Interactive Advertising: The Case of Skippable Ads

Anthony Dukes* Qihong Liu† Jie Shuai‡

May 2, 2018

Abstract

The skippable ad format, commonly used by online content platforms, requires viewers to see a portion of an advertiser's message before having the option to skip directly to the intended content and avoid viewing the entire ad. Under what conditions do viewers forego this option and what are its implications for advertisers and the platform? We develop a dynamic model of a viewer receiving incremental information from the advertiser. This model identifies conditions under which the viewer (i) skips the ad or (ii) engages with the advertiser. Our model incorporates the advertising market and assess implications of skippable ads on the platform's profit and advertisers' surplus. Relative to the traditional ad format, we find that there are unambiguously more advertisements and viewers on the platform with skippable ads. Under reasonable conditions, the skippable ad format is a strict Pareto improvement, which raises the surplus of advertisers and the profit of the platform. The source of the additional surplus is that skippable ads allow the viewers to use private information about the advertiser to make more efficient decisions about their ad viewing choices.

Keywords: Advertising, Commercial Media, Two-Sided Markets, Skippable Ads

1 Introduction

As is becoming increasingly common on content platforms is the notion of the skippable advertisement (hereafter: skippable ad). In a skippable ad, the viewer receives some brief information about the advertiser and then decides whether to continue viewing the ad or to move directly to the intended content. By moving directly to content, the viewer avoids the remaining portion of the ad. Consider, for instance, the video platform YouTube, which implements the skippable ad format known as *TrueView* ads. Before a YouTube user can watch a video, she is often required to view the first part of an ad. After about 5 seconds, the user has the option to "Skip Ad" and proceed directly to the

*Marshall School of Business, University of Southern California, Los Angeles CA 90089. E-mail: dukes@marshall.usc.edu.

†Department of Economics, University of Oklahoma, Norman, OK 73019. E-mail: qliu@ou.edu.

‡Wenlan School of Business, Zhongnan University of Economics and Law, 182 Nanhu Ave., East Lake High-tech Development Zone, Wuhan 430073, China. Email: shuaijie@gmail.com.

desired video. By the end of 2018, YouTube plans to implement the skippable format exclusively for all ads 30 seconds or longer.[1]

One defining aspect of the skippable ad is that the consumer receives some information about the advertiser's product and, after assessing that information, determines whether to pursue more information from the advertiser or to proceed with the intended content (e.g., a video). The novel feature of the skippable ad is the viewer's interaction with that initial information. Specifically, the viewer is forced to decide whether or not to continue attending to the ad based on her assessment of the first part. Prior work in marketing and media economics has modeled ad viewing as an "all or nothing" endeavor. That is, the viewer either sees each ad entirely or none of them (c.f. Anderson et al. 2015). In the former, the consumer is assumed to see all ads provided by the platform, even though it imposes nuisance costs that detract from the content. In the latter, which has tended to focus on ad-avoidance technologies, a consumer initially commits to seeing no ads so that there is no possibility of assessing whether any particular ad contains relevant information.

The other defining aspect of the skippable ad in online settings is the ability of content platforms to attribute payments from advertisers based on the viewers engagement with the ad. For instance, YouTube does not charge advertisers if the viewer skips to the desired video before finishing the advertisement. This aspect distinguishes our setting from other forms of *interactive advertising*, such as a print ad in a magazine or traditional newspaper, in which payment attribution is not possible.

The skippable ad is not restricted to YouTube's *TrueView* ads. This format is increasingly used by online and mobile content platforms. Some online news sites, for instance, deliver ads that intrude on content, but can be removed by clicking on an appropriate icon. Pop-ups on content sites ensure the reader sees some aspect of the ad, but the reader can choose whether to spend more time reading the ad or to close it and move on to the intended content. Even a non-pop-up ad along the side of content, known as the *rollover* or *mouse-over* ad, gives the viewer some initial information about the advertiser with the option to pursue more information by rolling the cursor over on the ad. Given the rise of this interactive ad format, our research poses the question: *What determines the consumer's decidion to skip an ad or watch it entirely?* This question provokes us to further ask: *What are the implications of skippable ads for advertisers and content platforms?*

A potential benefit of the skippable ad is that a consumer has more options relative to traditional ads. Any consumer can view the entire ad, as with traditional ads, but now also has the option to skip to the content sooner, without watching the ad in full. Does this option hurt the advertiser's ability to entice a potential customer? Or, does it help the platform to attract more potential customers for the advertiser? To address these questions, we examine the impact of the skippable ad format on advertising demand on the content platform. Specifically, we study the viewer's dynamic interaction with the information contained in the advertisement. We embed this interaction in a model of a two-sided market with a commercial platform selling skippable ads to advertisers, which

[1] See "YouTube is Getting Rid of Unskippable 30-Second Ads" in *Variety*, February 17, 2017.

are viewed by consumers interested in the platform's content. This model allows us to determine the impact of the skippable ad format on viewer demand for the platform and on the likelihood of *viewer conversion* (e.g., a purchase or favorable impression) as a result of the advertiser's message.

We treat a skippable advertisement as a series of informative signals about the advertiser's product arriving sequentially (Branco et al. 2012). The consumer is obliged to receive one signal, which is used to assess the value of acquiring another signal. Not surprisingly, we find that a consumer will skip to the desired content (e.g., the video), thereby ignoring the second signal, whenever she expects that the advertiser is not relevant for her. But we also find that a consumer may skip to content when she is immediately certain the product is relevant. As such, the advertiser need not pay for the ad even though the advertiser converts the viewer.

Our model also has two implications for the platform and the advertising market. First, we find that the introduction of skippable ads increases demand for the platform. Skippable ads give the consumer more flexibility about viewing an advertisement and the possibility of skipping to content when she believes the advertisement does provide decisive information. Therefore, relative to the traditional ad format, consumers achieve higher utility from participating on the platform. We refer to this as the *demand-enhancing effect* of skippable ads. While the demand-enhancing effect provides a direct benefit to the platform, it also induces reactions in the advertising market. Specifically, with higher demand from viewers, the platform sells to more advertisers. Consumers have a further, indirect benefit from the increased advertising because they have a better chance of finding a relevant advertiser.

Second, relative to the traditional ad format, we find that the skippable ad changes how likely a viewer will be converted. We call this change the *viewer conversion effect*. Our model shows that the consequences of the viewer conversion effect depend crucially on the viewer's predisposition toward the advertiser. If a viewer is positively inclined toward the advertiser, then skippable ads actually raise the likelihood of conversion. This follows from the intuition that a viewer, who has a high prior on an advertiser, is more willing to skip the ad because she is relatively certain that she likes the advertiser's product and does not need to wait for another signal. In this case, she can skip directly to her content without delay and the advertiser does not need to pay the platform. Interestingly, we find that the platform benefits in this case as well because it raises ad rates to account for the viewer's expressed interest in receiving an additional signal. As a result, when the consumer's predisposition (or prior) toward the advertiser is positive, then skippable ads are a Pareto improvement over traditional ads.

This process works in reverse when the viewer has a negative prior for the advertiser. Even if the signal from the advertiser is encouraging, it may not be enough to overcome the viewer's prior or prevent her from skipping to the content. However, under traditional ads, the viewer must receive the second signal, which may be unexpectedly large and enough to convert her. Thus, the traditional ad format may be preferable by advertisers and the platform when consumers have a negative predisposition toward advertisers.

Prior work on commercial media and two-sided media markets has tended to treat consumers as passive in receiving information through advertising (e.g. Dukes 2004, Anderson and Coate 2005, Gal-Or and Gal-Or 2005). That is, consumers passively receive all information from the advertiser. With skippable ads, however, consumers pay active attention to part of an advertisement and immediately evaluate whether to acquire additional informational content. This dynamic interaction implies fundamental differences in how ads are valued and sold. In particular, as we show, viewers who choose to receive more information about an advertiser are potentially more valuable to that advertiser, which affects ad pricing and the number of ads on the platform. Furthermore, in prior work, any advertisement is assumed to intrude on content as represented by a nuisance cost to the viewer. (Dukes and Gal-Or 2003, Godes, Ofek, and Savary 2009, and Kaiser and Song 2009). In our model, nuisance cost is specified as the viewer's discounted utility from waiting for content. Therefore, and in contrast to the earlier research, skippable ads imply that the amount of nuisance costs is determined by the viewer's decision whether or not to skip the ad. In this way, the incurrence of nuisance cost is directly linked to the relevance of the advertiser's information. If the viewer values the additional information in an advertisement, then she is willing to incur the nuisance of delayed content. This distinction implies a source of efficiency gains created by the skippable ad format, which has not been identified in prior work.

The ability of consumers to avoid an advertisement has traditionally been a worry for marketers. One reason marketers have been concerned about ad avoidance is that it confounds the advertiser's ability to communicate to consumers. For example, if a viewer is uninterested in a commercial, she can engage in *behavioral skipping*, such as leaving the room, socializing, or switching channels (known also as "zapping"). Prior work has examined the conditions affecting behavioral skipping and how to account for it when assessing advertising effectiveness (Siddarth and Chattopadhyay 1998 and Wilbur et al. 2013). Ad skipping in the context of interactive ads differs from behavioral skipping because (i) the consumer has the ability to avoid delay in accessing the desired content; and (ii) the content provider can charge the advertiser according to whether the ad was viewed or skipped. As we show, these two distinguishing features have new implications for viewers and advertisers.[2]

The inclination for some viewers to avoid advertising has inspired versioning options that include "ad-free" options, such as *YouTube Red*, and the creation of ad-avoidance technology (AAT), such as *Tivo*. Offering the consumer the option to pay for an ad-free option can be viewed as a form of quality segmentation via versioning (e.g. Halbheer et al. 2014) in the context of the product-line design literature and, therefore, is not the focus of our work. More connected to our research, by contrast, is AAT, which is aimed at blocking ads entirely from commercial content. For example, digital video recorders (DVR's) allow the user to bypass all television advertisements and browser-based ad blockers eliminate all internet ads. Much of the research on AAT examines the consequences of such technology for advertisers and the content provider. This research suggests that AAT reduces the platform's advertising revenue and alters the composition of consumers. Furthermore, the technology

[2] Marketers can also make an advertisement worthy of watching (e.g. Gustafson and Siddarth 2007 and Teixeira et al. 2010). We do not consider strategic aspects of the advertiser to affect viewer attention to a commercial.

induces the platform to increase advertising volumes, choose lower content quality and more mass-market content through consumers' self-segmentation (e.g Wilbur 2008, Ghosh and Stock 2010, Anderson and Gans 2011, and Johnson 2013). Unlike with skippable ads, consumers employing AAT decide to skip *all* ads without considering whether any individual ad might be relevant. Consequently, with skippable ads, consumers make an informed choice about whether a particular ad is worth considering more seriously. This distinction implies that viewers have more individual control on what ads to watch and how advertisers are charged.[3] Consequently, and in contrast to the literature on AAT, we find that skippable ads can have benefits to both the platform and the advertisers.

2 Model Basics

The fundamental element of our model is the viewer's dynamic interaction with a skippable ad. This forms the foundation of a broader model of a two-sided market that includes a monopoly platform and a set of advertisers. Advertisers communicate with viewers through commercial advertisements sold by the platform. Advertisements are a sequence of noisy signals that viewer uses to infer the benefit of engaging with the advertiser (e.g. purchase). The viewer's interaction with the sequence of signals mimics the gradual learning process of Branco et al. (2012).

We consider a three period model. In period 0, the platform decides advertising rate and advertisers decide whether or not to show an advertisement on the platform. Consumers observe how many advertisements have been sold and then decide whether or not to join the platform and engage in content. In period 1, the consumer, if exposed to an ad, receives a noisy signal about a random advertiser's product. If the platform utilizes the skippable ad format, then the consumer decides whether to skip the rest of the ad. If she skips, then she enjoys content immediately. Otherwise, she decides to delay the content in order to acquire a second signal from that advertiser in period 2. After acquiring the second signal, the viewer enjoys the content at end of period 2. At any point in this process, the viewer can decide to engage with the advertiser.

We are agnostic with regard to how the viewer engages with the advertiser. Generally, we refer to a *conversion* process, which defines the point at which the advertiser obtains some value from the viewer's engagement. This can mean the viewer purchases the advertiser's product or simply that the viewer obtained a favorable impression of the advertiser. This approach allows us to consider a wide variety of motivations for advertising without specifying the details of how advertiser profits from consumer engagement.

We start with viewer behavior, which is the major component of our analysis. At the end of period 1, the viewer decides whether to skip the ad or to continue receiving information. If the viewer skips ad and enjoys the content (e.g. watch video) right away, the viewer obtains a utility of

[3] Recent work in economics and marketing studies the mechanisms designed to improve attribution methods (Berman 2015, Lewis and Rao 2016). While payment attribution is used with skippable ads and accounted for in our model, it is not our focus.

$w > 0$. The value w is the same for all viewers.[4] If the viewer continues watching the ad in period 2 (e.g. the video is delayed until the end of period 2), then the value of watching content will be discounted to δw, where $\delta \in (0, 1)$.

The true match value for a viewer, if converted by the advertisement, is $v_c \sim N(\mu, s^2)$. The mean μ is the consumer's prior on the value of the product, which can be positive or negative – a parameter that provides crucial conditions for our results. An ad is defined by two noisy signals about v_c arriving sequentially and independently. The signals help inform the viewer about the value of engaging with the advertiser at some later time (e.g. seeing the advertised movie). Each signal v_i is generated after watching period $i = 1, 2$ of the ad, with

$$v_i = v_c + \varepsilon_i, \quad i = 1, 2,$$

where $\varepsilon_i \sim N(0, \sigma_i^2)$. We assume for simplicity that $\sigma_i^2 = \sigma^2$, which implies $v_i \sim N(\mu, s^2 + \sigma^2)$. Under the *traditional ad format*, the viewer must receive both signals.

Under the *skippable ad format*, the viewer may skip signal v_2 but not the first signal v_1. At the end of period 1, the viewer decides whether or not to skip to the desired content. If the viewer does not skip, then she receives the second signal v_2 in period 2. An advertiser has a willingness to pay t for a viewer who converts. An interval of advertisers of measure 1 is distributed with CDF $G(t)$ and density function $g(t)$. We assume that $G(t)$ is invertible on its support.

The platform chooses advertising rate r. This is the rate per viewer under traditional ads and the rate per non-skipping viewer under skippable ads. Note that under skippable ads, if the viewer receives signal v_1, but not v_2, then the advertiser will not pay to the platform. In our model, when a viewer does not skip an ad, two things happen: (1) the viewer receives a second signal; (2) the viewer will have to wait for the requested video content so w is discounted. In practice, a viewer may become interested in the advertising content and click the advertiser link. In this case, the viewer will obtain more information about the advertiser (second signal), and the requested video will also be delayed. Therefore, clicking through an advertiser link is similar to our case of not skipping ads.[5]

In the next section, we analyze the viewer's interaction with a skippable ad. Specifically, we study the viewer's decision of whether to skip the ad and enjoy the intended content, or to wait for another signal v_2. In Section 4, we then embed that model into the general two-sided market model with advertiser decisions before comparing the equilibrium outcomes with the skippable ad format and the traditional ad format.

[4]Though we focus on ads appearing at the start of content consumption, we can interpret it as occurring at any point. For instance, longer videos on YouTube may include periodic interruption with skippable ads after the video begins. In this case, our model applies at the start of the ad so long as there is additional content to be consumed afterwards.

[5]YouTube charges the advertiser if the viewer watches the whole ad or clicks on the advertiser link.

