	0.06%	0.10%	0.26%	0.11%
Exponential	0.16%	0.21%	0.85%	–

Table 1 The relative gap $\frac{R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2})-R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{x,Exp})}{R_{X,\mathbf{D}_1,\mathbf{D}_2}(\lambda^*_{X,\mathbf{D}_1,\mathbf{D}_2})}\times 100(\%)$

We note that the revenue depends on the full-state probability, $\mathbb{P}_n$, a small revenue gap means that $\mathbb{P}_n$ may not be sensitive to the forms of $\mathbf{D}_1$ and $\mathbf{D}_2$. One possible explanation for the $\mathbb{P}_n$'s robustness is that its structure is somewhat similar to Erlang loss formula, $\mathbb{P}_n^{Er} = \frac{(rx)^n}{n!} / \sum_{j=0}^{n} \frac{(rx)^j}{j!}$. In fact, when $x \to \infty$ or when $n \to \infty$, both formulas converge together as both approach either 1 or 0. In addition, Erlang has this property that its structure is independent of the form of underlying service distribution, $\mathbf{D}_2$. Nevertheless, we note that while this structural similarity is helpful in explaining $\mathbb{P}_n$'s possible insensitivity to $\mathbf{D}_2$, it cannot not explain the reason for its insensitivity to both $\mathbf{D}_1$ and $\mathbf{D}_2$.

Providing Bounds for the Publisher's Revenue When X is a random variable, the publisher's revenue function is not analytically tractable. Thus, one way to proceed is to develop bounds for such a function. Bounds are useful in that they provide a worst-case or a best-case revenue scenario given the model parameters. In this section, we develop an analytical bound for the publisher's revenue when X is a random variable. In order to start, we first notice that the publisher's expected revenue rate with the random variable X is:

$$R(\lambda; X) = \lambda(1 - \mathbb{P}_S(\lambda; X, n, \mu))p(\lambda, X, n)X. \tag{11}$$

Note that $R(\lambda; X)$ is a *random variable* because it is a function of X. In other words, $R(\lambda; X)$ has a probability distribution and not a fixed outcome. Thus, we can only give a bound for the *probability* that, for any observed $X, R(\lambda; X)$ is *sufficiently close* to the deterministic revenue used in Section 3 (i.e., $R(\lambda; x)$, $x = \mathbb{E}(X)$). The next proposition provides this bound.

Proposition 11 *Let $R(\lambda^*; x^*)$ be the maximum of $R(\lambda, x)$ when optimized over both λ and x. Then*

$$\Pr(|R(\lambda, X) - R(\lambda, x)| \geq \epsilon) \leq 2\exp(-\frac{2\epsilon^2}{R(\lambda^*, x^*)^2}).$$

the impressions' request, X (together with various distributions for the advertisers and viewers' arrivals) and show that the optimal revenue tends to be affected minimally.

The proof relies on two well-known theorems in probability theory. The first is McDiarmid's Inequality (McDiarmid (1989)), which says that if X is a random variable and $\psi(\cdot)$ is a function that satisfies $\sup_{x_0}|\psi(x_0)-\psi(\widehat{x})|\leq c$ (i.e., replacing the optimal value x_0 by some other value $\widehat{x}$ changes the value of ψ by at most c) then $\Pr(|\psi(X)-\mathbb{E}(\psi(X))|\geq\epsilon)\leq 2\exp(-2\epsilon^2c^{-2})$. The second is Jensen Inequality, which says that if X a random variable and $\psi(\cdot)$ a concave function, then $\mathbb{E}(\psi(X))\leq\psi(\mathbb{E}(X))$. Clearly, $\psi(X):=R(\lambda,X)$ satisfies McDiarmid's Theorem because:

$$\sup_{x_0}|R(\lambda,x_0)-R(\lambda,\widehat{x})|\leq\sup_{x_0,\lambda}\left|R(\lambda,x_0)-R(\widehat{\lambda},\widehat{x})\right|=\max_{x_0,\lambda}R(\lambda,x_0)=R(\lambda^*,x^*).$$

Thus, we have $\Pr(|R(\lambda,X)-\mathbb{E}(R(\lambda,X))|\geq\epsilon)\leq 2\exp(-\frac{2\epsilon^2}{R(\lambda^*,x^*)^2})$. In addition, by (7), $R(\lambda,x_0)$ is concave with respect to x_0 and λ. Thus, by Jensen's Inequality, $\mathbb{E}(R(\lambda,X))\leq R(\lambda,\mathbb{E}(X))=R(\lambda,x)$. Therefore, $\Pr(|R(\lambda,X)-R(\lambda,x)|\geq\epsilon)\leq\Pr(|R(\lambda,X)-\mathbb{E}(R(\lambda,X))|\geq\epsilon)$, and the result follows. For example, with the numerical information in this section (i.e., $p(\lambda)=0.02-0.2\lambda^{0.8}-10^{-7}X$, $\mu=1,\mathbb{E}(X)=x=1000$, and $n=4$), we have $R(\lambda^*,x^*)=0.066$. Thus, $\Pr(|R(\lambda,X)-R(\lambda,x)|\geq\epsilon=0.1)\leq 0.02$. ■

5.3. Fixed Advertising Campaign Length

Some ad networks allow the advertisers to request a certain advertising campaign length instead of delivering a certain 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 random impressions request (RIR) system analyzed above. The reason is that the number of impressions received by each advertiser with ad campaign κ_h during a horizon T_{κ_h}, is a random variable $X_{\kappa_h}^{T_{\kappa_h}}\sim Poisson(\mu_{\kappa_h}T_{\kappa_h})$, where $\mu_{\kappa_h}=\sum\mu_{v_j},v_j\in V(\kappa_h)$, is the overall viewers' arrival rate for an ad with campaign κ_h, 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 $\mathbb{E}(X_{\kappa_h}^{T_{\kappa_h}})=\mu_{\kappa_h}T_{\kappa_h}$ in our base model.

In addition, we can extend the fixed campaign length system to incorporate not a single horizon T_{κ_h} but multiple horizon values that the advertisers can choose from. We define the choice set as $\Omega_{\kappa_h}=\left\{T_{\kappa_h}^1,...,T_{\kappa_h}^m\right\}$. We can argue that this system is equivalent to the RIR system. We let $\tau_{\kappa_h}^i\in[0,1]$ be the percentage of the advertisers preferring to stay in the system for $T_{\kappa_h}^i\in\Omega_{\kappa_h}$ time units. Since the viewers' interarrival times are exponential each advertiser choosing $T_{\kappa_h}^i$ is served with $X_{\kappa_h}^{T_i}\sim Poisson(\mu_{\kappa_h}T_{\kappa_h}^i)$ impressions. This system of multiple campaign lengths can be approximated with a deterministic request system with $x_{\kappa_h}=\sum_{i=1}^{m}\tau_{\kappa_h}^i\mu_{\kappa_h}T_{\kappa_h}^i$ impressions. The continuous version of the multiple campaign lengths system, in which the service time T_{κ_h} is a continuous random variable can be approximated by a deterministic request system with $x_{\kappa_h} = \mathbb{E}(X_{\kappa_h}^{T_{\kappa_h}}) = \int_0^\infty \mu_{\kappa_h} t_{\kappa_h} h_{\kappa_h}(t_{\kappa_h}) dt_{\kappa_h}$ impressions where T_{κ_h} follows the probability density function $h_{\kappa_h}(T_{\kappa_h})$.

6. Conclusion

Optimal pricing of display ads while considering the uncertainties involved in demand from advertisers and supply of visits from viewers has received minimal attention in the marketing and operations research literature. This paper attempts to bridge this gap. We consider a revenue optimization model for a web publisher selling its advertising space through an ad network. The web publisher generates revenue by displaying ads on its website and charges according to the CPM pricing scheme, which it needs to optimize. We model the web publisher's operation with a queuing system, where the arrival process corresponds to the advertisers sent by the ad network and the service process corresponds to viewers visiting the website. Given the fact that all advertisers on display pay once a viewer uploads the webpage, the advertisers whose ads are displayed are served in a "synchronized" manner. We derive a closed-form solution for 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 them in detail.

On the managerial side, we show that the optimal price to charge per impression may increase 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 were pleased to offer a theoretical explanation that they were seeking. In addition, we provide further insights by showing that the optimal CPM price may increase in the number of rotating ads on the page. This behavior may not seem intuitive compared to our common intuition from the supply-demand relationship, since an increase in the number of advertising spaces can be interpreted as an increase in the system's service capacity. In addition, with random ad rotation, we show that the optimal CPM price may increase as the publisher adds a service capacity to serve more ads.

The framework for the web publisher's operations can be extended in several directions. First, even though cost-per-impression is a common pricing scheme, others exist such as cost-per-click or even a mix of the two. Second, in this paper, we have not considered the competition between publishers. Exploring targeted pricing with competing publishers would be an interesting direction. Finally, studying how the publisher's optimal CPM pricing may change when advertisers or publishers become risk-averse would be an interesting direction.

We hope that the modeling approach in this paper can serve as a basis for many promising research directions beyond this work, and in doing so, stimulate future research on online advertising in operations research and marketing.

References

Adida, E., V. DeMiguel. 2011. Supply chain competition with multiple manufacturers and retailers. *Oper. Res.* **1** 156–172.

Akella, R, A Broder, V Josifovski. 2009. Introduction to computational advertising. Course Lectures, University of California, Santa Cruz.

Alaei, S., E. Arcaute, S. Khuller, W. Ma, A. Malekian, J. Tomlin. 2009. Online allocation of display advertisements subject to advanced sales contracts. *Proceeding, ADKDD '09*. ACM, 69–77.

Anandan, R. 2012. 3 predictions for display ads this year. *Business Standard* .

Araman, V. F., K. Fridgeirsdottir. 2011. A uniform allocation mechanism and cost-per-impression pricing for online advertising. Working Paper.

Araman, V. F., I. Popescu. 2010. Media revenue management with audience uncertainty: Balancing upfront and spot market sales. *Manufacturing Service Oper. Management* **12**(2) 190–212.

Arpita, G., P. McAfee, K. Papineni, S. Vassilvitskii. 2009. Bidding for representative allocations for display advertising. *Internet and Network Economics*. 208–219.

Balseiro, S., O. Besbes, G.Y. Weintraub. 2012. Auctions for online display advertising exchanges: Approximations and design. Columbia Business School Working Paper.

Balseiro, S., J. Feldman, V. Mirrokni, S. Muthukrishnan. 2014. Yield optimization of display advertising with ad exchange. Forthcoming.

Balseiro, S.R., O. Besbes, G.Y. Weintraub. 2013. Auctions for online display advertising exchanges: approximations and design. Tech. rep., Working Paper.

Baye, M. R., J. Morgan. 2000. A simple model of advertising and subscription fees. *Economics Letters* **69**(3) 345–351.

Bharadwaj, V., W. Ma, M. Schwarz, J. Shanmugasundaram, E. Vee, J. Xie, J. Yang. 2010. Pricing guaranteed contracts in online display advertising. *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 399–408.

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, Springer, NY.

Celis, L.E., G. Lewis, M.M. Mobius, H. Nazerzadeh. 2012. Buy-it-now or take-a-chance: Price discrimination through randomized auctions. National Bureau of Economic Research.

Chakrabarti, D., E. Vee. 2012. Traffic shaping to optimize ad delivery. *Proceedings of the 13th ACM Conference on Electronic Commerce*. ACM, 302–318.

Chen, Y.J. 2011. Optimal dynamic auctions for display advertising. Tech. rep., UC Berkeley Working Paper.

Chickering, D. M., D. Heckerman. 2003. Targeted advertising on the web with inventory management interfaces. *Interfaces* **33** 71–77.

Ciocan, D.F., V. Farias. 2012. Model predictive control for dynamic resource allocation. *Math of Oper. Res.* **37**(3).

Danaher, P. J. 2007. Modeling page views across multiple websites with an application to internet reach and frequency prediction. *Marketing Sci.* **26**(3) 422–437.

DCP. 2010. Double-click for publishers: Ad serving and management software developed by google. www.google.com/admanager/.

Edelman, B., M. Ostrovsky, M Schwarz. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American Economic Review* **97**(1) 242–259.

Fjell, K. 2009. Online advertising: Pay-per-view versus pay-per-click - a comment. *Journal of Revenue and Pricing Management* **8**(2/3) 200–206.

Fredricksen, C. 2011. Online advertising market poised to grow 20http://www.emarketer.com/.

Gross, D., C.M. Harris. 1998. *Fundamentals of Queueing Theory*. Wiley, New York, NY.

IAB. 2011. Internet advertising revenue report: 2011 second quarter and first six months results www.iab.net.

J. Yang, Jian, E. Vee, S. Vassilvitskii, J. Tomlin, J. Shanmugasundaraml, T. Anastasakos, O. Kennedy. 2010. Inventory allocation for online graphical display advertising. *Yahoo Technical Report* .

