Delusion in Attribution: Caveats in Using Attribution for Multimedia Budget Allocation

Peter J. Danaher and Harald J. van Heerde

Abstract
Media attribution is the assignment of a percentage weight to each media touchpoint a consumer is exposed to prior to purchasing. Many firms consider using attribution to allocate media budgets, particularly for digital media, but an important question is whether this is appropriate. An initial hurdle when answering this question is that, despite the surge in interest for media attribution in marketing academia and practice, attribution does not have an agreed-on formal definition. Therefore, this article proposes an attribution formulation based on the relative incremental contribution that each medium makes to a purchase, taking into account advertising carryover and interaction effects. The formulation shows that attribution is proportional to the marginal effectiveness of a medium times its number of exposures. This means that often-used media will have high attribution weights. However, the profit-maximizing allocation for a fixed budget is a function of advertising effectiveness, but not a function of past exposure levels. By offering analytical derivations and studying simulated and empirical data, the paper shows how attribution can offer misleading insights on how to allocate resources across media. Moreover, the empirical example demonstrates that substantial gains in purchase probability can be made using profit-maximizing allocation compared with attribution-based allocation.

Keywords
Advertising carryover, advertising response, attribution modeling, interaction effects, path-to-purchase, probit model, profit-maximizing media allocation

Many firms nowadays use individual-level data to study the trajectory that a consumer follows prior to an eventual purchase. For example, a consumer may be exposed to a banner ad followed by an email before (s)he decides to purchase a product. The increasing availability of individual-level data containing all the media touchpoints, together with purchase behavior, has led to a surge of interest in media attribution for marketing practice (Google Analytics 2012). Attribution modeling is defined as the science of using advanced analytics to allocate appropriate credit for a desired customer action to each marketing touchpoint across all online and offline channels (Kannan, Reinarz, and Verhoef 2016; Moffett, Pilecki, and McAdams 2014). The importance of understanding multiple touchpoint attribution in today’s digitized business environment is underscored by attribution being identified as the number one research priority for marketing managers (MSI Research Priorities 2016–2018).

Marketing practitioners use a variety of attribution methods. One common method is last touchpoint attribution (LTA), which assigns a purchase exclusively to the most recent touchpoint to which a consumer was exposed (Li and Kannan 2014). Although the limitations of LTA have been highlighted in both the academic and trade literatures (see, e.g., Li and Kannan 2014; Singer 2013), LTA remains by far the most popular attribution method used in practice, with 54% of ad agencies and advertisers making media allocation decisions using LTA (Google Analytics 2012). Since LTA is a rather crude measure, alternative attribution methods have been proposed, which take a more balanced (frequency-based) account of all the touchpoints prior to purchase (Google Analytics 2012).

Marketing academia has also started to tackle the attribution question (e.g., Barajas 2015; Li and Kannan 2014; Li et al. [2023]).
2017). For example, Li and Kannan (2014, p. 41) criticize the attribution methods used in practice, because they “do not take into account the timing and sequence of earlier communication and the resulting carryover and spillover effects’ relative impact in leading to website visits and conversions.” Consequently, Li and Kannan (2014) estimate these carryover and spillover effects, allowing measurement of the incremental contribution of multiple channels. Similarly, Xu, Duan, and Whinston (2014) develop econometric models linking media exposure to purchase behavior, and use simulations to assess the impact of each medium. Attribution models are increasingly complete, accounting for aspects such as the long-term effectiveness of advertising media (De Haan, Wiesel, and Pauwels 2016), within- and across-channel effects (Anderl et al. 2016), different types of online search behavior (Ghose and Todri 2016), funnel effects (Abhishek, Fader, and Hosanagar 2015) and endogeneity due to targeting (Zantedeschi, Feit, and Bradlow 2017). We refer to Kannan, Reinarz, and Verhoeft (2016) for an excellent overview of the extant attribution literature.

Despite these advances, a key question remains unanswered: Should attribution be used for allocation? If a medium obtains an 80% weight in attribution, does it mean it should get 80% of the media budget? With the advent of attribution, it is only natural that marketing practitioners “are looking to attribution to understand how different media perform so they can adjust the media mix and improve performance” (Google Analytics 2012, p. 4).

Our formalized expressions for attribution and FBPM allocation permit us to address the focal question of the appropriateness of attribution for allocation decisions. A major new insight of this paper is that attribution is proportional to the incremental contribution that each medium makes to a purchase probability. By expressing attribution in mathematical terms, we provide more precision to the attribution concept. It also facilitates an understanding of the commonalities and differences between attribution and allocation.

Importantly, we develop a profit-maximization schema so that it is consistent with the attribution schema. This means that both schema are based on the same fixed (and known) media budget. We also choose to focus on time-invariant attribution and time-invariant profit-maximizing media allocation. Henceforth, we label our solution to the fixed budget, time-invariant allocation problem as Fixed Budget Profit Maximization allocation (FBPM allocation). This solution is a restricted version of the more general profit maximization problem (with flexible budget and time-varying allocation), as studied by Raman et al. (2012).

In formalizing attribution, we need to account for advertising carryover effects (Leone 1995; Sethuraman, Tellis, and Briesch 2011), whereby a current purchase may be partly due to a medium that a consumer was exposed to previously, and the attribution method should give credit to this medium. While carryover effects seem an obvious requirement, the most-commonly used method in practice (LTA) does not account for these effects.

The measure also must accommodate potential interaction effects between advertising media (e.g., Naik and Raman 2003). The effectiveness of one medium (e.g., a catalog) may be enhanced if a consumer is also exposed to a second medium (e.g., email), leading to synergistic or positive interaction effects between media. Conversely, overexposure to marketing communication in one medium may decrease the effectiveness of another medium, leading to a negative interaction effect (e.g., Burmester et al. 2015). Therefore, a key purpose of this study is to derive a formal expression for attribution that accommodates both advertising carryover and interaction effects.

Our formalized expressions for attribution and FBPM allocation permit us to address the focal question of the appropriateness of attribution for allocation decisions. A major new insight of this paper is that attribution is proportional to the marginal effectiveness of a medium times its number of exposures. While FBPM allocation does not depend on the number of times a consumer is exposed to a medium, attribution does; the higher the number of exposures, the higher the attribution to the medium. This implies that a simplistic use of the proposed attribution metric creates the illusion that some media, especially the frequently-used ones, are very effective and hence they should receive more advertising investment. However, when profit-maximization rules are used, media budgets should be allocated as a function of response (i.e., effectiveness) parameters, not in proportion to (potentially vastly different) attribution weights.

This article proceeds by formally deriving attribution based on a probit model where the purchase decision is driven by
current and past exposure to different advertising media. The model accounts for advertising carryover using an ad stock approach and includes interaction effects. Using the same probit purchase model, we derive the FBPM allocation rule. Next, based on simulated and actual data, we show that allocation decisions that are simplistically based on attribution weights—even if they are measured correctly—lead to a substantial drop in purchase rates compared with FBPM allocation. This is the core message of this article. We find that across the three brands examined in our empirical example, the average percent increase in purchase probability from the current purchase probability for the LTA, probit model attribution, and FBPM allocation are, respectively, 3%, 12%, and 25%.

It is also important to clarify what we do not do in this article. We do not propose another (complex) model for the way consumers step from one touchpoint to another on their journey to a purchase. Rather, our intention is to develop a broad method that links media touchpoints to a purchase, then apportion weights to each of these touchpoints. We also do not dwell much on endogeneity issues or on response parameter heterogeneity issues. While these are obviously important in an empirical setting, the main point of this article is that even in a setting with no endogeneity and homogeneous consumers, attribution and allocation are related but distinct concepts and that media allocation based on formal optimization is superior to allocation based on attribution.

**Attribution: Developing a Measure**

**Last Touchpoint Attribution: A Simple Example**

To explain attribution, we start with a simple example that tracks one consumer as (s)he is exposed to two media, for example, email (A1) and banner ads (A2). Media exposure and purchase incidence are shown in Figure 1. There are seven exposures to email and three to banner ads and for each exposure we observe an associated (non-)purchase. Out of ten total exposures to both media, there are three purchases and seven nonpurchases. Attribution of media touchpoints to a purchase is about assigning a medium (and possibly several media) to a purchase event. That is, given that a purchase occurs, what are the relative contributions of media A1 and A2 in driving the purchase?

A commonly used method for attribution in the digital media industry is to assign 100% weight to the medium to which a person is most recently exposed prior to a purchase, known as last touchpoint attribution (LTA), defined as:

$$\text{Attribution}_{m}^{\text{LTA}} = \frac{n_{pm}}{n_{p}},$$

where $n_{pm}$ is the number of instances where a purchase directly follows a touchpoint via medium m and $n_{p}$ is the total number of touchpoints where a purchase occurs ($n_{p} = \sum_{m} n_{pm}$). In Figure 1, there are three purchases, two of which are immediately preceded by an email (exposures 1 and 6) and one by a banner ad (exposure 10). Hence, LTA suggests a 66.7% attribution weight to email and 33.3% to banner ads.

While the LTA method is simple, an obvious weakness is that it ignores well-established advertising carryover (e.g., Leone 1995; Sethuraman, Tellis, and Briesch 2011) and media interaction effects (e.g., Naik and Raman 2003). Another characteristic of LTA is that it is influenced by the frequency with which different advertising messages are sent to consumers. For example, email has a high frequency as it is almost costless to send. Even if email is completely ineffective, LTA will give it significant weight only because email is usually sent so frequently that it happens to be close in time to a purchase, mostly by coincidence.