3 The Viewer Model: Interaction with the Skippable Ad

The decision of whether to skip the ad after receiving the first signal, v_1 is the central aspect of the viewer model. The viewer must assess the expected value of having both signals, conditional on the first, balanced with the cost of delaying content. That is, the viewer determines whether she expects to engage with the advertiser because of the second signal, given the first. This engagement, or conversion from the advertiser's perspective, is defined by whether her expectation of v_c exceeds zero. That is, the expected conversion condition is $E(v_c|v_1, v_2) > 0$. As we show below, this expectation depends on the distribution of the sum $v_1 + v_2$. The following lemma specifies both the distribution of viewer's value at the end of period 1 and a necessary and sufficient condition for the viewer to be converted.

Lemma 1 *Let $v_c \sim N(\mu, s^2)$ and $v_i = v_c + \varepsilon_i$, where $\varepsilon_i \sim N(0, \sigma^2)$ are i.i.d., $i = 1, 2$. Then*

(i) The sum of the two signals, conditional on the first, follows a normal distribution:

$$(v_1 + v_2)|v_1 \sim N(A, B^2)$$

where

$$A = v_1 + \mu + \frac{s^2}{s^2 + \sigma^2} \cdot (v_1 - \mu), \quad B^2 = \frac{\sigma^2(2s^2 + \sigma^2)}{s^2 + \sigma^2}.$$

(ii) With two signals v_1 and v_2, the conversion condition is

$$E(v_c|v_1, v_2) \geq 0 \Leftrightarrow v_1 + v_2 \geq -\mu \cdot \frac{\sigma^2}{s^2}.$$

Proof. See the Appendix for proofs of all formal claims. ■

Lemma 1 (i) establishes that the distribution of the advertiser's value to the viewer conditional on the information from the first part of the ad is normally distributed. This distribution features prominently in the subsequent analysis of the viewer's skipping decision.

The condition for viewer conversion in Lemma 1 (ii) says that, after receiving both signals, the viewer will convert with the advertiser exactly when the conditional mean of $E(v_c|v_1, v_2)$ exceeds zero. As we will see in Section 4, this condition also applies to the case of traditional ads because the viewer receives both signals. Finally, Lemma 1 shows that with a strong prior (e.g. $\mu >> 0$), the viewer may be converted even if $v_1 + v_2 < 0$.

With skippable ads, the viewer can forego the information supplied by v_2, thereby deciding whether or not to purchase the product with only one signal. The viewer can also watch the second period ad and then use both signals to make purchasing decisions. There are three options for her to consider. The first two options involve the viewer skipping the second part of the ad after getting the first signal v_1. If she converts, her expected utility as a function of her first signal v_1 is

$$U_1^{skip}(v_1) = E(v_c|v_1) + w = E(v_1 - \varepsilon_1|v_1) + w = v_1 - E(\varepsilon_1|v_1) + w$$

$$\Rightarrow U_1^{skip} = v_1 - \frac{\sigma^2}{s^2+\sigma^2}(v_1 - \mu) + w = \frac{s^2 v_1 + \sigma^2 \mu}{s^2+\sigma^2} + w. \tag{1}$$

If she does not convert, her utility is simply

$$U_2^{skip}(v_1) = w, \tag{2}$$

which is the utility gained by directly viewing the desired content and not engaging with the advertiser. This value is independent of her first signal.

A third option for the viewer is to continue watching the ad, obtain another signal v_2, and then decide whether or not to buy. Once the viewer receives the second signal, the optimal conversion decision is given in Lemma 1 (ii). The corresponding expected utility is

$$U_3^{skip}(v_1) = E\left[v_c \,\middle|\, v_1, v_1 + v_2 \geq -\mu \cdot \frac{\sigma^2}{s^2}\right] \cdot Pr\left[v_1 + v_2 \geq -\mu \cdot \frac{\sigma^2}{s^2} \,\middle|\, v_1\right] + \delta w.$$

As is evident from this expression, the viewer's expected value of the advertised product conditional on v_1 has a truncated Normal distribution. This truncation distribution is defined by the condition in Lemma 1 where the viewer assesses whether, upon receiving the second signal v_2, the signals' sum will exceed the purchase threshold $-\mu(\sigma^2/s^2)$. The endowed properties of the Normally distributed signals imply we can derive a closed form expression for the viewer's utility from viewing the second part of the ad, given in Lemma 2.

Lemma 2 *The viewer's expected utility from viewing two signals, conditional on the first signal,* v_1*, is*

$$U_3^{skip}(v_1) = \frac{s^2 v_1 + \sigma^2 \mu}{s^2+\sigma^2} \cdot [1 - \Phi(\alpha)] + \frac{1}{2}\left(B - \frac{C}{B}\right) \cdot \phi(\alpha) + \delta w, \tag{3}$$

where

$$\alpha = -\frac{\mu(\sigma^2/s^2) + A}{B} \quad \text{and} \quad C = \frac{\sigma^4}{s^2+\sigma^2}$$

with A *and* B *given in Lemma 1 and* $\phi(\cdot)$ *and* $\Phi(\cdot)$ *are the pdf and CDF of standard normal distribution. Furthermore,* U_3^{skip}*:*

(i) increases with δ, w*;*

(ii) increases with μ*;*

(iii) is in general non-monotonic in s *and* σ *except under some special cases:*

(iii-a) increases with s *if* $v_1 > \mu > 0$*;*

(iii-b) decreases with σ *if* $v_1 > \mu > 0$ *and* $\sigma^4 > 2s^4$*, but increases with* σ *if* $\sigma^4 < 2s^4$ *and* $\mu > v_1 > 0$.

The first two terms in the expression U_3^{Skip} are the expected value of the advertiser's product, conditioned on it being acceptable to the viewer. The variable α is the expected value of the sum $v_1 + v_2$ conditional v_1 (which is A), accounting the truncation cutoff. (The term C is the covariance of the sums $v_1 + v_2$ and $\varepsilon_1 + \varepsilon_2$.) An important property of equation (3) is that it strictly increases with v_1. Recall from Lemma 1 that $E[(v_1 + v_2)|v_1]$ is increasing in v_1. As a result, the viewer's expected utility from conversion after two signals, as dictated by option 3, increases.

Lemma 2 also describes several comparative properties of U_3^{skip}. Part (i) is immediate as δ and w enter into U_3^{skip} only positively in the form of δw. As either parameter increases, the viewer has greater benefit from eventually viewing the content. Part (ii) indicates that a higher prior on the advertiser increases the expected value of conversion after viewing two signals. Recall that s is a measure of the dispersion of consumer's true value, v_c about μ. For part (iii-a), $v_1 > \mu$ indicates a good signal (better than the prior). If s is small, then most likely this occurs because of the noise in the signal generating process. By contrast, if s is large, then the prior becomes less informative and thus a good signal becomes more valuable. For Part (iii-b), having a good signal ($v_1 > \mu$) is more beneficial if the signal becomes less noisy (σ decreases), while having a bad signal ($v_1 < \mu$) is less damaging if the signal becomes more noisy (σ increases).

The characterization of U_3^{skip} in (3) permits us to derive the viewer's optimal skipping decision as a function of v_1. Define the indifference conditions $U_3^{skip} = U_2^{skip}$ and $U_3^{skip} = U_1^{skip}$ and $v_1 = v_L, v_H$ as solutions to each equation, respectively. In order to derive meaningful solutions to these equations, we impose the following condition:

Assumption 1: $U_3^{skip}(v_1 = -\mu \cdot \sigma^2/s^2) > w$.

This assumption simply states that, for the "average" first signal, the viewer prefers to continue watching the ad. As indicated in the following result, it assures us that v_L and v_H have meaningful interpretations.[6]

Lemma 3 *There exist unique values $v_1 = v_L$ and $v_1 = v_H$ such that $U_2^{skip}(v_L) = U_3^{skip}(v_L)$ and $U_1^{skip}(v_H) = U_3^{skip}(v_H)$. Moreover, $v_H > -\mu(\sigma^2/s^2) > v_L$ holds if and only if Assumption 1 holds.*

Figure 1 plots the utilities for the 3 options under skippable ads, as well as the threshold values v_L and v_H. The uniqueness of v_L and v_H is implied by the following single-crossing conditions, that is, U_3^{skip} crosses U_1^{skip} and U_2^{skip} each exactly once. This is because,

$$\frac{\partial U_1^{skip}}{\partial v_1} > \frac{\partial U_3^{skip}}{\partial v_1} > \frac{\partial U_2^{skip}}{\partial v_1}. \tag{4}$$

The intuition for (4) follows from the fact that v_1 and v_c are positively correlated. Having a higher v_1 implies, on average, a larger v_c which raises the expected value of purchasing the product. Having a second signal will not change this. Thus we have $\frac{\partial U_1^{skip}}{\partial v_1} > 0$. On the other hand, U_3^{skip} is not as responsive to changes in v_1 as U_1^{skip} is. In U_1^{skip}, the viewer has only one signal and buys the product. In contrast, in U_3^{skip}, there are two signals. The second signal v_2 is positively correlated to v_1 but only imperfectly. Furthermore, the viewer has the option to not purchase the product in which case having lower v_1 and v_2 won't affect the viewer utility.

[6] Assumption 1 implies that the skipping rate is positive. As $U_3^{skip}(v_1 = -\mu \cdot \sigma^2/s^2) \downarrow w$, skipping rate converges to 1. Conversely, as $U_3^{skip}(v_1 = -\mu \cdot \sigma^2/s^2) \uparrow \infty$, skipping rate converges to 0.

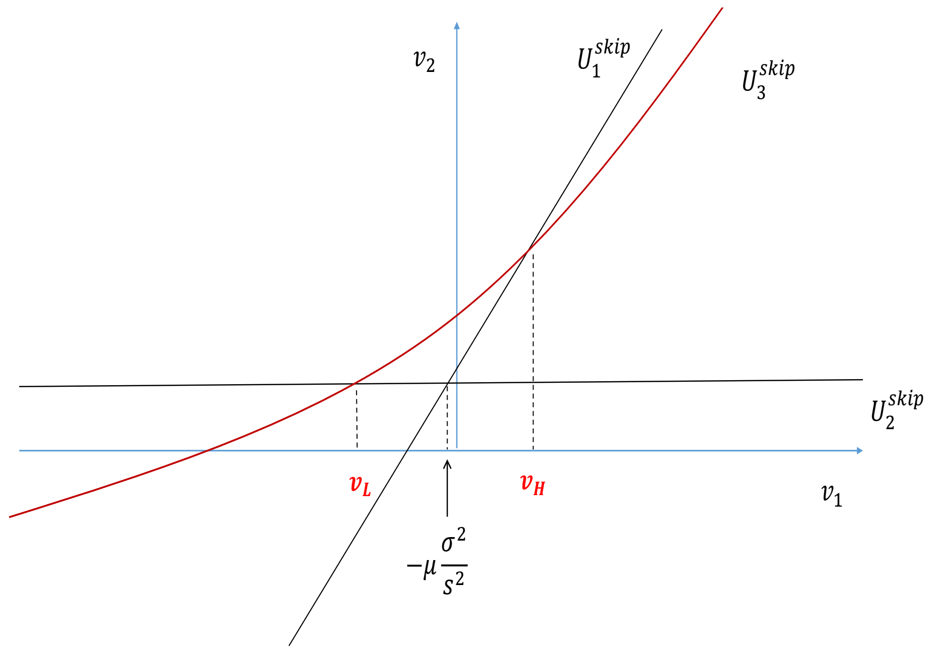

Figure 1: Viewer options under skippable ads

The definitions of v_L and v_H and Lemma 3 imply the following result.

Proposition 1 *Under the skippable ads format, after a viewer obtains the first signal v_1, the viewer's optimal viewing decision is characterized as follows:*

(i) If $v_1 \leq v_L$, the viewer skips the ad and does not convert;

$$Pr(v_1 \leq v_L) = \Phi\left(\frac{v_L - \mu}{\sqrt{s^2 + \sigma^2}}\right).$$

(ii) If $v_1 \geq v_H$, the viewer skips the ad and converts;

$$Pr(v_1 \geq v_H) = 1 - \Phi\left(\frac{v_H - \mu}{\sqrt{s^2 + \sigma^2}}\right).$$

(iii) If $v_1 \in (v_L, v_H)$, the viewer continues viewing the ad to obtain another signal v_2, and converts if and only if $v_1 + v_2 \geq -\mu(\sigma^2/s^2)$;

$$Pr(v_L < v_1 < v_H) = \Phi\left(\frac{v_H - \mu}{\sqrt{s^2 + \sigma^2}}\right) - \Phi\left(\frac{v_L - \mu}{\sqrt{s^2 + \sigma^2}}\right).$$

Proposition 1 fully specifies the viewer's reaction to the first signal v_1. Furthermore, it reports the probability of each option being chosen. The viewer will skip the second part of the ad for large and small values of v_1. In these situations, the viewer is relatively certain that another signal will not be decisive in changing her mind about converting. For intermediate values of v_1, the viewer seeks another signal from the advertiser to become better informed about making her conversion decision. The opportunity cost of acquiring another signal is the delay in viewing her desired content, $(1-\delta)w$.

It is worth mentioning that the result in Proposition 1 shows how skippable ads offer a novel perspective on the notion of *nuisance costs* used in prior models of commercial media (e.g Masson et al. 1990). Earlier models of two-sided platforms with informative advertising focus on the traditional format and assume information is passively accepted by the viewer and that nuisance costs are incurred automatically when joining the platform (e.g. Dukes 2004, Anderson and Coate 2005, Peitz and Valletti 2008). The proposition summarizes how the consumer trades off the informational benefits of an advertisement with the cost of delayed content - the nuisance cost. In particular, skippable ads give the viewer the option to pay a nuisance cost in exchange for information that can be useful in making a decision about the advertiser.

The following proposition establishes several properties of the thresholds v_L and v_H.

Proposition 2 *The cutoff values v_L and v_H have the following properties.*

(i) $v_L + v_H = -2\mu(\sigma^2/s^2)$.

(ii) v_L decreases with μ, δ, increases with w, but can increase or decrease with s and σ.

(iii) v_H decreases with μ, w, increases with δ, but can increase or decrease with s and σ.

From part (i), we see that $\frac{v_L+v_H}{2}$, the average value of the decision thresholds for v_1, depends only on the distributional parameters equals the conversion threshold given in Lemma 1. Under Assumption 2, these thresholds lie on either side of the conversion threshold $-\mu(\sigma^2/s^2)$. As μ increases, thresholds v_L and v_H shift leftward in Figure 1 regardless of the sign. An increase in $(1-\delta)w$ tends to squeeze these values, thereby shrinking the range of values for v_1 that induce watching the entire ad. The ambiguous comparatives statics given in parts (ii) and (iii) can be seen as follows. First, suppose that $\sigma/s \to 0+$, which implies that signals become more accurate indicators of v_c. This implies that the second signal is mostly a repetition of the first signal and there is little to gain from waiting for the second signal. Now, suppose that $\sigma/s \to +\infty$. In this case the signal is uninformative, and consequently an additional signal adds little value. It is easy to see that when σ/s is some intermediate ranges, the value of the second signal will be higher.

4 Comparing Traditional and Skippable Ads

Having developed the viewer model and its important properties, we are now in a position to compare ad formats. Our ultimate objective is to compare how a change from the traditional format to the skippable format affects viewers, the platform, and advertisers. To that end, our first task is to assess how the ad format affects different rates of conversion. Conversion rates affect the benefit derived by advertisers and viewers when deciding whether to participate on the platform, which we explore in Subsections 4.2 and 4.3. Finally, in Subsection 4.4, we derive the overall impact of skippable ads on advertisers' surplus and platform profits.

We use the traditional format as the benchmark. Implementing the traditional format in our model is to assume that the viewer is obligated to receive both signals from the advertiser. This is equivalent to option 3 under skippable ads, that is,

$$U^{trad}(v_1) = U_3^{skip}(v_1), \quad \forall v_1,$$

where the U_3^{skip} expression is given earlier in equation (3).