Kumar, S., V. S. Jacob, C. Sriskandarajah. 2006. Scheduling advertisements on a web page to maximize revenue. *Eur. J. Oper. Res.* **173**(3) 1067–1089.

Kumar, S., S. P. Sethi. 2009. Dynamic pricing and advertising for web content providers. *Eur. J. Oper. Res.* **197**(3) 924–944.

Lewis, R.A., D.H. Reiley. 2011. Does retail advertising work? measuring the effects of advertising on sales via a controlled experiment on yahoo! Yahoo Research Working Paper.

Mangàni, A. 2003. Online advertising: Pay-per-view versus pay-per-click. *Journal of Revenue & Pricing Management* **2** 295–302.

McAfee, P., P. Papineni, S. Vassilvitskii. 2010. Maximally representative allocations for guaranteed delivery advertising campaigns. Yahoo Research Working Paper.

McDiarmid, C. 1989. On the method of bounded differences. *Surveys in combinatorics* **141**(1).

Mookerjee, R., S. Kumar, V.S. Mookerjee. 2012. To show or not show: Using user profiling to manage internet advertisement campaigns at chitika. *Interfaces* **42**(5) 449–464.

Muthukrishnan, S. 2009. Ad exchanges: Research issues. Stefano Leonardi, ed., *Internet and Network Economics*, *Lecture Notes in Computer Science*, vol. 5929. Springer Berlin / Heidelberg, 1–12.

Najafi-Asadolahi, Sami, Kristin Fridgeirsdottir. 2014. Cost-per-click pricing for display advertising. *Manufacturing Service Oper. Management* **16**(4) 482–497.

Novak, T.P., D.L. Hoffman. 2000. Advertising and pricing models for the web. D. Hurley, B. Kahin, H. Varian, eds., *Internet Publishing and Beyond*. MIT Press, Cambridge, MA.

Peterson, T. 2011. Display ad spending to hit 200b dollars, says google vp. Available at http://www.dmnews.com/display-ad-spending-to-hit-200b-says-google-vp/article/204870/.

Prasad, A., V. Mahajan, B. Bronnenbert. 2003. Advertising versus pay-per-view in electronic media. *International Journal of Research in Marketing* **20** 13–30.

Savin, S. V., M. A. Cohen, N. Gans, Z. Katalan. 2005. Capacity management in rental businesses with two customer bases. *Oper. Res.* **53** 617–631.

Talluri, K., G. van Ryzin. 2004. *The theory and practice of revenue management*. Kluwer Academic Press.

Turner, J., A. Scheller-Wolf, S. Tayur. 2011. Scheduling of dynamic in-game advertising. *Oper. Res.* **59** 1–16.

Van Mieghem, J.A. 2000. Price and service discrimination in queuing systems: Incentive compatibility of gc scheduling. *Management Sci.* **46** 1249–1267.

Appendix. Proofs

Proof of Proposition 3 We consider a Markov chain where the state of the system is defined to be the vector $\mathbf{k} \triangleq (k_1, k_2, ..., k_n) = \sum_{j=1}^{n} k_j \mathbf{e}_j^T \in (\mathbb{N} \cup \{0\})^n$ where $\mathbf{e}_j^T$ is the jth unit vector and k_j is the number of remaining impressions in a slot. We define the sets $\mathcal{G}_h(\mathbf{k}) \triangleq \{j \mid \langle \mathbf{k}, \mathbf{e}_j \rangle = h\}$ and $\mathcal{G}_{>0}(\mathbf{k}) \triangleq \{j \mid \langle \mathbf{k}, \mathbf{e}_j \rangle > 0\}$ in which $\langle \mathbf{k}, \mathbf{e}_j \rangle$ is the inner product of the two vectors $\mathbf{k}$ and $\mathbf{e}_j$, the jth unit vector. In addition, we define $|\mathcal{G}_h(\mathbf{k})|$ and $|\mathcal{G}_{>0}(\mathbf{k})|$ to be the sizes of the sets $\mathcal{G}_h(\mathbf{k})$ and $\mathcal{G}_{>0}(\mathbf{k})$, respectively (i.e., the number slots with h, and with positive remaining impressions, respectively). To prove the proposition, we first show that the solution is of the structural form

$$\pi_{\mathbf{k}} = Ar^{|\mathcal{G}_{>0}(\mathbf{k})|}(1+r)^{n-|\mathcal{G}_{>0}(\mathbf{k})|-1}\mathbf{1}_{|\mathcal{G}_{>0}(\mathbf{k})|<n} + Ar^n \mathbf{1}_{|\mathcal{G}_{>0}(\mathbf{k})|=n}, \tag{A.1}$$

Next, by summing over the relevant $\pi's$, we show that the steady-state probability that there are i advertisers in the system is of the form:

$$\mathbb{P}_i = \binom{x+i-1}{i} Ar^i (1+r)^{n-i-1}\mathbf{1}_{i<n} + \binom{x+n-1}{n} Ar^n \mathbf{1}_{i=n}, \qquad i = 0, 1, 2, ..., n. \tag{A.2}$$

Finally, using the fact that $\sum_{i=0}^{n} \mathbb{P}_i = 1$, we obtain the coefficient A and show that the closed-form result holds as introduced in the proposition.

In order to show $\pi_{\mathbf{k}}$ has the structural form in (A.1), we need to identify all possible states of the system and obtain the flow balance equation for every state. We note that each transition equation is a complex multidimensional difference equation for which there is no standard mathematical approach to solve, while in order to find the steady-state probabilities, we need to consider and solve all transition equations in a single system. Therefore, we use the verification approach, which in this problem is identical to mathematical induction, to show that the closed-form results hold. The CPM system has in general 5 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 $(\mathbf{k}+x\mathbf{e}_1^T)$ with rate λ, or it can come from any of the states $\mathbf{v}_i^T \triangleq \sum_{j=1}^{i} \mathbf{e}_j^T, 1\le i\le n$, with rate μ. As a result, the flow balance equation becomes: $r\pi_{\mathbf{0}} = \sum_{i=1}^{n} \pi_{\mathbf{v}_i^T}$ where $r=\lambda/\mu$.

ii) Define $\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|} \triangleq \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} \mathbf{e}_j^T$ and $\mathbf{w}_q^T \triangleq \sum_{j=|\mathcal{G}_{>0}(\mathbf{k})|+1}^{q} \mathbf{e}_j^T$, $|\mathcal{G}_{>0}(\mathbf{k})|+1\le q\le n$. If $\mathbf{k}=\sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j\mathbf{e}_j^T$, $k_j>0$ with $|\mathcal{G}_{>0}(\mathbf{k})|<n$ and $|\mathcal{G}_x(\mathbf{k})|=0$ then $\mathbf{k}$ can either go to the state $\mathbf{k}-\ \mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|}$ with rate μ or to the state $\mathbf{k}+x\mathbf{e}^T_{|\mathcal{G}_{>0}(\mathbf{k})|+1}$ with rate λ. It can also comes from either $\mathbf{k}+\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|}$ or any of the states $\mathbf{k}+\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|}+\mathbf{w}_q^T$, $|\mathcal{G}_{>0}(\mathbf{k})|+1\le q\le n$ with rate μ. Hence, the balance equation would become

$$(1+r)\pi_{\mathbf{k}} = \pi_{\mathbf{k}+\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|}} + \sum_{q=|\mathcal{G}_{>0}(\mathbf{k})|+1}^{n} \pi_{\mathbf{k}+\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|}+\mathbf{w}_q^T}, \tag{A.3}$$

For example, if $\mathbf{k}=(4,0,0)$ and $x=5$ we have $\mathcal{G}_{>0}(\mathbf{k})=\{1\}$, $\mathcal{G}_0(\mathbf{k})=\{2,3\}$, $|\mathcal{G}_{>0}(\mathbf{k})|=1$, $\mathbf{v}_1^T=(1,0,0)$, $\mathbf{w}_2^T=(0,1,0)$, and $\mathbf{w}_3^T=(0,1,1)$. 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 $\mathbf{k}=\sum_{j=1}^{|\mathcal{G}_x(\mathbf{k})|} x\mathbf{e}_j^T + \sum_{j=|\mathcal{G}_x(\mathbf{k})|+1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j\mathbf{e}_j^T$ 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 $(1+r)\pi_{\mathbf{k}} = r\pi_{(\mathbf{k}-x\mathbf{e}_1^T)^T\mathbf{D}_{n\times n}}$, where $\mathbf{D}_{n\times n}\triangleq[\mathbf{e}_2 \vdots \mathbf{e}_3 \vdots ... \vdots \mathbf{e}_{n-1} \vdots \mathbf{0}_{1\times n}]_{n\times n}$. For example, if we take $x=5$ and $n=4$ and $\mathbf{k}=(5,5,4,0)$ then $\mathbf{k}-x\mathbf{e}_1^T=(0,5,4,0)$ 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, $(\mathbf{k}-x\mathbf{e}_1^T)^T\mathbf{D}_{n\times n}=(5,4,0,0)$. Therefore, the flow balance equation becomes: $(1+r)\pi_{(5,5,4,0)}=r\pi_{(5,4,0,0)}$.

iv) If $\mathbf{k}=\sum_{j=1}^{|\mathcal{G}_x(\mathbf{k})|} x\mathbf{e}_j^T + \sum_{j=|\mathcal{G}_x(\mathbf{k})|+1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j\mathbf{e}_j^T$, 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:

$$\pi_{\mathbf{k}} = r\pi_{(\mathbf{k}-x\mathbf{e}_1^T)^T\mathbf{D}_{n\times n}}, \ \textit{where}\ \mathbf{D}_{n\times n}=[\mathbf{e}_2 \vdots \mathbf{e}_3 \vdots ... \vdots \mathbf{e}_{n-1} \vdots \mathbf{0}_{1\times n}]_{n\times n}. \tag{A.4}$$

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, we have: $\pi_{(5,5,4,3)}=r\pi_{(5,4,3,0)}$.

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:

$$\pi_{\mathbf{k}} = \pi_{\mathbf{k}+\mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|}}, \text{ where } \mathbf{v}^T_{|\mathcal{G}_{>0}(\mathbf{k})|} = \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} \mathbf{e}_j^T. \tag{A.5}$$

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 Equation (A.1) satisfies the Flow Balance Equations (i) - (v). Due to space limitation, here we only verify the solution for item (i) as the rest items are verified similarly.

For item (i) by inserting Equation (A.1) into the flow balance equation we obtain a left hand side of $r\pi_{\mathbf{0}} = Ar(1+r)^{n-1}$ and a right hand side of $\sum_{j=1}^{n} \pi_{\mathbf{v}_j^T} = A(\sum_{j=1}^{n-1} r^j(1+r)^{n-j-1} + r^n)$. We use induction to show that both sides are equal. 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 $\sum_{j=1}^{k} r^j(1+r)^{k-j} + r^{k+1} = r(1+r)^k$. It is easy to see that $\sum_{j=1}^{k} r^j(1+r)^{k-j} + r^{k+1}$ is equal to $(1+r)\sum_{j=1}^{k-1} r^j(1+r)^{k-j-1} + r^k + r^{k+1}$. Using the induction assumption we obtain $(1+r)[r(1+r)^{k-1} - r^k] + r^k + r^{k+1}$ that simplifies to $r(1+r)^k$, which completes the induction proof.

In order to obtain A, we note that $\mathbb{P}_0 = \pi_{\mathbf{0}} = A(1+r)^{n-1}$. Let us then consider the state of having i $(1 \le i \le n)$ advertisers in the publisher's system where each advertiser has k_j impressions left to satisfy with $\mathbf{k} = \sum_{j=1}^{|\mathcal{G}_{>0}(\mathbf{k})|} k_j \mathbf{e}_j^T,\ |\mathcal{G}_{>0}(\mathbf{k})| = i \le n,\ |\mathcal{G}_x(\mathbf{k})| = 0$. Without loss of generality, let k_j be increasing in j, i.e., $k_1 \le k_2 \le ... \le k_i$. It is easy to see that $\mathbb{P}_i = \sum_{k_1=1}^{x} \sum_{k_2=k_1}^{x} ... \sum_{k_i=k_{i-1}}^{x} \pi_{\mathbf{k}}$, where by Lemma 1 $\mathbb{P}_i$ reduces to $\mathbb{P}_i = \binom{x+i-1}{i}\pi_{\mathbf{k}}, i \le n$. Moreover, since $\sum_{i=0}^{n} \mathbb{P}_i = 1$, using (A.1), we have $\sum_{i=0}^{n} \binom{x+i-1}{i}\pi_{\mathbf{k}} = 1$, which gives $A = \left(\sum_{j=0}^{n-1} \binom{x+j-1}{j} r^j(1+r)^{n-j-1} + \binom{x+n-1}{n} r^n\right)^{-1}$. Finally, using Lemma 3, A reduces to $A = \left(\sum_{j=0}^{n} \binom{x+n-1}{j} r^j\right)^{-1}$. Dividing the numerator and denominator of $\mathbb{P}_i$ by $(1+r)^{x+n-1}$, gives

$$\mathbb{P}_i = \frac{\binom{x+i-1}{i}\left(\frac{r}{1+r}\right)^i \left(\frac{1}{1+r}\right)^x \mathbf{1}_{i<n} + \binom{x+n-1}{n}\left(\frac{r}{1+r}\right)^n \left(\frac{1}{1+r}\right)^{x-1} \mathbf{1}_{i=n}}{\sum_{j=0}^{n} \binom{x+n-1}{j}\left(\frac{r}{1+r}\right)^j \left(\frac{1}{1+r}\right)^{x+n-1-j}}. \tag{A.6}$$

Hence, the proof is complete. ■

Proof of Proposition 4 (i) In order to show $\mathbb{P}_S$ is increasing in r we show its derivative with respect to r is always positive. By differentiating $\mathbb{P}_S$ with respect to r and simplifying we get:

$$\frac{\partial \mathbb{P}_S}{\partial r} = \frac{\partial \mathbb{P}_S}{\partial \overline{r}} \frac{\partial \overline{r}}{\partial r} = (\frac{S}{n}) \frac{\sum_{i=0}^{S} \binom{x+S-1}{S}\binom{x+S-1}{i} \overline{r}^{S+i-1}(S-i)}{\sum_{i=0}^{S} \binom{x+S-1}{i}\overline{r}^i} \ge 0, \qquad \overline{r} = \frac{S}{n} r \tag{A.7}$$

which is always positive. Hence, $\mathbb{P}_S$ is increasing in both r and $\overline{r}$.

