Beyond the Last Touch: Attribution in Online Advertising

Ron Berman[1]

March 5, 2015

[1]University of Pennsylvania; The Wharton School. email: ronber@wharton.upenn.edu
I would like to thank Zsolt Katona, Ganesh Iyer and Shachar Kariv for their dedicated support during this project. I am also indebted to J. Miguel Villas-Boas, Przemyslaw Jeziorski, Ben Hermalin and Haluk Ergin. Finally, I would like to acknowledge helpful comments by participants at the Marketing in Israel 12 and Marketing Science 2013 conference, as well as seminar participants at CKGSB, UC Riverside, UMD, UofW, UIUC, CMU, Kellogg, Wharton, LBS, NYU, INSEAD, Bocconni, Wash U and UCSD.

Abstract

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

Our results show that when no attribution is being used CPM compensation is more beneficial to the advertiser than CPA payment as a result of free-riding on other's efforts. When an attribution process is added to the campaign, it creates a contest between the publishers and as a result has potential to improve the advertiser's profits when no baseline exists. Specifically, we show that last-touch attribution can be beneficial for CPA campaigns when the process is not too accurate or when advertising exhibits concavity in its effects on consumers. When baseline effects exist we show that last-touch methods become inefficient and a method based on the Shapley value is more profitable for a wide range of campaign parameters.

Using data from a large scale online campaign we apply the model's insights and show evidence for baseline exploitation. An estimate of the publishers' Shapley value is then used to distinguish effective publishers from the exploiting ones, and can be used to aid advertisers to better optimize their campaigns.

1 Introduction

Digital advertising campaigns in the U.S. commanded US $42.8 Billion in revenues during 2013 with an annual growth rate of 18% in the past 10 years,[1] surpassing all other media spending including broadcast TV for the first time in history. In many of these online campaigns advertisers choose to deliver ads through multiple publishers with different media technologies (*e.g.* Banners, Videos, etc.) that can reach overlapping target populations.

This paper analyzes the *attribution* process that online advertisers perform to compensate publishers following a campaign in order to elicit efficient advertising. Although this process is commonly used to benchmark publisher performance, when asked about how the publishers compare, advertisers' responses range from "We don't know" to "It looks like publisher X is best, but our intuition says this is wrong." In a recent survey[2], for example, only 26% of advertisers claimed they were able to measure their social media advertising effectiveness while only 37% of advertisers agreed that their facebook advertising is effective. In a time when consumers shift their online attention towards social media, it is surprising to witness such low approval of its effectiveness.

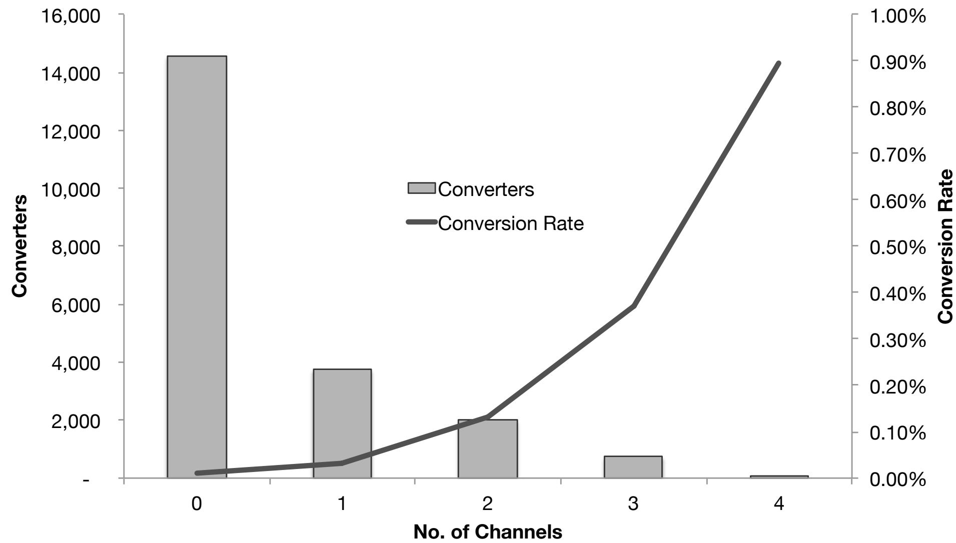
To illustrate the potential difficulties in attribution from multiple publisher usage, Figure 1 depicts the performance of a car rental campaign exposed to more than 13 million online consumers in the UK, when the number of converters[3] and conversion rates are broken down by the number of advertising publishers that consumers were exposed to. As can be seen, a large number of converters were exposed to ads by more than one publisher; it also appears that the conversion rate of consumers increases with the number of publishers they were exposed to.

An important characteristic of such multi-publisher campaigns is that the advertisers do not know a-priori how effective each publisher may be. Such uncertainty may arise, *e.g.*, when publishers can target consumers based on prior information, when using new untested ads or because consumer visit patterns shift over time. Given that online campaigns collect detailed browsing and ad-exposure history from consumers, we ask what obstacles this uncertainty may create to the advertiser's ability to properly mount a campaign.

The first obstacle that the advertiser faces during multi-publisher campaigns is that the ads interact in a non-trivial manner to influence consumers. From the point of view of the advertiser,

[1]Source: 2013 IAB internet advertising revenue report.
[2]Source: "2013 Social Media Marketing Industry Report", www.socialmediaexaminer.com
[3]Converters are car renters in this campaign. Conversion rate is the rate of buyers to total consumers.

Figure 1: Converters and Conversion Rate by Publisher Exposure



getting consumers to respond to advertising constitutes a team effort by the publishers. In such situations a classic result in the economics literature is that publishers can piggyback on the efforts of other publishers, thus creating moral hazard (Holmstrom 1982). If the advertiser tries to base its decisions solely on the measured performance of the campaign, such free-riding may prevent it from correctly compensating publishers to elicit efficient advertising.

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

To address these issues advertisers use contracts that compensate the publishers based on the data collected during a campaign. We commonly observe two types of contracts in the industry: effort based and performance based contracts. In an effort based contract, publishers receive payment based on the number of ads they showed during a campaign. These schemes, commonly known as cost per mille (CPM) are popular for display (banner) advertising, yet their popularity is declining in favor of performance based payments.

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

This paper uses a game theoretical model to focus on the impact of different incentive schemes and attribution processes on the decision of publishers to show ads and the resulting profits of the advertisers. Although attribution is commonly considered solely a measurement issue, our results show that the method used to allocate compensation will heavily influence the ability to create efficient campaign and will bias measurements that do not take the strategic behavior of publishers into consideration.

Our goal is to develop payment and measurement schemes that alleviate the effects of moral-hazard and asymmetric information and yield improved results to the advertiser. To this end Section 3 introduces a static model of consumers, two publishers and an advertiser engaged in an advertising campaign. Consumers in our model belong to one of two segments: a baseline and a non-baseline segment which we call the "affected" segment. Baseline consumers are not impacted by ads yet purchase products regardless. In contrast, exposure to ads from multiple publishers has a positive impact on the purchase probabilities of affected consumers. Our model allows for a flexible specification of advertising impact, including increasing returns (convex effects) and decreasing returns (concave effects) of multiple ad exposures. The choice to use a static game is influenced partly by this flexibility in advertising impact specification, as well as by an attempt to focus on long term steady state impact of attribution and not on the impact during the campaign itself.

The publishers in our model may have private information about whether consumers belong to the baseline and make a choice regarding the number of ads to show to every consumer in each segment. The advertiser, in its turn, designs the payment scheme to be used after the campaign as well as the measurement process that will determine publisher effectiveness.

Section 4 uses a benchmark fixed share compensation scheme to show that moral-hazard is more detrimental to advertiser profits than using effort based compensation. We find that CPM campaigns outperform CPA campaigns for every type of conversion function and under quite general conditions. As ads from multiple publishers affect the same consumer, each publisher experiences an externality from actions by other publishers and can reduce its advertising effort, raising a question about the industry's preference for this method. We give a possible explanation for this behavior by focusing on single publisher campaigns in which CPA may outperform CPM for convex conversion functions.