A limitation of the LTA method and its variants is that they are not based on linking the probability of purchase to the advertising media. Previous studies have shown that media differ in their ability to elicit a purchase (e.g., Danaher and Dagger 2013; Dinner, van Heerde, and Neslin 2014) and so an important building block for our attribution method is a model that explicitly links the probability of purchase to different media, in the current and prior periods. The model allows for the differing strengths of each media at effecting a purchase and allows for advertising carryover effects as well as interaction effects among media.

**Proposed Measure for Attribution**

We now develop an attribution measure. Given that a purchase is observed, attribution measures the contribution to this purchase from medium m relative to all media. We start with a case in which the purchase outcome (y) is captured by a binary (0/1) variable and exposure to medium m ($A_{m}$) is binary as well. We generalize this later to other measurement scales for
the purchase outcome (e.g., a continuous measure such as dollar amount) and the medium (e.g., a continuous measure such as the number of emails received).

We use a purchase probability model to calculate the relative importance of each medium at influencing a purchase. For each purchase occasion we then calculate the increment in purchase probability due to the actual exposure to each medium, and then apportion these increments to get a weight for each medium. To formalize this notion, we reprise the example from the previous section and introduce a probit purchase probability model that is a function of exposure to two media. The initial model is straightforward with just current effects, but later we add lagged and interaction effects and examine other models.

Case 1: Two Media, Only Current Effects, No Lagged or Interaction Effects

Suppose that consumers are exposed to just email and banner ads. We use a probit model to link these exposures to purchase incidence, using the following latent model:

\[
y_{it}^* = \beta_0 + \beta_1 A_{1it} + \beta_2 A_{2it} + \epsilon_{it},
\]

where the observed outcome is \(y_{it} = 1\) if \(y_{it}^* > 0\) and \(y_{it} = 0\) otherwise. \(A_{mit}\) is 1 if person \(i\) is exposed to medium \(m\) (\(m = 1, 2\)) at time \(t\) and zero otherwise. Let \(\epsilon_{it} \sim N(0, 1)\), where \(i = 1, 2, \ldots, N\) and \(t = 1, 2, \ldots, T\). The probit model for purchase incidence calculates the probability a purchase is made given covariates: \(Pr(y_{it}^* = 1) = \Phi(\beta_0 + \beta_1 A_{1it} + \beta_2 A_{2it})\), where \(\Phi(\cdot)\) is the standard normal cumulative distribution function.

Define \(\Delta_{mit}\) as the increment in purchase probability due to medium \(m\) for consumer \(i\) at time \(t\). This increment is zero if the consumer is not exposed to medium \(m\) at this purchase occasion \((A_{mit} = 0)\) because the medium does not contribute to this purchase. Thus,

\[
\Delta_{mit} = \begin{cases} 
Pr(y_{it} = 1|A_{mit} = 1) - Pr(y_{it} = 1|A_{mit} = 0) & \text{if } A_{mit} = 1 \\
0 & \text{if } A_{mit} = 0.
\end{cases}
\]

Using the fact that \(A_{mit}\) is an indicator variable for whether or not the consumer is exposed to medium \(m\), we can write Equation 3 more succinctly as

\[
\Delta_{mit} = [Pr(y_{it} = 1|A_{mit} = 1) - Pr(y_{it} = 1|A_{mit} = 0)] \times A_{mit}.
\]

We now define the marginal effect of a current exposure to medium \(m\) as:

\[
ME_{mit}^C = Pr(y_{it} = 1|A_{mit} = 1) - Pr(y_{it} = 1|A_{mit} = 0).
\]

Using this definition, we can restate Equation 4 as

\[
\Delta_{mit} = ME_{mit}^C \times A_{mit}.
\]

Hence, \(\Delta_{mit}\) is the marginal effect of medium \(m\) on purchase probability \((ME_{mit}^C)\) times an indicator function for whether the consumer is exposed to the medium \((A_{mit} = 1)\) or not \((A_{mit} = 0)\). Equation 6 is a crucial building block for our attribution measure as it captures the effectiveness of a medium and the exposure to it.

To illustrate the computation of \(\Delta_{mit}\) for the probit model and medium 1 we obtain

\[
\Delta_{1it} = ME_{1it}^C \times A_{1it} = [\Phi(\beta_0 + \beta_1 \times 1 + \beta_2 A_{2it}) - \Phi(\beta_0 + \beta_1 \times 0 + \beta_2 A_{2it})] \times A_{1it}.
\]

Similarly, we can define \(\Delta_{2it}\) for medium 2. We define attribution for medium \(m\) and individual \(i\) at the \(t\)th observation as

\[
Attribution_{mit} = \frac{\Delta_{mit}}{\Delta_{1it} + \Delta_{2it}}, \quad m = 1, 2.
\]

Conventionally, attribution has been calculated only for those situations where a purchase occurs (Li and Kannan 2014). Consequently, we now average across all observations where a purchase occurs to give an overall measure of attribution for medium \(m\), defined as

\[
Attribution_{m} = \frac{1}{n_p \{(i,t,y_{it}=1\}} \sum Attribution_{mit},
\]

where \(n_p = \sum_{t=1}^{T} \sum_{i=1}^{N} y_{it}\) is the total number of purchases in the observation period. Thus defined, attribution is the relative contribution that medium \(m\) makes to all purchases.

Computation. As Equation 7 shows, attribution depends on the parameters of the probit model \((\beta_1, \beta_2)\) and the observed advertising \((A_{1it}, A_{2it})\) for each observation. To calculate it empirically from the data, step through each observation in the individual-level data and select just those observations where a purchase is made. For these purchase occasions insert the observed values of \(A_{1it}\) and \(A_{2it}\) into Equation 7, and the corresponding equation for medium 2, and thus obtain empirical values of \(Attribution_{mit}\) in Equation 8 and \(Attribution_{m}\) in Equation 9.

Case 2: Two Media, Lagged Effects but No Interaction Effects

We now allow for carryover effects of media exposure. We replace the two current exposure terms \((A_{1it}, A_{2it})\) by ad stock specifications: \(AS_{mit} = (1 - \lambda_m)A_{mit} + \lambda_m AS_{mit-1}\), where \(AS_{mit}\) represents ad stock for medium \(m\) for consumer \(i\) in period \(t\), and \(\lambda_m\) is a decay parameter in the interval \([0, 1]\) (see, e.g., Braun and Moe 2013; Dinner, van Heerde, and Neslin 2014). If \(\lambda_m = 0\), there are no carryover effects and we revert to the current-effect only situation (i.e., \(AS_{mit} = A_{mit}\)). Using ad stock, the probit model becomes

\footnote{In Web Appendix A we discuss an alternative ad stock formulation that is sometimes used in the literature: \(AS_{mit}^* = \gamma_m A_{mit} + \lambda_m AS_{mit-1}\). We show that this is mathematically equivalent to our approach.}
\[ y_{it}' = \beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it} + \epsilon_{it}. \] (10)

The contribution of medium m to the purchase probability, \(\Delta_{mit}\), is the difference between purchase probability with the observed exposure (current or past) to medium m and without it. For medium 1, we can write \(\Delta_{1it} = \Pr(y_{it} = 1|AS_1, AS_2) - \Pr(y_{it} = 1|AS_1 = 0, AS_2)\). Using the probit model, we can calculate this as

\[
\Delta_{1it} = \Pr(y_{it} = 1|AS_1, AS_2) - \Pr(y_{it} = 1|AS_1 = 0, AS_2) = \Phi(\beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it}) - \Phi(\beta_0 + \beta_1 \times 0 + \beta_2 AS_{2it}).
\] (11)

We now link \(\Delta_{mit}\) to exposure levels by using a Taylor series expansion to approximate the probit function, leading to

\[
\Delta_{1it} = \Phi(\beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it}) - \Phi(\beta_0 + \beta_1 \times 0 + \beta_2 AS_{2it}) \approx \phi(\beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it}) \times \beta_1 \times AS_{1it},
\] (12)

where \(\phi(.)\) is the pdf of a standard normal distribution. Equation 12 shows that the purchase utility contribution of medium 1 is approximately equal to (i) the pdf of the normal distribution times (ii) the marginal effectiveness of the medium times (iii) the level of ad stock present for that medium. This means that the contribution grows with a more effective medium (term ii) and with more exposures to the medium (higher ad stock levels; term iii). The pdf term \(\phi(.)\) (term i) is the same for all media and acts as a normalizing constant.

**Sequencing issues.** Since we compute the increment \(\Delta_{mit}\) at each time t, our attribution measure implicitly accounts for the sequencing of exposures. Suppose a consumer is exposed to medium 1 at time \(t - 1\) and to medium 2 at time \(t\), and both media have carryover effects. The lift in probability due to medium 2 at \(t\) is calculated relative to the base probability that results from the prior exposure of medium 1 at \(t - 1\). For probability models that are typically nonlinear, this lift will be different than when the exposure order is reversed. If we only knew that a consumer was exposed to media 1 and 2 and did not know the order, we would have to cycle through all possible sequences (2 in this case) and assign attribution based on the Shapley (1953) value, as explained in Berman (2017) and Li and Kannan (2014).