(ii) After some calculations we get

$$\mathbb{P}_S(x+1) - \mathbb{P}_S(x) = \frac{\overline{r}^S \sum_{i=0}^{S} \overline{r}^i \binom{x+S-1}{S} \binom{x+S}{i} \binom{S-i}{x}}{\sum_{j=0}^{S} \binom{x+S-1}{j} \overline{r}^j \sum_{k=0}^{S} \binom{x+S}{k} \overline{r}^k} \geq 0.$$

Hence, $\mathbb{P}_S$ is increasing in x.

(iii) We prove $\mathbb{P}_{S+1} \leq \mathbb{P}_S$ using contradiction. Let us assume $\mathbb{P}_{S+1} > \mathbb{P}_S$, which identically can be expressed as

$$\overline{r}_S \frac{x+S}{S+1} \sum_{i=0}^{S} \frac{\overline{r}_S^i (x+S-1)!}{i!(x+S-1-i)!} > \sum_{i=0}^{S+1} \frac{\overline{r}_{S+1}^i (x+S)!}{i!(x+S-i)!}, \tag{A.8}$$

where $\overline{r}_S = \frac{S}{n} r$. Reindexing the sum on the right hand side by setting $i = j+1$ and simplifying gives

$$\sum_{i=0}^{S} \frac{\overline{r}_S^i (x+S)!}{(S+1)i!(x+S-1-i)!} > \frac{1}{\overline{r}_S} + \sum_{j=0}^{S} \frac{\left(\frac{\overline{r}_{S+1}}{\overline{r}_S}\right) \overline{r}_{S+1}^j (x+S)!}{(j+1)!(x+S-1-j)!}. \tag{A.9}$$

By comparing the sums term by term, and noting that $\overline{r}_{S+1} > \overline{r}_S$, we see that each term on the left hand side is smaller than the corresponding one on the right hand side, which contradicts the assumption of $\mathbb{P}_{S+1} > \mathbb{P}_S$. Hence, we must have $\mathbb{P}_{S+1} \leq \mathbb{P}_S$. ■

Proof of Proposition 6 (i) We know that $R = \sum\limits_{v_j \in \mathcal{V}} R_{v_j}(\lambda_{v_j})$ where $R_{v_j}(\lambda_{v_j}) = \lambda_{v_j} x_{v_j} (1 - \mathbb{P}_{S_{v_j}}(\lambda_{v_j})) p(\lambda_{v_j})$. Thus, the FOC becomes $\frac{\partial R}{\partial \lambda_{\kappa_h}} = \sum\limits_{v_j \in \mathcal{V}} \frac{\partial R}{\partial \lambda_{v_j}} \frac{\partial \lambda_{v_j}}{\partial \lambda_{\kappa_h}} = 0$ where $\frac{\partial \lambda_{v_j}}{\partial \lambda_{\kappa_h}} = \frac{\mu_{v_j}}{\mu_{\kappa_h}} 1_{v_j \in V(\kappa_h)}$. Hence, $\frac{\partial R}{\partial \lambda_{\kappa_h}} = \sum\limits_{v_j \in V(\kappa_h)} \frac{\partial R}{\partial \lambda_{v_j}} \frac{\mu_{v_j}}{\mu_{\kappa_h}} = \mu_{\kappa_h} \sum\limits_{v_j \in V(\kappa_h)} \frac{\partial R}{\partial \lambda_{v_j}} \mu_{v_j}$. Therefore, the FOC reduces to $\sum\limits_{v_j \in V(\kappa_h)} \frac{\partial R}{\partial \lambda_{v_j}} \mu_{v_j} = 0$. (ii) Next, to show R is concave with respect to $\Lambda = (\lambda_{\kappa_1}, ..., \lambda_{\kappa_{|\mathcal{K}|}})$, as R is twice continuously differentiable, it is enough to show that:

$$H^R_{|\mathcal{K}|\times|\mathcal{K}|} = \begin{bmatrix} \frac{\partial^2 R}{\partial \lambda_{\kappa_1}^2} & \frac{\partial^2 R}{\partial \lambda_{\kappa_1} \partial \lambda_{\kappa_2}} & \cdots & \frac{\partial^2 R}{\partial \lambda_{\kappa_1} \partial \lambda_{\kappa_{|\mathcal{K}|}}} \\ \frac{\partial^2 R}{\partial \lambda_{\kappa_2} \partial \lambda_{\kappa_1}} & \frac{\partial^2 R}{\partial \lambda_{\kappa_2}^2} & \cdots & \frac{\partial^2 R}{\partial \lambda_{\kappa_2} \partial \lambda_{\kappa_{|\mathcal{K}|}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 R}{\partial \lambda_{\kappa_{|\mathcal{K}|}} \partial \lambda_{\kappa_1}} & \frac{\partial^2 R}{\partial \lambda_{\kappa_{|\mathcal{K}|}} \partial \lambda_{\kappa_2}} & \cdots & \frac{\partial^2 R}{\partial \lambda_{\kappa_{|\mathcal{K}|}}^2} \end{bmatrix}_{|\mathcal{K}|\times|\mathcal{K}|}$$

is negative semidefinite, where H_R is the Hessian matrix of R. It can be observed that H_R is a diagonal matrix because for all off-diagonal elements we have $\frac{\partial^2 R}{\partial \lambda_{\kappa_h} \partial \lambda_{\kappa_{h'}}} = 0$. Thus, to show H_R is negative semidefinite, it is enough only to show that the diagonal elements are negative, i.e., $\frac{\partial^2 R}{\partial \lambda_{\kappa_h}^2} \leq 0$, $h = 1, ..., |\mathcal{K}|$. However, taking the second derivative of R gives

$$\frac{\partial^2 R}{\partial \lambda_{\kappa_h}^2} = \sum_{v_j \in \mathcal{V}} \frac{\partial^2 R}{\partial \lambda_{v_j}^2} (\frac{\mu_{v_j}}{\mu_{\kappa_h}})^2 1_{v_j \in V(\kappa_h)} = \sum_{v_j \in V(\kappa_h)} \frac{\partial^2 R_{v_j}(\lambda_{v_j})}{\partial \lambda_{v_j}^2} (\frac{\mu_{v_j}}{\mu_{\kappa_h}})^2.$$

Thus, if $\frac{\partial^2 R_{v_j}(\lambda_{v_j})}{\partial\lambda^2_{v_j}} \le 0$ holds for all $v_j \in V(\kappa_h)$, then $\frac{\partial^2 R}{\partial\lambda^2_{\kappa_h}} \le 0$. To show $\frac{\partial^2 R_{v_j}(\lambda_{v_j})}{\partial\lambda^2_{v_j}} \le 0$, re-expressing $R_{v_j}(\lambda_{v_j})$ as

$$R_{v_j}(\lambda_{v_j}) = \mu_{v_j}\frac{n_{v_j}}{S_{v_j}} L_{v_j}(\lambda_{v_j}) p_{v_j}(\lambda_{v_j}), L_{v_j}(\lambda_{v_j}) = \widehat{r}_{v_j} x_{v_j}(1-\mathbb{P}_{S_{v_j}}),$$

where $\widehat{r}_{v_j} = \frac{\lambda_{v_j}}{\mu_{v_j}}\frac{S_{v_j}}{n_{v_j}} = r_{v_j}\frac{S_{v_j}}{n_{v_j}}$ gives

$$\frac{\partial^2 R_{v_j}(\lambda_{v_j})}{\partial\lambda^2_{v_j}} = \mu_{v_j}\frac{n_{v_j}}{S_{v_j}}\left(\frac{\partial^2 L_{v_j}(\lambda_{v_j})}{\partial\lambda^2_{v_j}} p_{v_j}(\lambda_{v_j}) + L_{v_j}(\lambda_{v_j})\frac{\partial^2 p_{v_j}(\lambda_{v_j})}{\partial\lambda^2_{v_j}} + 2\frac{\partial L_{v_j}(\lambda_{v_j})}{\partial\lambda_{v_j}}\frac{\partial p_{v_j}(\lambda_{v_j})}{\partial\lambda_{v_j}}\right). \quad \text{(A.10)}$$

Knowing that $p_{v_j}(\lambda_{v_j})$ is positive and concave decreasing and L_{v_j} is concave increasing we have that $\frac{\partial^2 R_{v_j}(\lambda_{v_j})}{\partial\lambda^2_{v_j}} \le 0$. Thus, H_R is negative semidefinite and R is concave in Λ. (iii) Showing the average number of advertisers is $L_{v_j}(\lambda_{v_j}) = r_{v_j}\frac{S_{v_j}}{n_{v_j}} x_{v_j}(1-\mathbb{P}_{S_{v_j}})$ is immediate from Little's Law. ■

Proof of Proposition 7 (i) Let $\Lambda^* = [\lambda^*_{\kappa_h}]_{1\times|\mathcal{K}|}$ be the optimal vector. First, taking the first derivative using the chain rule, we find $dR(\Lambda^*)/dS_{v_j} = \partial R(\Lambda^*)/\partial S_{v_j} + \sum_{\kappa_h\in\mathcal{K}} \partial R(\Lambda^*)/\partial\lambda_{\kappa_h} \times \partial\lambda_{\kappa_h}/\partial S_{v_j}$. However, as Λ^* satisfies $\partial R(\Lambda^*)/\partial\lambda_{\kappa_h} = 0$, we find $dR(\Lambda^*)/dS_{v_j} = \partial R(\Lambda^*)/\partial S_{v_j}$ (*Envelope Theorem)*. Second, we note that $\partial R(\Lambda^*)/\partial S_{v_j} = \partial R_{v_j}(\lambda^*_{v_j})/\partial S_{v_j}$. As $R_{v_j}(\lambda^*_{v_j}) = \mu_{v_j}(\frac{n_{v_j}}{S_{v_j}}) L_{v_j} p_{v_j}(\lambda^*_{v_j}, S_{v_j}, x_{v_j})$, we get $\partial R_{v_j}(\lambda^*_{v_j})/\partial S_{v_j} = \mu_{v_j}(\frac{n_{v_j}}{S_{v_j}}) \times \partial(L_{v_j}P_{v_j})/\partial S_{v_j} - \frac{1}{S^2_{v_j}} n_{v_j}\mu_{v_j} L_{v_j} p_{v_j}$. We show that $\partial R_{v_j}(\lambda^*_{v_j})/\partial S_{v_j} = 0$ has a single root. First, let $A_{v_j} := \partial(L_{v_j}P_{v_j})/\partial S_{v_j}$. Then from $\partial R_{v_j}(\lambda^*_{v_j})/\partial S_{v_j} = 0$, it is clear that: $\frac{1}{S^2_{v_j}}(A_{v_j}S_{v_j} - L_{v_j}P_{v_j}) = 0$. Thus, $\partial^2 R_{v_j}(\lambda^*_{v_j})/\partial S^2_{v_j} = \frac{1}{S_{v_j}}\frac{\partial A_{v_j}}{\partial S_{v_j}} - \frac{2}{S^3_{v_j}}(A_{v_j}S_{v_j} - L_{v_j}P_{v_j})$. In addition, since $\partial L_{v_j}/\partial S_{v_j} > 0$ and $\partial^2 L_{v_j}/\partial S^2_{v_j} \le 0$, $\partial p_{v_j}/\partial S_{v_j} \le 0$, $\partial^2 p_{v_j}/\partial S^2_{v_j} \le 0$, we have that $\partial A_{v_j}/\partial S_{v_j} < 0$. Thus, $\partial^2 R_{v_j}(\lambda^*_{v_j})/\partial S^2_{v_j} < 0$, i.e., $\partial R_{v_j}(\lambda^*_{v_j})/\partial S_{v_j}$ is strictly decreasing in S_{v_j}. Therefore, it crosses the zero-line at a unique point, namely, $S^*_{v_j}$. If $S^*_{v_j} \ge 0$ then for all $n_{v_j} \le S^*_{v_j}$, $R_{v_j}(\lambda^*_{v_j})$ (or $R(\Lambda^*)$) is increasing in S_{v_j} and for all $S_{v_j} \ge S^*_{v_j}$ it is decreasing in S_{v_j}. If $S^*_{v_j} < 0$ then for all $S_{v_j} \ge 0$, $R_{v_j}(\lambda^*_{v_j})$ (or $R(\Lambda^*)$) is decreasing in S_{v_j}.