Any impact on advertisers is a consequence of how viewers react to the skippable ad format. Therefore, we start our comparison by examining how skippable ads affect viewer decisions using the model developed above. There are two effects. First, the type of ad format affects the utility a potential viewer obtains from joining the platform. Consequently, a platform that switches to the skippable ad format can induce more viewers on the platform and thereby increase the reach of an advertiser. We call this effect the *demand enhancing effect*. Second, the type of ad format affects the consumers' expected benefit from the advertiser's product or service. We use model described above to assess the degree to which consumers are more or less likely to convert to the advertiser, which we call the *viewer conversion effect*. Obviously, any change in the viewer conversion rate will affect the demand for advertising on the platform. We start with the impact on viewer conversion.

4.1 Viewer Conversion

We first compare viewer conversion rate under traditional ads and skippable ads respectively. Recall that viewer's ad skipping decision depends on the single signal v_1 and how it compares to v_L and v_H. This defines three conditions as follows:

$$C_{1a} : v_1 \geq v_H \quad \text{(Skip Ad and Convert)}$$

$$C_{1b} : v_1 \in (v_L, v_H) \quad \text{(View Entire Ad)}$$

$$C_{1c} : v_1 \leq v_L \quad \text{(Skip Ad and Not Convert)}.$$

The viewer conversion condition, in the presence of two signals, defines two more conditions:

$$C_{2a} : v_1 + v_2 \geq -\mu\frac{\sigma^2}{s^2} \quad \text{(Convert after Entire Ad)}$$

$$C_{2b} : v_1 + v_2 < -\mu\frac{\sigma^2}{s^2} \quad \text{(Not Convert after Entire Ad)}.$$

The conditions above are graphically portrayed in Figure 2. The conditions for skipping behavior, C_{1a}, C_{1b}, and C_{1c}, are depicted by the sets partitioning all realizations of the first signal v_1. These are represented by the vertical strips to the right of v_H (C_{1c}), to the left of v_L (C_{1a}), and in between

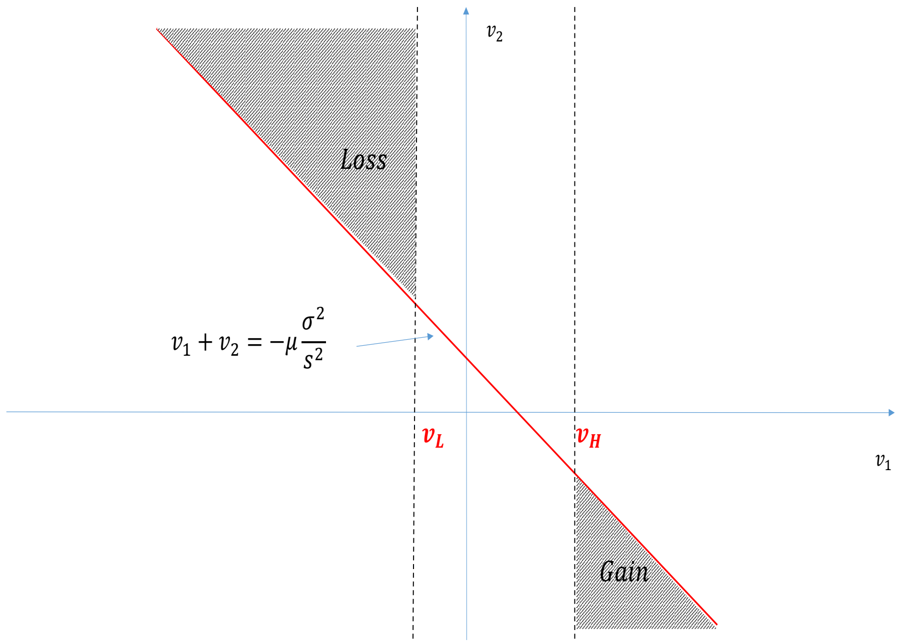

Figure 2: Viewer Conversion Effect

v_L and v_H (C_{1b}). The conditions for conversion with two signals are depicted by the upper and lower planes defined by the diagonal line $v_1 + v_2 = -\mu(\sigma^2/s^2)$ (C_{2a} and C_{2b}, respectively).

Denote by VCR^{trad} the *viewer conversion rate conditional on seeing a traditional ad* and by VCR^{skip} the *viewer conversion rate conditional on seeing a skippable ad* as VCR^{skip}. Then

$$VCR^{trad} = Pr(C_{2a}) \quad \text{and} \quad VCR^{skip} = Pr(C_{1a}) + Pr(C_{1b} \cap C_{2a}). \tag{5}$$

Moving from traditional ads to skippable ads, there is a gain in the viewer conversion rate when $v_1 \geq v_H$. Under skippable ads, the viewer will skip the ad and buy the product. Under traditional ads, however, the viewer may decide not to buy the product after getting the second signal. By contrast, when $v_1 \leq v_L$, traditional ads can lead to a purchase that would never happen under a skippable ad. Specifically, with traditional ads, the viewer may obtain a very large value of v_2 that more than offsets the low initial signal. The two impacts of the skippable ad format on viewer conversion can be seen graphically in Figure 2. The shaded region in upper portion of the figure is the loss in conversion due to skippable ads. In this region, a viewer would have converted in a traditional ad because she received a very large second signal that offsets her pessimistic first signal ($v_1 < v_L$). The shaded region in the lower portion of the figure is the gain in conversion due to skippable ads. In this region, the viewer is converted immediately due to the strong first signal ($v_1 > v_H$) without having to waiting for the second signal, which would have been sufficiently negative to make her unfavorable to the advertiser. As the following proposition establishes, the relative strength of these two impacts is precisely determined by the viewer's prior belief of an advertiser's product.

Proposition 3 *(Viewer Conversion)* $VCR^{skip} - VCR^{trad}$ *has the same sign as the viewer's prior* μ*. Therefore, relative to traditional ads, viewer conversion is more likely under skippable ads if and only if* $\mu > 0$*.*

Under either format, the first signal is necessary for the viewer to be aware of the advertiser. The condition $\mu > 0$ means that she is predisposed favorably toward that advertiser and that first signal is likely to be supportive of conversion. Suppose that $v_1 > v_H$ so that the viewer is convinced of the advertiser without the second signal. With skippable ads, she converts immediately and foregoes the second signal v_2. With traditional ads, however, she is forced to wait for the second signal, which with some probability may be so low that $v_2 < -v_1 - \mu(\sigma^2/s^2)$ and ends up *not* converting. That is, when $\mu > 0$, traditional ads reduce the conversion rate. The opposite is the case for when $\mu < 0$. This intuition reflects the consumer information search problem studied in Branco et al. (2012).

4.2 Platform Participation by Viewers

Next, we characterize viewer's platform participation decision under each ad format. We determine the measure of viewers who will join the platform and in turn be reached by advertisers. For now, fix the measure of advertisers who will advertise on the platform and denote it by $N > 0$.[7] Let $h(N)$ denote the probability of each viewer seeing an ad. We restrict $h(N) < 1$ and rule out the possibility that a viewer may see more than one ad.

Under traditional ads, the marginal consumer d, which also represents the demand for the platform, is determined by

$$U^{trad}(d) = (1 - h(N)) \cdot w + h(N) \cdot U_{ad}^{trad} - d,$$

where

$$U_{ad}^{trad} \equiv \delta w + E(v_c | C_{2a}) \cdot Pr(C_{2a}), \tag{6}$$

is the expected utility of the viewer conditional on her seeing a traditional ad. Similarly, under skippable ads, the marginal consumer d is determined by

$$U^{skip} = (1 - h(N)) \cdot w + h(N) \cdot U_{ad}^{skip} - d$$

where

$$\begin{aligned} U_{ad}^{skip} &= [w + E(v_c | C_{1a})] \cdot Pr(C_{1a}) + w \cdot Pr(C_{1c}) \\ &+ [\delta w + E(v_c | C_{1b} \cap C_{2a}) \cdot Pr(C_{2a} | C_{1b})] \cdot Pr(C_{1b}). \end{aligned} \tag{7}$$

is the expected utility of the viewer conditional on seeing a skippable ad.

Consumers have the option of not joining the platform and obtaining zero utility. Suppose consumers are characterized by their value of d, uniformly distributed on the interval $[0, \bar{d}]$ with

[7] Because our emphasis is on the viewer's dynamic interaction with a skippable ad, we have treated the advertising market with less detail. For instance, our model does not account for important competitive interactions of advertisers or the platform's strategic allocation of advertisers across viewers.

density 1. Then platform demand under the two different ad regimes is defined by the thresholds $d^i(N)$, $i \in \{skip, trad\}$, for which joining gives zero utility:

$$d^{trad}(N) = (1 - h(N)) \cdot w + h(N) \cdot U_{ad}^{trad}, \tag{8}$$

$$d^{skip}(N) = (1 - h(N)) \cdot w + h(N) \cdot U_{ad}^{skip}. \tag{9}$$

A comparison of (6) and (7) gives us that $U_{ad}^{skip} > U_{ad}^{trad}$ and, therefore, the ranking of (8) and (9) for fixed N.

Lemma 4 *(Direct Demand-Enhancing Effect) For any fixed $N > 0$, we have $d^{skip}(N) > d^{trad}(N)$.*

Skippable ads are unambiguously beneficial for consumers because they give control to the viewer in what ads they see and do not see. In particular, viewers can avoid delay in accessing the desired content when it is unlikely that the advertiser will be relevant. We refer to it as the *direct* demand-enhancing effect because the result is obtained under the condition that N is the same under the two ad formats. The condition that N is constant is equivalent to assuming that the advertising market does not react to a switch in the advertising regime. This is considered in the next section.

4.3 Impact on the Advertising Market

To understand the reaction of the advertising market, we first derive an individual advertiser's decision to buy an ad and then determine the platform's optimal number of ads. Recall from Section 2, each advertiser is characterized by a $t > 0$, which represents the value of a conversion. The distribution of advertisers is described by the cdf $G(t)$, which has density $G'(t) = g(t) > 0$ and is everywhere invertible.

Given any advertising rate $r > 0$, advertiser t advertises with traditional ads if and only if $t \cdot VCR^{trad} \geq r$. This defines a threshold for advertiser participation: $t \geq t^{trad} \equiv r/VCR^{trad}$. Then advertising demand is measured by the number advertisers who participate on the platform: $N = 1 - G(t^{trad})$. Therefore, the advertising rate as a function of the number of advertisers is

$$r^{trad}(N) = VCR^{trad} \cdot G^{-1}(1 - N). \tag{10}$$

Because VCR^{trad} is independent of r and N, and G is monotonic, there is a one-to-one correspondence between r and N. Therefore, we can simplify the platform's decision as choosing N to maximize its expected advertising profit (assuming zero cost):

$$N^{trad} \equiv \underset{N}{\operatorname{argmax}} \underbrace{VCR^{trad} \cdot G^{-1}(1 - N)}_{r^{trad}(N)} \cdot h(N) \cdot d^{trad}(N). \tag{11}$$

With skippable ads the advertiser needs to pay only when the viewer does not skip, which occurs with probability $Pr(C_{1b})$. Advertiser t advertises if and only if $t \cdot VCR^{skip} \geq r \cdot Pr(C_{1b})$. Therefore, the threshold for advertiser participation is $t \geq t^{skip} \equiv r \cdot Pr(C_{1b})/VCR^{skip}$, which defines an advertiser demand facing the platform $N = 1 - G(t^{skip})$. The corresponding advertising rate for skippable ads is then

$$r^{skip}(N) = \frac{VCR^{skip}}{Pr(C_{1b})} \cdot G^{-1}(1 - N). \tag{12}$$

The platform's problem with skippable ads is to choose N to maximizes its expected advertising profit

$$N^{skip} \equiv \underset{N}{\text{argmax}} \underbrace{VCR^{skip} \cdot G^{-1}(1 - N)}_{r^{skip}(N) \cdot Pr(C_{1b})} \cdot h(N) \cdot d^{skip}(N). \tag{13}$$

The advertising rate r^{skip} under skippable ads is the advertising rate per non-skipping viewer. On a *per platform viewer* basis, the advertising rate is $Pr(C_{1b}) \cdot r^{skip}$. Thus, the direct impact of skippable ads on per viewer ad rates, holding viewer participation constant is

$$Pr(C_{1b}) \cdot r^{skip} - r^{trad} = \left(VCR^{skip} - VCR^{trad}\right) \cdot G^{-1}(1 - N).$$

Thus, the direct impact of skippable ads on advertising rate is determined precisely by the sign of μ, as determined by the viewer conversion effect in Proposition 2.

The following proposition summarizes the impact of skippable ads on the equilibrium levels of advertisers, viewers, and the total number of advertisements shown on the platform.

Proposition 4 *If the platform uses skippable ads instead of traditional ads, then there are*

(i) more advertisers on the platform, $N^{skip} > N^{trad}$ *with* $t^{skip} < t^{trad}$*; and*

(ii) more viewers on the platform, $d^{skip}(N^{skip}) > d^{skip}(N^{trad}) > d^{trad}(N^{trad})$.

This result says generally that skippable ads generate more activity on the platform, relative to traditional ads. Part (i) says the platform attracts more advertisers with the skippable format. Part (ii) shows that viewers benefit further from this format because they have a better chance of seeing a relevant ad. The first inequality $d^{skip}(N^{skip}) > d^{skip}(N^{trad})$ is what we refer to as *indirect demand enhancing effect* of skippable ads. To see this, recall from (6) and (7) that $U_{ad}^{skip} > \max\{w, U_{ad}^{trad}\}$ so that the marginal impact of N on d^{skip} is always positive and bigger than on d^{trad}. Therefore, $N^{skip} > N^{trad}$ implies viewers have a better chance that they will be converted by an advertiser. The second inequality in part (ii) is from Lemma 4.

4.4 Impact on Profits

The above results make clear that skippable ads have direct and indirect benefits to viewers. What has not been evaluated is the impact of skippable ads on the platform's profit and advertisers' surplus. As we show in this section, switching to the skippable ad format can have positive and negative impacts on advertisers and the platform. To better understand these opposing impacts, we decompose the demand-enhancing and the viewer conversion effects.

- *Viewer Conversion Effect*: Viewer conversion rate is higher (lower) under skippable ads if and only if $\mu > 0$ ($\mu < 0$).
- *Demand-Enhancing Effect*: As long as $d^{trad} < \bar{d}$, the platform has unambiguously more viewers with skippable ads.
  - Direct impact: the option to skip ads directly raises viewer's utility from the platform, which leads to greater participation.
  - Indirect impact: the direct impact implies that the platform adjusts its advertising rate to induce more advertiser participation, which induces further viewer participation.

For ad regime $i \in \{skip, trad\}$, equilibrium platform profit is given by

$$\pi_P^i \equiv VCR^i \cdot G^{-1}(1 - N^i) \cdot h(N^i) \cdot d^i(N^i). \tag{14}$$

Next, we show that equilibrium aggregate advertiser surplus is given by

$$\Pi_A^i = d^i \cdot VCR^i \int_{t^i}^{\infty} (t - t^i) \cdot \frac{h(N)}{N} g(t)dt. \tag{15}$$

Starting with traditional ads,

$$\begin{aligned}
\Pi_A^{trad} &= \int_{t^{trad}}^{+\infty} t \cdot VCR^{trad} \cdot d^{trad} \cdot \frac{h(N)}{N} g(t)dt - r^{trad} \cdot d^{trad} \cdot h(N), \\
&= \int_{t^{trad}}^{+\infty} t \cdot VCR^{trad} \cdot d^{trad} \cdot \frac{h(N)}{N} g(t)dt - t^{trad} \cdot VCR^{trad} \cdot d^{trad} \cdot h(N), \\
&= d^{trad} \cdot VCR^{trad} \cdot \int_{t^{trad}}^{+\infty} (t - t^{trad}) \cdot \frac{h(N)}{N} g(t)dt.
\end{aligned}$$

Similarly, aggregate advertiser surplus under skippable ads can be derived in the form of (15), noting that the payment to the platform under skippable ads, accounting for the probability a viewer watches the entire advertisement, is $r^{skip} \cdot Pr(C_{1b}) \cdot d^{skip} \cdot h(N)$.