**Case 3: Two Media, Lagged Effects and Interaction Effects**

We now augment Equation 10 with the interaction between the two media:

\[ y_{it}' = \beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it} + \beta_3 AS_{1it} AS_{2it} + \epsilon_{it} \] (13)

We now define the contribution of medium 1 as

\[
\Delta_{1it} = \Phi(\beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it} + \beta_3 AS_{1it} AS_{2it}) - \Phi(\beta_0 + \beta_1 \times 0 + \beta_2 AS_{2it} + \beta_3 \times 0 \times AS_{2it})
\] (14)

Based on a Taylor series expansion, this contribution can be approximated by

\[
\Delta_{1it} \approx \phi(\beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it} + \beta_3 AS_{1it} AS_{2it}) \times \Phi(\beta_1 AS_{1it} AS_{2it} AS_{1it}) \times \Phi(\beta_0 + \beta_1 \times 0 + \beta_2 AS_{2it} + \beta_3 \times 0 \times AS_{2it}).
\] (15)

We define \(ME_{1it} = \beta_1 + \beta_3 AS_{2it}\) as the relevant moderated marginal effect. The R superscript highlights that it may not be constant but depends on relevant interaction effects and other covariates for consumer \(i\) at time \(t\). Using \(ME_{1it}\), we can write Equation 15 as:

\[
\Delta_{1it} \approx \phi(\beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it} + \beta_3 AS_{1it} AS_{2it}) \times ME_{1it} AS_{1it}.
\] (16)

Equation 16 shows that the probability lift due to medium 1 is approximately proportional to its marginal effectiveness times the exposure level to medium 1 captured through ad stock.

**Proposed Attribution Measure for General Case**

For a general case, the purchase outcome (\(y\)) may be binary (e.g., yes/no purchase), continuous (e.g., dollars spent), discrete (e.g., number of items bought), or any other purchase measure at the individual level. Medium \(m\) can be binary (e.g., exposure to advertising) or continuous (e.g., number of emails received). The model linking exposure to purchase can be probit (binary \(y\)), linear regression (continuous \(y\)), or any other model suitable for the context. The dependence of \(y\) on the stock of exposure to the \(M\) media is represented by \(y = F(AS_{mit}, AS_{jit} j \neq m)\), where \(F(.)\) is a function that links the covariates to the dependent variable (e.g., \(F = cdf\) of a standard normal distribution in the probit example).

We define \(\Delta_{mit}\) as the contribution of medium \(m\) to a purchase, which can be measured as the difference in \(y\) for the observed stock of medium \(m\) versus zero stock:

\[
\Delta_{mit} = F(AS_{mit} = AS_{mit} \text{obs}, AS_{jit} j \neq m) - F(AS_{mit} = 0, AS_{jit} j \neq m),
\] (17)

where \(AS_{mit} \text{obs}\) are the observed values for \(AS_{mit}\).

Using a Taylor series expansion, we can approximate \(\Delta_{mit}\) by

\[
\Delta_{mit} \approx \Phi(\beta_0 + \beta_1 AS_{mit} + \beta_2 AS_{2mit}) - \Phi(\beta_0 + \beta_1 \times 0 + \beta_2 AS_{2mit}) \approx \phi(\beta_0 + \beta_1 AS_{mit} + \beta_2 AS_{2mit}) \times \beta_1 \times AS_{mit}.
\]
\[
\Delta_{mit} \approx f(AS_{1it} = AS_{2it} = \ldots, AS_{Mit} = AS_{Mit}^o) \times ME_{mit}^R \times AS_{mit},
\]

where \(f(x)\) is the derivative of \(F(x)\) with respect to \(x\). Using Equation 18, we can approximate the (exact) overall attribution in Equation 9 by:

\[
\text{Attribution}_{in} = \frac{1}{n_p} \sum_{i \in \{1, \ldots, n\}} \frac{\Delta_{mit}}{\sum_{j=1}^{M} \Delta_{jit}} \approx \frac{1}{n_p} \sum_{i \in \{1, \ldots, n\}} \frac{\sum_{j=1}^{M} f(AS_{1it} = AS_{2it} = \ldots, AS_{Mit} = AS_{Mit}^o) \times ME_{mit}^R \times AS_{mit}}{\sum_{j=1}^{M} \sum_{k=1}^{M} f(AS_{1it} = AS_{2it} = \ldots, AS_{Mit} = AS_{Mit}^o) \times ME_{mit}^R \times AS_{jit}} = \frac{1}{n_p} \sum_{i \in \{1, \ldots, n\}} \frac{\sum_{j=1}^{M} ME_{mit}^R \times AS_{mit}}{\sum_{j=1}^{M} ME_{mit}^R \times AS_{jit}}
\]

(19)

Properties of the Proposed Attribution Measure

The proposed attribution measure in Equation 19 has the following properties:

1. **Attribution grows in the marginal effect of a medium on the purchase outcome.** For medium \(m\), the larger the marginal effect \(ME_{mit}^R\), the larger the attribution measure in Equation 19. In other words, a more effective medium will get more credit.

2. **When a medium has no direct or interactive effect on the purchase outcome, it obtains a zero attribution weight.** If medium \(m\) is ineffective (\(ME_{mit}^R = 0\)), Equation 19 implies that \(\text{Attribution}_{in} = 0\). It is noteworthy that this basic requirement does not apply for the LTA method, in which a medium can be associated with a purchase just because it happened to be the most recent medium to which a consumer is exposed.

3. **The attribution measure increases in the number of exposures to a medium.** Suppose there are two media with exactly the same, constant, marginal effect and decay parameter, so that \(ME_{1it}^R = ME_{2it}^R = ME^R\). Hence, whenever a consumer is exposed to these two media, their contribution to a purchase is the same. However, suppose the consumer is exposed twice as often to medium 1 as to medium 2, so that \(AS_1 = 2AS_2\). This implies that the overall contribution of medium 1 will be twice the contribution of medium 2, thereby doubling the attribution measure in Equation 19 for medium 1 versus medium 2.

4. **Attribution allows for ad carryover and interaction.** We allow for these effects through the use of ad stock and the purchase probability model in Equation 13.

5. **If there is no carryover effect and no interaction effect, the attribution measure reduces to LTA.** When there are no carryover or interaction effects but there is a nonzero current effect, we are back to Case 1 described previously. If consumer \(i\) purchases at time \(t\), only the medium at that time is relevant. Assuming it is medium \(m\), the contribution of this medium is \(\Delta_{mit} = ME_{mit}^R \times \Delta_{mit} = ME_{mit}^R\), but the other media (\(j \neq m\)) have zero contribution. Hence, \(\text{Attribution}_{in} = \Delta_{mit} / \sum_{j=1}^{M} \Delta_{jit} = 1\) while \(\text{Attribution}_{in} = 0\) for \(j \neq m\). Thus, the purchase at time \(t\) by consumer \(i\) is fully attributed to medium \(m\) and \(\text{Attribution}_{in} = n_{mit} / n_p\), where \(n_{mit}\) is the number of instances where a purchase results from an exposure to medium \(m\). This attribution expression is the same as LTA in Equation 1.

Alternative measures for attribution. Li and Kannan (2014) (hereinafter, LK) and Xu, Duan, and Whinston (2014) (hereinafter, XDW) propose alternative approaches to attribution, albeit without explicit expressions, which makes it difficult to make exact comparisons to ours. A similarity across the three approaches is that they are all based on the relative marginal contribution of each medium to the purchase probability. Thus, LK and XDW satisfy properties 1 (grows in the marginal effect) and 2 (zero when marginal effect is zero). The approaches of LK and XDW account for carryover effects, which is achieved in XDW by calculating the contribution of each medium for each observation, as in our approach, and achieved by LK by using the Shapley value based on all possible sequences. However, neither the LK nor XDW methods are specifically developed to capture interaction effects (part of property 4), and it is not clear whether they reduce to LTA in the absence of carryover and interaction effects (property 5).

A key difference arises for property 3 (attribution increases in the number of exposures), which LK and our method share. Thus, LK and our measures capture overall attribution, which is the percentage of purchases that can be attributed to each medium summed across all observations. On the other hand, XDW yield an average conversion probability for each medium, averaged across the number of exposures for each medium.

The conclusion is that our measure is most similar to LK. In that sense, our article’s contribution is not to offer a superior measure. Instead, our intention is to offer a formal expression for attribution (new to the literature) and, also new to the literature, contrast it with FBPM allocation as shown in the next section.

---

3 Note that computing the weights as ratios of increments is similar in spirit to the Shapley Value used in attribution literature (Berman 2017; Li and Kannan 2014). The difference is that we do not have to use permutations over different orders of touchpoints because we observe the exact order of touchpoints.

4 We base the discussion of these properties mostly on the approximation in Equation (19) because these properties can be directly seen from Equation 19, but the same logic applies to the exact Equation (17).
**Fixed Budget Profit-Maximizing Media Allocation**

As explained earlier, it is quite natural to assume that attribution can be used for multimedia allocation purposes. Indeed, our elaborate derivation of an attribution measure may create the suggestion that attribution is suitable for deciding what fraction of a media budget goes to each medium. However, even our formal expression for attribution is not suited for allocation purposes, as we explain next.