(ii) Invoking the Envelope Theorem as in Part (i) gives $dR(\Lambda^*)/dx_{v_j} = \partial R(\Lambda^*)/\partial x_{v_j}$. As $\partial R(\Lambda^*)/\partial x_{v_j} = \partial R_{v_j}(\lambda^*_{v_j})/\partial x_{v_j}$, we have $dR(\Lambda^*)/dx_{v_j} = \partial R_{v_j}(\lambda^*_{v_j})/\partial x_{v_j}$. As in Part i, we can show $\partial R(\lambda^*_{v_j})/\partial x_{v_j}$ is strictly decreasing in x_{v_j}, i.e., $\partial^2 R_{v_j}(\lambda^*_{v_j})/\partial x^2_{v_j} < 0$ (because $L_{v_j}/\partial x_{v_j} > 0$ and $\partial^2 L_{v_j}/\partial x^2_{v_j} < 0$, $\partial p_{v_j}/\partial x_{v_j} \le 0$, and $\partial^2 p_{v_j}/\partial x^2_{v_j} \le 0$). Thus, $\partial R(\lambda^*_{v_j})/\partial x_{v_j}$ has a single root, namely $x^*_{v_j}$. If $x^*_{v_j} \ge 0$ then for all $x_{v_j} \le x^*_{v_j}$, $R_{v_j}(\lambda^*_{v_j})$ (or $R(\Lambda^*)$) is increasing in x_{v_j} and for all $x_{v_j} \ge x^*_{v_j}$ it is decreasing in x_{v_j}. If $x^*_{v_j} < 0$ then for all $x_{v_j} \ge 0$, $R_{v_j}(\lambda^*_{v_j})$ (or $R(\Lambda^*)$) is decreasing in x_{v_j}.

(iii) Using Envelop Theorem, we have $dR(\Lambda^*)/d\mu_{v_j} = \partial R(\Lambda^*)/\partial\mu_{v_j}$. Since $\partial R(\Lambda^*)/\partial\mu_{v_j} = \partial R_{v_j}(\lambda^*_{v_j})/\partial\mu_{v_j}$, we have $dR(\Lambda^*)/d\mu_{v_j} = \partial R_{v_j}(\lambda^*_{v_j})/\partial\mu_{v_j} = -\partial\mathbb{P}_{S_{v_j}}(\lambda^*_{v_j})/\partial\mu_{v_j}$. We note that $\mathbb{P}_{S_{v_j}}(\lambda^*_{v_j})$ depends on $\lambda^*_{v_j}$ and μ_{v_j} only through $r^*_{v_j} = \lambda^*_{v_j}/\mu_{v_j}$, not each of them separately. Thus, by the chain rule $\partial\mathbb{P}_{S_{v_j}}(\lambda^*_{v_j})/\partial\mu_{v_j} = \partial\mathbb{P}_{S_{v_j}}(\lambda^*_{v_j})/\partial r^*_{v_j} \times 1/\mu_{v_j} \ge 0$, which is positive by Part (i) of Proposition 4. Hence, $dR(\Lambda^*)/d\mu_{v_j} \le 0$ and the result follows. ■

Proof of Proposition 9 **(i)** We need to show that $\frac{\partial \lambda^*_{v_j}}{\partial S_{v_j}} \leq 0$. By Implicit Function Theorem, we have

$$\frac{\partial \lambda^*_{v_j}}{\partial S_{v_j}} = -\frac{\frac{\partial^2 R_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j} \partial S_{v_j}}}{\frac{\partial^2 R_{v_j}(\lambda^*_{v_j})}{\partial \lambda^2_{v_j}}}$$

$$= -\frac{\left\{\begin{array}{c} \mu_{v_j}\frac{n_{v_j}}{S_{v_j}}\Big(\frac{\partial^2 L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}\partial S_{v_j}} p(\lambda^*_{v_j}, x_{v_j}, S_{v_j}) + \frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial S_{v_j}} \frac{\partial p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}} \\ + L_{v_j}(\lambda^*_{v_j}) \frac{\partial^2 p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial S_{v_j} \partial \lambda_{v_j}} + \frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} \frac{\partial p(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial S_{v_j}}\Big) - \mu_{v_j}\frac{n_{v_j}}{S^2_{v_j}} L_{v_j}(\lambda^*_{v_j}) p(\lambda^*_{v_j}, x_{v_j}, S_{v_j}) \end{array}\right\}}{\frac{\partial^2 L_{v_j}(\lambda^*_{v_j})}{\partial \lambda^2_{v_j}} p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j}) + 2\frac{\partial \Gamma_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} \frac{\partial p(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}} + L_{v_j}(\lambda^*_{v_j}) \frac{\partial^2 p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda^2_{v_j}}},$$

in which by First Order Necessary Condition we have $\frac{\partial R_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} = \frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j}) + L_{v_j}(\lambda^*_{v_j}) \frac{\partial p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}} = 0$. Based on Proposition 6, it is clear that $\frac{\partial^2 R_{v_j}(\lambda^*_{v_j})}{\partial \lambda^2_{v_j}} \leq 0$. Based on the proposition assumptions and Lemma 5, it is sufficient to show that $\frac{\partial^2 L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j} \partial S_{v_j}} \leq 0$. For convenience, let $\rho_{v_j} = \frac{\lambda^*_{v_j} x_{v_j}}{\mu_{v_j}} \frac{S_{v_j}}{n_{v_j}}$. If $x_{v_j} >> S_{v_j}$ then it is easy to see that $\mathbb{P}_{S_{v_j}}$ is simplified to $\mathbb{P}_{S_{v_j}} = \frac{\rho_{v_j}^{S_{v_j}}(1-\rho_{v_j})}{1-\rho_{v_j}^{S_{v_j}+1}}$.

Thus, $L_{v_j} = \rho_{v_j}(1 - \mathbb{P}_{S_{v_j}}) = \frac{\rho_{v_j}(1-\rho_{v_j}^{S_{v_j}})}{1-\rho_{v_j}^{S_{v_j}+1}}$. Next, by the chain rule, we have $\frac{\partial L_{v_j}}{\partial \lambda_{v_j}} = \frac{\partial L_{v_j}}{\partial \rho_{v_j}} \frac{\partial \rho_{v_j}}{\partial \lambda_{v_j}} = \frac{\partial L_{v_j}}{\partial \rho_{v_j}} (\frac{S_{v_j}}{n_{v_j}})(\frac{x_{v_j}}{\mu_{v_j}})$. Thus, obtaining the second derivative gives $\frac{\partial^2 L_{v_j}}{\partial \lambda_{v_j} \partial S_{v_j}} = \frac{x_{v_j}}{n_{v_j}\mu_{v_j}} (\frac{\partial^2 L_{v_j}}{\partial \rho_{v_j} \partial S_{v_j}} S_{v_j} + \frac{\partial L_{v_j}}{\partial \rho_{v_j}})$. In order to determine $\frac{\partial^2 L_{v_j}}{\partial \lambda_{v_j} \partial S_{v_j}}$, we need to characterize $\frac{\partial L_{v_j}}{\partial \rho_{v_j}}$ and $\frac{\partial^2 L_{v_j}}{\partial \rho_{v_j} \partial S_{v_j}}$. First, from the formulation for L_{v_j}, we find $\frac{\partial L_{v_j}}{\partial \rho_{v_j}} = \frac{1-(1+S_{v_j})\rho_{v_j}^{S_{v_j}} + S_{v_j}\rho_{v_j}^{S_{v_j}+1}}{(1-\rho_{v_j}^{1+S_{v_j}})^2} \geq 0$. In addition, calculating the derivative of $\frac{\partial L_{v_j}}{\partial \rho_{v_j}}$ with respect to S_{v_j} gives:

$$\frac{\partial^2 L_{v_j}}{\partial \rho_{v_j} \partial S_{v_j}} = -\frac{\rho_{v_j}^{S_{v_j}}\left((1-\rho_{v_j})\left(1-\rho_{v_j}^{1+S_{v_j}}\right) + \left(1-\rho_{v_j}\left(2-\rho_{v_j}^{S_{v_j}}\right) + S_{v_j}(1-\rho_{v_j})\left(1+\rho_{v_j}^{1+S_{v_j}}\right)\right)\ln(\rho_{v_j})\right)}{\left(1-\rho_{v_j}^{1+S_{v_j}}\right)^3}.$$

Replacing the values of $\frac{\partial L_{v_j}}{\partial \rho_{v_j}}$ and $\frac{\partial^2 L_{v_j}}{\partial \rho_{v_j} \partial S_{v_j}}$ in $\frac{\partial^2 L_{v_j}}{\partial \lambda_{v_j} \partial S_{v_j}}$ gives:

$$\frac{\partial^2 L_{v_j}}{\partial \lambda_{v_j} \partial S_{v_j}} = \frac{x_{v_j}}{n\mu} \frac{\left\{\begin{array}{c} \left(1-\rho_{v_j}^{1+S_{v_j}}\right)(1+2S_{v_j})\rho_{v_j}^{S_{v_j}}(\rho_{v_j}-1) \\ -S_{v_j}\rho_{v_j}^{S_{v_j}}\left(1-2\rho_{v_j}+\rho_{v_j}^{1+S_{v_j}} + S_{v_j}(1-\rho_{v_j})\left(1+\rho_{v_j}^{1+S_{v_j}}\right)\right)\ln(\rho_{v_j}) \end{array}\right\}}{\left(1-\rho_{v_j}^{1+S_{v_j}}\right)^3} \tag{A.11}$$

It can be seen that $\lim_{\lambda_{v_j}\to 0} \frac{\partial^2 L_{v_j}}{\partial \lambda_{v_j} \partial S_{v_j}} = \frac{x_{v_j}}{n\mu}$. In addition for any $\rho_{v_j} \geq \exp(1)$ (i.e., $\lambda^*_{v_j} \geq \frac{\exp(1)\mu_{v_j} n_{v_j}}{x_{v_j} S_{v_j}}$) the numerator of (A.11) is always negative. Thus, $\frac{\partial^2 L_{v_j}}{\partial \lambda_{v_j} \partial S_{v_j}} \leq 0$. In addition, at $\rho_{v_j} = 1$, the numerator is zero. Therefore, by Rolle's theorem there exists some $w_{v_j} \in [\frac{\mu n}{x_{v_j} S_{v_j}}, \frac{\exp(1)\mu n}{x_{v_j} S_{v_j}}]$, such that for any

$\lambda^*_{v_j} \geq w_{v_j}$, the numerator is negative while the denominator is positive. Hence, $\frac{\partial \lambda^*_{v_j}}{\partial S_{v_j}} > 0$ in the neighborhood of $\lambda^*_{v_j}$. Note that in the worst case, if there is no interior value for w_{v_j}, w_{v_j} would merely coincide the upper-bound. That is $w_{v_j} = \frac{\exp(1)\mu n}{x_{v_j} S_{v_j}}$.

(ii) By Implicit Function Theorem:

$$\frac{dp(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{dS_{v_j}} = \underbrace{\frac{\partial p(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial S_{v_j}}}_{<0} + \underbrace{\frac{\partial p(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}}}_{<0} \frac{\partial \lambda^*_{v_j}}{\partial S_{v_j}}.$$

It is clear that when $\frac{\partial \lambda^*_{v_j}}{\partial S_{v_j}} > 0$ then $\frac{dp(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{dS_{v_j}} < 0$. That is, if the publisher decides to attract more advertisers it optimally lowers its price. But, if $\frac{\partial \lambda^*_{v_j}}{\partial S_{v_j}} < 0$ then $\frac{dp(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{dS_{v_j}}$ may become either positive or negative. That is, if the publisher sees that it needs fewer advertisers it may decide to increase or decrease the price. The increase or decrease of the price depends on whether the impact of the reduction of λ_{v_j} on the price increase is greater or the impact of the price penalty due to an added slot. ■

Electronic Companion

Cost-Per-Impression Pricing for Display Advertising

EC.1. Proofs of Propositions 5 and 8

Lemma 1 *Given any natural numbers $x \in \mathbb{N}$ and $n \in \mathbb{N}$,*

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

Proofs of all lemmas are provided in the Technical Supplement.