Since CPA campaigns suffer from under-provision of effort by publishers, we observe that advertisers try to make these campaigns more efficient by employing an attribution process such as last-touch. By adding this process advertisers effectively create a contest among the publishers to receive a commission, and can counteract the effects of free-riding by incentivizing publishers to increase their advertising efforts closer to efficient amounts. We include attribution in our model through a function that allocates the commission among publishers based on the publishers' efforts and performance and has the following four requirements: Efficiency, Symmetry, Pay-to-play and Marginality. To model Last-Touch attribution with these requirements, we notice that publishers are unable to exactly predict whether they will receive attribution for a conversion because of uncertainty about the consumer's behavior in the future. As a result, our model admits last-touch attribution as a noisy contest between the publishers. The magnitude of the noise serves as a measurement of the publisher's ability to predict the impact of showing an additional ad on receiving attribution and depends on the technology employed by the publisher. Our analysis of this noisy process shows that in CPA campaigns with last-touch attribution, publishers increase their equilibrium efforts and yield higher profits to the advertiser when the noise is not too small. When the attribution process is too discriminating or the conversion function too convex, however, no pure strategy equilibrium exists, and publishers are driven to overexert effort. Cases of low noise level can occur, for example, when publishers are sophisticated and can predict future consumer behavior with high accuracy.

The negative properties of last-touch attribution under low noise levels as well as adverse selection[4] has motivated us to search for an alternative attribution method that resolves these issues. The Shapley value is a cooperative game theory solution concept that allocates value among players in a cooperative game, and has the advantage of admitting the four requirements mentioned above along with uniqueness over the space of all conversion functions with the addition of an additivity property. Intuitively, the Shapley value (Shapley 1952) has the economic impact of allocating the average marginal contribution of each publisher as a commission, and this paper proposes its use as an improved attribution scheme. In equilibrium we find that the Shapley attribution scheme increases profit for the advertiser compared to regular CPA schemes regardless of the structure of the conversion function, while it improves over last-touch attribution for small noise ranges.

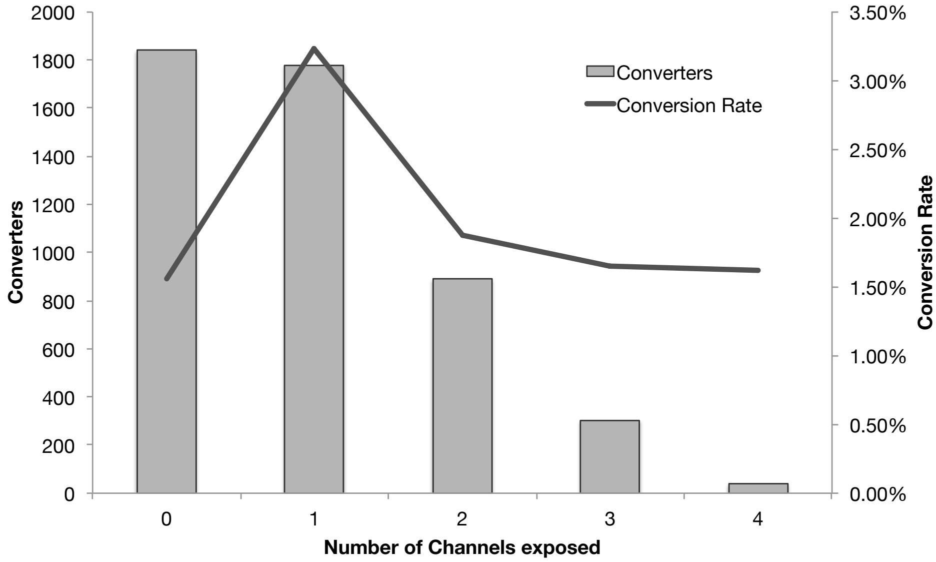
Section 6 analyzes the impact of private information the publisher may have about the baseline conversion rate of consumers when targeting ads. In equilibrium, we show that using both CPM and CPA campaigns without attribution yield less profit to the advertiser than just ignoring the information completely. Adding last-touch attribution improves the advertiser's profit somewhat, but the Shapley value yields higher profit for a wide range of parameters, as it explicitly controls for the effect of the baseline.

Having explored the impact of free-riding and baseline exploitation, Section 7 seeks initial evidence for these phenomena. The data we analyze comes from a car rental campaign in the UK that was exposed to more than 13.4 million consumers. We observe that the budgets allocated to publishers exhibit significant heterogeneity and their estimates of effectiveness are highly varied when using last-touch methods. A simple OLS estimate of publisher effectiveness when interacting with other publishers, however, gives an indication for baseline exploitation as predicted by our model, and lends credibility to the focus on the baseline in our analysis. Evidence for such exploitation can be gleaned from Figure 2, which describes the conversion behavior of consumers who were exposed to advertising only after visiting the car rental website without purchasing (*i.e.*, retargeting ads). If we compare the conversion rate of consumers who were exposed to two or more publishers post-visit, it would appear that the advertising had little effect compared to no exposure post-visit.

We posit that the publishers target consumers with high probability of buying in order to be credited with the sale which is a by-product of the attribution method used by advertisers. To

[4] These results are presented in Section 6.

Figure 2: Converters and Conversion Rate of Visitors by Publisher Exposure



try and identify publishers who free-ride on others, we calculate an estimate of average marginal contributions of publishers based on the Shapley value, and use these estimates to compare the performance of publishers to last-touch methods. The results show that a few publishers operate at efficient levels, while others target high baseline consumers to game the compensation scheme. Consequently, the advertiser for that campaign has made changes to its budget allocations and reduced its costs per acquisition by a substantial margin.

The discussion in Section 8 examines the impact of heterogeneity in consumer behavior on publisher behavior. We conclude with consideration of the managerial implications of proper attribution.

2 Industry Description and Related Work

Online advertisers have a choice of multiple ad formats including Search, Display/Banners, Mobile, Digital Video and more. Among these formats, search advertising commanded 43% of the online advertising expenditures in the U.S. followed by 19% of spending going to display/banner ads. Mobile advertising, which had virtually no budgets allocated to it in 2009, has grown to 17% of total ad expenditures in 2013. The market is concentrated with the top 10 providers commanding more than 70% of the entire industry revenue.

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

When contracting with publishers, advertisers make decisions on the compensation mechanism that will be used to pay the publishers. The two major forms of compensation are performance based payment known as Cost Per Action (CPA) and impression based payment known as Cost Per Mille (CPM). Click based pricing, known as Cost Per Click (CPC), is a performance based scheme for the purpose of our discussion. In 2013 performance based pricing took 65% of industry revenue compared to 41% in 2005. The growth has overshadowed impression based models that have declined from 46% to 33% of industry revenue. Part of this shift can be attributed to auction based click pricing pioneered by Google for its search ads. This shift resulted in significant research attention given to ad auction mechanisms from both an empirical and theoretical perspective which is not covered in this study. It is interesting to note that hybrid models based on both performance and impressions commanded only 2% of ad revenues in 2013. The results of Zhu and Wilbur (2011) and Hu et al. (2014) consider which payment schemes advertisers and publishers should use. A contributing factor to this decision is the market power and asymmetric information held by the publisher about the effectiveness of ads, and the risk averseness and competition faced by the advertiser.

In the past few years, the advertising industry has shown increased interest in improved attribution methods. In a recent survey[5] 54% of advertisers indicated they used a last-touch method, while

[5]Source: "Marketing Attribution: Valuing the Customer Journey" by EConsultancy and Google.

42% indicated that being "unsure of how to choose the appropriate method/model of attribution" is an impediment to adopting an attribution method. Research focusing on the advertiser's problem of measuring and compensating multiple publishers is quite recent, however, with the majority focusing on empirical applications to specific campaign formats. Tucker (2012) analyzes the impact of better attribution technology on campaign decisions by advertisers. The paper finds that improved attribution technology lowered the cost per attributed converter. The paper also overviews theoretical predictions about the impact of refined measurement technology on advertising prices and makes an attempt to verify these claims using the campaign data. Shao and Li (2011), Kireyev et al. (2013), Li and Kannan (2013), Anderl et al. (2014) and Abhishek et al. (2012) build specific attribution models for online campaign data using a conversion model of consumers and interaction between publishers. They find that publishers have strong interaction effects between one another which are typically not picked up by traditional measurements. In addition, Dalessandro et al. (2012) shows through simulation that using the Shapley value can approximate the causal effect of different channels. The work by Jordan et al. (2011) has similar goals to ours, yet analyzes a different aspect of the attribution problem and shows how using last-touch attribution will cause inefficient allocation of ads.

On the theory side, classic mechanism design research on team compensation closely resembles the problem an advertiser faces. Among the voluminous literature on cooperative production and team compensation the classic work by Holmstrom (1982) analyzes team compensation under moral hazard when team members have no private information. Our contribution is the analysis of the interaction of such a team with a profit maximizing principal that can design both a measurement and compensation mechanism. We show that using purely observational data to estimate effectiveness of publishers may lead to strong adverse selection effects, while ignoring the incentive mechanism used to affect the data generation will lead to moral hazard when multiple publishers are involved.