For profit-maximizing media allocation, the objective is to allocate a fixed media budget across available media platforms to maximize profit (see, e.g., Danaher and Dager 2013; Rust 1986). We want to make the comparison between attribution and FBPM allocation as clean and fair as possible. This means that both are based on (i) the same econometric model; (ii) the same set of N customers; (iii) the same horizon of T time periods; (iv) the same fixed and known budget B spent on advertising, leading to (v) constant percentages (both for attribution and allocation) across media.

At a micro level, a manager’s media allocation decision involves $A_{it}$ and $A_{2it}$ (spend on media 1 and 2 for person $i$ at time $t$). However, recall that attribution yields the percentage of the observed purchases that are attributable to each medium, $m = 1, \ldots, M$, summing to 100%. Therefore, to put the optimization method on a level footing with attribution, we need to set up the optimization method to establish a constant percentage allocation across media, as opposed to setting weekly or daily advertising exposure targets for each individual. Such tactical targeting decisions are also important of course, but they are secondary to the first major decision, which is how much of a fixed (and known) total budget B should be apportioned to each medium. Consequently, we define the optimization problem as fixed (and known) total budget $B$ should be apportioned to each targeting decision $m$ where a manager seeks to find the profit-maximizing constant percentage allocation across media, subject to a budget constraint. 

As Web Appendix B explains, an appealing feature of the ad stock formulation $AS_{mit}$ is that for constant advertising levels ($A_{mit} = A_m$, $\forall i, t$), the resulting ad stock level is also the same: $AS_{mit} = AS_m = A_m$, $\forall i, t$. Thus, the decision variables $A_1$ and $A_2$ can be directly used as arguments in the probit model (Equation 13) and the optimization setup can be formally written as:

$$\text{Max } \Pi(A_1, A_2) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left( K \times \text{Pr}(y_{it} = 1|A_1, A_2) \right) - B$$

subject to

$$\sum_{i=1}^{N} \sum_{t=1}^{T} (c_1A_1 + c_2A_2) = B$$

The FBPM solution, derived in Web Appendix B, is:

$$A_1^* = \frac{c_2\beta_1 - c_1\beta_2 + \beta_3(B/NT)}{2c_1\beta_3},$$

$$A_2^* = \frac{c_1\beta_2 - c_2\beta_1 + \beta_3(B/NT)}{2c_2\beta_3}.$$  

This solution depends on the response parameters $(\beta_1, \beta_2, \beta_3)$ but not on the ad stock parameters $(\lambda_1, \lambda_2)$. This is because the ad stock formulation implies that $\beta_m$ captures the full long-term effect including carryover effects. This can be seen by cumulating the impact of a unit shock in advertising for medium $m$ in period $t$ on utility in periods $t$, $t+1$, $t+2$, ... ($\lambda_1 \lambda_2 \ldots \lambda_{T}$, where $\lambda_T = 1$). Obvi-ously, when estimating the model, the carryover parameters and response parameters are determined jointly, and hence the solution implicitly depends on the carryover parameters.

**Comparison between FBPM allocation and attribution.** To understand the difference between FBPM allocation and attribution, we express the FBPM allocation in relative terms:

**FBPM allocation to medium 1**

$$A_1^* = \frac{A_1}{A_1^* + A_2^*} = \frac{c_2\beta_1 - c_1\beta_2 + c_2\beta_3(B/NT)}{c_1\beta_2 + c_2\beta_2 - c_1c_2(\beta_1 + \beta_2) + (c_1 + c_2)\beta_3(B/NT)}.$$  

Using Equation 19, the attribution fraction for medium 1 is equal to:

**Attribution to medium 1**

$$= \frac{1}{n_p} \sum_{(i,t) \mid y_{it}=1} \beta_1 AS_{1it} + \beta_2 AS_{1it} AS_{2it}$$

Apart from the difference in functional form of Equations 24 and 25, one key difference is that the FBPM allocation in

---

5 Naik and Raman (2003) derive optimal allocation decisions with an unknown budget across an infinite horizon using an aggregate sales model with multimedia interactions, using a common carryover parameter across media. We cannot use their results directly because a known and fixed budget, a finite horizon, individual-level data and we want to allow for media-specific carryover parameters.

6 Another observation is that given that the total budget is fixed, the decision problem of finding the optimal advertising share for a medium is equivalent to one where the manager needs to find the optimal advertising level for this medium.
Equation 24 does not depend on ad stock levels, whereas attribution in Equation 25 does. If, for example, exposure to medium 1 increases, the ad stock level $AS_{1it}$ for medium 1 also increases, which increases the attribution to medium 1 (using Equation 25) but does not affect the FBPM allocation in Equation 24. Another key difference is that FBPM allocation in Equation 24 depends on media costs ($c_1$, $c_2$), but attribution in Equation 25 does not.

**Generalizing to other functional forms.** Table 1 showcases a variety of well-known functional forms for advertising response models, including the linear model with and without interactions, a model with square roots and the multiplicative model. For each model we offer expressions for both attribution and FBPM allocation (the derivations are in Web Appendix C). The key takeaway from Table 1 is that for each model these two metrics diverge. Whereas attribution depends on exposure (captured by ad stock) but not on cost, FBPM allocation depends on cost but not on exposure. Thus, more exposure leads to higher attribution, but does not affect the FBPM allocation, whereas higher cost for a medium does not affect attribution but does affect FBPM allocation.

**Relationship between attribution and advertising elasticity.** A traditional measure of advertising response in the literature is advertising elasticity (Sethuraman, Tellis, and Briesch 2011). An elasticity is defined as the percentage increase in an outcome variable when an input variable (ad stock for medium $m$ in this case) increases by 1%:

$$\eta_{mit} = \frac{\partial y_{it}}{\partial AS_{mit}} \frac{AS_{mit}}{y_{it}}.$$  \hspace{1cm} (26)

For the probit model in Equation 13, this elasticity can be calculated as:

$$\eta_{mit} = \frac{\partial \Pr(y_{it} = 1)}{\partial AS_{mit}} \frac{AS_{mit}}{\Pr(y_{it} = 1)} = \frac{\phi(\beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it} + \beta_3 AS_{1it} AS_{2it}) \times ME^R_{mit} \times AS_{mit}}{\Pr(y_{it} = 1)},$$  \hspace{1cm} (27)

where $ME^R_{mit} = \beta_m + \beta_3 AS_{jit,i \neq m}$. By substituting Equation 27 into Equation 16 we can formally link the marginal probability $\Delta_{mit}$ to the elasticity $\eta_{mit}$ for the probit model:

$$\Delta_{mit} = \phi(.) \times ME^R_{mit} \times AS_{mit} = \eta_{mit} \times \Pr(y_{it} = 1).$$  \hspace{1cm} (28)

Equation 28 holds at the individual level of observation for consumer $i$ and time $t$. It says that the contribution $\Delta_{mit}$ of medium $m$ to a purchase can be approximated by the product of the elasticity $\eta_{mit}$ and the probability of a purchase for consumer $i$ at time $t$. Thus, in this particular instance, there is a close link between elasticity and attribution, but this is not necessarily the case for other functional forms. This is shown in Web Appendix D, which compares the elasticity and the attribution contribution $\Delta_{mit}$ for the same models as in Table 1.

**Results from a Simulation Study**

We now discuss a simulation study with the objective being to illustrate how different FBPM allocation is from attribution. We simulate purchase incidence using a probit model that incorporates ad stock and interaction between two media:

$$y_{it}^* = \alpha_i + \beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it} + \beta_3 AS_{1it} AS_{2it} + \epsilon_{it},$$  \hspace{1cm} (29)

where the observed outcome is $y_{it} = 1$ if $y_{it}^* > 0$ and $y_{it} = 0$ otherwise. Compared with the earlier stylized examples, this model is a bit more complete by including a random intercept $\alpha_i \sim N(0, \sigma^2_{\alpha})$ to capture individual-level heterogeneity in purchase propensity.

We set up the media exposure and response parameters so that the first medium is (relatively) high exposure frequency, low cost, low effectiveness and the second medium is low frequency, high cost, high effectiveness. For illustration purposes, we label the first medium “online display ads” and the second “social media ads.”

We simulate daily media exposures and purchases for 2,000 consumers across three months (90 days). We simulate ad exposures using a multinomial distribution such that each day delivers at most one ad exposure to either medium or no exposure. We set the daily multinomial probability of exposure to media 1, 2, and neither, as, respectively, .8, .1, and .1. As a result, display ads are more frequent than social media ads, in a ratio of .8:1. Full details are given in Web Appendix E.

Table 2 gives the parameters used for simulating purchases, ensuring online display ads are less effective than social media ads ($\beta_1 = .35 < \beta_2 = .7$). Carryover effects range from no carryover ($\lambda = 0$) to a high level of carryover ($\lambda = .8$). The interaction effect ($\beta_3 = .25$) is set to be sufficiently large to generate some synergy between the media but is smaller than the main effects. Conversion rates vary between 4% and 5%, similar to what we observe in the empirical data discussed in the next section.

We obtain the true attribution weights based on the probit model using

$$\text{Attribution}_{mit}^{probit} = \frac{1}{\eta_{mit}} \sum_{j=1}^{\Lambda_{mit}} \Delta_{jmit},$$  \hspace{1cm} (30)

where $\Delta_{jmit} = \phi((\beta_0 + \beta_1 AS_{1it} + \beta_2 AS_{2it} + \beta_3 AS_{1it} AS_{2it})/\sqrt{1 + \sigma^2_{\epsilon}}) - \phi((\beta_0 + \beta_1 \times 0 + \beta_2 AS_{2it} + \beta_3 \times 0 \times AS_{2it})/\sqrt{1 + \sigma^2_{\epsilon}})$, and similarly for the second medium.