Lemma 2 *Let $Q(x) = Q_N(x)/Q_D(x)$, where*

$$Q_N(x) = \left(\sum_{i=0}^{S}\binom{x+S-1}{i}\binom{x+S}{S}\overline{r}^{S+i} + \sum_{i=0}^{S}\binom{x+S-1}{S}\binom{x+S}{i}\overline{r}^{S+i}(S-i)\right) \tag{EC.2}$$

and

$$Q_D(x) = \left(\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^{i}\sum_{i=0}^{S}\binom{x+S}{i}\overline{r}^{i}\right). \tag{EC.3}$$

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

Lemma 3 *Given any natural numbers $x \in \mathbb{N}$ and $S \in \mathbb{N}$,*

$$\sum_{i=0}^{S}\sum_{j=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{j}\overline{r}^{i+j} \geq \sum_{i=0}^{S}\binom{x+S-1}{S}\binom{x+S-1}{i}\overline{r}^{S+i}(S+1-i). \tag{EC.4}$$

Lemma 4 *Given any $x, S \in \mathbb{N} \cup \{0\}$, and $\overline{r} \in \mathbb{R}_+$*

$$\sum_{i=0}^{S}\sum_{j=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{j}\overline{r}^{i+j}(S+1-i)(S+i-2j) \geq 0. \tag{EC.5}$$

Proof of Proposition 5 (i) We set $L_S(x) = \overline{r}x(1-\mathbb{P}_S(x))$ where $\overline{r} = rS/n$. We need to show that $\Delta L_S(x) = L(x+1) - L(x) \geq 0$. We have $\Delta L_S(x) = \overline{r}x(\mathbb{P}_S(x) - \mathbb{P}_S(x+1)) + \overline{r}(1-\mathbb{P}_S(x+1))$. Focusing on the first term in $\Delta L_S(x)$ we get

$$x(P_S(x) - P_S(x+1)) = \frac{x\overline{r}^S\left[\binom{x+S-1}{S}\sum_{i=0}^{S}\binom{x+S}{i}\overline{r}^i - \binom{x+S}{S}\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i\right]}{\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i\sum_{i=0}^{S}\binom{x+S}{i}\overline{r}^i}. \tag{EC.6}$$

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

$$x(\mathbb{P}_S(x) - \mathbb{P}_S(x+1)) = -\frac{\overline{r}^S\binom{x+S-1}{S}\sum_{i=0}^{S}\binom{x+S}{i}\overline{r}^i(S-i)}{\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i\sum_{i=0}^{S}\binom{x+S}{i}\overline{r}^i}. \tag{EC.7}$$

The result in (EC.7) also shows that the full state probability is increasing in x. In addition, we can see that $1-\mathbb{P}_S(x+1)=\sum_{i=0}^{S-1}\binom{x+S}{i}\overline{r}^i/\sum_{i=0}^{S}\binom{x+S}{i}\overline{r}^i$. Therefore we can simplify $\Delta L_S(x)$ as

$$\Delta L_S(x)=\overline{r}\frac{-\sum_{i=0}^{S}\binom{x+S-1}{S}\binom{x+S}{i}\overline{r}^{S+i}(S-i)+\sum_{i=0}^{S-1}\binom{x+S}{i}\overline{r}^i\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i}{\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i\sum_{i=0}^{S}\binom{x+S}{i}\overline{r}^i}. \tag{EC.8}$$

Now to show $\Delta L_S(x)$ is positive we need to show its numerator is always positive. But, this is always true according to Lemma 1. Therefore, $L_S(x)$ is increasing in x.

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

$$\Delta L_S(x)=1-\frac{\sum_{i=0}^{S}\binom{x+S-1}{i}\binom{x+S}{S}\overline{r}^{S+i}+\sum_{i=0}^{S}\binom{x+S-1}{S}\binom{x+S}{i}\overline{r}^{S+i}(S-i)}{\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i\sum_{i=0}^{S}\binom{x+S}{i}\overline{r}^i}, \tag{EC.9}$$

is decreasing in x. But this is always true based on Lemma 2.

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

$$\frac{\partial(\overline{r}\mathbb{P}_S)}{\partial\overline{r}}=\frac{\binom{x+S-1}{S}\overline{r}^S[\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i(S+1-i)]}{[\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i]^2}. \tag{EC.10}$$

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

$$[\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i]^2-[\binom{x+S-1}{S}\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^{i+S}(S+1-i)]\geq 0, \tag{EC.11}$$

which is true according to Lemma 3. Hence, $L_S(x)$ is increasing in $\overline{r}$. Now we show that $\frac{d^2L_S(x)}{d\overline{r}^2}\leq 0$. Note that showing $\frac{d^2L_S(x)}{dr^2}\leq 0$ is equivalent to showing $\frac{\partial^2(\overline{r}\mathbb{P}_S)}{\partial\overline{r}^2}\geq 0$. So we work with the latter one. From the relation (54) in the paper we have

$$\frac{\partial(\overline{r}\mathbb{P}_S)}{\partial\overline{r}}=\frac{\binom{x+S-1}{S}\overline{r}^{S+1}[\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^{i-1}(S+1-i)]}{[\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i]^2}. \tag{EC.12}$$

Furthermore,

$$\frac{\partial^2(\overline{r}\mathbb{P}_S)}{\partial\overline{r}^2}=\frac{\binom{x+S-1}{S}\overline{r}^{S-1}}{[\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}^i]^3}[\sum_{i=0}^{S}\sum_{j=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{j}\overline{r}^{i+j}(S+1-i)(S+i-2j)]. \tag{EC.13}$$

Based on Lemma 4 we have that $\frac{\partial^2(\overline{r}\mathbb{P}_S)}{\partial\overline{r}^2}\geq 0$. Hence, $L_S(x)$ is concave increasing in both r and $\overline{r}$.

(iv) The proof is immediate from part (iii) and part (iii) of Proposition 2. ■

Proof of Proposition 8 (i) We need to show that $\frac{\partial \lambda^*_{v_j}}{\partial x_{v_j}} \leq 0$. By Implicit Function Theorem, we have

$$\frac{\partial \lambda^*_{v_j}}{\partial x_{v_j}} = -\frac{\frac{\partial^2 R_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j} \partial x_{v_j}}}{\frac{\partial^2 R_{v_j}(\lambda^*_{v_j})}{\partial \lambda^2_{v_j}}} = -\frac{\left\{\begin{array}{c} \frac{\partial^2 L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j} \partial x_{v_j}} p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j}) + \frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial x_{v_j}} \frac{\partial p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}} \\ + L_{v_j}(\lambda^*_{v_j}) \frac{\partial^2 p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial x_{v_j} \partial \lambda_{v_j}} + \frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} \frac{\partial p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial x_{v_j}} \end{array}\right\}}{\left\{\begin{array}{c} \frac{\partial^2 L_{v_j}(\lambda^*_{v_j})}{\partial \lambda^2_{v_j}} p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j}) \\ + 2\frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} \frac{\partial p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}} + L_{v_j}(\lambda^*_{v_j}) \frac{\partial^2 p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda^2_{v_j}} \end{array}\right\}}, \tag{EC.14}$$

in which by First Order Necessary Condition we have

$$\frac{\partial R_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} = \mu_{v_j}(\frac{n_{v_j}}{S_{v_j}})(\frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j}) + L_{v_j}(\lambda^*_{v_j}) \frac{\partial p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}}) = 0.$$

Note that since x_{v_j} is discrete we are slightly abusing the Implicit Function Theorem. Consider x_{v_j} to be continuous rather than discrete. It is clear that if $\lambda^*_{v_j}$ is increasing in real-valued x_{v_j}, it is increasing in discrete values of x_{v_j}. Similarly, if $L_{v_j}(\lambda^*_{v_j})$ and $\frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}}$ are increasing (/ decreasing) in any increasing sequence of real values x_{v_j} then the monotonicity holds for any increasing sequence of integer values x_{v_j}. As $\lambda^*_{v_j}$ is a maximizer of a concave function we have $\frac{\partial^2 R_{v_j}(\lambda^*_{v_j})}{\partial \lambda^2_{v_j}} \leq 0$ (Proposition 6), i.e., the denominator is negative. In addition, since $\frac{\partial p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial x_{v_j}} \leq 0$ and $\frac{\partial^2 p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial x_{v_j} \partial \lambda_{v_j}} \leq 0$, we are left with showing $\frac{\partial^2 L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j} \partial x_{v_j}} p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j}) + \frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial x_{v_j}} \frac{\partial p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}} \leq 0$. Using the FONC, $\frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j}) + L_{v_j}(\lambda^*_{v_j}) \frac{\partial p_{v_j}(\lambda^*_{v_j}, x_{v_j}, S_{v_j})}{\partial \lambda_{v_j}} = 0$, we are are left with showing that

$$g(\lambda^*_{v_j}) \triangleq \frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial x_{v_j}} \frac{\partial L_{v_j}(\lambda^*_{v_j})}{\partial \lambda_{v_j}} - \frac{\partial^2 L_{v_j}(\lambda^*_{v_j})}{\partial x_{v_j} \partial \lambda_{v_j}} L_{v_j}(\lambda^*_{v_j}) \geq 0. \tag{EC.15}$$

Without loss of generality we set $\mu_{v_j} = 1$ and thus $\lambda^*_{v_j} = r_{v_j}$. Now we have $L_{v_j} = \overline{r}_{v_j} x_{v_j}(1 - \mathbb{P}_{x_{v_j}})$ and then

$$\frac{\partial L_{v_j}}{\partial x_{v_j}} = \overline{r}_{v_j}(x_{v_j} + 1)(1 - \mathbb{P}_{x_{v_j}+1}) - \overline{r}_{v_j} x_{v_j}(1 - \mathbb{P}_{x_{v_j}}). \tag{EC.16}$$

(We denote $\mathbb{P}_{S_{v_j}}$ with $\mathbb{P}_{x_{v_j}}$ to emphasize the dependence on x_{v_j}.) Also from the proof of Proposition 3 we have that $\frac{\partial L_{v_j}}{\partial \lambda_{v_j}} = x_{v_j}(1 - f_{x_{v_j}})$ where $f_{x_{v_j}}$ is

$$f_{x_{v_j}} \triangleq \frac{\sum_{i=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}-1}{i} \binom{x_{v_j}+S_{v_j}-1}{S_{v_j}} \overline{r}_{v_j}^{S_{v_j}+i} (S_{v_j} - i + 1)}{\sum_{i=0}^{S_{v_j}} \sum_{j=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}-1}{i} \binom{x_{v_j}+S_{v_j}-1}{j} \overline{r}_{v_j}^{i+j}}. \tag{EC.17}$$

Hence, we get

$$\frac{\partial^2 L_{v_j}(\lambda^*_{v_j})}{\partial x_{v_j} \partial \lambda_{v_j}} = 1 - x_{v_j}(f_{x_{v_j}+1} - f_{x_{v_j}}) - f_{x_{v_j}+1}. \tag{EC.18}$$

Using (EC.16) and (EC.18) in (EC.15) and after some algebra we get

$$g(\lambda^*_{v_j}) = (1-\mathbb{P}_{x_{v_j}})(f_{x_{v_j}+1} - f_{x_{v_j}}) - (1-f_{x_{v_j}})(\mathbb{P}_{x_{v_j}+1} - \mathbb{P}_{x_{v_j}}). \tag{EC.19}$$

Next we calculate each term in $g(\lambda^*_{v_j})$ by inserting the relevant functions. Using the relation (5) in the paper as well as (EC.17) in (EC.19) we get

$$\begin{aligned}
&f_{x_{v_j}+1} - f_{x_{v_j}} && \text{(EC.20)}\\
&= \frac{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}}{S_{v_j}}\binom{x_{v_j}+S_{v_j}}{k}\overline{r}_{v_j}^{\,i+j+S_{v_j}+k}(S_{v_j}-k+1)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}}{i}\binom{x_{v_j}+S_{v_j}}{j}\binom{x_{v_j}+S_{v_j}-1}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\overline{r}_{v_j}^{\,i+j+k+l}} && \text{(EC.21)}\\
&- \frac{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}}{i}\binom{x_{v_j}+S_{v_j}}{j}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{k}\overline{r}_{v_j}^{\,i+j+S_{v_j}+k}(S_{v_j}-k+1)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}}{i}\binom{x_{v_j}+S_{v_j}}{j}\binom{x_{v_j}+S_{v_j}-1}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\overline{r}_{v_j}^{\,i+j+k+l}}.
\end{aligned}$$