Now suppose that there is no demand-enhancing effect. That is, assume $d^{trad} = \bar{d}$ so that the change in ad regimes has no impact on the number of advertisers: $N^{skip} = N^{trad}$ and $t^{trad} = t^{true}$. By Proposition 3 we immediately have the following lemma.

Lemma 5 *(Viewer Conversion Effect only) Suppose $d^{trad} = \bar{d}$. Skippable ads imply higher platform profits and higher aggregate advertiser profits ($\pi_P^{skip} > \pi_P^{trad}$ and $\Pi_A^{skip} > \Pi_A^{trad}$) if and only if $\mu > 0$.*

When the demand-enhancing effect is off, the viewer conversion effect completely determines the impact on advertisers and the platform. Recall that when $\mu > 0$, skippable ads are more likely to induce conversion with the first signal. Advertisers are thus better off. When $\mu < 0$, this logic implies reduced advertiser surplus and platform profit. In either case, when there is no demand-enhancing effect, Lemma 5 implies the platform's incentive for implementing skippable ads is perfectly aligned with advertisers.

We now turn to the more nuanced case in which the demand-enhancing effect is active. To understand these nuances, we shut down the viewer conversion effect by setting $\mu = 0$. Then, by Proposition 3, the viewer conversion rate is the same under the two ad formats ($VCR^{trad} = VCR^{skip}$). We know from Proposition 4, part (ii) that it is optimal for the platform to raise ad levels under the skippable regime. Specifically, the optimizations in (11) and (13) imply

$$\frac{d}{dN}\left[G^{-1}(1-N)\cdot h(N)\cdot d^{i}(N)\right]\Big|_{N=N^{trad}} > 0,$$

so that the marginal return to advertising volume is higher under the skippable ad regime due to the demand-enhancing effect. Hence, $\pi_P^{skip} > \pi_P^{trad}$ for $\mu = 0$.

To understand how changes in the market setting affect aggregate advertiser surplus, we look into equation (15). We already know that $VCR^{trad} = VCR^{skip}$ and $d^{skip} > d^{trad}$. Therefore, a sufficient condition for $\Pi_A^{true} > \Pi_A^{trad}$ is to have the integral in (15) increase with N. The impact of a marginal advertiser on aggregate advertiser surplus is calculated by the following derivative,

$$\begin{aligned}\frac{d}{dN}\int_{t(N)}^{\infty}(t-t(N))\cdot\frac{h(N)}{N}g(t)dt &= \left\{-(t-t(N))\cdot\frac{h(N)}{N}g(t)|_{t=t(N)}\right.\\ &- \left.\int_{t(N)}^{\infty}\frac{h(N)}{N}\cdot g(t)dt\right\}\frac{dt(N)}{dN}\\ &+ \int_{t(N)}^{\infty}(t-t(N))\frac{h'(N)\cdot N-h(N)}{N^2}\cdot g(t)dt.\end{aligned}$$

The expression above provides a decomposition of the three effects when N increases slightly. This occurs when the advertising rate changes. A slight increase in the number of advertisers, dN earns these new advertisers positive surplus, which is captured by the first term. These advertisers are at the margin and this impact goes toward zero for an infinitesimal change in N. Larger N also corresponds to a lower advertising rate, leaving the infra-marginal advertisers paying less for advertising, an effect captured by the second term. The third term is more subtle and depends on the allocation of advertisers to viewers, which is not explicitly modeled, but is subsumed in our interpretation of $h(N)$.

In general, more ads N increase the probability $h(N)$ that a viewer sees an advertisement. If viewing probability $h(N)$ increases only slightly when N increases ($h'(N)\cdot N < h(N)$), then we interpret this as if marginal advertisers were crowding out some of the ads from the existing, infra-marginal advertisers. And, because marginal advertisers have lower t values, any crowd-out effect may hurt the advertisers as a whole.[8] Alternatively, if $h(N)$ increases at a higher rate ($h'(N)\cdot N > h(N)$), then any viewer is not less likely to see a given advertiser's ad as the number of ads N increases. This

[8] If, instead of random allocation, ad spots are allocated efficiently with advertisers with high t values shown to viewers first, then there is no crowd out effect, and an increase in N unambiguously benefits all advertisers.

latter condition can also be interpreted as if the platform had a sufficient amount of content that much of it does not show ads. Any increase in the number of ads simply expands into the volume of content without ads.

By Proposition 4, we know that $N^{skip} > N^{trad}$. Therefore, a sufficient condition for advertisers to be better off ($\Pi_A^{skip} > \Pi_A^{trad}$) is to have $d\Pi_A^i/dN > 0$. This is implied by the following condition:

Assumption 2: $h(N) \leq h'(N) \cdot N$.

Note that the specification $h(N) = N < 1$ satisfies Assumption 2. We assume this condition for the remainder of the analysis. The above arguments for $\mu = 0$ imply the following result.

Lemma 6 *(Demand-Enhancing Effect only) Let $\mu = 0$. When the platform uses skippable ads instead of traditional ads,*

(i) the platform is strictly better off: $\pi_P^{skip} > \pi_P^{trad}$; and

(ii) under Assumption 2, advertisers, as a whole, are better off: $\Pi_A^{skip} > \Pi_A^{trad}$.

The condition in part (ii) of this lemma is a sufficient condition for improved advertiser surplus when the platform implements skippable ads. Thus, under this condition, the platform's incentive for switching to skippable ads is aligned with advertiser's interests when $\mu = 0$. We are unable to establish this alignment outside of the condition in part (ii) of Lemma 6.

As is evident in Lemmas 5 and 6, there are broad conditions under which the platform will implement skippable ads precisely when it is beneficial for advertisers. This is a reasonable outcome given that the platform collects no direct fees from viewers. All revenue for the platform come from the advertiser-side. Prior work in media economics notes how more viewer benefits lead to higher demand for the platform and, consequently, more advertising revenue. This is the result in Lemma 6. What is new for skippable ads relative to that prior work is seen in Lemma 5. It says that the ad format benefits advertisers by helping them convert viewers better, thus raising the share of the platform's revenue.

Combining the results in the two lemmas above, we deduce the general results in the following proposition.

Proposition 5 *Suppose the platform switches from traditional ads to skippable ads.*

(i) If $\mu > 0$, then the platform earns more profits and aggregate advertiser surplus increases. The skippable ad format is a Pareto improvement over the traditional format.

(ii) Suppose $\mu < 0$ and the demand-enhancing effect is sufficiently small ($d^{trad} \uparrow \bar{d}$). Then advertisers and the platform are worse off with skippable ads.

(iii) Suppose $\mu < 0$ and the viewer conversion effect is sufficiently small ($\mu \to 0^-$). Then advertisers and the platform are better off with skippable ads.

This general result implies that unless the demand-enhancing effect is small, the skippable ad format is an equilibrium outcome. Furthermore, the availability of skippable ad technology leads to a Pareto improvement over the traditional format. Because skippable ads lead to a better viewer experience, the increased viewership is a benefit to advertisers and the platform.

5 Conclusion

The growing use of interactive advertising on mobile and online content platforms is changing the way viewers acquire information about products and services. The skippable ad format, in particular, seems to make up a large part of the interactive advertising category and is used by YouTube and many online news sites. This paper examined the skippable ad and focused on the viewer's interaction with this format.

The fact that viewers decide how much of the advertiser's information to acquire in a skippable ad is fundamentally different than in the traditional ad format seen in conventional media. Another distinctive feature of this ad format is that platforms can charge advertisers only when viewers see the entire ad. As we argued, these distinctions have important implications on the demand for advertising.

We constructed a model of the viewer's sequential interaction with advertiser's information when deciding whether to continue watching an ad or to skip. This model showed that a viewer watches more of the advertisement when she expects the subsequent information to be helpful in making a decision about the advertiser. If she is relatively certain about the advertiser, positively or negatively, then she can move on to her desired content without the cost of delay.

We embedded this model into a two-sided market in which the platform can sell ads that are served to viewers. We found that, relative to the traditional ad format, the use of skippable ads leads to more viewers on the platform. With more viewers under the skippable ad, the platform sells more advertising. Under certain conditions, the movement from the traditional to the skippable ad format is a Pareto improvement. The source of the efficiency is that the skippable ad guides the viewer to more meaningful information relative to her preferences. As a result, she obtains direct benefit from participating on the platform. When viewers are predisposed to advertised products, then advertisers are more likely to convert viewers because viewers skip more often and are less likely, relative to traditional ads, to receive information that reverses their prior. Platforms subsequently benefit from higher advertising demand.

This paper has focused on a representative viewer and her interaction with the skippable ad. This approach afforded us the ability to build a detailed model of the viewer's information processing and skipping decision. However, it restricted our ability to examine how viewer heterogeneity affects the platform's strategy vis-á-vis the advertising market. Furthermore, we were also unable to draw out more detailed implications for advertising strategies. For instance, if an advertiser pays only for ads that are watched in entirety, then an advertiser may strategically modify its message sequencing or the content of the initial signal relative to traditional ads. Finally, we were unable to study the optimal dynamics of ad placement when it can occur outside of the first part of the content (e.g. Zhou 2003). We hope that some of these issues can be undertaken in future research.

A Appendix

In order to prove the results in the main text, it is first helpful to have the following intermediate results.

Lemma A1 *Let $v_c \sim N(\mu, \sigma^2)$ and $v_i = v_c + \varepsilon_i$, where $\varepsilon_i \sim N(0, \sigma^2)$ are i.i.d., $i = 1, 2$. Then* $\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ v_c \\ v_2 \end{pmatrix} | v_1 \sim (\bar{\mu}, \bar{\mathbf{\Sigma}})$, *where*

$$\bar{\mu} = \begin{pmatrix} \frac{\sigma^2}{s^2+\sigma^2}(v_1 - \mu) \\ 0 \\ \mu + \frac{s^2}{s^2+\sigma^2}(v_1 - \mu) \\ \mu + \frac{s^2}{s^2+\sigma^2}(v_1 - \mu) \end{pmatrix}, \quad \bar{\mathbf{\Sigma}} = \begin{pmatrix} \frac{s^2\sigma^2}{s^2+\sigma^2} & 0 & -\frac{s^2\sigma^2}{s^2+\sigma^2} & -\frac{s^2\sigma^2}{s^2+\sigma^2} & \varepsilon_1 \\ 0 & \sigma^2 & 0 & \sigma^2 & \varepsilon_2 \\ -\frac{s^2\sigma^2}{s^2+\sigma^2} & 0 & \frac{s^2\sigma^2}{s^2+\sigma^2} & \frac{s^2\sigma^2}{s^2+\sigma^2} & v_c \\ -\frac{s^2\sigma^2}{s^2+\sigma^2} & \sigma^2 & \frac{s^2\sigma^2}{s^2+\sigma^2} & \frac{\sigma^2(2s^2+\sigma^2)}{s^2+\sigma^2} & v_2 \\ \varepsilon_1 & \varepsilon_2 & v_c & v_2 & \end{pmatrix}.$$

The last row and column in $\bar{\Sigma}$ indicate the position of the random variables, and are not part of the covariance matrix.

Proof of Lemma A1

The 5 random variables associated with the advertising signals are jointly distributed normally, as follows:

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ v_c \\ v_2 \\ v_1 \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \\ \mu \\ \mu \\ \mu \end{pmatrix}, \mathbf{\Sigma} \right),$$

where

$$\mathbf{\Sigma} = \left(\begin{array}{cccc|c} \sigma^2 & 0 & 0 & 0 & \sigma^2 \\ 0 & \sigma^2 & 0 & \sigma^2 & 0 \\ 0 & 0 & s^2 & s^2 & s^2 \\ 0 & \sigma^2 & s^2 & s^2+\sigma^2 & s^2 \\ \hline \sigma^2 & 0 & s^2 & s^2 & s^2+\sigma^2 \end{array}\right) \equiv \left(\begin{array}{c|c} \mathbf{\Sigma_{11}} & \mathbf{\Sigma_{12}} \\ \hline \mathbf{\Sigma_{21}} & \mathbf{\Sigma_{22}} \end{array}\right).$$

Note that $\mathbf{\Sigma_{22}} = s^2 + \sigma^2$, and $\mathbf{\Sigma_{11}}$ is the 4×4 matrix obtained by deleting the last row and column in $\mathbf{\Sigma}$. $\mathbf{\Sigma_{21}} = (\sigma^2, 0, s^2, s^2) = (\mathbf{\Sigma_{12}})^T$.

Let $Y \equiv (\varepsilon_1, \varepsilon_2, v_c, v_2)^T$. The distribution of Y conditional on v_1 is given by

$$Y|v_1 \equiv \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ v_c \\ v_2 \end{pmatrix} |v_1 \sim (\bar{\mu}, \bar{\mathbf{\Sigma}}),$$

where

$$\bar{\mu} = \begin{pmatrix} 0 \\ 0 \\ \mu \\ \mu \end{pmatrix} + \mathbf{\Sigma_{12}} \mathbf{\Sigma_{22}^{-1}} (v_1 - \mu),$$

$$\bar{\mathbf{\Sigma}} = \mathbf{\Sigma_{11}} - \mathbf{\Sigma_{12}} \mathbf{\Sigma_{22}^{-1}} \mathbf{\Sigma_{21}}.$$

Substituting the covariance matrix, we have

$$\bar{\mu} = \begin{pmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ E(v_c) \\ E(v_2) \end{pmatrix} |v_1 = \begin{pmatrix} 0 \\ 0 \\ \mu \\ \mu \end{pmatrix} + \begin{pmatrix} \sigma^2 \\ 0 \\ s^2 \\ s^2 \end{pmatrix} \frac{1}{s^2 + \sigma^2} (v_1 - \mu) = \begin{pmatrix} \frac{\sigma^2}{s^2+\sigma^2}(v_1 - \mu) \\ 0 \\ \mu + \frac{s^2}{s^2+\sigma^2}(v_1 - \mu) \\ \mu + \frac{s^2}{s^2+\sigma^2}(v_1 - \mu) \end{pmatrix}.$$

$$\begin{aligned}
\bar{\mathbf{\Sigma}} &= \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & \sigma^2 \\ 0 & 0 & s^2 & s^2 \\ 0 & \sigma^2 & s^2 & s^2 + \sigma^2 \end{pmatrix} - \begin{pmatrix} \sigma^2 \\ 0 \\ s^2 \\ s^2 \end{pmatrix} \frac{1}{s^2 + \sigma^2} (\sigma^2, 0, s^2, s^2) \\
&= \begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & \sigma^2 \\ 0 & 0 & s^2 & s^2 \\ 0 & \sigma^2 & s^2 & s^2 + \sigma^2 \end{pmatrix} - \frac{1}{s^2 + \sigma^2} \begin{pmatrix} \sigma^4 & 0 & s^2\sigma^2 & s^2\sigma^2 \\ 0 & 0 & 0 & 0 \\ s^2\sigma^2 & 0 & s^4 & s^4 \\ s^2\sigma^2 & 0 & s^4 & s^4 \end{pmatrix} \\
&= \begin{pmatrix} \frac{s^2\sigma^2}{s^2+\sigma^2} & 0 & -\frac{s^2\sigma^2}{s^2+\sigma^2} & -\frac{s^2\sigma^2}{s^2+\sigma^2} & \varepsilon_1 \\ 0 & \sigma^2 & 0 & \sigma^2 & \varepsilon_2 \\ -\frac{s^2\sigma^2}{s^2+\sigma^2} & 0 & \frac{s^2\sigma^2}{s^2+\sigma^2} & \frac{s^2\sigma^2}{s^2+\sigma^2} & v_c \\ -\frac{s^2\sigma^2}{s^2+\sigma^2} & \sigma^2 & \frac{s^2\sigma^2}{s^2+\sigma^2} & \frac{\sigma^2(2s^2+\sigma^2)}{s^2+\sigma^2} & v_2 \\ \varepsilon_1 & \varepsilon_2 & v_c & v_2 & \end{pmatrix}.
\end{aligned}$$

The last row and column indicate the position of the random variables, and are not part of the covariance matrix. Note that the conditional covariance matrix is independent of μ. ■

The following lemma builds off of the previous lemma and establishes the distributions needed to derive the expected benefit of viewing the advertiser's second signal conditional on the first.