**Estimated and Approximate Attribution Weights**

Equation (30) assumes the parameters are known, but in a practical situation, these parameters have to be estimated. Table 3 gives the estimated parameters resulting from data
Table 1. Expressions for Attribution and FBPM allocation for Various Models Linking Ad Stock to a Dependent Variable.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Functional Form</th>
<th>Attribution for Medium 1: $\Delta_{1e} / (\Delta_{1e} + \Delta_{2e})$</th>
<th>FBPM Allocation to Medium 1 of Budget B (Cost Per Medium: $c_1, c_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Linear regression</td>
<td>$y_e = \beta_0 + \beta_1 AS_{1e} + \beta_2 AS_{2e} + \epsilon_e$</td>
<td>$\frac{\beta_1 AS_{1e}}{\beta_1 AS_{1e} + \beta_2 AS_{2e}}$</td>
<td>$A_1^* = \frac{(B/NT)}{c_1}$ if $\frac{\beta_1}{c_1} &gt; \frac{\beta_2}{c_2}$, else $A_1^* = 0$</td>
</tr>
<tr>
<td>2: Linear regression with square roots</td>
<td>$y_e = \beta_0 + \beta_1 \sqrt{AS_{1e}} + \beta_2 \sqrt{AS_{2e}} + \epsilon_e$</td>
<td>$\frac{\beta_1 \sqrt{AS_{1e}}}{\beta_1 \sqrt{AS_{1e}} + \beta_2 \sqrt{AS_{2e}}}$</td>
<td>$A_1^* = \frac{c_2 \beta_2^2 (B/NT)}{c_1 \beta_1^2 + c_2 \beta_2^2}$</td>
</tr>
<tr>
<td>3: Linear-log regression</td>
<td>$y_e = \beta_0 + \beta_1 \ln AS_{1e} + \beta_2 \ln AS_{2e} + \epsilon_e^2$</td>
<td>$\frac{\beta_1 \ln(AS_{1e} + 1)}{\beta_1 \ln(AS_{1e} + 1) + \beta_2 \ln(AS_{2e} + 1)}$</td>
<td>$A_1^* = \frac{\beta_2 (B/NT)}{c_1 (\beta_1 + \beta_2)}$</td>
</tr>
<tr>
<td>4: Multiplicative regression</td>
<td>$y_e = \beta_0 AS_{1e}^\beta_1 AS_{2e}^\beta_2 e^{\epsilon_e^2}$</td>
<td>$\frac{\beta_0 AS_{1e}^\beta_1 AS_{2e}^\beta_2 - \beta_0 AS_{1e}^\beta_1}{2 \beta_0 AS_{1e}^\beta_2 AS_{2e}^\beta_1 - \beta_0 AS_{1e}^\beta_1 - \beta_0 AS_{2e}^\beta_2}$</td>
<td>$A_1^* = \frac{\beta_2 (B/NT)}{c_1 (\beta_1 + \beta_2)}$</td>
</tr>
<tr>
<td>5: Linear regression with interaction</td>
<td>$y_e = \beta_0 + \beta_1 AS_{1e} + \beta_2 AS_{2e} + \beta_3 AS_{1e} AS_{2e} + \epsilon_e$</td>
<td>$\frac{(\beta_1 + \beta_2 AS_{2e}) AS_{1e}}{\beta_1 AS_{1e} + \beta_2 AS_{2e} + 2 \beta_3 AS_{1e} AS_{2e}}$</td>
<td>$A_1^* = \frac{(B/NT)}{c_1}$ if $\frac{\beta_1}{c_1} &gt; \frac{\beta_2}{c_2}$, else $A_1^* = 0$</td>
</tr>
<tr>
<td>6: Probit model</td>
<td>$Pr(y_e = 1) = \Phi(\beta_0 + \beta_1 AS_{1e} + \beta_2 AS_{2e})$</td>
<td>$\frac{\Phi(\beta_0 + \beta_1 AS_{1e} + \beta_2 AS_{2e}) - \Phi(\beta_0 + \beta_2 AS_{2e})}{2\Phi(\beta_0 + \beta_1 AS_{1e} + \beta_2 AS_{2e}) - \Phi(\beta_0 + \beta_1 AS_{1e}) - \Phi(\beta_0 + \beta_2 AS_{2e})}$</td>
<td>$A_1^* = \frac{(B/NT)}{c_1}$ if $\frac{\beta_1}{c_1} &gt; \frac{\beta_2}{c_2}$, else $A_1^* = 0$</td>
</tr>
<tr>
<td>7: Probit model with interaction</td>
<td>$Pr(y_e = 1) = \Phi(\beta_0 + \beta_1 AS_{1e} + \beta_2 AS_{2e} + \beta_3 AS_{1e} AS_{2e})$</td>
<td>$\frac{\Phi(\beta_0 + \beta_1 AS_{1e} + \beta_2 AS_{2e} + \beta_3 AS_{1e} AS_{2e}) - \Phi(\beta_0 + \beta_2 AS_{2e})}{2\Phi(\beta_0 + \beta_1 AS_{1e} + \beta_2 AS_{2e} + \beta_3 AS_{1e} AS_{2e}) - \Phi(\beta_0 + \beta_1 AS_{1e}) - \Phi(\beta_0 + \beta_2 AS_{2e})}$</td>
<td>$A_1^* = \frac{(B/NT)}{c_1}$ if $\frac{\beta_1}{c_1} &gt; \frac{\beta_2}{c_2}$, else $A_1^* = 0$</td>
</tr>
</tbody>
</table>

*As is common in log-log and multiplicative models, we need to add 1 to AS to ensure that the model can be used when AS = 0.
Table 2. Parameter Values for Purchase Simulations.

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Parameter Notation</th>
<th>No Carryover</th>
<th>Low Carryover</th>
<th>High Carryover</th>
<th>High Carryover in Just the 2nd Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No Interaction</td>
<td>Interaction</td>
<td>Interaction</td>
<td>Interaction</td>
</tr>
<tr>
<td>Intercept</td>
<td>$\beta_0$</td>
<td>$-2.3$</td>
<td>$-2.3$</td>
<td>$-2.3$</td>
<td>$-2.3$</td>
</tr>
<tr>
<td>Current exposure</td>
<td>$\beta_1$ (display)</td>
<td>$.35$</td>
<td>$.35$</td>
<td>$.35$</td>
<td>$.35$</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$ (social)</td>
<td>$.70$</td>
<td>$.70$</td>
<td>$.70$</td>
<td>$.70$</td>
</tr>
<tr>
<td>Advertising carryover</td>
<td>$\lambda_1$ (display)</td>
<td>$0$</td>
<td>$.1$</td>
<td>$.2$</td>
<td>$0$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_2$ (social)</td>
<td>$0$</td>
<td>$.4$</td>
<td>$.8$</td>
<td>$.8$</td>
</tr>
<tr>
<td>Interaction effect</td>
<td>$\beta_3$ (display × social)</td>
<td>$0$</td>
<td>$.25$</td>
<td>$.25$</td>
<td>$.25$</td>
</tr>
<tr>
<td>Variance of random</td>
<td>$\sigma^2$</td>
<td>$1/3$</td>
<td>$1/3$</td>
<td>$1/3$</td>
<td>$1/3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. observations</td>
<td></td>
<td>$180,000$</td>
<td>$180,000$</td>
<td>$180,000$</td>
<td>$180,000$</td>
</tr>
<tr>
<td>No. Purchases</td>
<td></td>
<td>$7,781$</td>
<td>$8,182$</td>
<td>$8,240$</td>
<td>$8,276$</td>
</tr>
<tr>
<td>Conversion Rate</td>
<td></td>
<td>$4.32%$</td>
<td>$4.55%$</td>
<td>$4.58%$</td>
<td>$4.60%$</td>
</tr>
</tbody>
</table>

Table 3. Probit Model Parameter Estimates for Simulated Data.

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>True Parameter Values</th>
<th>No Carryover</th>
<th>Low Carryover</th>
<th>High Carryover</th>
<th>High Carryover in Just the 2nd Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No Interaction</td>
<td>Interaction</td>
<td>Interaction</td>
<td>Interaction</td>
</tr>
<tr>
<td>Intercept</td>
<td>$\beta_0 = -2.3$</td>
<td>$-2.304 (.027)^a$</td>
<td>$-2.305 (.029)$</td>
<td>$-2.301 (.032)$</td>
<td>$-2.326 (.030)$</td>
</tr>
<tr>
<td>Current exposure</td>
<td>$\beta_1 = .35$</td>
<td>$.346 (.024)$</td>
<td>$.342 (.027)$</td>
<td>$.342 (.031)$</td>
<td>$.364 (.027)$</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = .70$</td>
<td>$.696 (.028)$</td>
<td>$.682 (.047)$</td>
<td>$.664 (.112)$</td>
<td>$.721 (.104)$</td>
</tr>
<tr>
<td>Interaction effects</td>
<td>$\beta_3 = .25$</td>
<td>$-$</td>
<td>$.287 (.097)$</td>
<td>$.273 (.144)$</td>
<td>$.225 (.121)$</td>
</tr>
<tr>
<td>Variance of random</td>
<td>$\sigma^2 = 1/3$</td>
<td>$.317 (.015)$</td>
<td>$.323 (.015)$</td>
<td>$.328 (.015)$</td>
<td>$.329 (.015)$</td>
</tr>
</tbody>
</table>

$^a$Standard errors are given in parentheses.

generated by Equation (29) using the true parameters given in Table 2. All of the estimates are close to their corresponding true parameters. Estimated values of $\Delta_{mit}$ are substituted into Equation (30) to obtain estimated probit attribution weights, denoted by $Attribution_{\text{Probit}}$. A third variant of Equation (30) is to replace the true value of $\Delta_{mit}$ with its Taylor series approximation given in Equation (15), denoted by $Attribution_{\text{Probit-Approx}}$.  