3 Model of Advertiser and Publishers

Consider a market with three types of players: an advertiser, two publishers and a continuous mass of homogenous consumers with measure 1. Our interest is in the analysis of the interplay between the advertiser and publishers through the number of ads shown to consumers and allocation of payment to publishers. We assume advertisers do not have direct access to online consumers, rather they have to invest money and show ads through publishers in order to encourage consumers to purchase their products.

3.1 Consumers

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

The response of consumers to advertising depends on the effectiveness of advertising as well as on the propensity of consumers to convert without seeing any ads which we call the baseline conversion rate. The baseline captures the impact of various states of consumers resulting from exogenous factors such as brand preference, frequency of purchase in steady state and effects of offline advertising prior to the campaign. When each publisher $i \in \{1, 2\}$ shows q_i ads, we let $(q_1 + q_2)^\rho$ denote the conversion rate of consumers who have a zero baseline.[6] We call these consumers "affected" consumers. The symmetric effect of advertising across channels allows us to focus on the effect from using multiple channels in a campaign and the effects of free-riding. Asymmetric publisher effectiveness is discussed in Appendix A.

By denoting the baseline probability of conversion as s, the advertiser expects to observe the following conversion rate after the campaign:

$$x(q_1, q_2) = s + (q_1 + q_2)^\rho (1 - s) \tag{1}$$

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

[6] Additivity of advertising effects is not required but simplifies exposition. Normalization of x to be within $[0, 1]$ yields similar results.

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

Finally, we let the baseline s have an expected value of $\mathbb{E}[s] = \mu$. The flexible structure will let us understand the impact of various campaign environments on the incentives of advertisers and publishers. Initially we assume s is fixed at zero, and then we examine the baseline impact starting in section 6.

3.2 Publishers

Publishers in the model make a simultaneous choice about the number of ads q_i to show to each consumer and try to maximize their individual profits. When showing these ads publishers incur a cost resulting from their efforts to attract consumers to their websites and from the alternative cost of not showing ads from other advertisers. We define the cost of showing q_i ads as $\frac{q_i^2}{2}$.[8]

Both publishers have complete information about the values of ρ and s, as well as the conversion function x and the cost functions.

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

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

3.3 The Advertiser

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

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

$$\pi = x(q_1, q_2) - b_1(q_1, q_2, x) - b_2(q_1, q_2, x) \tag{3}$$

[8] Any cost structure where costs are more convex than advertising effectiveness is sufficient for the results, *e.g.*, linear costs with advertising returns that turn concave.

3.4 Types of Contracts - CPM and CPA

The advertising industry primarily uses two types of contracts - performance based contracts (CPA) in which publishers are compensated on the outcome of a campaign, and effort based contracts (CPM) in which publishers receive payment based on the amount of ads they show.[9] As noted in the introduction, hybrid contracts that make use of both types of payments are uncommon. As shown by Zhu and Wilbur (2011), in environments that allow hybrid campaigns, rational publishers expectations will rule out hybrid strategies by advertisers.

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

$$u_i = q_i p_i^M - \frac{q_i^2}{2} \tag{4}$$

CPA contracts (cost per action) are performance based contracts.[10] In these contracts the advertiser designates a target action to be carried out by a consumer, upon which time a price p_i^A will be paid to the publishers involved in causing the action. The prices are defined as a share of the revenue x, yielding the following publisher profit:

$$u_i = x(q_i, q_{-i}) p_i^A - \frac{q_i^2}{2} \tag{5}$$

The timing of the game is illustrated in Figure 3. The advertiser first decides on a compensation scheme that will be based on the observed efforts q_i, performance x or both. The publishers in turn learn the value of the baseline s and make a decision about how many ads q_i to show to the consumers. Consumers respond to ads and convert according to $x(q_1, q_2)$. Finally, the advertiser observes q_i and x, compensates each publisher with b_i and payouts are realized.

Several features of the model make the analysis interesting and are considered in the next sections. The first is that the interaction among the publishers is essentially of a team generating conversions. A well known result by Holmstrom (1982) shows that no fixed allocation of output

[9] When auctions are used to sell impressions the mechanism is still CPM, while when clicks are sold, the mechanism is akin to a CPA mechanism. The publisher still controls which consumers are exposed to which ads.

[10] Cost per click (CPC) contracts are performance based and fall under CPA.

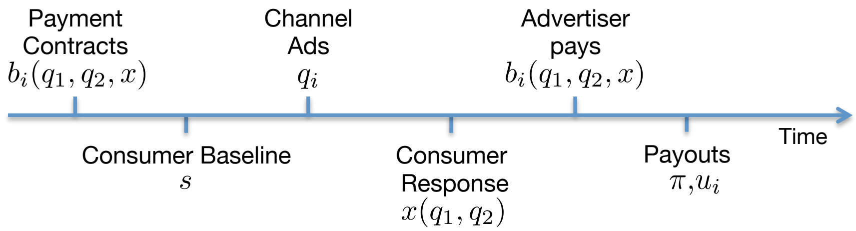


Figure 3: Timing of the Campaign

among team members can generate efficient outcomes without breaking the budget. In our model, however, a principal is able to break the budget, yet its goal is profit maximization rather than efficiency. Nonetheless, the externality that one publisher causes on another by showing ads will create moral hazard under a CPA model as will be presented in the next section.

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

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

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

4 CPM vs. CPA and the Role of Attribution

We start by developing a benchmark that assumes the advertiser is integrated with the publishers and $s = 0$. The optimal allocation of ads is found by solving $\max_{q_1,q_2}(q_1+q_2)^\rho - \frac{q_1^2}{2} - \frac{q_2^2}{2}$ yielding

$$q_1^* = q_2^* = \left(\rho \cdot 2^{\rho-1}\right)^{\frac{1}{2-\rho}} \tag{6}$$

which is strictly increasing in ρ.

In a non-vertically integrated market, when using CPM based payments, the publisher will choose to show $q_i^M = p_i^M$ ads. Because of symmetry, in equilibrium $q^M = p^M = p_1^M = p_2^M$ and the number of ads displayed is:

$$q^M = p^M = \arg\max_p (2p)^\rho - 2p^2 = \frac{\rho^{\frac{1}{2-\rho}}}{2} \tag{7}$$

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

$$q^A = \left(\rho 2^{\rho-1} p^A\right)^{\frac{1}{2-\rho}} \tag{8}$$

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

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

Proposition 1. *When* $0 < \rho < 2$*:*

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

Proposition 1 shows that using CPA causes the publishers to free-ride and not provide enough effort to generate sales in the campaign. The intuition is that the externality each publisher receives from the other publisher gives an incentive to lower efforts, which consequently lowers total output of the campaign. Under CPM payment, however, publishers do not experience this externality and cannot piggyback on efforts by other publishers. By properly choosing a price for an impression, the advertiser can then incentivize the publishers to show a higher number of ads.

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

The final part of Proposition 1 gives one explanation to the market observation that campaigns predominantly use CPA schemes. When the publishers have market power to determine the payment scheme, *e.g.* the case of Google in the search market, the publishers should prefer a CPA based payment when ρ is small, *i.e.*, when publishers are extreme strategic substitutes. In this case, the possibility for free-riding is at its extreme, and even minute changes in efforts by competing publishers increase the profits of each publisher significantly. For example, if consumers are extremely prone to advertising and a single ad is enough to influence them to convert, any publisher that shows an ad following the first one immediately receives "free" commission. If a search engine which typically arrives later in the conversion funnel of a consumer is aware of that, it will prefer to use CPA payment to free-ride on previous publisher advertising.

A question that arises is about the motivation of advertisers, in contrast to publishers, to prefer CPA campaigns over CPM ones. The following corollary shows that when advertisers do not take into account the interaction between the publishers, CPA campaigns are also profitable for the advertiser.

Corollary 1. *When there is one publisher in a campaign and* $0 < \rho < 2$*:*

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

Corollary 1 reverses some of the results of Proposition 1 for the case of one publisher campaigns.

Since free-riding is not possible in these campaigns, we find that CPA campaigns better coordinate the publisher and the advertiser when ads have increasing marginal returns, while CPM campaigns are more efficient for decreasing marginal returns.

4.1 The Role of Attribution

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

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

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

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

$$u_i^A = f_i(q_i, q_{-i}, x)x(q_1, q_2)p^A - \frac{q_i^2}{2} \tag{9}$$

An initial observation is that the process creates a contest between the two publishers for credit. Once ads have been shown, the investment has been sunk yet credit depends on delayed attribution. It is well known (see, *e.g.*, Sisak (2009) and Konrad (2007)) that contests will elicit the agents to overexert effort in equilibrium compared to a non-contest situation. As a result the attribution process can be used to incentivize the publishers to increase their efforts and show a number of ads closer to the vertically integrated market levels.