Lemma A2 *The sums $v_1 + v_2$ and $\varepsilon_1 + \varepsilon_2$, conditional on v_1, follow a joint normal distribution:*

$$\begin{pmatrix} v_1 + v_2 \\ \varepsilon_1 + \varepsilon_2 \end{pmatrix} | v_1 \sim N\left(\begin{pmatrix} A \\ \Delta \end{pmatrix}, \begin{pmatrix} B^2 & C \\ C & B^2 \end{pmatrix} \right),$$

where

$$A = v_1 + \mu + \frac{s^2}{s^2 + \sigma^2} \cdot (v_1 - \mu), \quad B^2 = \frac{\sigma^2(2s^2 + \sigma^2)}{s^2 + \sigma^2},$$
$$C = \frac{\sigma^4}{s^2 + \sigma^2}, \quad \Delta = \frac{\sigma^2}{s^2 + \sigma^2}(v_1 - \mu).$$

Proof of Lemma A2

Derive the joint distribution of $v_1 + v_2$ and $\varepsilon_1 + \varepsilon_2$, conditional on v_1. We divide the analysis into 3 steps.

Step 1: Calculate the conditional distribution $\varepsilon_1 + \varepsilon_2|_{v_1}$

$\varepsilon_1 + \varepsilon_2 | v_1$ is joint normal with mean

$$E(\varepsilon_1 + \varepsilon_2 | v_1) = E(\varepsilon_1 | v_1) + E(\varepsilon_2 | v_1) = \frac{\sigma^2}{s^2 + \sigma^2}(v_1 - \mu),$$

and variance

$$\begin{aligned} var(\varepsilon_1 + \varepsilon_2 | v_1) &= var(\varepsilon_1 | v_1) + var(\varepsilon_2 | v_1) + 2cov(\varepsilon_1 | v_1, \varepsilon_2 | v_1) \\ &= \frac{s^2 \sigma^2}{s^2 + \sigma^2} + \sigma^2 + 0 = \frac{\sigma^2(2s^2 + \sigma^2)}{s^2 + \sigma^2} \equiv B^2. \end{aligned}$$

Step 2: Calculate the conditional distribution $v_1 + v_2|_{v_1}$

Similarly, $v_1 + v_2 | v_1$ is joint normal with mean

$$E(v_1 + v_2 | v_1) = v_1 + E(v_2 | v_1) = v_1 + \mu + \frac{s^2}{s^2 + \sigma^2} \cdot (v_1 - \mu) \equiv A,$$

and variance

$$var(v_1 + v_2 | v_1) = var(v_2 | v_1) = \frac{\sigma^2(2s^2 + \sigma^2)}{s^2 + \sigma^2} = B^2,$$

with B^2 defined earlier. Note that μ enters the A term but not the B^2 term.

Step 3: Calculate the conditional covariance $Cov(v_1 + v_2, \varepsilon_1 + \varepsilon_2|_{v_1})$

Let $\Delta \equiv E(\varepsilon_1 + \varepsilon_2 | v_1) = \frac{\sigma^2}{s^2 + \sigma^2}(v_1 - \mu)$. Then

$$\begin{aligned} cov(\varepsilon_1 + \varepsilon_2, v_1 + v_2 | v_1) &= E\left[(\varepsilon_1 + \varepsilon_2 - \Delta) \cdot (v_1 + v_2 - E(v_1 + v_2 | v_1)) | v_1 \right] \\ &= E\left[(\varepsilon_1 + \varepsilon_2 - \Delta) \cdot (v_2 - E(v_2 | v_1)) | v_1 \right] \\ &= E(v_2 \cdot \varepsilon_1 | v_1) + E(v_2 \cdot \varepsilon_2 | v_1) - E(v_2 \cdot \Delta | v_1) \\ &- E[E(v_2 | v_1) \cdot \varepsilon_1 | v_1] - E[E(v_2 | v_1) \cdot \varepsilon_2 | v_1] + E[E(v_2 | v_1) \cdot \Delta | v_1]. \end{aligned}$$

There are 6 terms in the covariance expression above:

- The 1st term is

$$E(v_2 \cdot \varepsilon_1 | v_1) = E[(v_1 - \varepsilon_1 + \varepsilon_2) \cdot \varepsilon_1 | v_1) = E(v_1 \varepsilon_1 | v_1) - E(\varepsilon_1^2 | v_1) + 0$$
$$= v_1 \cdot E(\varepsilon_1 | v_1) - [var(\varepsilon_1 | v_1) + E(\varepsilon_1 | v_1)^2]$$
$$= \frac{\sigma^2}{s^2 + \sigma^2} \cdot v_1(v_1 - \mu) - \left[\frac{s^2 \sigma^2}{s^2 + \sigma^2} + \left(\frac{\sigma^2}{s^2 + \sigma^2}\right)^2 \cdot (v_1 - \mu)^2\right].$$

- The 2nd term is

$$E((v_c + \varepsilon_2) \cdot \varepsilon_2 | v_1) = E(v_c \cdot \varepsilon_2 | v_1) + E(\varepsilon_2^2 | v_1) = 0 + \sigma^2 = \sigma^2.$$

The 3rd term and the 6th term cancel out.

- The 4th term is

$$E(v_2 | v_1) \cdot E(\varepsilon_1 | v_1) = \left(\mu + \frac{s^2}{s^2 + \sigma^2}(v_1 - \mu)\right) \cdot \frac{\sigma^2}{s^2 + \sigma^2}(v_1 - \mu)$$
$$= \frac{\sigma^2}{s^2 + \sigma^2} \cdot \mu \cdot (v_1 - \mu) + \left(\frac{s \cdot \sigma}{s^2 + \sigma^2} \cdot (v_1 - \mu)\right)^2,$$

based on our earlier derivations.

- The 5th term is zero.

The covariance term is then

$$\begin{aligned} cov(\varepsilon_1 + \varepsilon_2, v_1 + v_2 | v_1) &= \frac{\sigma^2}{s^2 + \sigma^2} \cdot v_1(v_1 - \mu) - \left[\frac{s^2 \sigma^2}{s^2 + \sigma^2} + \left(\frac{\sigma^2}{s^2 + \sigma^2}\right)^2 \cdot (v_1 - \mu)^2\right] \\ &+ \sigma^2 - \frac{\sigma^2}{s^2 + \sigma^2} \cdot \mu \cdot (v_1 - \mu) - \left(\frac{s \cdot \sigma}{s^2 + \sigma^2} \cdot (v_1 - \mu)\right)^2 \\ &= \frac{\sigma^2}{s^2 + \sigma^2}(v_1 - \mu)^2 - \frac{s^2 \sigma^2}{s^2 + \sigma^2} + \sigma^2 - \frac{\sigma^4 + s^2 \sigma^2}{(s^2 + \sigma^2)^2} \cdot (v_1 - \mu)^2 \\ &= \sigma^2 - \frac{s^2 \sigma^2}{s^2 + \sigma^2} = \frac{\sigma^4}{s^2 + \sigma^2} \equiv C. \quad \blacksquare \end{aligned}$$

Proof of Lemma 1

(i) This follows directly from Lemma A2.

(ii) We first derive the distribution of v_c conditional on v_1 and v_2. From the above results, we have the following joint distribution:

$$\begin{pmatrix} v_c \\ v_2 \\ v_1 \end{pmatrix} \sim \left(\begin{pmatrix} \mu \\ \mu \\ \mu \end{pmatrix}, \bar{\mathbf{\Sigma}} \right) =$$

where

$$\bar{\mathbf{\Sigma}} = \left(\begin{array}{c|cc} s^2 & s^2 & s^2 \\ \hline s^2 & s^2+\sigma^2 & s^2 \\ s^2 & s^2 & s^2+\sigma^2 \end{array}\right) = \left(\begin{array}{c|c} \mathbf{\Sigma_{11}} & \mathbf{\Sigma_{12}} \\ \hline \mathbf{\Sigma_{21}} & \mathbf{\Sigma_{22}} \end{array}\right).$$

The conditional mean of v_c given v_1 and v_2 is

$$\begin{aligned} E(v_c|v_1,v_2) &= \mu + \mathbf{\Sigma_{12}}\mathbf{\Sigma_{22}^{-1}} \begin{pmatrix} v_1-\mu \\ v_2-\mu \end{pmatrix} \\ &= \mu + (s^2,s^2)\frac{1}{(s^2+\sigma^2)^2-s^4}\begin{pmatrix} s^2+\sigma^2 & -s^2 \\ -s^2 & s^2+\sigma^2 \end{pmatrix}\begin{pmatrix} v_1-\mu \\ v_2-\mu \end{pmatrix} \\ &= \mu + \frac{1}{\sigma^2(2s^2+\sigma^2)}(s^2\sigma^2, s^2\sigma^2)\begin{pmatrix} v_1-\mu \\ v_2-\mu \end{pmatrix} \\ &= \mu + \frac{s^2}{2s^2+\sigma^2}(v_1+v_2-2\mu) \\ &= \frac{\sigma^2}{2s^2+\sigma^2}\mu + \frac{s^2}{2s^2+\sigma^2}(v_1+v_2). \end{aligned}$$

Conversion is optimal for the viewer if and only if $E(v_c|v_1,v_2) \geq 0$, which is equivalent to the condition $v_1+v_2 \geq -\mu\cdot(\sigma^2/s^2)$. ■

Proof of Lemma 2

$$\begin{aligned} U_3^{skip} &= E\left(v_c \middle| v_1, v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\right)\cdot Pr\left[v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\middle| v_1\right] + \delta w \\ &= E\left(\frac{v_1+v_2-\varepsilon_1-\varepsilon_2}{2} \middle| v_1, v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\right)\cdot Pr\left[v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\middle| v_1\right] + \delta w \\ &= \frac{1}{2}E\left(v_1+v_2 \middle| v_1, v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\right)\cdot Pr\left[v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\middle| v_1\right] + \delta w \\ &- \frac{1}{2}E\left(\varepsilon_1+\varepsilon_2 \middle| v_1, v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\right)\cdot Pr\left[v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\middle| v_1\right] + \delta w. \end{aligned} \tag{16}$$

The first component of (16) is calculated as follows. $E\left(v_1+v_2 \middle| v_1, v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\right)$ is the expectation in the case of one-sided truncation and it equals (see Wikipedia Truncated normal one-sided truncation of lower tail `https://en.wikipedia.org/wiki/Truncated_normal_distribution`.)

$$A + \frac{B\phi(\alpha)}{1-\Phi(\alpha)},$$

where A and B^2 are defined in Lemma 1 and $\alpha = -\frac{\mu\frac{\sigma^2}{s^2}+A}{B}$.

Previously we have shown that $(v_1+v_2)|v_1$ follows Normal distribution with mean A and variance B^2. Thus

$$Pr\left[v_1+v_2 \geq -\mu\cdot\frac{\sigma^2}{s^2}\middle| v_1\right] = 1-\Phi(\alpha).$$

The first component of (16) is then

$$\frac{1}{2}\cdot\left(A+\frac{B\phi(\alpha)}{1-\Phi(\alpha)}\right)\cdot(1-\Phi(\alpha)).$$

To calculate the second component of (16), recall the conditional joint distribution is

$$\left(\begin{array}{c} v_1+v_2 \\ \varepsilon_1+\varepsilon_2 \end{array}\right)\Bigg| v_1 \sim N\left(\left(\begin{array}{c} A \\ \Delta \end{array}\right),\left(\begin{array}{cc} B^2 & C \\ C & B^2 \end{array}\right)\right).$$

Then

$$\left(\begin{array}{c} \frac{v_1+v_2-A}{B} \\ \frac{\varepsilon_1+\varepsilon_2-\Delta}{B} \end{array}\right)\Bigg| v_1 \sim N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right),\left(\begin{array}{cc} 1 & \frac{C}{B^2} \\ \frac{C}{B^2} & 1 \end{array}\right)\right).$$

Then

$$E\left(\frac{\varepsilon_1+\varepsilon_2-\Delta}{B}\Bigg| v_1, \frac{v_1+v_2-A}{B}>\alpha\right)=\frac{C}{B^2}\cdot\frac{\phi(\alpha)}{1-\Phi(\alpha)}.$$

This comes from the link `https://en.wikipedia.org/wiki/Multivariate_normal_distribution#Bivariate_conditional_expectation`.

Then, the second component of (16) becomes

$$E\left(\varepsilon_1+\varepsilon_2\Bigg| v_1, v_1+v_2>-\mu\frac{\sigma^2}{s^2}\right)=\Delta+\frac{C}{B}\cdot\frac{\phi(\alpha)}{1-\Phi(\alpha)}. \tag{17}$$

Substituting this into the U_3^{skip} expression, we can obtain

$$U_3^{skip}=\frac{1}{2}\left[A+\frac{B\phi(\alpha)}{1-\Phi(\alpha)}\right]\cdot[1-\Phi(\alpha)]-\frac{1}{2}\left[\Delta+\frac{C}{B}\cdot\frac{\phi(\alpha)}{1-\Phi(\alpha)}\right]\cdot[1-\Phi(\alpha)]+\delta w.$$

The U_3^{skip} expression can be simplified to that in (3). We now establish the comparative static properties of U_3^{skip}.

(i) It is straightforward to see that

$$\frac{dU_3^{skip}}{d\delta}=w>0, \quad \frac{dU_3^{skip}}{dw}=\delta>0.$$

(ii) Consider any two μ values of a and b with $b>a$. We transform a series of variables as follows:

$$v_1'=v_1, \quad v_2'=v_2-\frac{\sigma^2}{s^2+\sigma^2}(b-a), \quad v_c'=v_c-\frac{\sigma^2}{s^2+\sigma^2}(b-a)$$

$$\varepsilon_1'=\varepsilon_1+\frac{\sigma^2}{s^2+\sigma^2}(b-a), \quad \varepsilon_2'=\varepsilon_2.$$

This transformation increases the conditional means of v_2, v_c and ε_1, but not the conditional mean of ε_2 or the whole conditional variance/covariance matrix. Thus, the following two conditional random vectors are equal in distribution:

$$(v_1, v_2, v_c, \varepsilon_1, \varepsilon_2 | v_1, \mu = a) \stackrel{D}{=} (v_1', v_2', v_c', \varepsilon_1', \varepsilon_2' | v_1, \mu = b).$$

Therefore,

$$\begin{aligned} U_3^{skip}(v_1, \mu = a) &= E\left(v_c \middle| v_1, \mu = a, v_1 + v_2 > -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 + v_2 > -\mu\frac{\sigma^2}{s^2} \middle| v_1, \mu = a\right) \\ &= E\left(v_c' \middle| v_1, \mu = b, v_1' + v_2' > -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1' + v_2' > -\mu\frac{\sigma^2}{s^2} \middle| v_1, \mu = b\right) \\ &< E\left(v_c \middle| v_1, \mu = b, v_1' + v_2' > -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1' + v_2' > -\mu\frac{\sigma^2}{s^2} \middle| v_1, \mu = b\right) \\ &\le E\left(v_c \middle| v_1, \mu = b, v_1 + v_2 > -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 + v_2 > -\mu\frac{\sigma^2}{s^2} \middle| v_1, \mu = b\right) \\ &= U_3^{skip}(v_1, \mu = b). \end{aligned}$$

The first = follows from the equality of the original and transformed distributions. The first inequality holds because $v_c' < v_c$ and the second due to viewer optimality. This establishes that U_3^{skip} increases in μ.