**Comparison of Attribution Measures**

Table 4 gives the true probit model attribution weights as well as estimated and approximate attribution weights defined above. It also gives the probit model attribution weights when we permit the advertising response parameters to be heterogeneous (see Web Appendix F for details) as well as the LTA weights. The final column in Table 4 gives the FBPM allocation percentages using Equations (22) and (23).

Table 4 shows that for the no-carryover situation, all five attribution methods (true, estimated, approximate and heterogeneous probit models, and LTA) agree precisely. In this case online display ads gets 82.1% of the attribution weight and social media ads 17.9%. We note that the social media ad weight (17.9%) is higher than the incidence of social media ads among advertising exposures (11%). This reflects the fact that although social media ads are less frequent than online display ads, social media is a more effective advertising medium, lifting its attribution weight.

The next three panels of Table 4 show that for increasing advertising carryover levels, the attribution weight for online display ads diminishes (from 82.1% to 74.0%), with social media attribution weight increasing correspondingly (from 17.9% to 26.0%). This change in attribution weight is captured very accurately by the estimated and heterogeneous probit models, while the approximate probit model attribution captures these changes reasonably (but not perfectly). However, as carryover increases, the attribution weights for the LTA method trend in the opposite direction to probit model attribution, which highlights that the LTA method ignores carryover and interaction effects, with LTA strongly favoring the most frequently occurring medium. Clearly, the LTA method gives very biased attribution weights when there are carryover effects and interaction effects, whereas all the probit-model based attribution methods stay very close to the true attribution weights.

The final column of Table 4 shows the FBPM allocation percent weights. Importantly, FBPM allocation weights are completely different from the attribution weights. Whereas attribution favors display ads (with a 74.0–82.1% weight in the true probit attribution), FBPM allocation puts much less weight
on display ads (0–18.5%) and instead emphasizes social media ads (81.5–100%) due to its higher response parameter. This reinforces once more that FBPM allocation and attribution are very different metrics with very different purposes.

Comparison of Purchase Probabilities

A key question is whether it makes much difference to the purchase outcome when using attribution weights rather than FBPM allocation weights. For a meaningful comparison, we derive the number of ads, not the proportion. 8

We now use the probit model in Equation (29) to calculate the expected purchase probability under current, LTA, probit attribution and FBPM allocation. For illustration purposes, we use the “Low-Carryover + Interaction” case, with similar insights for higher levels of carryover. Table 5 shows that current purchase probability is 4.63%, which serves as the baseline for comparison. This is achieved with a 89% weight for medium 1. As we saw in Table 4, both the LTA and probit attribution methods still favor the first medium (between 80.4 and 87.0% weight) but nevertheless result in higher estimated purchase probabilities of 4.78% and 4.92%, respectively. In contrast, FBPM allocation (with no heterogeneity in ad response parameters) reduces medium 1 (19% weight) in favor of medium 2 (81% weight), and this appears to be the better option, as this results in a 17% increase in purchase probability from the baseline, attaining 5.42%. Permitting the ad response parameters to be heterogeneous increases the purchase probability further, to 5.55%.

Table 4. Comparison of Attribution Methods and FBPM allocation Using Individual-Level Simulated Advertising Exposure Data.

<table>
<thead>
<tr>
<th>Media</th>
<th>True Probit Model Attribution Probit</th>
<th>Estimated Probit Model Attribution Probit</th>
<th>Approximate Probit Model Attribution Probit</th>
<th>True Probit Model Attribution Probit with Heterogeneity in Response Parameters</th>
<th>Last Touch Point Attribution</th>
<th>Profit-maximizing Allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Carryover + No interaction</td>
<td>Display 82.1%</td>
<td>82.1%</td>
<td>82.1%</td>
<td>82.1%</td>
<td>82.1%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Social</td>
<td>17.9%</td>
<td>17.9%</td>
<td>17.9%</td>
<td>17.9%</td>
<td>17.9%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Low Carryover + Interaction</td>
<td>Display 80.4%</td>
<td>80.4%</td>
<td>81.3%</td>
<td>80.7%</td>
<td>87.0%</td>
<td>18.5%</td>
</tr>
<tr>
<td>Social</td>
<td>19.6%</td>
<td>19.6%</td>
<td>18.7%</td>
<td>19.3%</td>
<td>13.0%</td>
<td>81.5%</td>
</tr>
<tr>
<td>High Carryover + Interaction</td>
<td>Display 77.7%</td>
<td>77.8%</td>
<td>80.4%</td>
<td>78.0%</td>
<td>91.2%</td>
<td>18.5%</td>
</tr>
<tr>
<td>Social</td>
<td>22.3%</td>
<td>22.2%</td>
<td>19.6%</td>
<td>22.0%</td>
<td>8.8%</td>
<td>81.5%</td>
</tr>
<tr>
<td>High Carryover in just the 2nd medium + Interaction</td>
<td>Display 74.0%</td>
<td>74.2%</td>
<td>77.3%</td>
<td>74.0%</td>
<td>91.9%</td>
<td>18.5%</td>
</tr>
<tr>
<td>Social</td>
<td>26.0%</td>
<td>25.8%</td>
<td>22.7%</td>
<td>26.0%</td>
<td>8.1%</td>
<td>81.5%</td>
</tr>
</tbody>
</table>

Table 5. Comparison of Purchase Probabilities for Alternative Allocation Methods.

<table>
<thead>
<tr>
<th>Media</th>
<th>Current Number of Exposures</th>
<th>LTA</th>
<th>Probit Model Attribution Probit with Heterogeneity in Response Parameters</th>
<th>Probit Model Attribution Probit with Heterogeneity in Response Parameters</th>
<th>FBPM allocation. No Heterogeneity in Response Parameters</th>
<th>FBPM allocation. With Heterogeneity in Response Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display exposures</td>
<td>144,785 (89%)</td>
<td>138,805 (87%)</td>
<td>124,158 (80%)</td>
<td>125,353 (80%)</td>
<td>22,530 (19%)</td>
<td>23,016 (19%)</td>
</tr>
<tr>
<td>Social exposures</td>
<td>17,517 (11%)</td>
<td>21,504 (13%)</td>
<td>31,269 (20%)</td>
<td>30,472 (20%)</td>
<td>99,020 (81%)</td>
<td>98,696 (81%)</td>
</tr>
<tr>
<td>Total exposures</td>
<td>162,302 (100%)</td>
<td>160,309 (100%)</td>
<td>155,427 (100%)</td>
<td>155,825 (100%)</td>
<td>121,550 (100%)</td>
<td>121,712 (100%)</td>
</tr>
<tr>
<td>Cost</td>
<td>$17,106</td>
<td>$17,106</td>
<td>$17,106</td>
<td>$17,106</td>
<td>$17,106</td>
<td>$17,106</td>
</tr>
<tr>
<td>Purchase probability</td>
<td>4.63%</td>
<td>4.78%</td>
<td>4.92%</td>
<td>4.94%</td>
<td>5.42%</td>
<td>5.55%</td>
</tr>
<tr>
<td>Percent change</td>
<td>–</td>
<td>3%</td>
<td>6%</td>
<td>7%</td>
<td>17%</td>
<td>20%</td>
</tr>
</tbody>
</table>

---

8 It is easily shown that $A_{m}^{LTA}$, the number of ads in medium m under LTA for the budget constraint in Equation 21, can be computed as $A_{m}^{LTA} = B \times Attribution_{m}^{LTA} / \sum_{j=1}^{M} c_{j} Attribution_{j}^{LTA}$, where $Attribution_{m}^{LTA}$ is given in Equation 1. Similarly, $A_{m}^{Probit}$, the number of ads in medium m under the probit model definition of attribution can be calculated as $A_{m}^{Probit} = B \times Attribution_{m}^{Probit} / \sum_{j=1}^{M} c_{j} Attribution_{j}^{Probit}$, where $Attribution_{m}^{Probit}$ is given by Equation 30. Note that when using these expressions, we need to divide B by NT, analogous to Equations 22 and 23, to get the number of exposures per person per day.
This simulation example underscores the shortcomings of attribution, no matter how it is measured. Any reasonable measure of attribution favors the volume of advertising too much and understates the effectiveness of a medium at generating an advertising response. In contrast, FBPM allocation is not unduly influenced by the volume of advertising and allocates based on effectiveness (including heterogeneity in effectiveness), interaction, carryover and cost. As the results of Table 5 demonstrate, FBPM allocation can produce substantially higher purchase outcomes than attribution allocation.