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

5 Last-Touch and Shapley Value Attribution

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

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

$$f_i^{LT}(q_i, q_{-i}) = Pr(q_i\varepsilon_i > q_{-i}\varepsilon_{-i}) = \int_1^d G\left(\frac{q_i}{q_{-i}}\varepsilon\right) g(\varepsilon)d\varepsilon \tag{10}$$

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

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

Two noticeable properties of last-touch attribution are due discussion. The first is that the more ads a publisher will show, the higher probability it has of being the last one to show an ad before a consumer's purchase. Last-touch attribution therefore has the Marginality property described above. It also trivially has the 3 other properties. The second property is that last-touch attribution makes use of the number of ads shown only in a trivial manner. The credit given to the publisher only depends on the last ad shown to a consumer and whether the consumer had converted.

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

$$(2q)^{\rho-1}\rho p^A = q \tag{11}$$

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

$$(2q)^{\rho-1}\left(2f_1'(1)+\frac{1}{2}\rho\right)p^A = q \tag{12}$$

where $f'(1)$ is the marginal increase in the share of attribution when showing an additional ad when $q_1 = q_2$. Comparing equations (11) and (12) we see that if $\left(2f_1'(1)+\frac{1}{2}\rho\right) > \rho$, then the publisher faces a higher marginal revenue for the same amount of effort. As a result it will have an incentive to increase its effort in equilibrium when the conversion function is concave compared to the case when no attribution was used. Gershkov et al. (2009) show conditions under which such a tournament can achieve Pareto-optimal allocation when symmetric team members use a contest to allocate the revenue among themselves. Whether this contest is sufficient to compensate for free-riding in online campaigns remains yet to be seen.

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

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

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

- *In a CPA campaign there exists a value ρ^d with $\frac{\sqrt{17}-1}{2} > \rho^d > 0$ for $d > 3$ such that if $\rho \leq \rho^d$, there exists a symmetric pure strategy equilibrium. ρ^d is increasing in d.[11] In this equilibrium $q^{A-LT} = \left(\frac{\rho}{2}2^{\rho-1}(\frac{d+1}{d-1} + \frac{\rho}{2})\right)^{\frac{1}{2-\rho}}$.*
- *For any noise level $d > 3$, $q^A < q^M < q^{A-LT}$.*

Proposition 2 shows surprising findings about the impact of last-touch attribution on different campaign types. The contest among the publishers has a symmetric pure strategy equilibrium in a CPA campaign when ρ is low enough or when the noise d is high enough. In these cases, more advertising is being shown in equilibrium compared to regular CPM and CPA campaigns, and more revenue will be generated by the campaign. As a result, the advertiser may make higher profit compared to the case of no attribution as well as for the case of CPM campaigns with no attribution.

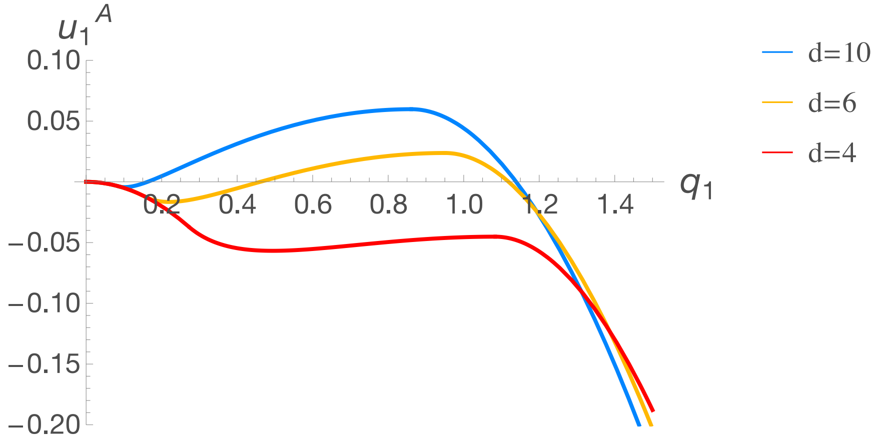
The upper limit on the value of ρ is impacted by two factors. The first is the requirement that publishers make a positive profit in equilibrium. If, for example, the noise is extremely low ($d < 3$), the contest is too discriminating and the effort required from the publishers in equilibrium is too high to make positive profit. In such a case publishers would prefer not to participate in the campaign. The second limit comes from a a technical requirement to ensure quasi-concavity of the profit function around the symmetric equilibrium.

Figure 4 illustrates the best-response of publisher 1 to publisher's 2 equilibrium strategy to give intuition for this result. When the noise becomes small and the contest too discriminating, the

[11] $\rho^d = \min(2 - \frac{4}{d-1}, \frac{\sqrt{1+14d+17d^2}-7-d}{2(d-1)})$

best-response function loses the property of having a maximum point at q^{A-LT} as a result of too strong competition for attribution.

Figure 4: Best Response of Player 1 Under Last-Touch Attribution



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

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

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

5.1 The Shapley Value as an Attribution Scheme

The Shapley value (Shapley 1952) is a cooperative game theory solution concept that allocates value among players in a cooperative game. A cooperative game is defined by a characteristic function $x(q_1, \ldots, q_M)$ that assigns for each coalition of players and their contribution q_i the value they created. For a set of M publishers, the Shapley value is defined as following:[12]

$$\phi_i(x) = \sum_{S \subset (M \setminus i)} \frac{|S|!(|M| - |S| - 1)!}{|M|!}(x_{S \cup i} - x_S) \tag{13}$$

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

The value has the four properties mentioned in the previous section: Efficiency, Symmetry, Null Player and Marginality.[13] In addition, it is the unique allocation function that has these properties

[12]This is a continuous version of the value.

[13]Some of these properties can be shown to be derived from others.

with the addition of an additivity property over the space of cooperative games defined by the conversion function $x(\cdot)$. For the case of two publishers $M = 2$ the Shapley value reduces to:

$$\phi_1 = \frac{x(q1+q2)-x(q_2)+x(q_1)-0}{2} \quad \phi_2 = \frac{x(q1+q2)-x(q_1)+x(q_2)-0}{2} \tag{14}$$

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

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

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

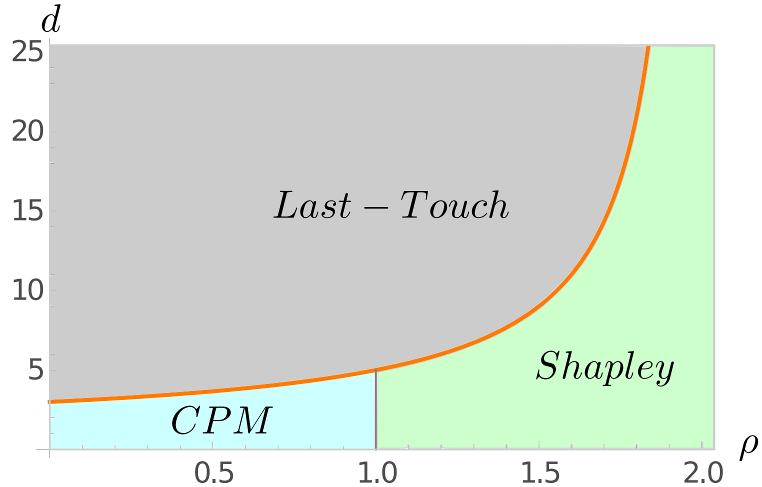
Proposition 3.

- *When $0 < \rho < 2$, using the Shapley value for attribution yields $q^{A-S} = \left(\frac{\rho^2}{4}(2^{\rho-1}+1)\right)^{\frac{1}{2-\rho}}$.*
- *For $\rho < \rho^d$, $q^A < q^{A-S} < q^{A-LT}$.*
- *The profit of the advertiser is higher under Shapley value attribution than under Last-Touch attribution iff $q^{A-S} > q^{A-LT}$,* i.e. $d < \frac{4}{2-\rho} + 1$.
- *The profit of the publisher is higher under the Shapley value attribution than under regular CPM pricing iff $\rho > 1$.*

Proposition 3 is a major result of this paper, showing that the Shapley value can be more profitable when publishers are strategic complements. Contrary to Last-touch attribution, a symmetric pure strategy equilibrium exists for any value of ρ, including very convex functions. When considering lower values of ρ for which Last-Touch attribution improves the efficiency of the campaign, we see that when the noise level d is low enough, the Shapley value will yield better results for the advertiser if $\rho > 1$, while CPM will be better when $\rho < 1$. Figure 5 depicts for which values of ρ and d are each attribution and compensation schemes more profitable.