(iii) We now show that U_3^{skip} is non-monotonic in s and δ. Using the $U_3^{skip}(v_1)$ expression, taking derivatives and simplifying, we can obtain,

$$\text{sign}\left(\frac{\partial U_3^{skip}}{\partial s}\right) = -\text{sign}\left(\Psi_1 \cdot (\Psi_2 + 1)\sqrt{\pi}(s^2 + \sigma^2)(\mu - v_1) - \Psi_3 \cdot \sigma^4(3s^2 + 2\sigma^2)\right),$$

$$\text{sign}\left(\frac{\partial U_3^{skip}}{\partial \sigma}\right) = \text{sign}\left(\Psi_1 \cdot (\Psi_2 + 1)\sqrt{\pi}(s^2 + \sigma^2)(\mu - v_1) - \Psi_3 \cdot \frac{\sigma^2}{2}(2\sigma^4 - 4s^4)\right),$$

where

$$\Psi_1 = 2\left(\frac{\sigma^2(2s^2 + \sigma^2)}{s^2 + \sigma^2}\right)^{\frac{3}{2}}, \quad \Psi_2 = erf\left(\frac{1}{2}\frac{(2s^2 + \sigma^2)(\mu\sigma^2 + s^2 v_1)\sqrt{2}}{s^2(s^2 + \sigma^2)\sqrt{\frac{\sigma^2(2s^2+\sigma^2)}{s^2+\sigma^2}}}\right),$$

$$\Psi_3 = \sqrt{2} \cdot e^{\frac{1}{2}\frac{(\mu\sigma^2 + v_1 s^2)^2(2s^2+\sigma^2)}{\sigma^2 s^4(s^2+\sigma^2)}}.$$

Note that $\Psi_1 > 0$ and $\Psi_3 > 0$. For Ψ_2, where $erf(x)$ is the Gauss error function, it is positive if and only if the argument is positive. Thus, $\Psi_2 > 0$ if and only if $\mu\sigma^2 + v_1 s^2 > 0$, or simply $v_1 > -\mu(\sigma^2/s^2)$. It follows that $\partial U_3^{skip}/\partial s$ and $\partial U_3^{skip}/\partial \sigma$ can be both positive or negative, depending on the signs of $\mu - v_1$ and Ψ_2. Thus, U_3^{skip} is non-monotonic in each s and σ.

(iii-a) In the special case of $v_1 > \mu > 0$, we have $\mu - v_1 < 0$ and $\Psi_2 > 0$. This implies $\partial U_3^{skip}/\partial s > 0$.

(iii-b) In the special case of $v_1 > \mu > 0$ and $\sigma^4 > 2s^4$, then $\partial U_3^{skip}/\partial\sigma < 0$. If $\sigma^4 < 2s^4$ and $\mu > v_1 > 0$, then $\partial U_3^{skip}/\partial\sigma > 0$. ■

Proof of Lemma 3

We first show the existence and uniqueness of v_L and v_H and then establish their ordering.

Existence. This is established using the continuity of $U_i^{skip}(v_1)$ in v_1, for $i = 1, 2$ and 3 as well as the fact that $U_1^{skip} > U_2^{skip}$ if and only if $v_1 > -\mu(\sigma^2/s^2)$. First observe that for v_1 large enough

$$U_3^{skip}(v_1) \approx \frac{s^2 v_1 + \sigma^2\mu}{s^2+\sigma^2} + \delta w \in (U_2^{skip}, U_1^{skip}),$$

which holds because $\alpha \to +\infty$ implies $1 - \Phi(\alpha) \to 0$ and $\phi(\alpha) \to 0$. Similarly,

$$\lim_{v_1 \to -\infty} U_3^{skip}(v_1) = \delta w \in (U_1^{skip}, U_2^{skip}),$$

which holds because $\alpha \to -\infty$ implies $1 - \Phi(\alpha) = 1$ and $\phi(\alpha) = 0$. Therefore, by continuity of $U_i^{skip}(v_1)$, there exists values v_L and v_H such that $U_3^{skip}(v_L) = U_2^{skip}(v_L)$ and $U_3^{skip}(v_H) = U_1^{skip}(v_H)$.

Uniqueness. A sufficient condition for single-crossing is the condition in (4). It is straightforward to see

$$\frac{\partial U_1^{skip}}{\partial v_1} = \frac{s^2}{s^2+\sigma^2}, \quad \frac{\partial U_2^{skip}}{\partial v_1} = 0.$$

We now show that

$$\frac{\partial U_3^{skip}}{\partial v_1} > 0.$$

We do this by showing that $U_3^{skip}(v_1 = b) > U_3^{skip}(v_1 = a)$ must hold for any $b > a$. Note that an increase of v_1 from a to b increases the conditional mean of v_1, v_2, v_c and ε_1, but not the conditional mean of ε_2 or the whole conditional variance/covariance matrix. Define a set of new variables using the following transformation:

$$v_1' = v_1 - (b-a), \quad v_2' = v_2 - \frac{s^2}{s^2+\sigma^2}(b-a), \quad v_c' = v_c - \frac{s^2}{s^2+\sigma^2}(b-a)$$
$$\varepsilon_1' = \varepsilon_1 - \frac{\sigma^2}{s^2+\sigma^2}(b-a), \quad \varepsilon_2' = \varepsilon_2.$$

Then the following two conditional distributions are exactly the same:

$$(v_1, v_2, v_c, \varepsilon_1, \varepsilon_2 | v_1 = a) \overset{D}{=} (v_1', v_2', v_c', \varepsilon_1', \varepsilon_2' | v_1 = b).$$

Therefore,

$$\begin{aligned}
U_3^{skip}(v_1 = a) &= E\left(v_c | v_1 = a, v_1 + v_2 > -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 = a, v_1 + v_2 > -\mu\frac{\sigma^2}{s^2}\right) \\
&= E\left(v_c' | v_1 = b, v_1' + v_2' > -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 = b, v_1' + v_2' > -\mu\frac{\sigma^2}{s^2}\right) \\
&< E\left(v_c | v_1 = b, v_1' + v_2' > -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 = b, v_1' + v_2' > -\mu\frac{\sigma^2}{s^2}\right) \\
&\leq E\left(v_c | v_1 = b, v_1 + v_2 > -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 = b, v_1 + v_2 > -\mu\frac{\sigma^2}{s^2}\right) \\
&= U_3^{skip}(v_1 = b).
\end{aligned}$$

The first $=$ follows from the equality of the original and transformed distributions. The first inequality holds because $v_c' < v_c$ and the second due to viewer optimality. Because $a < b$ are arbitrary, we have $U_3^{skip}(v_1)$ increasing for all v_1. That is, $\frac{\partial U_3^{skip}}{\partial v_1} > 0$.

It remains to show that

$$\frac{\partial U_1^{skip}}{\partial v_1} > \frac{\partial U_3^{skip}}{\partial v_1}.$$

If we ignore the discounting of w, then U_1^{skip} is the same as getting the second signal but always buys the product anyway (i.e., not using either signal). That is

$$U_1^{skip} - w = U_3^{skip} - \delta w + \Lambda(v_1),$$

where

$$\Lambda(v_1) \equiv E\left(v_c | v_1, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) < 0.$$

Then the condition $\partial U_1^{skip}/\partial v_1 > \partial U_3^{skip}/\partial v_1$ is implied by $\Lambda'(v_1) > 0$. Note that $\Lambda(v_1) < 0$ for all v_1 so $\Lambda(v_1)$ increasing in v_1 means a smaller $|\Lambda(v_1)|$. We prove this next.

Consider again the transformation of $(v_1, v_2, v_c, \varepsilon_1, \varepsilon_2 | v_1 = a)$ from above with $a > b$.

We need to show that $\Lambda(v_1 = b) > \Lambda(v_1 = a)$.

$$\begin{aligned}
\Lambda(v_1 = a) &= E\left(v_c | v_1 = a, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 = a, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \\
&= E\left(v_c' | v_1 = b, v_1' + v_2' \leq -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 = b, v_1' + v_2' \leq -\mu\frac{\sigma^2}{s^2}\right) \\
&= E\left(v_c' | v_1 = b, v_1' + v_2' \leq -\mu\frac{\sigma^2}{s^2}, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \\
&\times \ Pr\left(v_1 = b, v_1' + v_2' \leq -\mu\frac{\sigma^2}{s^2}, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \\
&+ \ E\left(v_c' | v_1 = b, v_1' + v_2' \leq -\mu\frac{\sigma^2}{s^2}, v_1 + v_2 > -\mu\frac{\sigma^2}{s^2}\right) \\
&\times \ Pr\left(v_1 = b, v_1' + v_2' \leq -\mu\frac{\sigma^2}{s^2}, v_1 + v_2 > -\mu\frac{\sigma^2}{s^2}\right) \\
&< E\left(v_c' | v_1 = b, v_1' + v_2' \leq -\mu\frac{\sigma^2}{s^2}, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \\
&\times \ Pr\left(v_1 = b, v_1' + v_2' \leq -\mu\frac{\sigma^2}{s^2}, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \\
&= E\left(v_c' | v_1 = b, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 = b, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \\
&< E\left(v_c | v_1 = b, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \cdot Pr\left(v_1 = b, v_1 + v_2 \leq -\mu\frac{\sigma^2}{s^2}\right) \\
&= \Lambda(v_1 = b).
\end{aligned}$$

We now explain this chain of conditions. The second equality follows from the equivalence of distributions. The next equality is a simple decomposition on complementary conditional events: $v_1 + v_2 \leq -\mu(\sigma^2/s^2)$ and $v_1 + v_2 > -\mu(\sigma^2/s^2)$. The term conditioned on $v_1 + v_2 > -\mu(\sigma^2/s^2)$ is negative. To see this, consider a tighter conditional event: $v_1' + v_2' = -\mu(\sigma^2/s^2)$, which satisfies $v_1 + v_2 > -\mu(\sigma^2/s^2)$. Clearly

$$E\left(v_c' \middle| v_1 = b, v_1' + v_2' = -\mu\frac{\sigma^2}{s^2}\right) = 0.$$

Therefore, conditioning on $v_1' + v_2' \leq -\mu(\sigma^2/s^2)$, makes this expectation negative. The subsequent equality follows from the redundancy of the conditioning events. That is, $v_1 + v_2 \leq -\mu(\sigma^2/s^2)$ implies $v_1' + v_2' \leq -\mu(\sigma^2/s^2)$. And, the final inequality is due to the fact that $v_c' < v_c$.

Ordering. Observe that $U_2^{skip}(v_1) = w$ holds for any v_1. At $v_1 = -\mu(\sigma^2/s^2)$, we also have $U_1^{skip}(v_1) = w$.

If $U_3^{skip}(v_1 = -\mu(\sigma^2/s^2)) = w$, then U_i^{skip}'s, $i = 1, 2, 3$ cross at $v_1 = -\mu(\sigma^2/s^2)$, so that $v_L = v_H = -\mu(\sigma^2/s^2)$. Now suppose that $U_3^{skip}(v_1 = -\mu(\sigma^2/s^2)) > w$. Using the property (4) derived earlier, we immediately have $v_H > -\mu(\sigma^2/s^2) > v_L$. ■

Proof of Proposition 1

This follows directly from Lemma 3 and the definitions of v_L and v_H. ■.

Proof of Proposition 2

(i) To establish this property, we exploit the dependence of U_3^{skip} on δ. In particular, U_i^{skip}, $i = 1, 2$ is independent of δ where as U_3^{skip} as a function of $\delta \in \mathbb{R}$ spans $\mathbb{R}$. (It does not matter for this argument that the model restricts $\delta \in [0, 1]$ because we are establishing general result about the function U_3^{skip}.)

For any constellation of μ, s, σ, and w, there exists a unique value $\tilde{\delta}$ such $U_3^{skip}(v_1 = -\mu(\sigma^2/s^2)) = w$. At this value $v_1 = -\mu(\sigma^2/s^2)$, we also have $U_1^{skip} = U_2^{skip} = w$. Thus, $v_L = v_H = -\mu \cdot (\sigma^2/s^2)$ so the property $v_L + v_H = -2\mu \cdot (\sigma^2/s^2)$ holds for $\tilde{\delta}$ and $v_1 = -\mu \cdot (\sigma^2/s^2)$. Now consider an arbitrary $\delta > \tilde{\delta}$. A necessary and sufficient condition for $v_L + v_H = -2\mu \cdot (\sigma^2/s^2)$ is

$$\frac{d(v_L + v_H)}{d\delta} = 0. \tag{18}$$

The proof of (i) is completed by establishing equation (18). Recall the definition of v_L is the solution to $U_3^{skip} - U_2^{skip} = 0$. Taking total differentiation, we have

$$\left(\frac{\partial U_3^{skip}}{\partial v_1} - \frac{\partial U_2^{skip}}{\partial v_1}\right) dv_1 + \left(\frac{\partial U_3^{skip}}{\partial \delta} - \frac{\partial U_2^{skip}}{\partial \delta}\right) d\delta = 0$$

$$\Rightarrow \left(\frac{\partial U_3^{skip}}{\partial v_1}\right) dv_1 + \left(\frac{\partial U_3^{skip}}{\partial \delta}\right) d\delta = 0 \Rightarrow \frac{dv_L}{d\delta} = -\left.\frac{\frac{\partial U_3^{skip}}{\partial \delta}}{\frac{\partial U_3^{skip}}{\partial v_1}}\right|_{v_1 = v_L}.$$

Similarly, v_H is determined by $U_3^{skip} - U_1^{skip} = 0$. Taking total differentiation, we have

$$\left(\frac{\partial U_3^{skip}}{\partial v_1} - \frac{\partial U_1^{skip}}{\partial v_1}\right) dv_1 + \left(\frac{\partial U_3^{skip}}{\partial \delta} - \frac{\partial U_1^{skip}}{\partial \delta}\right) d\delta = 0$$

$$\Rightarrow \left(\frac{\partial U_3^{skip}}{\partial v_1} - \frac{s^2}{s^2 + \sigma^2}\right) dv_1 + \left(\frac{\partial U_3^{skip}}{\partial \delta}\right) d\delta = 0 \Rightarrow \frac{dv_H}{d\delta} = -\left.\frac{\frac{\partial U_3^{skip}}{\partial \delta}}{\frac{\partial U_3^{skip}}{\partial v_1} - \frac{s^2}{s^2+\sigma^2}}\right|_{v_1 = v_H}.$$

Since $\frac{\partial U_3^{skip}}{\partial \delta} = w > 0$ does not depend on v_1, $\frac{d(v_L + v_H)}{d\delta} = 0$ if and only if the denominators in the $\frac{dv_L}{d\delta}$ and $\frac{dv_H}{d\delta}$ expressions add up to zero. Therefore, (18) is equivalent to the condition

$$\left.\frac{\partial U_3^{skip}}{\partial v_1}\right|_{v_1 = v_L} + \left.\frac{\partial U_3^{skip}}{\partial v_1}\right|_{v_1 = v_H} = \frac{s^2}{s^2 + \sigma^2}. \tag{19}$$