Multiplying the both sides of (EC.20) and simplifying the right side gives

$$\begin{aligned}
&(1-\mathbb{P}_{x_{v_j}})(f_{x_{v_j}+1} - f_{x_{v_j}})\\
&= \frac{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}-1} \binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}}{S_{v_j}}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\overline{r}_{v_j}^{\,i+j+S_{v_j}+k+l}(S_{v_j}-k+1)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}}\sum_{h=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}}{i}\binom{x_{v_j}+S_{v_j}}{j}\binom{x_{v_j}+S_{v_j}-1}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\binom{x_{v_j}+S_{v_j}-1}{h}\overline{r}_{v_j}^{\,i+j+k+l+h}}\\
&- \frac{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}-1} \binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\overline{r}_{v_j}^{\,i+j+S_{v_j}+k+l}(S_{v_j}-k+1)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}}\sum_{h=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}}{i}\binom{x_{v_j}+S_{v_j}}{j}\binom{x_{v_j}+S_{v_j}-1}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\binom{x_{v_j}+S_{v_j}-1}{h}\overline{r}_{v_j}^{\,i+j+k+l+h}}.
\end{aligned} \tag{EC.22}$$

Using (EC.17) we get

$$\mathbb{P}_{x_{v_j}+1} - \mathbb{P}_{x_{v_j}} = \frac{\frac{1}{x_{v_j}}\sum_{i=0}^{S_{v_j}}\binom{x_{v_j}+S_{v_j}}{i}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\overline{r}_{v_j}^{\,i+S_{v_j}}(S_{v_j}-i)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\binom{x_{v_j}+S_{v_j}}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\overline{r}_{v_j}^{\,i+j}}. \tag{EC.23}$$

Now we use (EC.17) and (EC.23) to obtain $(1-f_{x_{v_j}})(\mathbb{P}_{x_{v_j}+1} - \mathbb{P}_{x_{v_j}})$, the second term in $g(\lambda^*_{v_j})$. After some simplification we obtain

$$\begin{aligned}
&(1-f_{x_{v_j}})(\mathbb{P}_{x_{v_j}+1} - \mathbb{P}_{x_{v_j}})\\
&= \frac{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}}{l}\overline{r}_{v_j}^{\,i+j+S_{v_j}+k+l}\left(\frac{S_{v_j}-k}{x_{v_j}}\right)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}}\sum_{h=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\binom{x_{v_j}+S_{v_j}}{h}\overline{r}_{v_j}^{\,i+j+k+l+h}}\\
&- \frac{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}}{j}\binom{x_{v_j}+S_{v_j}}{k}\overline{r}_{v_j}^{\,i+j+k+2S_{v_j}}(S_{v_j}-i+1)\left(\frac{S_{v_j}-j}{x_{v_j}}\right)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}}\sum_{h=0}^{S_{v_j}} \binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\binom{x_{v_j}+S_{v_j}}{h}\overline{r}_{v_j}^{\,i+j+k+l+h}}.
\end{aligned} \tag{EC.24}$$

Adding Equations (EC.22) and (EC.24) gives (notice the denominators are in fact the same)

$$g(\lambda^*_{v_j}) \tag{EC.25}$$

$$= \frac{x_{v_j}\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}-1}\binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}}{S_{v_j}}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\overline{r}_{v_j}^{i+j+k+l}(S_{v_j}-k+1)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}}\sum_{h=0}^{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\binom{x_{v_j}+S_{v_j}}{h}\overline{r}_{v_j}^{i+j+k+l+h}}$$
$$- \frac{x_{v_j}\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}-1}\binom{x_{v_j}+S_{v_j}}{i}\binom{x_{v_j}+S_{v_j}}{j}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\overline{r}_{v_j}^{i+j+k+l}(S_{v_j}-k+1)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}}\sum_{h=0}^{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\binom{x_{v_j}+S_{v_j}}{h}\overline{r}_{v_j}^{i+j+k+l+h}}$$
$$- \frac{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}}{l}\overline{r}_{v_j}^{i+j+k+l}(S_{v_j}-k)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}}\sum_{h=0}^{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\binom{x_{v_j}+S_{v_j}}{h}\overline{r}_{v_j}^{i+j+k+l+h}}$$
$$+ \frac{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}}{j}\binom{x_{v_j}+S_{v_j}}{k}\overline{r}_{v_j}^{i+j+k+S_{v_j}}(S_{v_j}-i+1)(S_{v_j}-j)}{\sum_{i=0}^{S_{v_j}}\sum_{j=0}^{S_{v_j}}\sum_{k=0}^{S_{v_j}}\sum_{l=0}^{S_{v_j}}\sum_{h=0}^{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}}{k}\binom{x_{v_j}+S_{v_j}-1}{l}\binom{x_{v_j}+S_{v_j}}{h}\overline{r}_{v_j}^{i+j+k+l+h}}.$$

To show that $g(\lambda^*_{v_j}) \geq 0$ 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 $\overline{r}_{v_j}$. Let us assume that the power of $\overline{r}_{v_j}$ is z where $0 \leq z \leq 4S_{v_j}$. If $g(\lambda^*_{v_j}) \geq 0$ then the coefficient of $\overline{r}^z_{v_j}$ for each and every z, $0 \leq z \leq 4S_{v_j}$, needs be positive. We divide the range into four parts: $0 \leq z < S_{v_j}$, $S_{v_j} \leq z < 2S_{v_j}$, $2S_{v_j} \leq z < 3S_{v_j}$, $3S_{v_j} \leq z \leq 4S_{v_j}$. Here we will illustrate the proof for $0 \leq z < S_{v_j}$. The other ranges are proved similarly.

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

$$B(x_{v_j}, S_{v_j}, z) = \sum_{i=0}^{z}\sum_{j=0}^{z-i}\sum_{k=0}^{z-i-j} H_{i,j,k}(x_{v_j}, S_{v_j}) C_{i,j,k}(x_{v_j}, S_{v_j}), \tag{EC.26}$$

where $H_{i,j,k}(x_{v_j}, S_{v_j})$ is

$$H_{i,j,k}(x_{v_j}, S_{v_j}) \triangleq \binom{x_{v_j}+S_{v_j}-1}{i}\binom{x_{v_j}+S_{v_j}-1}{j}\binom{x_{v_j}+S_{v_j}-1}{S_{v_j}}\binom{x_{v_j}+S_{v_j}-1}{k}\binom{x_{v_j}+S_{v_j}-1}{z-i-j-k}(x_{v_j}+S_{v_j})^2 \geq 0, \tag{EC.27}$$

and $C_{i,j,k}(x_{v_j}, S_{v_j})$ is

$$C_{i,j,k}(x_{v_j}, S_{v_j}) \triangleq \left(\frac{S_{v_j}-k+1}{x_{v_j}+S_{v_j}-k}\right) - \frac{x_{v_j}(S_{v_j}-k+1)}{(x_{v_j}+S_{v_j}-i)(x_{v_j}+S_{v_j}-j)} - \left(\frac{S_{v_j}-k}{x_{v_j}+S_{v_j}-k}\right)\left(\frac{1}{x_{v_j}+S_{v_j}-(\underbrace{z-i-j-k}_{l})}\right). \tag{EC.28}$$

We can see that $H_{i,j,k}(x_{v_j}, S_{v_j}) \geq 0$. Hence, we only need to show $C_{i,j,k}(x_{v_j}, S_{v_j}) \geq 0$. After some simplification in (EC.28) we get

$$C_{i,j,k}(x_{v_j}, S_{v_j}) \geq \frac{ij(S_{v_j}-l-1) + S^2_{v_j}(S_{v_j}-i-j-l-1) + kx_{v_j}(S_{v_j}-l)}{(S_{v_j}+x_{v_j}-i)(S_{v_j}+x_{v_j}-j)(S_{v_j}+x_{v_j}-k)(S_{v_j}+x_{v_j}-k)} \tag{EC.29}$$

$$+\frac{S_{v_j}x_{v_j}(2S_{v_j}-2i-2j-l-2)+x_{v_j}^2(S_{v_j}-i-j-1)}{(S_{v_j}+x_{v_j}-i)(S_{v_j}+x_{v_j}-j)(S_{v_j}+x_{v_j}-k)(S_{v_j}+x_{v_j}-k)}$$
$$\geq 0.$$

Having in mind that $z=i+j+k+l$ and $0\leq z<S_{v_j}$ we notice that in the right side of (EC.29) each term in the numerator (and the denominator) is positive. Hence, $C_{i,j,k}(x_{v_j},S_{v_j})\geq 0$. Given that other ranges for z hold we have that $g(\lambda^*_{v_j})\geq 0$, which ensures $\frac{\partial\lambda^*_{v_j}}{\partial x_{v_j}}\leq 0$. As the price is decreasing in λ_{v_j}, we have proved that the optimal price is increasing in x_{v_j}.

(ii) The proof is immediate from part (i) and the chain rule $\frac{dp_{v_j}(\lambda^*_{v_j},x_{v_j},S_{v_j})}{dx_{v_j}}=\frac{\partial p_{v_j}(\lambda^*_{v_j},x_{v_j},S_{v_j})}{\partial\lambda_{v_j}}\frac{\partial\lambda^*_{v_j}}{\partial x_{v_j}}+\frac{\partial p_{v_j}(\lambda^*_{v_j},x_{v_j},S_{v_j})}{\partial x_{v_j}}$. ■

EC.2. Lemmas and Proofs

Lemma 1 *Given $x \in \mathbb{N}$, $i \in \mathbb{N}$, and $\kappa \in \mathbb{R}$, the following result holds:*

$$\sum_{k_1=1}^{x}\sum_{k_2=k_1}^{x}\dots\sum_{k_i=k_{i-1}}^{x}\kappa=\binom{x+i-1}{i}\kappa. \tag{EC.30}$$

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 $B_{s+1}=\binom{x+s-1}{s}+\binom{x+s-2}{s}+\binom{x+s-3}{s}+\dots+\binom{s}{s}$. By using Lemma 2 we obtain that this summation is equal to $\binom{x+s}{s+1}$, which completes the proof. ∎

Lemma 2 *Given $k \in \mathbb{N}\cap[0,x-1]$ and $x \in \mathbb{N}\cup\{0\}$, the following result holds:*

$$\sum_{i=k}^{x+k-1}\binom{i}{k}=\binom{x+k}{k+1}. \tag{EC.31}$$

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 3 *Given $x \in \mathbb{N}\cap[S-1,\infty)$ with $S \in \mathbb{N}$ and $\overline{r}_{v_j} \in \mathbb{R}_+$,*

$$\sum_{i=0}^{S-1}\binom{x+i-1}{i}r^i(1+r)^{S-i-1}=\sum_{i=0}^{S-1}\binom{x+S-1}{i}r^i. \tag{EC.32}$$

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

$$D(k) \triangleq \sum_{i=0}^{k-1}\binom{x+i-1}{i} r^i (1+r)^{k-i-1} - \sum_{i=0}^{k-1}\binom{x+k-1}{i} r^i = 0. \tag{EC.33}$$

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

$$D(k+1) = \sum_{i=0}^{k}\binom{x+i-1}{i} r^i (1+r)^{k-i} - \sum_{i=0}^{k}\binom{x+k}{i} r^i = 0. \tag{EC.34}$$

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

$$D(k+1) = (1+r)\sum_{i=0}^{k-1}\binom{x+i-1}{i} r^i (1+r)^{k-i-1} + \binom{x+k-1}{k} r^k - \sum_{i=0}^{k}\binom{x+k}{i} r^i. \tag{EC.35}$$

Using the induction assumption we get

$$\begin{aligned} D(k+1) &= (1+r)\sum_{i=0}^{k-1}\binom{x+k-1}{i} r^i + \binom{x+k-1}{k} r^k - \sum_{i=0}^{k-1}\binom{x+k}{i} r^i - \binom{x+k}{k} r^k \qquad \text{(EC.36)} \\ &= \sum_{i=0}^{k-1}\left[\binom{x+k-1}{i} - \binom{x+k}{i}\right] r^i - \sum_{i=0}^{k-1}\binom{x+k-1}{i} r^{i+1} + \binom{x+k-1}{k} r^k - \binom{x+k}{k} r^k, \end{aligned}$$

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

$$\begin{aligned} D(k+1) &= \sum_{j=0}^{k-2}\binom{x+k-1}{j} r^{j+1} + \binom{x+k-1}{k-1} r^k - \sum_{i=0}^{k-1}\binom{x+k-1}{i} r^{i+1} \qquad \text{(EC.37)} \\ &= \sum_{j=0}^{k-1}\binom{x+k-1}{j} r^{j+1} - \sum_{i=0}^{k-1}\binom{x+k-1}{i} r^{i+1} = D(k) = 0, \end{aligned}$$

which completes the proof. ■

Lemma 4 *Given any natural numbers $x \in \mathbb{N}$, and $S \in \mathbb{N}$,*

$$\sum_{i=0}^{S}\sum_{j=0}^{S-1}\binom{x+S-1}{i}\binom{x+S}{j}\overline{r}_{v_j}^{i+j} \geq \sum_{i=0}^{S}\binom{x+S-1}{S}\binom{x+S}{i}\overline{r}_{v_j}^{S+i}(S-i). \tag{EC.38}$$

Proof The Lemma can be proved using the same approach as in the proof of Lemma 3. ■