The intuition behind this result can be illustrated best for extreme values of ρ. When $\rho < 1$ and is extremely low, the initial ads have the most impact on the consumer. As a result, there will be significant free-riding which Last-touch is best suited to solve, while the marginal increase that the Shapley value allocates is not too high. When $\rho > 1$, however, if the noise is low enough, the

Figure 5: Profitability of Each Compensation Scheme



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

publishers will be inclined to show too many ads because of the low uncertainty about their success of being the last one to show an ad. In essence, the competition is too strong and overcompensates for free-riding. The Shapley value in this case is better suited to incentivize the players as the marginal increase from one to the second player is highest with a convex function.

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

The next section discusses the baseline effect on the results of the campaign when using the two attribution methods.

6 Baseline Exploitation

In this Section we relax the assumption that the baseline $s = 0$ and examine its impact on the performance of the attribution schemes. When the baseline is non-zero, the advertiser cannot discern from conversions whether they were caused by advertising effects or simply because consumers had other reasons for converting. As mentioned before, baseline effects can be the result of offline advertising, brand preference of consumers and other sources. As publishers have more information about consumers reaching their sites, this private information may cause adverse selection -

publishers can target consumers with high baselines to receive credit for those conversions.

To illustrate the intuition behind the reason baselines cause a problem for attribution, if we consider equation (12), the first order condition of an advertiser showing q ads to all consumers now becomes:

$$\left[(2q)^{\rho-1}\left(2f_1'(1)+\frac{1}{2}\rho\right)(1-s)+f'(1)s\right]p^A=q \tag{15}$$

In the extreme case of $s=1$, the publishers will elect to show advertising to baseline consumers instead of affected consumers and be attributed credit, without the ads having any impact.

We will perform the full analysis of the impact of attribution with adverse selection while denoting $\mathbb{E}[s]=\mu$ and assuming that $\rho=1$ to simplify the analysis. In a vertically integrated market, if the advertiser had full information it could choose not to target baseline consumers at all yielding the profit $\pi=s+(1-s)(q_1+q_2)-(1-s)\left(\frac{q_1^2}{2}-\frac{q_1^2}{2}\right)$ with an expected value of 1. When the advertiser cannot identify which consumers are in the baseline,[14] it will have to show ads to both segments of the population at the same rate. To determine this rate the advertiser maximizes $\mathbb{E}[\pi]$ yielding an expected profit of $1-\mu+\mu^2$. When $\mu=0$ or $\mu=1$ the advertiser will reach the full information profit, while when $\mu=\frac{1}{2}$, the expected profit will be minimal at $\frac{3}{4}$.

The following result performs a comparison between CPM and CPA schemes without attribution in a fashion similar to section 4. It will serve a benchmark to results that incorporate the two attribution mechanisms considered in this paper:

Proposition 4. *When s is private information of the publishers:*

- *CPA campaigns are only profitable for the advertiser when $\mu<\frac{1}{2}$, yielding profit $\frac{1}{4(1-\mu)}$.*
- *When $\mu<\frac{1}{2}$, CPM campaigns are always more profitable than CPA campaigns, yielding profit $\frac{1+\mu^2}{2}$.*
- *For every value of μ, the profit in an incomplete information vertically integrated market is higher than under both CPM or CPA.*

Proposition 4 shows that adverse selection may impact the profits of advertisers substantially. The first part shows that under some conditions, regular CPA campaigns become non-profitable to

[14] s can be interpreted as a baseline conversion rate of the entire population having one segment impacted by ads only above the baseline. Results will remain the same.

a point where running them is not desired completely. The reason for this stems from the ability of both publishers to free-ride on the exogenous baseline, without the previous effect of a strategic response. Consequently, when there is a high enough expected baseline, it is not advisable to use a CPA campaign at all without a proper attribution mechanism.

The second part reiterates the result from Proposition 1 showing that without any attribution mechanism in place, CPM campaigns are more profitable for advertisers. The last part is somewhat surprising, as it shows that when the publishers have complete information about the advertising environment, the advertiser is worse off than if it had no information and had just shown a number of ads targeting the expected population. In essence, adverse selection exacerbates the impact of moral-hazard.

Given the previous result that attribution helps to alleviate the effects of free riding, we proceed to analyze CPA campaigns with last-touch and Shapley value attribution when $s > 0$. When using last-touch attribution, the advertiser would pay the commission p^{A-LT} only for converting customers which were "touched" by any of the publishers. Consequently, publishers would prefer to show ads to baseline consumers rather than not expose them to ads in order to receive this commission. These ads will have no impact on the conversion rate observed by the advertiser, but the attribution mechanism does not allow the advertiser to determine which of the ads were effective or not. In equilibrium, each publisher will show a different number of ads to each population. The baseline population will be shown $q^b = \sqrt{p^{A-LT} f_1'(1)}$ ads by each publisher, while the affected population will be shown $q = p^{A-LT} (f(1) + 2f_1'(1))$ as before.

Solving for the equilibrium commission p^{A-LT} offered by the advertiser, we find that:

Proposition 5. *When $s > 0$ and last-touch attribution is used:*

- *Running a CPA campaign is profitable for $\mu < \frac{1+3d}{4d}$ when $d > 3$.*
- *There exists a value of the baseline μ^d such that if $\mu > \mu^d$ the channels show more ads to baseline consumers than to affected consumers.*

Proposition 5 shows that the range of values of the baseline μ for which the campaign is profitable for the advertiser is larger than before, for every value of $d > 3$. This increase in profit, however, comes at the cost of showing an inefficient number of ads to baseline consumers, a number of ads which may be higher than the number of ads shown to affected consumers. This means, for example, that performing an analysis using observational data generated using last-touch attribution and assuming more ads were probably shown to the affected consumers is incorrect.

The effect of the baseline and in general additional asymmetric information the publisher may have on the efficacy of last-touch attribution campaigns is profound. When comparing these results to the effect on Shapley value attribution, we find that the impact is much less pronounced:

Proposition 6. *When $s > 0$ and Shapley value attribution is used:*

- *Running a CPA campaign is always profitable.*
- *There exists a range of baseline values μ s.t. Shapley value attribution is more profitable than Last-touch. This range widens as d increases.*
- *Consumers in the baseline will not be shown ads under this scheme.*

Proposition 6 ties up the results from the previous section and the current one and compares the profits and efficiency of the two attribution methods. Since the Shapley value controls for the baseline when calculating the average marginal contribution, the publisher cannot make use of the additional information it has to adversely select the population it will show ads to. Consequently, using a CPA campaign can always be profitable, and in addition, under a wide range of environments, the Shapley value will be more profitable than last-touch campaigns. It should be noted however that when μ is very small and the baseline effect is negligible, last-touch campaigns can still remain relatively profitable.

7 An Application to Online Campaigns

This section applies the insights from Sections 4, 5 and 6 to data from a large scale advertising campaign for car rental in UK.

The campaign was run during April and May 2013 and its total budget exceeded US $\$65,000$ while utilizing 8 different online publishers. These publishers include two online magazines, two display (banner) ad networks, two travel search websites, an online travel agency and a media exchange network. During the campaign more than 13.4 million online consumers[15] were exposed to more than 40.4 million ads.

[15] An online consumer is measured by a unique cookie file on a computer.

The summary of the campaign results in Table 1 shows that the campaign more than quadrupled conversion rates for the exposed population.

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

Table 1: Performance of Car Rental Campaign in the UK

To associate the return of the campaign the advertiser computed last-touch attribution for the publishers based on the last ad they displayed to consumers. Table 2 shows the attributed performance alongside the average cost per attributed conversion. We see that the allocation of budgets correlates with the attributed performance of the publishers, while the cost per conversion can be explained by different average sales through each publisher and quantity discounts.[16]

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

Table 2: Last Touch Attribution for Car Rental Campaign

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

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

[16] Publisher number 3 targets business travelers and yields more profit per attributed conversion.

Publishers	logdiff coef	se
1	-0.657	(0.849)
2	-2.175***	(0.693)
3	-1.960***	(0.703)
4	-0.986	(0.751)
5	-1.559**	(0.691)
6	-1.689**	(0.744)
7	-0.588	(0.748)
8	-0.539	(0.813)
R^2	0.650	
Observations	88	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Logit Estimates of Publisher Effectiveness

7.1 Evidence of Baseline Exploitation and Detection of Free-Riding

Section 6 shows that publishers can target high baseline consumers to deceive the advertiser regarding their true effectiveness. To test the hypothesis that publishers target high baseline consumers, Table 3 shows the results of the logit estimates on the market share differences of each publisher combination in our data.[17] The estimate shows that no publisher adds a statistically significant increase in utility for consumers compared to the baseline. More surprising is the result that a few publishers seem to decrease the response of consumers, thus providing some evidence to the adverse selection theory. It should be noted, however, that the analysis uses aggregate data on consumer exposures and may also suffer from not taking heterogeneity and endogeneity into account. Although these issues may exist, since publishers could favorably choose which consumers to expose the ads to, we should not have received negative estimated effects.