Using the U_3^{skip} expression in (3), we have

$$\frac{\partial U_3^{skip}}{\partial v_1} = \frac{s^2}{s^2 + \sigma^2}[1 - \Phi(\alpha)] - \frac{s^2 v_1 + \sigma^2 \mu}{s^2 + \sigma^2}\phi(\alpha)\frac{\partial \alpha}{\partial v_1} + \frac{1}{2}\left(B - \frac{C}{B}\right)\phi'(\alpha)\frac{\partial \alpha}{\partial v_1}, \tag{20}$$

so that (19) can be expanded as follows:

$$\left.\frac{\partial U_3^{skip}}{\partial v_1}\right|_{v_1=v_L} + \left.\frac{\partial U_3^{skip}}{\partial v_1}\right|_{v_1=v_H} = \left\{ \left.\frac{s^2}{s^2+\sigma^2}[1-\Phi(\alpha)]\right|_{v_L} + \left.\frac{s^2}{s^2+\sigma^2}[1-\Phi(\alpha)]\right|_{v_H} \right\}$$
$$- \left\{ \left.\frac{s^2 v_1 + \sigma^2\mu}{s^2+\sigma^2}\phi(\alpha)\frac{\partial\alpha}{\partial v_1}\right|_{v_L} + \left.\frac{s^2 v_1 + \sigma^2\mu}{s^2+\sigma^2}\phi(\alpha)\frac{\partial\alpha}{\partial v_1}\right|_{v_H} \right\}$$
$$+ \left\{ \left.\frac{1}{2}\left(B-\frac{C}{B}\right)\phi'(\alpha)\frac{\partial\alpha}{\partial v_1}\right|_{v_L} + \left.\frac{1}{2}\left(B-\frac{C}{B}\right)\phi'(\alpha)\frac{\partial\alpha}{\partial v_1}\right|_{v_H} \right\}.$$

We show that the second and third bracketed terms are each zero when $v_L + v_H = -2\mu\cdot(\sigma^2/s^2)$. The second term is simplified by noting the following facts: $\phi(\alpha|v_L) = \phi(\alpha|v_H)$ and $\frac{\partial\alpha}{\partial v_1}|_{v_L} = \frac{\partial\alpha}{\partial v_1}|_{v_H}$. The latter fact is verified directly by differentiating A and we know $\phi(\alpha|v_L) = \phi(\alpha|v_H)$ because of the symmetry implied by

$$\begin{aligned}
\alpha(v_L) + \alpha(v_H) &= -\frac{\mu\frac{\sigma^2}{s^2} + A}{B}|_{v_L} - \frac{\mu\frac{\sigma^2}{s^2} + A}{B}|_{v_H} \\
&= 2\mu\frac{\sigma^2}{s^2} + A|_{v_L} + A|_{v_H} \\
&= 2\mu\frac{\sigma^2}{s^2} + \frac{2s^2}{s^2+\sigma^2}(v_L + v_H) + 2\frac{\sigma^2}{s^2+\sigma^2} \\
&= 2\mu\frac{\sigma^2}{s^2} + \frac{2s^2}{s^2+\sigma^2}\left(-2\mu\cdot\frac{\sigma^2}{s^2}\right) + 2\frac{\sigma^2}{s^2+\sigma^2} \\
&= 0.
\end{aligned}$$

These simplifications imply that the second term in (20) is zero if and only if

$$\frac{s^2 v_1 + \sigma^2\mu}{s^2+\sigma^2}|_{v_1=v_L} + \frac{s^2 v_1 + \sigma^2\mu}{s^2+\sigma^2}|_{v_1=v_H} = 0.$$

Evaluating the LHS of the above condition gives

$$\frac{s^2}{s^2+\sigma^2}(v_L + v_H) + 2\frac{\sigma^2\mu}{s^2+\sigma^2},$$

which is zero precisely when $v_L + v_H = -2\mu\cdot(\sigma^2/s^2)$. The third term in equation (20) is also zero. To see this, note that $\frac{1}{2}\left(B - \frac{C}{B}\right)\cdot\frac{\partial\alpha}{\partial v_1}$ is independent of v_1. Therefore, the term is zero if and only $\phi'(\alpha|v_L) = -\phi'(\alpha|v_H)$. The analysis above verified that $\alpha(v_L) = -\alpha(v_H)$. The symmetry of the standard Normal pdf therefore implies $\phi'(\alpha|v_L) = \phi'(-\alpha|v_H) = -\phi'(\alpha|v_H)$. Hence, the third term in (20) is zero.

Give that the second and third terms of (20) are zero, the condition in (19) is finally established as follow.

$$\begin{aligned}\frac{\partial U_3^{skip}}{\partial v_1}|v_L + \frac{\partial U_3^{skip}}{\partial v_1}|v_H &= \frac{s^2}{s^2+\sigma^2}\cdot[(1-\Phi(\alpha|v_L)) + (1-\Phi(\alpha|v_H))] \\ &= \frac{s^2}{s^2+\sigma^2}\cdot[(1-\Phi(\alpha|v_L)+\Phi(\alpha|v_L)] \\ &= \frac{s^2}{s^2+\sigma^2}.\end{aligned}$$

The second equality is because

$$\alpha(v_L) = -\alpha(v_H) \Rightarrow \Phi(\alpha|v_L) = 1-\Phi(\alpha|v_H).$$

So equation (19) holds, and $\frac{d(v_L+v_H)}{d\delta} = 0$. Therefore, $v_L + v_H = -2\mu\cdot(\sigma^2/s^2)$ always holds. Combined with the single-crossing condition, $v_L + v_H = -2\mu\cdot(\sigma^2/s^2)$ is the only solution.

(ii) Because $U_2^{skip} = w, \forall v_1$, the definition of v_L implies the implicit function $U_3^{skip}(v_L) = w$. To determine the sign of $\frac{dv_L}{d\mu}$, we use the Implicit Function Theorem. That is,

$$\frac{dv_L}{d\mu} = -\frac{\frac{\partial U_3^{skip}}{\partial \mu}}{\frac{\partial U_3^{skip}}{\partial v_1}} < 0,$$

because $\frac{\partial U_3^{skip}}{\partial v_1} > 0$ and $\frac{\partial U_3^{skip}}{\partial \mu} > 0$.

Similarly,

$$\frac{dv_L}{d\delta} = -\frac{\frac{\partial U_3^{skip}}{\partial \delta}}{\frac{\partial U_3^{skip}}{\partial v_1}} < 0, \quad \frac{dv_L}{dw} = -\frac{\frac{\partial U_3^{skip}}{\partial w} - 1}{\frac{\partial U_3^{skip}}{\partial v_1}} > 0,$$

because $\frac{\partial U_3^{skip}}{\partial \delta} = w > 0$ and $\frac{\partial U_3^{skip}}{\partial w} - 1 = \delta - 1 < 0$.

The signs of

$$\frac{dv_L}{ds} = -\frac{\frac{\partial U_3^{skip}}{\partial s}}{\frac{\partial U_3^{skip}}{\partial v_1}}, \quad \frac{dv_L}{d\sigma} = -\frac{\frac{\partial U_3^{skip}}{\partial \sigma}}{\frac{\partial U_3^{skip}}{\partial v_1}},$$

are uncertain because, in general, $\frac{\partial U_3^{skip}}{\partial s}$ and $\frac{\partial U_3^{skip}}{\partial \sigma}$ can take either sign.

(iii) The definition of v_H implies the implicit function $U_3^{skip}(v_H) - U_1^{skip}(v_H) = 0$. To determine the sign of $\frac{dv_H}{d\mu}$, we use the Implicit Function Theorem. That is,

$$\frac{dv_H}{d\mu} = -\frac{\frac{\partial U_3^{skip}}{\partial \mu} - \frac{\partial U_1^{skip}}{\partial \mu}}{\frac{\partial U_3^{skip}}{\partial v_1} - \frac{\partial U_1^{skip}}{\partial v_1}}.$$

The numerator, $\frac{\partial U_3^{skip}}{\partial \mu} - \frac{\partial U_1^{skip}}{\partial \mu}$, can be simplified as

$$\frac{1}{2}\frac{(erf(z)-1)\sigma^2}{s^2+\sigma^2},$$

where $erf(z)$ is an error function and always less than 1 (converges to 1 when $z \to +\infty$). Therefore, the numerator is negative. The denominator, $\frac{\partial U_3^{skip}}{\partial v_1} - \frac{\partial U_1^{skip}}{\partial v_1}$ is also negative because $\frac{\partial U_3^{skip}}{\partial v_1} - \frac{\partial U_1^{skip}}{\partial v_1} < 0$ (as shown earlier). Therefore, $\frac{dv_H}{d\mu} < 0$ always holds.

Similarly,

$$\frac{dv_H}{d\delta} = -\frac{\frac{\partial U_3^{skip}}{\partial \delta}}{\frac{\partial U_3^{skip}}{\partial v_1} - \frac{\partial U_1^{skip}}{\partial v_1}} > 0, \quad \frac{dv_H}{dw} = -\frac{\frac{\partial U_3^{skip}}{\partial w} - 1}{\frac{\partial U_3^{skip}}{\partial v_1} - \frac{\partial U_1^{skip}}{\partial v_1}} < 0,$$

because $\frac{\partial U_3^{skip}}{\partial \delta} = w > 0$ and $\frac{\partial U_3^{skip}}{\partial w} - 1 = \delta - 1 < 0$.

Finally, the signs of

$$\frac{dv_H}{ds} = -\frac{\frac{\partial U_3^{skip}}{\partial s}}{\frac{\partial U_3^{skip}}{\partial v_1} - \frac{\partial U_1^{skip}}{\partial v_1}}, \quad \frac{dv_H}{d\sigma} = -\frac{\frac{\partial U_3^{skip}}{\partial \sigma}}{\frac{\partial U_3^{skip}}{\partial v_1} - \frac{\partial U_1^{skip}}{\partial v_1}},$$

are ambiguous because in general $\frac{\partial U_3^{skip}}{\partial s}$ and $\frac{\partial U_3^{skip}}{\partial \sigma}$ can take either sign. ■

Proof of Proposition 3

In order to evaluate $VCR^{skip} - VCR^{trad}$, we make the following simplification:

$$\begin{aligned} VCR^{skip} - VCR^{trad} &= Pr(C_{1a}) + Pr(C_{1b} \cap C_{2a}) - Pr(C_{2a}) \\ &= Pr(C_{1a} \cap C_{2a}) + Pr(C_{1a} \cap C_{2b}) + Pr(C_{1b} \cap C_{2a}) \\ &- [Pr(C_{1a} \cap C_{2a}) + Pr(C_{1b} \cap C_{2a}) + Pr(C_{1c} \cap C_{2a})] \\ &= \underbrace{Pr(C_{1a} \cap C_{2b})}_{Gain} - \underbrace{Pr(C_{1c} \cap C_{2a})}_{Loss}. \end{aligned}$$

The first term above represents the *Gain* in viewer conversion, while the second the *Loss*. Hence,

$$Gain \equiv Pr(C_{1a} \cap C_{2b}) = \int_{v_H}^{+\infty} Pr(v_2 < X - v_1 | v_1) \cdot f_1(v_1) dv_1, \quad (21)$$

where $f_1(v_1)$ is the density function of v_1 and $X \equiv -\mu \cdot (\sigma^2/s^2)$. Similarly,

$$Loss \equiv Pr(C_{1c} \cap C_{2a}) = \int_{-\infty}^{v_L} Pr(v_2 > X - v_1 | v_1) \cdot f_1(v_1) dv_1. \quad (22)$$

Using the above simplification, we establish the proposition's claim as follows. We first show that $Gain = Loss$ at $\delta = 1$. Then we show that $Gain - Loss$ is strictly monotone in $\delta \in (0, 1)$. Specifically, we show that $Gain - Loss$ is strictly increasing (decreasing) when $\mu < 0$ ($\mu > 0$). The proposition's claim is immediately established once we have shown this monotonicity property.

Step 1: $\delta \to 1 \Rightarrow Gain - Loss \to 0$

If $\delta \to 1$, then there is no cost for the viewer to wait for the second signal. Therefore, no viewer will ever skip ad, regardless of their v_1. (That is $\delta \to 1 \Rightarrow v_L \to -\infty$ and $v_H \to +\infty$.) Therefore, viewer conversion rate must be the same under traditional and skippable ads, and $Gain - Loss \to 0$.

Step 2: $\frac{d(Gain-Loss)}{d\delta}$ and μ have opposite signs.

To evaluate this derivative, we exploit the results of Lemma A1, which established that $v_2|v_1$ follows normal distribution with a mean $\Theta(v_1) = \mu + \frac{s^2}{s^2+\sigma^2}(v_1 - \mu)$, and variance $\Gamma^2 = \frac{\sigma^2(2s^2+\sigma^2)}{s^2+\sigma^2}$. This yields the following simplification of the integrands in (21) and (22):

$$\begin{aligned}
Pr(v_2 < X - v_H | v_1 = v_H) &= Pr\left(\frac{v_2 - \Theta(v_H)}{\Gamma} < \frac{X - v_H - \left(\mu + \frac{s^2}{s^2+\sigma^2}(v_H - \mu)\right)}{\Gamma} \middle| v_1 = v_H \right) \\
&= Pr\left(\frac{v_2 - \Theta(v_H)}{\Gamma} < \frac{X - \frac{2s^2+\sigma^2}{s^2+\sigma^2} v_H - \frac{\sigma^2}{s^2+\sigma^2}\mu}{\Gamma} \middle| v_1 = v_H \right) \\
&= \Phi\left(\frac{X - \frac{2s^2+\sigma^2}{s^2+\sigma^2} v_H - \frac{\sigma^2}{s^2+\sigma^2}\mu}{\Gamma} \right) \\
&= 1 - \Phi\left(\frac{X - \frac{2s^2+\sigma^2}{s^2+\sigma^2} v_L - \frac{\sigma^2}{s^2+\sigma^2}\mu}{\Gamma} \right) \\
&= Pr(v_2 > X - v_L | v_1 = v_L) > 0.
\end{aligned}$$

The fourth line above follows from the fact $v_L + v_H = 2X$ (by Proposition 2) and the symmetry property $\Phi(x) = 1 - \Phi(-x)$. The last step can be replicated by expanding $Pr(v_2 > X - v_L | v_1 = v_L)$ analogously as done for $Pr(v_2 < X - v_H | v_1 = v_H)$ above.

The above simplification enables the following derivation:

$$\begin{aligned}
\frac{d(Gain - Loss)}{d\delta} &= -\frac{dv_H}{d\delta} \cdot Pr(v_2 < X - v_H | v_1 = v_H) \cdot f_1(v_H) \\
&\quad -\frac{dv_L}{d\delta} \cdot Pr(v_2 > X - v_L | v_1 = v_L) \cdot f_1(v_L) \\
&= \frac{dv_L}{d\delta} \cdot Pr(v_2 > X - v_L | v_1 = v_L) \cdot [f_1(v_H) - f_1(v_L)]
\end{aligned}$$

The '$-$' sign in front of $\frac{dv_H}{d\delta}$ is because v_H is in the lower bound of the integral in (21). Note also that the $Pr(\cdot)$ and $f_1(\cdot)$ terms in (21) and (22) are independent of δ. The above derivation also employs the fact that $\frac{dv_H}{d\delta} = -\frac{dv_L}{d\delta}$ (implied by Proposition 2). Hence, the above derivative has the same sign as that of $\frac{dv_L}{d\delta} \cdot [f_1(v_H) - f_1(v_L)]$. And because $\frac{dv_L}{d\delta} < 0, \forall\delta \in (0,1)$, the sign of the difference $f_1(v_H) - f_1(v_L)$ fully determines the sign of the derivative above. The density $f_1(v_1)$ is normal and therefore centered around and peaked at $v_1 = \mu$. Thus, $f_1(v_H) - f_1(v_L) > 0$ if and only if $|v_H - \mu| < |v_L - \mu|$.