Incorporating Endogeneity into Simulated Data

We now study how endogeneity affects the results. An obvious manifestation of endogeneity is where a firm uses information from past purchases to adjust the number of advertising messages delivered to a customer. For example, a person who purchased from a brand in the previous month might receive more emails in the current month or be served more banner ads. Indeed, the marketing manager of the firm that supplied the data used in our subsequent empirical application is on record as saying that their primary basis for targeting customers with advertising is whether a customer has purchased recently or whether they have opened any emails delivered to them. In line with this, we simulate varying levels of targeting, where more ads are targeted towards those who purchased recently or had received more ads in the first medium in the previous 30 days. Web Appendix F gives the details.

Table 6 gives the parameter estimates for the probit model with random effects in Equation 29 fit to the simulated data with targeted ads to see how well it can recover the original parameters in the presence of endogeneity. As expected, with no endogeneity present the estimated parameters for the probit model with random effects are very close to the true values and these values are contained within the 95% confidence interval for their estimates. Even for increasing levels of endogeneity the estimated advertising parameter values for the probit model continue to be close to the true values, with each of the 95% confidence intervals still containing the true parameter value. It is worth noting that if we fit the probit model without random effects then bias is observed in the parameter estimates. Therefore, the inclusion of random effects helps make the probit model robust against endogeneity due to targeting of individuals based on their advertising and purchase history.

Table 7 contrasts the attribution weights for no, mild, and strong endogeneity. As can be seen, attribution weights are relatively stable as endogeneity increases. Attribution weights in the presence of endogeneity are almost all within one
percentage point of those calculated when no endogeneity is present. This indicates that the model-based measure of attribution is largely robust against endogeneity, which is not surprising given that the underlying parameter estimates are also robust to endogeneity.

**Empirical Application**

We now fit a probit model to real data and calculate then compare attribution and allocation weights. The data are for three different retail brands and comprise individual-level information on advertising exposure for up to six media together with purchase incidence.

The empirical data are from the Wharton Customer Analytics Initiative and pertain to a North American specialty retailer who wishes to remain anonymous. The retailer operates three brands, largely in the apparel sector. Although all three brands are owned by one firm, they are managed independently from three separate headquarters. Table 8 contrasts the brands (labeled B1, B2, and B3) with regard to purchase channel, purchase amount, and customer profile. It shows that the customers of all three brands are predominantly female, with B1 and B3 appealing more to younger customers and B2 to slightly older customers.

The data contain information on both online and in-store purchases. The online sessions data have the entire website visit history regardless of whether a purchase is made, which allows for a complete set of online touchpoint measurements. The in-store data are limited to store visits where a purchase occurred. To ensure fully observed covariate data, we include in-store purchases if they are made on the same day an online session occurred. The combined purchase incidence variable comprises either an online or an in-store purchase.

We use the year of 1 July 2011 to 30 June 2012, and restrict the data to include only those people that made at least one online purchase in this year. The panel sample sizes are reported in Table 8, and are reasonably large, ranging from 2696 for B3 up to 7663 for B1.

The data records a history of online activity from up to six media touchpoints prior to the eventual purchase (or non-purchase). One touchpoint medium is referral marketing, whereby a third-party website directs potential customers to the retailer’s websites. We also observe organic and paid search queries and website visits resulting from a social media touchpoint (primarily Facebook). The data also record the emails and physical catalogs sent to customers, which are respectively, the fifth and sixth touchpoint media.

**Model**

The latent utility (probit) model we fit to the empirical data, separately for each brand, mirrors that used previously for simulated data in Equation (29):
\begin{equation}
\begin{aligned}
y_{it}^* &= \alpha_i + \beta_0 + \sum_{m=1}^{6} \beta_m \text{AS}_{mit} + \sum_{m \in \{3, 4, 5\}} \beta_{m, 6} \text{AS}_{mit} \text{AS}_{6it} \\
&+ \sum_{l=2}^{12} \mu_l \mathbb{I}_{\text{Month}=l} + \pi \log(\text{PageViews}_{it}) + \epsilon_{it},
\end{aligned}
\end{equation}

for person \(i\) on day \(t\), and \(t\) ranges from 1 to \(T_i\), where \(T_i\) is the number of daily online sessions for an individual in the study period. Equation 31 includes monthly dummy variables (January is the baseline month) to capture seasonal changes in clothing and Christmas effects. By capturing the associated demand shocks, these monthly dummies also help to reduce possible endogeneity effects (see Web Appendix G). Another covariate is the log of the number of daily page views a person has at the retailer’s website when they have an online session on day \(t\). Page views are linked to online purchase incidence (Bucklin and Sismeiro 2003; Danaher, Mullarkey, and Essegaier 2006).

There are potentially 15 pairwise interactions for these 6 media but some initial model fitting indicated that the majority of these are statistically insignificant. We therefore markedly reduce these to just three pairwise interactions, comprising ad stock for email, paid search and social media interacted with catalog ad stock (medium 6), as shown in Equation 31.

**Model parameter estimates.** We fit the probit model with random effects in Equation 31 separately for each brand via maximum likelihood using the procedure Nlmixed in SAS. We estimated the six advertising carryover parameters by undertaking a grid search and selecting the model which gave the highest likelihood.

A potential concern for models that link advertising to sales is endogeneity due to targeting. As mentioned before, the parent firm’s primary basis for targeting decisions is past purchases and past emails that have been clicked or opened. As the simulations showed (and Web Appendix G shows in more detail), even in the presence of strong endogeneity, the proposed model and estimation method recovers the true parameters very well. Since model 31 includes an individual-level random effect, plus monthly dummy variables, it is also robust to the separate effects of cross-sectional and longitudinal endogeneity.

To conserve space, Table 9 only reports parameter estimates that are significant at the 5% level, while insignificant parameters are reported as 0. All media generally have significant advertising effects. The ad stock parameters show that lagged advertising effects are prevalent only for social media and catalogs. We also observe some significant interaction effects. Although carryover and interaction effects are small, they nevertheless show the benefit of accommodating such effects. The monthly dummy variables show that spring and December are periods with higher purchases compared with January. More page views result in a higher purchase likelihood, as in Bucklin and Sismeiro (2003) and Danaher, Mullarkey, and Essegaier (2006).

**Comparison of Attribution Allocation and FBPM allocation**

As discussed previously, a primary (albeit unjustified) purpose for media attribution is to help allocate advertising spend across media. This can be achieved by using either the LTA method or the proposed probit model for attribution. Although it is tempting to use attribution for allocation purposes, the simulations showed that a superior approach is to use a formal budget allocation approach. We now extend this to a real life setting in which there are six rather than two media. The optimization problem becomes:

\[ \text{Max } \Pi(A_1, A_2, \ldots, A_6) \]

\[ = \sum_i \sum_t [K \times \Pr(y_{it} = 1 | A_1, A_2, \ldots, A_6)] - B \quad (32) \]

subject to \[ \sum_i \sum_t \sum_{j=1}^{6} c_j A_{jit} = B. \quad (33) \]
As there is no closed-form solution to Equations 32 and 33 with six (rather than two) media, we use the Solver function in Excel to obtain the profit-maximizing solution. Equation (33) requires relative costs, which we obtain from Danaher and Dagger’s (2013) Table 2. For example, they report a relative cost for paid search at 3 units, and this gives 39 GRPs; thus, we use the relative cost of \( \frac{3}{39} = 0.077 \). The last six rows of Table 8 give the relative cost for each of the media in the empirical application.

When solving the optimization problem we encounter some practical challenges. Email and organic search are essentially costless, unless we can account for the marketing cost of driving organic search (which we cannot for this empirical application). This means that Equations 32 and 33 will give very large allocations to such media so long as they have nonzero effectiveness (which they do in this case). This distorts the FBPM allocation in favor of these media, so we remove them from the FBPM allocation exercise and keep them at their existing levels. This is also reasonable because organic search touchpoints are not within the control of the advertiser; rather they depend on user search queries and the algorithm used to order search items. Also, although email is nearly costless, increasing email volumes unabated is highly likely to annoy consumers. We therefore believe it is reasonable to assume that current email volumes are set around levels customers will tolerate.

Tables 10–12 compare the percentage weight across the media touchpoints for the current allocation, LTA, probit...
model attribution, and FBPM allocation. Although the total cost of these methods is the same, there are noticeable differences in the touchpoint distributions across the media. FBPM allocation suggests many more touchpoints in paid search compared with the current allocation and those suggested by LTA and probit model attribution. The difference is also noticeable for catalogs, where FBPM allocation removes all expenditure in this media channel, enabling more funds for less expensive media channels. This increases the total number of touchpoints and it increases the percentage of touchpoints for paid search.

The redistribution is also depicted for B1 in Figure 2.

Tables 10–12 also show the effect on purchase probability of reallocating touchpoints under LTA, probit model attribution and FBPM allocation. Table 10 and Figure 2 show that the current predicted purchase probability for B1 is 5.75%. Under LTA, probit model attribution and FBPM allocation the predicted purchases probabilities are, respectively, 5.84%, 6.80%, and 7.18%. While these are all higher than the current purchase probability, the percentage increase is greatest for FBPM allocation (from 5.75 to 7.18%, or +24.9%). This pattern is repeated for the other two brands in Tables 11 and 12. This shows that FBPM allocation consistently outperforms allocation based on attribution. Moreover, the differences are not minor; across the three brands, the average percent increase in purchase probability from the current probability for the LTA, probit model attribution and FBPM allocation are, respectively, 3.2%, 12.4%, and 25.3%. Therefore, FBPM allocation is still the best way to allocate expenditure across media, as opposed to using the LTA method or the proposed probit-based attribution method that compensates for the deficiencies of the LTA method.