Section 5 predicts that using a last-touch method will lead publishers to strategically increase the number of ads shown, while attempting to free-ride on others. If publishers were not attempting to game the last-touch method, we would expect to see their marginal contribution estimates be close to their last-touch attribution in equilibrium. An issue that arises with using marginal estimates from the data, however, is that the timing of ads being displayed is endogenous and depends on a decision by the consumer to visit a publisher and by the publisher to display the ad. The advertiser cannot control this decisions and order, which may raise an endogeneity issue during analysis.

[17]The estimation technique is described in Appendix C.

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

We developed a modified Shapley value estimation procedure to handle these issues. The computational issues are addressed by using the specific structure of online campaign data as described in Saldanha et al. (2014).

Figure 6: Last Touch Attribution vs. Shapley Value

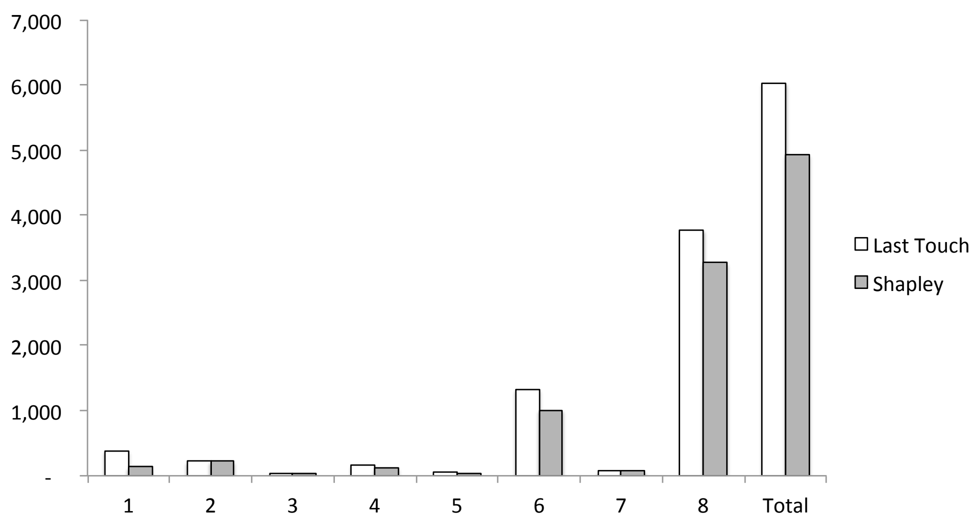


Figure 6 compares the results from a last-touch attribution process to the Shapley value estimation.

More than $1,000$ converters were reallocated to the baseline. In addition, a few publishers lost significant shares of their previously attributed contributions, showing evidence of baseline exploitation.

8 Conclusion

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

The method of last-touch attribution, as we have showed, has the potential to make CPA campaigns more efficient than CPM campaigns under some conditions. In contrast, attribution based on the Shapley value yields well behaved pure strategy equilibria that increase profits over last-touch attribution when the noise is not too small.

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

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

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

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

A key message of this paper is that performance based incentive schemes require a good attribution method to alleviate moral hazard issues. A proper estimate of marginal contributions creates a path for solving this complex problem and providing advertisers with better performing campaigns.

References

Abhishek, Vibhanshu, Peter Fader, Kartik Hosanagar. 2012. Media exposure through the funnel: A model of multi-stage attribution. *Working Paper* .

Anderl, Eva, Ingo Becker, Florian V Wangenheim, Jan Hendrik Schumann. 2014. Mapping the customer journey: A graph-based framework for online attribution modeling. *Working Paper* .

Blake, Thomas, Chris Nosko, Steven Tadelis. 2013. Consumer heterogeneity and paid search effectiveness: A large scale field experiment. *NBER Working Paper* 1–26.

Dalessandro, Brian, Claudia Perlich, Ori Stitelman, Foster Provost. 2012. Causally motivated attribution for online advertising. *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*. ACM, 7.

Dreze, Xavier, François-Xavier Hussherr. 2003. Internet advertising: Is anybody watching? *Journal of interactive marketing* **17**(4) 8–23.

Gershkov, Alex, Jianpei Li, Paul Schweinzer. 2009. Efficient tournaments within teams. *The RAND Journal of Economics* **40**(1) 103–119.

Ghose, Anindya, Sha Yang. 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science* **55**(10) 1605–1622.

Holmstrom, Bengt. 1982. Moral hazard in teams. *The Bell Journal of Economics* 324–340.

Hu, Yu Jeffrey, Jiwoong Shin, Zhulei Tang. 2014. Incentive problems in performance-based online advertising: Cost per click versus cost per action. *Working Paper* .

Jordan, Patrick, Mohammad Mahdian, Sergei Vassilvitskii, Erik Vee. 2011. The multiple attribution problem in pay-per-conversion advertising. *Algorithmic Game Theory*. Springer, 31–43.

Kireyev, Pavel, Koen Pauwels, Sunil Gupta. 2013. Do display ads influence search? attribution and dynamics in online advertising. *Working Paper* .

Konrad, Kai A. 2007. Strategy in contests: An introduction. Tech. rep., Discussion papers//WZB, Wissenschaftszentrum Berlin für Sozialforschung, Schwerpunkt Märkte und Politik, Abteilung Marktprozesse und Steuerung.

Lambrecht, Anja, Catherine Tucker. 2011. When does retargeting work? timing information specificity. *Timing Information Specificity (Dec 02, 2011)* .

Lewis, Randall A, Justin M Rao. 2012. On the near impossibility of measuring advertising effectiveness. Tech. rep., Working paper.

Li, Alice, P.K. Kannan. 2013. Modeling the conversion path of online customers. *Working Paper* .

Manchanda, Puneet, Jean-Pierre Dubé, Khim Yong Goh, Pradeep K Chintagunta. 2006. The effect of banner advertising on internet purchasing. *Journal of Marketing Research* 98–108.

McAfee, R. Preston, John McMillan. 1991. Optimal contracts for teams. *International Economic Review* 561–577.

Rutz, Oliver J., Randolph E Bucklin. 2011. From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research* **48**(1) 87–102.

Saldanha, Alexander, Ron Berman, Keshore Vummarao. 2014. Advertising conversion attribution. US Patent 8,775,248.

Shao, Xuhui, Lexin Li. 2011. Data-driven multi-touch attribution models. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 258–264.

Shapley, Lloyd S. 1952. *A Value for n-Person Games*. RAND Corporation.

Sherman, Lee, John Deighton. 2001. Banner advertising: Measuring effectiveness and optimizing placement. *Journal of Interactive Marketing* **15**(2) 60–64.

Sisak, Dana. 2009. Multiple-prize contests–the optimal allocation of prizes. *Journal of Economic Surveys* **23**(1) 82–114.

Tucker, Catherine. 2012. The implications of improved attribution and measurability for online advertising markets. *Working Paper* .

Yao, Song, Carl F. Mela. 2011. A dynamic model of sponsored search advertising. *Marketing Science* **30**(3) 447–468.

Zhu, Yi, Kenneth C Wilbur. 2011. Hybrid advertising auctions. *Marketing Science* **30**(2) 249–273.

A Extension - Asymmetric Publishers

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

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

When $\theta = 1$, we are back at the symmetric case. When $\theta < 1$, for example, publisher one is more effective as its costs of generating a unit of contribution to conversion are lower. [18]

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

Proposition 7. *When publishers are asymmetric:*

- *Under a CPM contract the same price* $p^M = \left(\frac{\rho(\theta+1)^{\rho-1}}{2}\right)^{\frac{1}{2-\rho}}$ *per impression will be offered to both publishers.*

- *Under a CPA contract, if* $\theta < 1$*, the advertiser will contract only with publisher one. If* $\theta > 1$*, the advertiser will only contract with publisher two.*

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

Performance based campaigns using CPA, in contrast, make the advertiser exclude the worst performing publisher from showing ads. The intuition is that because conversions are generated by symmetric "production" input units of both publishers, the advertiser may just as well buy all of

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

the input from the publisher who has the lowest cost of providing them. The only case when it is optimal for the advertiser to make use of both publishers is when $\theta = 1$ and they are symmetric.

Using a single publisher is significantly less efficient when two are available to the publisher. Adding an attribution process should create an opportunity for this shut-out publisher to compensate for its lower effectiveness with higher effort. This complex problem is left for future work.