First consider the case $\mu = 0$. The fact that $v_L + v_H = 0$ implies $f_1(v_H) = f_1(v_L)$. So $\frac{d(Gain-Loss)}{d\delta} = 0$ for all δ. Hence, $Gain = Loss$ holds for $\delta < 1$. That is $VCR^{skip} - VCR^{trad} = 0, \forall\delta \in (0,1)$.

Next suppose $\mu > 0$. In this case, $v_L + v_H = 2X < 0$, which implies $v_L < v_H$ and therefore $v_L < 0$. Then $|v_L - \mu| = \mu - v_L$. Furthermore,

$$v_L < v_H \Rightarrow \mu - v_L > \mu - v_H,$$

$$X < 0 < \mu \Rightarrow 2X < 2\mu \Rightarrow v_L + v_H < 2\mu \Rightarrow \mu - v_L > v_H - \mu.$$

With $\mu - v_L > \mu - v_H$ and $\mu - v_L > v_H - \mu$, it must be that

$$\mu - v_L > |v_H - \mu| \Rightarrow |v_L - \mu| > |v_H - \mu|.$$

Therefore, $f_1(v_H) > f_1(v_L)$ which establishes that $\frac{d(Gain-Loss)}{d\delta} < 0$, $\forall\delta < 1$. As $\delta \to 1$, $Gain - Loss > 0$ decreases from above and converges to zero. In other words, it must be that $VCR^{skip} - VCR^{trad} > 0$, $\forall\delta < 1$. The case of $\mu < 0$ analogously implies $VCR^{skip} - VCR^{trad} < 0$, $\forall\delta < 1$. ■

Proof of Proposition 4

To establish (i) and (ii), we evaluate the optimization conditions implied by (11) and (13). Noting that VCR^i is independent of N for $i \in \{skip, trad\}$, we know that N^i maximizes:

$$\Upsilon(N, u) \equiv G^{-1}(1 - N^i) \cdot h(N^i) \cdot d^i(N^i).$$

We first establish that neither $N^i = 0$ or 1 can be optimal for the platform. Obviously, $N = 0$ leads to zero profit. Assume that zero is the greatest lower bound for the support of t. To have $N = 1$, the advertising rate must be zero, leading to zero profit as well. Hence, we have $N^i \in (0,1)$, for $i \in \{skip, trad\}$. Therefore, any interior maximizer N^i solves

$$\begin{aligned} \frac{\partial\Upsilon}{\partial N} &= \frac{dG^{-1}}{dN} \cdot h \cdot d^i + G^{-1} \cdot h' \cdot d^i + G^{-1} \cdot h \cdot \left[(1-h')w + h' \cdot u\right] \\ &= \left[\frac{dG^{-1}}{dN} + G^{-1}\frac{h'}{h}\right] \cdot h \cdot w + \left[\frac{dG^{-1}}{dN} + 2 \cdot G^{-1}\frac{h'}{h}\right] \cdot h^2 \cdot (u - w) = 0, \end{aligned} \tag{23}$$

where $u = U^i_{ad}$ is given by (6) and (7) and the arguments of G, h, and h' have been dropped for notational convenience.

(i) We now establish $N^{skip} > N^{trad}$ by showing that any solution N^i to (23) is increasing in u. To show this, we need an intermediate result which states that the second square-bracketed term in (23) is positive.

Claim: $\frac{dG^{-1}}{dN} + 2 \cdot G^{-1} \frac{h'}{h} > 0$ always holds at N^i.

Even though for skippable ads, $u > w$ always holds, for traditional ads, $u < w$ is possible. Therefore, we must consider all possible orderings of u and w. First consider the case of $u = w$. Then the term in the first square brackets of (23) must be zero. Then the claim holds immediately.

Next, consider the case of $u > w$. The two terms in the two square brackets above must have opposite signs. With G^{-1} and $\frac{h'}{h}$ both being positive, it must be that $\frac{dG^{-1}}{dN} < 0$. In turn, we must have the first square bracket term being negative and the second positive. That is,

$$\frac{dG^{-1}}{dN} + 2 \cdot G^{-1} \frac{h'}{h} > 0, \quad \frac{dG^{-1}}{dN} + \cdot G^{-1} \frac{h'}{h} < 0.$$

Finally, assume that $u < w$. Rearranging the (23), we have

$$\left[\frac{dG^{-1}}{dN} + G^{-1} \frac{h'}{h} \right] \cdot h \cdot w = \left[\frac{dG^{-1}}{dN} + 2 \cdot G^{-1} \frac{h'}{h} \right] \cdot h^2 \cdot (w - u).$$

It should be clear from the above that the two square bracketed terms have the same sign (and cannot be zero). And because $h \in (0, 1)$ and $u < w$, $h \cdot w > h^2 \cdot (w - u) > 0$ must hold. Then the absolute value of the first square bracketed term must be smaller than the absolute value of the second one. Suppose that both of these terms were negative. Since $G^{-1} \frac{h'}{h} > 0$, it must be that,

$$\frac{dG^{-1}}{dN} + G^{-1} \frac{h'}{h} < \frac{dG^{-1}}{dN} + 2 \cdot G^{-1} \frac{h'}{h} < 0 \Rightarrow \left| \frac{dG^{-1}}{dN} + G^{-1} \frac{h'}{h} \right| > \left| \frac{dG^{-1}}{dN} + 2 \cdot G^{-1} \frac{h'}{h} \right|.$$

That is, the absolute value of the first square bracketed term must be smaller than the absolute value of the second one, a violation. Hence the claim is established.

Using the above claim, we now prove (i) by evaluating $\frac{dN^i}{du}$. From the Implicit Function Theorem, we have

$$\frac{dN^i}{du} = -\frac{\frac{\partial^2 \Upsilon(N,u)}{\partial N \partial u}}{\frac{\partial^2 \Upsilon(N,v)}{\partial N^2}}.$$

Second-order condition for maximization requires that the denominator is negative. Recall that

$$\Upsilon = G^{-1} \cdot h \cdot [(1 - h)w + hu] \Rightarrow \frac{\partial \Upsilon}{\partial u} = G^{-1} \cdot h^2 \Rightarrow$$

$$\frac{\partial^2 \Upsilon(N, u)}{\partial N \partial u} = \frac{\partial \left[G^{-1} \cdot h^2 \right]}{\partial N} = \left[\frac{dG^{-1}}{dN} + 2 \cdot G^{-1} \frac{h'}{h} \right] \cdot h^2,$$

which is positive because of the claim established above. Hence, $\frac{dN^i}{du} > 0$, which implies that $N^{skip} > N^{trad}$ because we know that u is larger for skippable ads: $U_{ad}^{skip} > U_{ad}^{trad}$.

And, lastly,

$$t^{skip} = G^{-1}(1 - N^{skip}) < G^{-1}(1 - N^{trad}) = t^{trad},$$

where the inequality follows from the first result in (i).

(ii) Lemma 4 immediately gives us $d^{skip}(N^{trad}) > d^{trad}(N^{trad})$. And, because $U_{ad}^{skip} > w$, we know $d^{skip}(N)$ is increasing in N. Part (i) above then implies $d^{skip}(N^{skip}) > d^{skip}(N^{trad})$. ■

Proof of Lemma 5

With no demand-enhancing effect, $d^{skip} = d^{trad}$, which further implies $N^{skip} = N^{trad}$ and $t^{skip} = t^{trad}$. This simplifies the comparisons of π_P^i and Π_A^i, for $i \in \{skip, trad\}$, given in (14) and (15):

$$\begin{aligned}
\pi_P^{skip} - \pi_P^{trad} &= VCR^{skip} \cdot G^{-1}(1 - N^{skip}) h(N^{skip}) d^{skip} \\
&- VCR^{trad} \cdot G^{-1}(1 - N^{trad}) h(N^{trad}) d^{trad} \\
&= \left(VCR^{skip} - VCR^{trad}\right) \cdot \left[G^{-1}(1 - N^{trad}) h(N^{trad}) d^{trad}\right],
\end{aligned}$$

and for advertisers' surplus

$$\begin{aligned}
\Pi_A^{skip} - \Pi_A^{trad} &= d^{skip} \cdot VCR^{skip} \int_{t^{skip}}^{\infty} (t - t^{skip}) \cdot \frac{h(N^{skip})}{N^{skip}} g(t) dt. \\
&- d^{trad} \cdot VCR^{trad} \int_{t^{trad}}^{\infty} (t - t^{trad}) \cdot \frac{h(N^{trad})}{N^{trad}} g(t) dt. \\
&= \left(VCR^{skip} - VCR^{trad}\right) \cdot \left[d^{trad} \cdot \int_{t^{trad}}^{\infty} (t - t^{trad}) \cdot \frac{h(N^{trad})}{N^{trad}} g(t) dt\right].
\end{aligned}$$

The impacts of skippable ads on platform profits and advertisers' surplus are precisely determined by $VCR^{skip} - VCR^{trad}$, which is given in Proposition 3.■

Proof of Lemma 6

Assuming $\mu = 0$ means that $VCR^{skip} = VCR^{trad}, \forall N$, by Proposition 3.

(i) Comparing platform profits,

$$\begin{aligned}
\pi_P^{skip} - \pi_P^{trad} &= VCR^{skip} \cdot \left[\Upsilon(N^{skip}, U_{ad}^{skip}) - \Upsilon(N^{trad}, U_{ad}^{trad})\right] \\
&> VCR^{skip} \cdot \left[\Upsilon(N^{skip}, U_{ad}^{skip}) - \Upsilon(N^{trad}, U_{ad}^{skip})\right] \\
&\geq VCR^{skip} \cdot \left[\Upsilon(N^{skip}, U_{ad}^{skip}) - \Upsilon(N^{skip}, U_{ad}^{skip})\right] = 0,
\end{aligned}$$

which follows from the fact that $\Upsilon(N, u)$ is increasing in u and that N^{skip} solves $\max_N \Upsilon(N, U_{ad}^{skip})$.

(ii) Apply a similar argument for advertisers' surplus. First define the integral

$$\Omega(N) \equiv \int_{t(N)}^{\infty} (t - t(N)) \cdot \frac{h(N)}{N} g(t) dt,$$

where $t(N) = G^{-1}(1-N)$. Under Assumption 2,

$$\begin{aligned}\frac{d\Omega(N)}{dN} &= \left\{-[t-t(N)]\cdot\frac{h(N)}{N}g(t)|_{t=t(N)} - \int_{t(N)}^{\infty}\frac{h(N)}{N}\cdot g(t)dt\right\}\frac{dt(N)}{dN} \\ &+ \int_{t(N)}^{\infty}[t-t(N)]\frac{h(N)-Nh'(N)}{N^2}\cdot g(t)dt,\end{aligned}$$

is positive. In particular, the first term is positive because the term in the curly bracket is negative (zero minus a positive integral), and $\frac{dt(N)}{dN} < 0$. And Assumption 2 implies the second term is positive. Combined, we have $\frac{d\Omega(N)}{dN} > 0$. Therefore, we can then write

$$\begin{aligned}\Pi_A^{skip} - \Pi_A^{trad} &= VCR^{skip}\cdot\left[d^{skip}\cdot\Omega(N^{skip}) - d^{trad}\cdot\Omega(N^{trad})\right] \\ &> VCR^{skip}\cdot d^{skip}\cdot\left[\Omega(N^{skip}) - \Omega(N^{trad})\right] \\ &> 0,\end{aligned}$$

where the first inequality is due to $d^{skip} > d^{trad}$ and the second inequality is because $N^{skip} > N^{trad}$ and $\frac{d\Omega(N)}{dN} > 0$. ■

References

[1] Anderson, S. and S. Coate (2005). "Market Provision of Broadcasting: A Welfare Analysis," *Review of Economic Studies* 72: 947-972.

[2] Anderson, S. and J. Gans (2011). "Platform Siphoning: Ad-Avoidance and Media Content," *American Economic Journal: Microeconomics* 3: 1-34.

[3] Anderson, S., J. Waldfogel and D. Stromberg (2015). *Handbook of Media Economics*, Elsevier: North Holland.

[4] Berman, R. (2015). "Beyond the Last Touch: Attribution in Online Advertising," working paper.

[5] Branco, F., M. Sun, J. Miguel Villas-Boas (2012). "Optimal Search for Product Information," *Management Science* 58: 2037-2056

[6] Dukes, A. (2004). "The Advertising Market in a Product Oligopoly," *The Journal of Industrial Economics* 52(3): 327-348.

[7] Dukes, A. and E. Gal-Or (2003). "Negotiations and Exclusive Contracts in Advertising," *Marketing Science* 22(2): 222-245.

[8] Gal-Or, E. and M. Gal-Or (2005). "Customized Advertising via a Common Media Distributor," *Marketing Science* 24: 241-253.

[9] Ghosh, B. and A. Stock (2010). "Advertising Effectiveness, Digital Video Recorders, and Product Market Competition," *Marketing Science* 29(4): 639-649.

[10] Godes, D., E. Ofek and M. Sarvary (2009). "Content vs. Advertising: The Impact of Competition on Media Firm Strategy," *Marketing Science*, 28(1): 20-35.

[11] Gustafson, P. and S. Siddarth (2007). "Describing the Dynamics of Attention to TV Commercials: A Hierarchical Bayes Analysis of the Time to Zap an Ad," *Journal of Applied Statistics* 34(5):585-609.

[12] Halbheer, D., F. Stahl, O. Koenigsberg, and D. Lehmann (2014). "Choosing a Digital Content Strategy: How Much Should be Free?" *International Journal of Research in Marketing*, 31(2):192-206.

[13] Johnson, J. (2013). "Targeted Advertising and Advertising Avoidance," *Rand Journal of Economics* 44: 128-144.

[14] Kaiser, U. and M. Song (2009). "Do Media Consumers Really Dislike Advertising? An Empirical Assessment of the Role of Advertising in Print Media Markets," *International Journal of Industrial Organization*, 27(2): 292-301.

[15] Lewis, R. and J. Rao (2015). "The Unfavorable Economics of Measuring the Returns to Advertising," *Quarterly Journal of Economics*, 130(4): 1941-1973.

[16] Masson, R., Mudambi R., and Reynolds, R. (1990), "Oligopoly in Advertiser-Supported Media," *Quarterly Review of Economics and Business*, 30(2): 3-16.

[17] Peitz, Martin and Tommaso Valletti (2008), "Content and Advertising in the Media: Pay-tv Versus Free-to-Air," *International Journal of Industrial Organization*, 26: 949-965.

[18] Siddarth, S. and A. Chattopadhyay (1998). "To Zap or Not to Zap: A Study of the Determinants of Channel Switching During Commercials," *Marketing Science*, 17(2):124-138.

[19] Theixeira, T., M. Wedel, and R. Pieters (2010). "Moment-to-Moment Optimal Branding in TV Commercials: Preventing Avoidance by Pulsing," *Marketing Science*, 29(5): 783-804.

[20] Wilbur, K. (2008). "A Two-sided, Empirical Model of Television Advertising and Viewing Markets," *Marketing Science* 27: 356-378.

[21] Wilbur, K., L. Xu, and D. Kempe (2013). "Correcting Audience Externalities in Television Advertising," *Marketing Science*, 32(6):892-912.

[22] Zhou, W. (2003). "The Choice of Commercial Breaks in Television Programs: The Number, Length, and Timing," *Journal of Industrial Economics*, 52(3): 315-326.