Lemma 5 *Let $Q(x) = Q_N(x)/Q_D(x)$, where*

$$Q_N(x) = \left(\sum_{i=0}^{S}\binom{x+S-1}{i}\binom{x+S}{S}\overline{r}_{v_j}^{S+i} + \sum_{i=0}^{S}\binom{x+S-1}{S}\binom{x+S}{i}\overline{r}_{v_j}^{S+i}(S-i)\right) \tag{EC.39}$$

and

$$Q_D(x) = \left(\sum_{i=0}^{S}\binom{x+S-1}{i}\overline{r}_{v_j}^{i}\sum_{i=0}^{S}\binom{x+S}{i}\overline{r}_{v_j}^{i}\right). \tag{EC.40}$$

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

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

$$
\begin{aligned}
A(x,S) \triangleq & \sum_{i=0}^{S}\sum_{j=0}^{S}\sum_{k=0}^{S} \binom{x+S-1}{i}\binom{x+S}{j}\binom{x+S}{k}\binom{x+S+1}{S}\overline{r}_{v_j}^{i+j+k} \\
&+ \sum_{i=0}^{S}\sum_{j=0}^{S}\sum_{k=0}^{S} \binom{x+S-1}{i}\binom{x+S}{j}\binom{x+S+1}{k}\binom{x+S}{S}\overline{r}_{v_j}^{i+j+k}(S-k) \\
&+ \sum_{i=0}^{S}\sum_{j=0}^{S}\sum_{k=0}^{S} \binom{x+S-1}{i}\binom{x+S}{j}\binom{x+S+1}{k}\binom{x+S}{S}\overline{r}_{v_j}^{i+j+k}(S-k) \\
&- \sum_{i=0}^{S}\sum_{j=0}^{S}\sum_{k=0}^{S} \binom{x+S}{i}\binom{x+S+1}{j}\binom{x+S-1}{k}\binom{x+S}{S}\overline{r}_{v_j}^{i+j+k} \\
&- \sum_{i=0}^{S}\sum_{j=0}^{S}\sum_{k=0}^{S} \binom{x+S}{i}\binom{x+S+1}{j}\binom{x+S}{k}\binom{x+S-1}{S}\overline{r}_{v_j}^{i+j+k}(S-k) \geq 0.
\end{aligned} \tag{EC.41}
$$

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

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

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

$$
\begin{aligned}
B(x,S,i,j,z) = & \binom{x+S-1}{i}\binom{x+S}{j}\binom{x+S-1}{z-i-j}\binom{x+S-1}{S}(x+S)^2(x+S+1) \\
& [\frac{1}{x(x+1)(x+S-z+i+j)} + \frac{(S-z+i+j)}{x(x+S-z+i+j)(x+S-z+i+j+1)} \\
& - \frac{1}{x(x+S-i)(x+S-j+1)} - \frac{(S-z+i+j)}{(x+S-i)(x+S+1-j)(x+S-z+i+j)}].
\end{aligned} \tag{EC.42}
$$

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

$$
\begin{aligned}
& \frac{1}{x(x+1)(x+S-z+i+j)} - \frac{1}{x(x+S-i)(x+S-j-1)} \\
& = \frac{(-i-iS-jS+S^2)+(Sx+zx-2ix-2jx)+(z-i-j)}{x(x+1)(x+S-z+i+j)(x+S-i)(x+S-j-1)} \geq 0.
\end{aligned} \tag{EC.43}
$$

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

$$
\begin{aligned}
& \frac{(S-z+i+j)}{x(x+S-z+i+j)(x+S-z+i+j+1)} - \frac{(S-z+i+j)}{(x+S-i)(x+S+1-j)(x+S-z+i+j)} \\
& = \frac{(S-z+i+j)((S-i)+(S^2-iS-jS)+(Sx+xz-2ix-2jx)+ij}{x(x+S-z+i+j)(x+S-z+i+j+1)(x+S-i)(x+S+1-j)} \geq 0.
\end{aligned} \tag{EC.44}
$$

Therefore, the coefficient of $\overline{r}_{v_j}^{z}$ is positive, which completes the proof for $0 \leq z < S$. ■

Lemma 6 *For $0 \leq j \leq i \leq S$ and $x \geq 1$ we have*

$$\binom{x+S-1}{i}\binom{x+S-1}{S+j-i} \geq \binom{x+S-1}{S}\binom{x+S-1}{j}. \tag{EC.45}$$

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

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

With further simplifications we get

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

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

Lemma 7 *Given any natural numbers $x \in \mathbb{N}$ and $S \in \mathbb{N}$,*

$$\sum_{i=0}^{S}\sum_{j=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{j}\overline{r}^{i+j} \geq \sum_{i=0}^{S}\binom{x+S-1}{S}\binom{x+S-1}{i}\overline{r}^{S+i}(S+1-i). \tag{EC.46}$$

Proof

We prove this lemma by selecting a few "convenient" terms from the double sum on the left hand side of the inequality 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 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 $i+j=2S$, then the term with $i+j=2S-1$, etc:

$$\begin{aligned}
&\sum_{i=0}^{S}\sum_{j=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{j}\overline{r}^{i+j} \qquad \text{(EC.47)}\\
&\geq \binom{x+S-1}{S}\binom{x+S-1}{S}\overline{r}^{2S} + [\binom{x+S-1}{S}\binom{x+S-1}{S-1}\\
&+ \binom{x+S-1}{S-1}\binom{x+S-1}{S}]\overline{r}^{2S-1} + [\binom{x+S-1}{S}\binom{x+S-1}{S-2}\\
&+ \binom{x+S-1}{S-1}\binom{x+S-1}{S-1} + \binom{x+S-1}{S-1}\binom{x+S-1}{S-1}\\
&+ \binom{x+S-1}{S-2}\binom{x+S-1}{S}]\overline{r}^{2S-2} + \ldots + [\binom{x+S-1}{S}\binom{x+S-1}{S-3}\\
&+ \binom{x+S-1}{S-1}\binom{x+S-1}{S-2} + \binom{x+S-1}{S-2}\binom{x+S-1}{S-1}
\end{aligned}$$

$$+\binom{x+S-1}{S-3}\binom{x+S-1}{S}]\overline{r}^{2S-3}+...+[\sum_{i=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{S-i}]\overline{r}^{S}.$$

After some algebra we obtain

$$\begin{aligned}&\sum_{i=0}^{S}\sum_{j=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{j}\overline{r}^{i+j} \\ &\geq[\sum_{i=S}^{S}\binom{x+S-1}{i}\binom{x+S-1}{2S-i}]\overline{r}^{2S}+[\sum_{i=S-1}^{S}\binom{x+S-1}{i}\binom{x+S-1}{2S-1-i}]\overline{r}^{2S-1} \\ &+[\sum_{i=S-1}^{S}\binom{x+S-1}{i}\binom{x+S-1}{2S-1-i}]\overline{r}^{2S-1}+[\sum_{i=S-2}^{S}\binom{x+S-1}{i}\binom{x+S-1}{2S-2-i}]\overline{r}^{2S-2} \\ &+...+[\sum_{i=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{S-i}]\overline{r}^{S} \\ &=\sum_{j=0}^{S}[\sum_{i=j}^{S}\binom{x+S-1}{i}\binom{x+S-1}{S+j-i}]\overline{r}^{j+S}.\end{aligned} \tag{EC.48}$$

Now we subtract the term $-\sum_{j=0}^{S}\binom{x+S-1}{S}\binom{x+S-1}{j}\overline{r}^{S+i}(S+1-j)$ from both sides of (EC.48) to get

$$\begin{aligned}&\sum_{i=0}^{S}\sum_{j=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{j}\overline{r}^{i+j}-\sum_{j=0}^{S}\binom{x+S-1}{S}\binom{x+S-1}{j}\overline{r}^{S+i}(S+1-j) \\ &\geq\sum_{j=0}^{S}\overline{r}^{S+j}[\sum_{i=j}^{S}\binom{x+S-1}{i}\binom{x+S-1}{S+j-i}-\binom{x+S-1}{S}\binom{x+S-1}{j}(S+1-j)].\end{aligned} \tag{EC.49}$$

On the other side Lemma 6 tells us that the below result is always correct:

$$\binom{x+S-1}{i}\binom{x+S-1}{S+j-i}\geq\binom{x+S-1}{S}\binom{x+S-1}{j} \text{ for } 0\leq\ j\leq i\leq S \text{ and } x\geq 1. \tag{EC.50}$$

Replacing (EC.50) in (EC.49) we get

$$\begin{aligned}&\sum_{j=0}^{S}\overline{r}^{S+j}[\sum_{i=j}^{S}\binom{x+S-1}{i}\binom{x+S-1}{S+j-i}-\binom{x+S-1}{S}\binom{x+S-1}{j}(S+1-j)] \\ &\geq\sum_{j=0}^{S}\overline{r}^{S+j}[\sum_{i=j}^{S}\binom{x+S-1}{S}\binom{x+S-1}{j}-\binom{x+S-1}{S}\binom{x+S-1}{j}(S+1-j)] \\ &=\sum_{j=0}^{S}\overline{r}^{S+j}[\binom{x+S-1}{S}\binom{x+S-1}{j}(S+1-j)-\binom{x+S-1}{S}\binom{x+S-1}{j}(S+1-j)]=0.\end{aligned} \tag{EC.51}$$

That shows the positivity of (EC.46) and completes the proof. ■

Lemma 8 *Given any $x,S\in\mathbb{N}\cup\{0\}$, and $\overline{r}\in\mathbb{R}_+$*

$$\sum_{i=0}^{S}\sum_{j=0}^{S}\binom{x+S-1}{i}\binom{x+S-1}{j}\overline{r}^{i+j}(S+1-i)(S+i-2j)\geq 0. \tag{EC.52}$$

Proof The lemma can be proved using a similar approach as in the proof of Lemma 3. ■

EC.3. Considering a Charge for 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). We assume no ad rotation, i.e., $S = n$. In addition, we 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[1]. 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:

$$R(\lambda, \mu, x, n) = Lp(\lambda, n)\mu + (n - L)e\mu, \tag{EC.53}$$

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 (EC.53). Our problem now reduces to

$$R(\lambda, \mu, x, n) = \lambda(1 - \mathbb{P}_n(\lambda, \mu, x))x(p(\lambda, n) - e) + ne\mu. \tag{EC.54}$$

Note that for a high value of e, i.e., $e \geq p(\lambda, n)$, the maximum of (EC.53) with respect to λ becomes $R^* = ne\mu$. This is because a large e makes the first term negative, 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 (EC.54) are both independent of λ. Hence, it is easy to see that the maximization of (EC.54) becomes equivalent to the maximization of

$$\max_{\lambda} \widehat{R}(\lambda, \mu, x, n) = \lambda(1 - \mathbb{P}_n(\lambda, \mu, x))x\widehat{p}(\lambda, n), \tag{EC.55}$$

where the price function is defined as $\widehat{p}(\lambda, n) = p(\lambda, n) - e \geq 0$.We can now see that (EC.55) has the same form as the basic model considered in our paper. In addition, since $p(\lambda, n)$ has the following

[1] 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.

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 (EC.55) 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 (EC.55), $p^*(\lambda^*(x),n)$ 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.

EC.4. More Detailed Explanation about Synchronization

We consider the basic model with n advertising slots, where each slot can display only one ad. Every time a visitor uploads the publisher's website all the ads in the system are displayed. An advertiser pays every time his ad is displayed regardless of whether the viewer *sees* the ads. One of the most well-known contracts in practice is called *Guaranteed Delivery* or *GD*. In the GD contracts the publisher promises each advertiser to deliver a certain number of impressions to the viewers. For each ad, the publisher keeps track of the number of times the ad is displayed. For example, when an advertiser arrives at the system and is promised to be delivered x impressions the counter's initial value for that advertiser would be set on x. Then each time the ad is displayed the counter's value drops by one unit. On a related issue, if there are multiple advertisers in the publisher's system, and a viewer arrives, the values of all the related counters drop by one unit at the same time (i.e., simultaneously). We refer to this property, in which all the counters count down simultaneously, as *synchronization*.

We recall that the publisher's system is formulated as a queuing model with the state vector $\mathbf{k}=(k_1,k_2,...,k_n)$, which indicates that the number of remaining impressions for one of the slots is k_1, for another slot it is k_2,..., and for the last one is k_n, without distinguishing between the slots. In order to observe the effect of synchronization on the state transitions consider the following example:

Consider that the state of the system is $\mathbf{k}=(k_1,k_2,...,k_i,\underbrace{0,...,0}_{n-i})$, where the first i components are positive and the rest are zero (empty slots). When a viewer arrives at the system (i.e., loads the publisher's webpage) the state of the system goes to the new state $(k_1-1,k_2-1,...,k_i-1,0,...,0)$ with rate μ. This is because when a viewer enters the system all the ads are displayed once to him. Therefore, the remaining impressions (i.e., the values of the counters) for all the positive components of the state vector (i.e., the active ads in the system) are reduced by one unit at the same time, while the zero-components (i.e., the empty slots) stay the same. As previously mentioned, synchronization is regarded as this simultaneous reduction of the counter values in the positive components.