B Appendix – Proofs

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

$$\rho(\rho-1)\left(q_i + q^A\right)^{\rho-2} p^A - 1 < 0 \tag{16}$$

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

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

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

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

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

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

Proof of Proposition 2 and Corollary 2. For completeness, we specify the resulting distribution function, $f_1(\frac{q_1}{q_2})$:

$$f_1(\frac{q_1}{q_2}) = \begin{cases} 1 & q_1 \geq dq_2 \\ -\frac{d^2q_2^2 - 2((d-1)d+1)q_1q_2 + q_1^2}{2(d-1)^2q_1q_2} & q_2 < q_1 < dq_2 \\ \frac{1}{2} & q_1 = q_2 \\ \frac{(q_2 - dq_1)^2}{2(d-1)^2q_1q_2} & \frac{q_2}{d} < q_1 < q_2 \\ 0 & q_1 \leq \frac{q_2}{d} \end{cases} \tag{17}$$

We first find the allocation of ads q_1 and q_2 as well as the CPA price p^A in a symmetric pure strategy equilibrium. We will then show these allocations are indeed an equilibrium.

The first order condition the publisher faces is $\rho(q_1+q_2)^{\rho-1}f_1(\frac{q_1}{q_2}) + (q_1+q_2)^{\rho}\frac{f'(\frac{q_1}{q_2})}{q_2} - q_1 = 0$. In a symmetric equilibrium $q_1 = q_2$, resulting in $f'(1) = \frac{1}{2}\frac{d+1}{d-1}$. Plugging in $q_1 = q_2$ and solving for the resulting value yields $q^{A-LT} = \left(2^{\rho-1}p\left[\frac{d+1}{d-1} + \frac{\rho}{2}\right]\right)^{\frac{1}{2-\rho}}$.

The advertiser's profit is $(2 * q^{A-LT})^{\rho}(1-p)$, which is proportional to $p^{\frac{\rho}{2-\rho}}(1-p)$, and its first order is proportional to $\frac{p^{-\frac{2(\rho-1)}{\rho-2}}(2p-\rho)}{\rho-2}$. The solution to the first order being zero is $p = \frac{\rho}{2}$. It can also be seen that the derivative is negative for $p < \frac{\rho}{2}$ and positive for $p > \frac{\rho}{2}$. The profit itself is non-negative for any $q^{A-LT} \geq 0$ and $0 \leq p \leq 1$. Hence this is the equilibrium price set by the advertiser.

Using the values of q^{A-LT} and p, the publisher's profit in equilibrium is $(2q^{A-LT})^{\rho}\frac{\rho}{4} - \frac{(q^{A-LT})^2}{2}$. For this allocation to be an equilibrium we require q^{A-LT} to yield positive profit and yield the maximum profit for each publisher. To show positive profit we solve the following inequality:

$$\begin{aligned}
(2q^{A-LT})^{\rho}\frac{\rho}{4} - \frac{(q^{A-LT})^2}{2} \geq 0 &\iff \\
\left(q^{A-LT}\right)^{2-\rho} \leq 2^{\rho}\frac{\rho}{2} &\iff \\
2^{\rho-1}\frac{\rho}{2}\left[\frac{d+1}{d-1} + \frac{\rho}{2}\right] \leq 2^{\rho}\frac{\rho}{2} &\iff \\
\frac{d+1}{d-1} + \frac{\rho}{2} \leq 2 &\iff \\
\rho \leq 4 - 2\frac{d+1}{d-1} = 2 - \frac{4}{d-1}
\end{aligned}$$

Since the first order is zero at q^{A-LT}, a necessary and sufficient condition for it to be a local maximum is having a negative second order at q^{A-LT}. For $q_1 < q_2$, solving the second order inequalities yields:

$$\begin{aligned}
\frac{2^{\rho-4}\rho((d-1)\rho(d(\rho+3)-\rho+5)+8)(q^{A-LT})^{\rho-2}}{(d-1)^2} - 1 < 0 \\
2^{\rho-4}\rho((d-1)\rho(d(\rho+3)-\rho+5)+8) < (q^{A-LT})^{2-\rho}(d-1)^2 \\
2^{\rho-4}\rho((d-1)\rho(d(\rho+3)-\rho+5)+8) < 2^{\rho-2}\rho\left[\frac{d+1}{d-1} + \frac{\rho}{2}\right](d-1)^2 \\
(d-1)\rho(d(\rho+3)-\rho+5)+8 < 4(d+1)(d-1) + 2\rho(d-1)^2 \\
\rho^2(d-1)^2 + \rho(d-1)(d+7) + 12 - 4d^2 < 0 \\
\rho < \frac{\sqrt{17d^2+14d+1} - d - 7}{2(d-1)}
\end{aligned}$$

When the last inequality follows from restricting $d > 3, 0 < \rho < 2$. Performing the same exercise for $q_1 \geq q_2$ yields a higher bound.

Setting $\rho^d = \min\left(2 - \frac{4}{d-1}, \frac{\sqrt{17d^2+14d+1}-d-7}{2(d-1)}\right)$ yields the condition for a local positive maximum of the publisher's profit.

To show this a global maximum, we partition the publisher's profit into 4 parts:

Part 1: $q_1 < \frac{q_2}{d}$: the profit is always negative.

Part 2: $\frac{q_2}{d} \leq q_1 \leq q_2$: We look at the third derivative of the revenue $(q_1+q_2)^{\rho} f_1(\frac{q_1}{q_2})p$. When plugging in $q_1 = a * q_2$ for $\frac{1}{d} \leq a < 1$ and equating to zero, we receive the following equality, after utilizing the fact that $a > 0$,$d > 0$,$\rho > 0$ and $q_2 > 0$:

$$a^3\rho^3(ad-1)^2 + 3a^2\rho^2(ad-1)(a(d+2)+1) + a\rho(a(a(11-ad((a+3)d+4))+15)+6) - 6(a+1)^3 = 0 \tag{18}$$

A solution to this equation only exists for high enough values of ρ. If a solution does not exist, then since the third derivative at $a = \frac{1}{d}$ is negative, the first derivative of the revenue is strictly concave. This allows for at most two intersection points of the marginal cost with marginal revenue. Since the one at $a = 1$ is a local maximum, the other can only be a local minimum, and $q_1 = q_2$ is a global maximum on $\frac{1}{d} \leq q_1 \leq q_2$

If a solution to Equation 18 exists, it is a solution a that solves:

$$-6 + (-18+6\rho])a + (-18+15\rho-3\rho^2)a^2 + (-6+11\rho-6\rho^2+\rho^3)a^3 + \tag{19}$$

$$(-4d\rho - 3d^2\rho + 6d\rho^2 + 3d^2\rho^2 - 2d\rho^3)a^4 + (-d^2\rho + d^2\rho^3)a^5 == 0 \tag{20}$$

under the condition $0 < \rho < 2$, $d > 3$ and $\frac{1}{d} < a \leq 1$. This solution is unique.

Consequently, the marginal revenue function switches between a convex and a concave shape once. It can be verified that at $q_1 = q_2$ the marginal revenue is locally concave, thus the marginal revenue starts as convex and then turns concave. Since the marginal revenue is positive at $\frac{1}{d}$, this implies that the marginal revenue intersects with the marginal cost at most twice. Using the same argument as before, the other extremum, if exists, is a minimum point.

Part 3: $q_2 < q_1 < dq_2$: Let

$$u' = \frac{d}{da}u_1(aq^{A-LT}, q^{A-LT}) \tag{21}$$

$$= \frac{\rho\left(-a^3(\rho+1) - a^2\left(1-2\left(d^2-d+1\right)\rho\right) - ad^2(\rho-1) + d^2\right)(a+1)^{\rho-1}(q^{A-LT})^{\rho}}{4a^2(d-1)^2} - a(q^{A-LT})^2 \tag{22}$$

Showing that $u' < 0$ for $1 < a < d$ will prove that $u'(a = 1)$ is the maximum point of this segment, as required.

For $a = 1$, $u'(a = 1) = \frac{2^{\rho-3}\rho(d(\rho+2)-\rho+2)(q^{A-LT})^{\rho}}{d-1} - (q^{A-LT})^2 = 0$

We can therefore write the following set of equations, the first is the inequality we would like to prove at $q_1 = aq_2$, the second is the first order condition at $q_1 = q_2$:

$$\rho\left(-a^3(\rho+1) - a^2\left(1-2\left(d^2-d+1\right)\rho\right) - ad^2(\rho-1) + d^2\right)(a+1)^{\rho-1}(q^{A-LT})^{\rho} < 4a^3(q^{A-LT})^2(d-1)^2 \tag{23}$$

$$\frac{2^{\rho-3}\rho(d(\rho+2)-\rho+2)(q^{A-LT})^{\rho}}{d-1} = (q^{A-LT})^2 \tag{24}$$

Dividing the first inequality by the second equality will maintain the inequality, resulting in:

$$\left(\frac{a+1}{2}\right)^{\rho-1}\frac{-a^3(\rho+1) - a^2\left(1-2\left(d^2-d+1\right)\rho\right) - ad^2(\rho-1) + d^2}{d(\rho+2)-\rho+2} < a^3(d-1) \tag{25}$$

Since $\frac{a+1}{2} > 1$ and $0 < \rho < 2$, $\left(\frac{a+1}{2}\right)^{\rho-1} < \frac{a+1}{2}$. Using this fact we can write the last inequality as:

$$(a+1)\frac{-a^3(\rho+1) - a^2\left(1-2\left(d^2-d+1\right)\rho\right) - ad^2(\rho-1) + d^2}{} < 2a^3(d-1)(d(\rho+2)-\rho+2) \tag{26}$$

It can be verified that this inequality holds for $0 < \rho < 2$,$d > 3$,$1 < a < d$, completing this part of the proof.