<table>
<thead>
<tr>
<th>Media</th>
<th>Current number of touchpoints</th>
<th>Current Allocation</th>
<th>Last Touchpoint Attribution</th>
<th>Probit Attribution</th>
<th>Optimal allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email</td>
<td>185,366</td>
<td>85.10%</td>
<td>83.70%</td>
<td>66.50%</td>
<td>61.40%</td>
</tr>
<tr>
<td>Organic Search</td>
<td>18,437</td>
<td>8.50%</td>
<td>8.30%</td>
<td>6.60%</td>
<td>6.10%</td>
</tr>
<tr>
<td>Catalog</td>
<td>5,361</td>
<td>2.50%</td>
<td>2.30%</td>
<td>0.00%</td>
<td>0%</td>
</tr>
<tr>
<td>Paid Search</td>
<td>4,754</td>
<td>2.20%</td>
<td>4.00%</td>
<td>20.20%</td>
<td>31.50%</td>
</tr>
<tr>
<td>Referral</td>
<td>2,766</td>
<td>1.30%</td>
<td>1.40%</td>
<td>6.70%</td>
<td>1.00%</td>
</tr>
<tr>
<td>Social</td>
<td>1,115</td>
<td>0.50%</td>
<td>0.20%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Media</th>
<th>Current number of touchpoints</th>
<th>Current Allocation</th>
<th>Last Touchpoint Attribution</th>
<th>Probit Attribution</th>
<th>Optimal allocation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email</td>
<td>185,366</td>
<td>85.10%</td>
<td>83.70%</td>
<td>66.50%</td>
<td>61.40%</td>
</tr>
<tr>
<td>Organic Search</td>
<td>18,437</td>
<td>8.50%</td>
<td>8.30%</td>
<td>6.60%</td>
<td>6.10%</td>
</tr>
<tr>
<td>Catalog</td>
<td>5,361</td>
<td>2.50%</td>
<td>2.30%</td>
<td>0.00%</td>
<td>0%</td>
</tr>
<tr>
<td>Paid Search</td>
<td>4,754</td>
<td>2.20%</td>
<td>4.00%</td>
<td>20.20%</td>
<td>31.50%</td>
</tr>
<tr>
<td>Referral</td>
<td>2,766</td>
<td>1.30%</td>
<td>1.40%</td>
<td>6.70%</td>
<td>1.00%</td>
</tr>
<tr>
<td>Social</td>
<td>1,115</td>
<td>0.50%</td>
<td>0.20%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

| Total touchpoints | 217,799 | 217,799 | 221,482 | 278,951 | 301,982 | Total | 217,799 | 221,300 | 278,841 | 301,947 |
| Cost            | $7,897  | $7,897  | $7,897  | $7,897  | $7,897  |       | $7,897  | $7,897  | $7,897  | $7,897  |
| Purchase probability | – | 5.75% | 5.84% | 6.80% | 7.18% | Percent change | – | – | 1.60% | 18.30% | 24.90% |

**Figure 2.** Multimedia allocation of the same budget for retailer B1.
Conclusion

Technological advances allow firms to track the customer journey along touchpoints prior to a purchase. This has led to a surge of academic and practitioner interest in attribution metrics, which measure the contribution of each touchpoint to a purchase relative to all touchpoints. The purpose of this paper is to point out that there are several caveats about the attribution concept, and we offer resolutions to each of these caveats.

The most important caveat in attribution modeling is the temptation to use attribution percentages as the basis for future media allocation. Both the academic marketing literature and practitioners suggest that allocation is a major purpose of attribution – but we highlight some concerns with this suggestion. We argue that attribution is an inherently descriptive approach capturing how much of a role each medium has played in driving (past) purchases. However, this does not mean that attribution should be used for allocating a budget across media to maximize profit. For that, formal allocation rules are superior, as we demonstrate.

To illustrate the main point of “delusion due to attribution”: suppose 80% of all purchases are attributed to medium A and 20% to medium B, and they are equally costly. A naïve conclusion is to use these percentages and spend four times as much on medium A than on medium B. We argue that this is a highly misleading conclusion because attribution weights are strongly driven by the number of past exposures. Instead, allocation decisions should be made using formal profit optimization. While both attribution and FBPM allocation increase in a medium’s marginal effectiveness, attribution grows with increasing ad exposure levels whereas FBPM allocation does not. Conversely, attribution does not depend on the cost of using a medium, whereas FBPM allocation decreases in the cost of a medium.

Another caveat is that many attribution metrics used in practice do not have a link to the incremental impact of the medium on the purchase outcome. Even if there is a zero purchase probability effect of a medium, LTA or other attribution methods that are based on counting the number of exposures prior to a purchase will attribute a significant portion of the purchase to that medium. To overcome this issue, this paper offers a new-to-the-literature formal equation for attribution: it is the marginal increment in the purchase outcome variable in the presence versus absence of a medium, relative to the increments due to all media. By making the measure and properties explicit, we add further academic rigor to the attribution concept. The underlying probability model also includes carryover and interaction effects.

In practice, we recommend proceeding as follows. First, ideally create experimental variation in media exposures to obtain valid response parameters. Next, use these response parameters to calculate attribution metrics (for understanding the contribution of each medium to the observed purchases) and FBPM allocation (for improved profit outcomes). Next, implement the FBPM allocation in practice. As the response of consumers could change due to the changed exposure levels, it would be wise to use the resulting data to recalibrate the model.

Limitations and Future Research

In line with Lehmann, McAlister, and Staelin (2011), we deliberately make the examples and models only as complicated as required to explain the difference between attribution and FBPM allocation as clearly as possible. While the results are robust to various forms of heterogeneity and endogeneity, other researchers may want to allow for more intricate forms of endogeneity (such as slope endogeneity).

A significant part of the attribution literature has taken the path-to-purchase perspective, capturing the complexities of how a consumer goes through different touchpoints to finally make a purchase (see, e.g., Abhishek, Fader, and Hosanagar 2015). In this paper, we take a step back and look at the core question a firm tries to answer when doing attribution analysis, namely, given that a purchase occurred, what is the relative contribution of each type of touchpoint to this purchase? Our approach boils down to linking the purchase decision at time t to all prior touchpoints, since (after repeated substitution) our Equation 13 can be restated as

\[ y_{it} = \beta_0 + \beta_1(1-\lambda_1)\sum_{s=0}^{t-1}\gamma_1^sA_{11,t-s} + \beta_2(1-\lambda_2)\sum_{s=0}^{t-1}\gamma_2^sA_{21,t-s} + \beta_3(1-\lambda_3)\sum_{s=0}^{t-1}\gamma_3^sA_{31,t-s} + \beta_4(1-\lambda_4)\sum_{s=0}^{t-1}\gamma_4^sA_{32,t-s} \]

Importantly, the equation does not assume or require randomness of the exposures \( \{A_{m1,t-s}\}_{s=0}^{t-1} \). Suppose medium 1 leads a customer to medium 2, and this leads to a purchase. This means that there are time series dependencies between the exposures to the two media. These dependencies will show up as correlations between the variables on the right-hand side of Equation 34, but that is what regression (type) models can handle unless the correlations become extreme. In the empirical study, none of the correlations are above .057 in absolute value and so there is no issue from this point of view.

Path-to-purchase models do offer insights into how different touchpoints play a role at different stages, but that is not the goal of this paper. Since our goal is to create consistency between the underlying econometric model for attribution and for FBPM allocation, we use the same econometric model for both.

In the empirical application, online consumer touchpoints were readily observed but for offline media we only observed catalogs. Future research may include additional offline touchpoints (to the extent they are observed) as well.

In determining FBPM allocation, we have made assumptions about costs per exposure as we do not have access to the true cost data. Future research should look at ways of providing realistic cost estimates for digital communication. Further
research could also investigate the possibly nonlinear (inverted) relationship between firm communication (e.g., emails) and purchase likelihood.

To allow for a balanced comparison with attribution, we set up a fixed budget profit maximization problem yielding a time-invariant allocation to each medium. We refrain from optimizing the overall budget and from specifying how the allocated amount is spent over time (cf., Raman et al. 2012). In principle, we could expand the set of feasible solutions by optimizing the total budget as well as the allocation across media, time and customers simultaneously. This would lead to an optimum that is better than or equal to when the advertising variable is fixed across time and customers, because the feasible set is larger. In that sense, the optimal allocation results in this paper provide a lower bound (conservative estimate) on what could be achieved.

We hope that this paper will inspire follow-up research. The measurement of multimedia attribution is an area where there much opportunity for marketing science principles to make an impact on practice.

Acknowledgments
The authors also thank seminar participants at Tilburg University, the University of Hamburg, the University of Zurich, the 2016 Marketing Dynamics Conference, and the 2017 Marketing Science Conference for their helpful comments.

Associate Editor
P. K. Kannan served as associate editor for this article.

Author Contribution
Peter J. Danaher and Harald J. van Heerde contributed equally to this research.

Declaration of Conflicting Interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors thank an anonymous retailer and the Wharton Customer Analytics Initiative for providing the data used in this study, as well as the Australian Research Council (grant DPI140102142) for partial financial support.

References