EC.5. Additional Simulations for Non-Poisson Arrivals

In this section we provide additional simulation results for the gap analysis of the different interarrival time distributions considered for advertisers' and viewers' arrivals. The numerical information of parameters is the same as in the paper. We note the gap is smaller with a larger number of impressions and fewer slots. To fully focus on the interarrival distributions, the number of impressions are assumed to be deterministic. The steps for each simulation process are as follows:

Step 1. We obtain the optimal advertisers' arrival rate, $\lambda^*_{D_1,D_2}$, when the advertisers' interarrival times follow the generic distribution D_1, and the viewers' interarrival times follow D_2. This involves 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 $R_{D_1,D_2}(\lambda^*_{D_1,D_2})$.

Step 2. We compute the optimal value for λ using the closed form solution provided in the paper. We represent this value 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 become $R_{D_1,D_2}(\lambda^*_{Exp})$.

Step 3. We compute the revenue gap using the following formula:

$$Gap = \frac{R_{D_1,D_2}(\lambda^*_{D_1,D_2}) - R_{D_1,D_2}(\lambda^*_{Exp})}{R_{D_1,D_2}(\lambda^*_{D_1,D_2})} \times 100(\%). \tag{EC.56}$$

Figure 1 shows a schematic presentation of how the revenue gap is obtained using the above steps.

We notice that the revenue gap is considerably higher (i.e., about 20.13%) when the viewers' and the advertisers' arrival processes are both deterministic. This suggests that the Poisson policy may not be a good approximation when there is no uncertainty in the model. However, we notice that when even either advertisers' or viewers' arrival process is not deterministic the Poisson policy tends to perform very well.

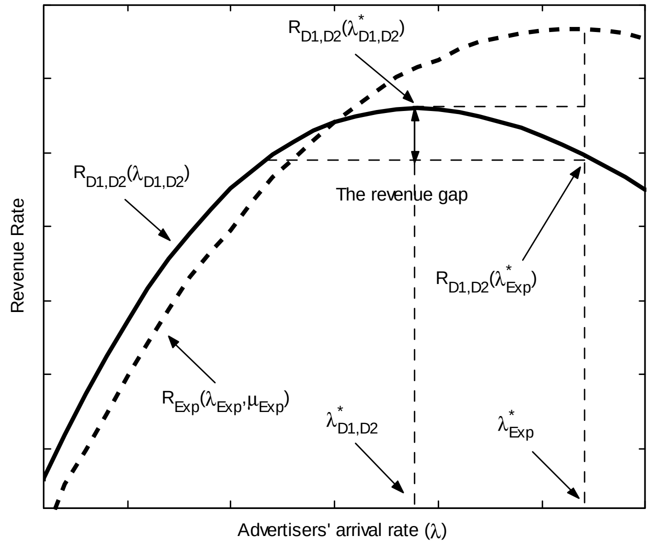

Figure 1 A schematic presentation of how the revenue gap is computed through the mentioned steps.

$n=2,\ x=500$	*Viewers* $(\mathbf{D}_2)$					
Advertisers $(\mathbf{D}_1)$	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Deterministic*	*Exponential*
Erlang-2	0.9%	0.19%	0.04%	0.64%	0.69%	0.42%
Erlang-4	1.02%	1.71%	1.36%	1.70%	0.70%	0.85%
Normal	1.24%	0.04%	0.24%	0.52%	1.16%	0.06%
Uniform	0.80%	0.69%	0.84%	0.63%	0.75%	0.58%
Deterministic	2.82%	2.80%	2.91%	2.88%	20.13%	3.02%
Exponential	0.26%	0.01%	0.08%	0.04%	0.46%	–

Table 2 The relative performance gap $\frac{R_{D_1,D_2}(\lambda^*_{D_1,D_2})-R_{D_1,D_2}(\lambda^*_{Exp})}{R_{D_1,D_2}(\lambda^*_{D_1,D_2})}\times 100(\%)$

$n=2,\ x=1000$	*Viewers* $(\mathbf{D}_2)$					
Advertisers $(\mathbf{D}_1)$	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Deterministic*	*Exponential*
Erlang-2	0.58%	0.35%	0.32%	0.37%	0.39%	1.05%
Erlang-4	1.09%	1.08%	0.95%	1.18%	0.50%	1.25%
Normal	0.22%	0.05%	0.37%	0.27%	0.17%	0.38%
Uniform	0.69%	1.01%	1.15%	0.40%	0.76%	0.73%
Deterministic	2.17%	1.95%	1.50%	2.32%	11.10%	2.23%
Exponential	0.35%	0.12%	0.33%	0.12%	0.14%	–

Table 3 The relative performance gap $\frac{R_{D_1,D_2}(\lambda^*_{D_1,D_2})-R_{D_1,D_2}(\lambda^*_{Exp})}{R_{D_1,D_2}(\lambda^*_{D_1,D_2})}\times 100(\%)$

EC.6. Additional Simulation Results for non-Poisson Arrivals and Random Requests

In this section we provide additional simulation results for the gap analysis of the different inter-arrival time distributions considered for advertisers' and viewers' arrivals as well as the various

$n=2,\ x=1500$	*Viewers* $(\mathbf{D}_2)$					
Advertisers $(\mathbf{D}_1)$	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Deterministic*	*Exponential*
Erlang-2	0.1%	1.13%	0.20%	1.11%	0.14%	0.64%
Erlang-4	1.27%	0.93%	0.86%	1.00%	0.73%	0.96%
Normal	0.12%	0.38%	0.07%	0.02%	0.05%	0.07%
Uniform	1.14%	1.4%	0.21%	1.13%	0.98%	1.06%
Deterministic	3.39%	3.54%	3.68%	3.26%	5.90%	3.70%
Exponential	0.32%	0.11%	0.03%	0.38%	0.53%	–

Table 4 The relative performance gap $\frac{R_{D_1,D_2}(\lambda^*_{D_1,D_2})-R_{D_1,D_2}(\lambda^*_{Exp})}{R_{D_1,D_2}(\lambda^*_{D_1,D_2})}\times 100(\%)$

$n=4,\ x=1500$	*Viewers* $(\mathbf{D}_2)$					
Advertisers $(\mathbf{D}_1)$	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Deterministic*	*Exponential*
Erlang-2	1.38%	0.88%	0.55%	0.03%	1.49%	0.84%
Erlang-4	0.97%	1.22%	0.46%	1.41%	1.04%	1.23%
Normal	0.19%	0.08%	0.06%	0.06%	0.11%	0.13%
Uniform	0.81%	0.18%	1.11%	1.65%	0.96%	0.31%
Deterministic	3.55%	3.34%	3.41%	3.49%	5.90%	3.16%
Exponential	0.73%	0.15%	0.22%	0.05%	0.35%	–

Table 5 The relative performance gap $\frac{R_{D_1,D_2}(\lambda^*_{D_1,D_2})-R_{D_1,D_2}(\lambda^*_{Exp})}{R_{D_1,D_2}(\lambda^*_{D_1,D_2})}\times 100(\%)$

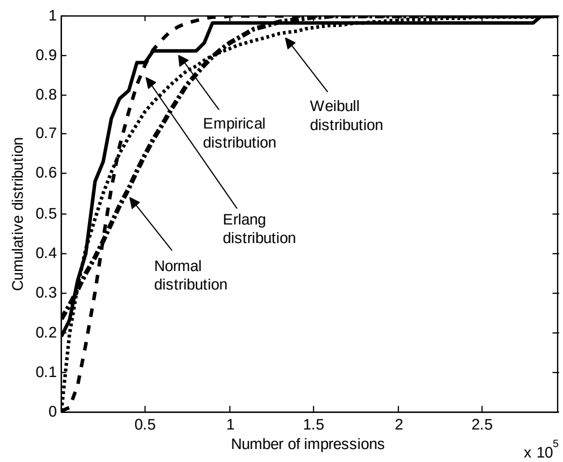

Figure 2 An illustration of the empirical distribution of the impressions requested by the advertisers of Aller Internett

distributions considered for the number of impressions, X. The numerical values of parameters are identical to the Extensions Section in the paper (non-Poisson Arrivals). Here, we consider multiple distributions for arrival processes as well as the impressions' request. The selection of the different distributions for X is based on our empirical observation on real data from Aller Internett (see Figure 2).

$D_2 \sim Exp$			Advertisers ($\mathbf{D}_1$)		
Impressions	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Exponential*
Erlang-2	1.14%	1.58%	0.01%	0.98%	1.31%
Erlang-4	0.87%	1.44%	0.23%	1.19%	0.34%
Normal	0.16%	0.86%	0.21%	0.85%	0.30%
Uniform	0.04%	0.66%	0.11%	0.40%	0.25%

Table 6 The relative gap $\frac{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})-R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{x,Exp})}{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})}\times 100(\%)$

$D_2 \sim Normal$			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 7 The relative gap $\frac{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})-R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{x,Exp})}{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})}\times 100(\%)$

$D_2 \sim Uniform$			Advertisers ($\mathbf{D}_1$)		
Impressions	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Exponential*
Erlang-2	0.31%	1.26%	0.38%	0.99%	0.21%
Erlang-4	0.49%	0.93%	0.39%	0.93%	0.55%
Normal	0.06%	0.66%	0.10%	0.26%	0.11%
Uniform	0.05%	1.65%	0.27%	0.46%	0.65%

Table 8 The relative gap $\frac{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})-R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{x,Exp})}{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})}\times 100(\%)$

$D_2 \sim Erlang$			Advertisers ($\mathbf{D}_1$)		
Impressions	*Erlang-2*	*Erlang-4*	*Normal*	*Uniform*	*Exponential*
Erlang-2	0.38%	1.00%	0.08%	1.29%	0.47%
Erlang-4	0.60%	0.73%	0.09%	1.45%	2.02%
Normal	0.22%	0.60%	0.40%	0.45%	0.26%
Uniform	0.11%	0.15%	0.20%	0.92%	0.39%

Table 9 The relative gap $\frac{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})-R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{x,Exp})}{R_{X,\mathbf{D_1},\mathbf{D_2}}(\lambda^*_{X,\mathbf{D_1},\mathbf{D_2}})}\times 100(\%)$

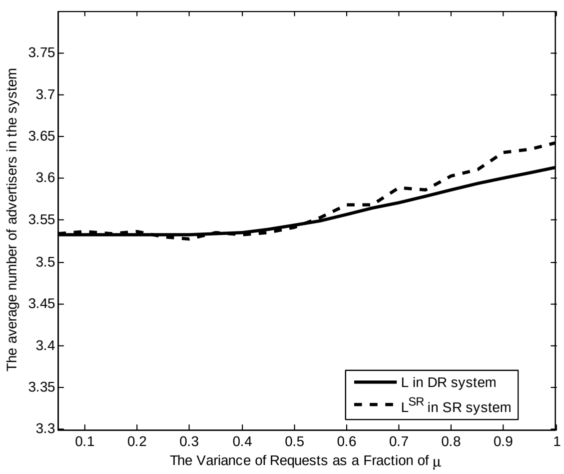

Figure 3 Comparison of L vs. simulated L^{RIR}

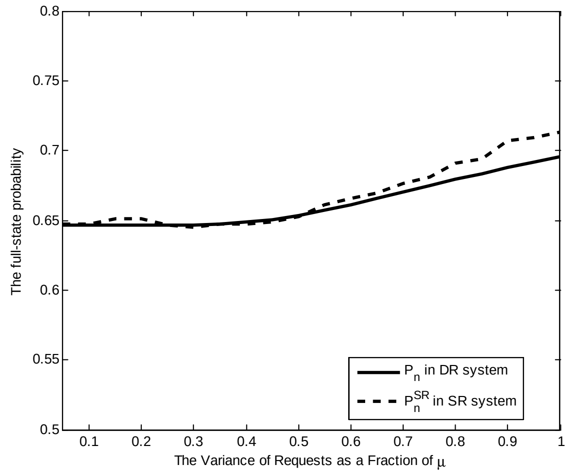

Figure 4 Comparison of $\mathbb{P}_n$ vs. simulated $\mathbb{P}_n^{RIR}$

EC.7. Simulating L and P_n with Random Numbers of Impressions

As mentioned in the paper, publishers usually display any number of impressions requested by advertisers or allow them to choose among several listed quantities. As mentioned in the Model's Section, 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 3). 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 random impressions request system (RIR) and the system where Assumption 3 applies by deterministic request system (DR). Solving this RIR system analytically does not appear to be tractable but to gain further insights, we perform additional 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 (we assume no rotation of ads). 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, $n = 4$ and there is no ad rotation, i.e., $s = 1$. 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 RIR 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 3 and 4 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 closed-form solution for the DR system with $x = \mathbb{E}(X) = \mu/(1-\Phi(\frac{-1}{\vartheta}))$.

Based on Figures 3 and 4 we can see that the performance measures considered for the RIR 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 RIR system's behavior even for low numbers of impressions.