Part 4: $q_1 \geq dq_2$: The probability of receiving the attribution in this case is 1. The profit function of player 1 becomes:

$$u_1 = (a+1)^{\rho}(q^{A-LT})^{\rho} - \frac{a^2(q^{A-LT})^2}{2} \tag{27}$$

Taking the derivative with respect to a and plugging in q^{A-LT} we receive:

$$u_1' = \rho(a+1)^{\rho-1}(q^{A-LT})^{\rho} - a(q^{A-LT})^2 < 0 \tag{28}$$

$$\rho(a+1)^{\rho-1} < \rho(a+1) < a(q^{A-LT})^{2-\rho} = a2^{\rho-2}\left(\frac{d+1}{d-1} + \frac{\rho}{2}\right) \tag{29}$$

$$a+1 < a\left(\frac{d+1}{d-1} + \frac{\rho}{2}\right) \tag{30}$$

Moving from the second to the third inequality uses the fact that $2^{\rho-2} > 1$. The last inequality holds whenever $d > 3$, $a > d$, $0 < \rho < 2 - \frac{4}{d-1}$. Hence, the maximum point in this segment is at

$a = d$, which we know from the previous segment is not a maximum of the function.

To prove the second part showing that $q^{A-LT} > q^M > q^A$, a simple comparison yields the result.

To prove the corollary, since $q^{A-LT} > q^M > q^A$ and the share of revenues given to the publishers under each scheme is equal, the profit under last touch attribution is higher. □

Proof of Proposition 3. The first order condition the publisher faces in a symmetric equilibrium is: $\frac{\rho(2q)^{\rho-1}+\rho q^{\rho-1}}{2}p^A = q$. The solution, after calculating the equilibrium share offered by the principal is: $q^{A-S} = \left(\frac{\rho^2}{4}(2^{\rho-1}+1)\right)^{\frac{1}{2-\rho}}$.

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

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

Proof of Proposition 4. In a CPA campaign, the publisher will maximize $u_i = s + (1-s)(q_1 + q_2)p_i^A - (1-s)\frac{(q_i)^2}{2}$, yielding $q_i = p_i^A$.

The advertiser's expected profit is:

$$\pi = \mathbb{E}[(s + (1-s)(q_1+q_2))(1-p_1^A-p_2^A)] = (\mu + (1-\mu)(p_1^A+p_2^A))(1-p_1^A-p_2^A) \tag{31}$$

Maximizing and letting $p_1^A = p_2^A$ yields $p^A = \frac{1-2\mu}{4(1-\mu)}$ with an expected advertiser's profit of $\pi^A = \frac{1}{4(1-\mu)}$. We omit the proof that the publishers and advertiser profits are strictly concave for $\rho \le 1$ for any value of μ. We notice that the price p^A is less than $\frac{1}{2}$ only for $\mu < \frac{1}{2}$ and this is a necessary condition for a non-negative advertiser profit.

In a CPM campaign the publisher has the opportunity of showing q_i^b ads to the baseline consumers and q_i ads to regular consumers, and being paid p^M for a regular impression and p_b^M for a baseline impression. The profit then becomes:

$$u_i = s\left(q_i^b p_b^M - \frac{(q_i^b)^2}{2}\right) + (1-s)\left(q_i p^M - \frac{(q_i)^2}{2}\right) \tag{32}$$

Maximizing the publisher's profit yields $q_i^M = p^M$ and $q_i^b = p_b^M$. The publisher's profit increases with the price per impression p, and as a result, the publisher will choose to show the same number of ads to both the baseline and affected consumers, and be paid the highest price offered to either of the segments. Consequently, the advertiser will offer just one price p^M per impression, and $q_i^b = q_i^M$. Solving for the optimal CPM price from the advertiser's point of view yields $p^M = \frac{1-\mu}{2}$ with an expected profit of $\pi^M = \frac{1+\mu^2}{2}$.

We notice that for $\mu < \frac{1}{2}$, π^M is strictly larger than π^A, proving that the advertiser's profit under CPM is higher than under CPA.

The expected profit in a vertically integrated incomplete information market is $1-\mu+\mu^2$, which is strictly higher than π^M under CPM, as claimed in the third item of the proposition. $\square$

Proof of Proposition 5. The profit of a publisher i when showing q_i^{b-LT} ads to baseline consumers and q_i^{LT} ads to affected consumers under last-touch attribution is

$$u_i = s\left(f\left(\frac{q_i^{b-LT}}{q_{-i}^{b-LT}}\right)p - \frac{(q_i^{b-LT})^2}{2}\right) + (1-s)\left((q_i + q_{-i})f\left(\frac{q_i^{b-LT}}{q_{-i}^{b-LT}}\right)p - \frac{(q_i^{b-LT})^2}{2}\right) \tag{33}$$

The symmetric equilibrium allocations are $q^{b-LT} = \sqrt{f'(1)p}$ and $q^{LT} = (\frac{1}{2} + 2f'(1)p)$.

The profit of the advertiser is $\pi^{LT} = (s+(1-s)(q_1+q_2))p$. Plugging in q^{LT} and maximizing the expected profit $\mathbb{E}[\pi^{LT}]$ with respect to p yields the price per conversion of $p = \frac{1}{2} - \frac{\mu}{2(1-\mu)(1+4f'(1))}$. Plugging in $f'(1) = \frac{d+1}{2(d-1)}$ we find that the price is only positive for $\mu < \frac{1+3d}{4d}$, and the profit is positive for every $d > 3$ for this range of expected baseline conversion rate μ.

To prove item 2, we check when $q^{b-LT} > q^{LT}$ and find that when $\mu > \mu^d = \frac{5d^2+6d+5}{8d^2+4d+4}$, more ads are being shown to the baseline consumers than to the affected consumers. The limit of $\lim_{d\to\infty} \mu^d = \frac{5}{8}$, with the range between the upper limit on μ and the lower limit μ^d widening as d increases. $\square$

Proof of Proposition 6. The profit of a publisher i when showing q_i^{b-SH} ads to baseline consumers and q_i^{SH} ads to affected consumers under last-touch attribution is

$$u_i = \frac{s+(1-s)(q_i^{SH}+q_{-i}^{SH}) - s + (1-s)(q_{-i}^{SH}) + s + (1-s)(q_i^{SH}) - s}{2}p - s\frac{(q_i^{b-SH})^2}{2} - (1-s)\frac{(q_i^{SH})^2}{2} \tag{34}$$

We notice that showing ads to baseline consumers does not increase the revenue of the publishers but costs them money. Therefore they will set $q^{b-SH} = 0$, proving part 3. Solving for the equilibrium we find that $q^{SH} = p$.

The profit of the advertiser is $\pi^{SH} = s + (1-s)(q_1 + q_2)(1-p)$, and its expectation is profitable for every $p > 0$ and every value of $0 < \mu < 1$, proving part 1.

Comparing $\mathbb{E}[\pi^{SH}]$ to $\mathbb{E}[\pi^{LT}]$ we find that the former is higher for

$$\frac{3d^2+4d+1}{5d^2+2d+1} - \frac{\sqrt{\frac{3d^4-8d^3+6d^2-1}{(5d^2+2d+1)^2}}}{\sqrt{2}} < \mu < \frac{3d^2+4d+1}{5d^2+2d+1} + \frac{\sqrt{\frac{3d^4-8d^3+6d^2-1}{(5d^2+2d+1)^2}}}{\sqrt{2}} \tag{35}$$

It is easy to show that the lower limit decreases and upper limit increases as d increases, resulting in the range of $\mu \in [0.35, 0.89]$ as $d \to \infty$, proving part 2. □

C Estimation of Publisher Effectiveness

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

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

We then have

$$\ln y_I - \ln(1 - y_I) = \sum_{i \in I} b_i x_i \tag{36}$$

which is estimated using OLS.