



Determining the Incremental Value of Marketing

In this whitepaper, OptiMine explores the tradeoffs of different measurement approaches, including more “modern” attribution methods, along with guidance on common pitfalls, risks and an overall comparison of the methods available to marketers today. The key take-aways are:

- ◆ Proper controlled tests are too expensive and time-consuming when thousands of digital ads are put into the mix
- ◆ Analysis paradigms that rely on tracking individual paths to purchase have severe practical limitations
- ◆ While statistical models may not be a marketer’s first choice for determining value, they have become the only viable option in today’s complex omni-channel environment



Determining the Incremental Value of Marketing

With the explosion of new ad channels, consumer devices and the rapid evolution of consumer behaviors, marketers find themselves back in uncharted waters struggling to understand the value of their advertising and marketing investments. Digital marketing was supposed to change all of this- with improved tracking, 1:1 marketing and advanced marketing capabilities that traditional marketers had only dreamed about. But the reality is that in today’s highly complex marketing ecosystem, understanding the value and contribution of each ad has actually gotten harder.

Prior to the advent of digital marketing, the incremental value of marketing or advertising was typically determined in one of two ways — controlled tests or econometric models. In a controlled test, there is a relatively straight-forward mechanism to determine who has “seen” and who has “not seen” a particular ad or campaign. Once a test has been conducted it is easy to compute the incremental value as the difference between the two groups. Econometric models have historically been used when a controlled test is not practical or possible. Instead of focusing on “who” has seen an ad, an econometric model focuses on whether or not a particular ad is likely to have caused an increase in value or activity. Of the two techniques, a controlled test puts the burden on the data collection and an econometric model puts the burden on the analytics.

Attribution modeling actually has two very different use cases, determining incremental ad value and targeting/personalization. Targeting/personalization use cases determine what prospect profiles should be targeted or what offers are likely to be appealing to a particular individual. While path analysis has been very successful for personalization and targeting (see Google, Facebook, and Amazon!), it has not been particularly successful for determining incremental ad value. The focus of this paper is on the incremental value use case, not the personalization use case.

Terminology Note

In this paper, the term “value” will be used to mean any event of interest to an advertiser - leads, orders, sales, conversions, engagements, accounts, revenue, profit - regardless of whether it occurs online or offline. Marketing and advertising will be used interchangeably to mean any ad, campaign, keyword, post, billboard, mail piece, or catalog regardless of whether it occurs in an online or offline context.

CONTROLLED TESTS

As digital marketing opportunities began to arise in the late 1990's, the apparent trackability of browsing behaviors made controlled testing the dominant choice for determining value. After all, if individual paths can be easily tracked dividing a population into treatment (aka "seen" or exposed) and control (aka "not-seen" or hold-out) groups should be fairly straight-forward. However, two major factors have largely prevented the success of what has become known as "attribution modeling" for determining the incremental value of digital marketing.

THOUSANDS OF ADS

Unlike traditional offline marketing where a handful of campaigns will be run over the course of a year, digital marketing typically involves thousands to millions of unique advertisements. A marketer's ability to run a proper controlled test where all other ads and conditions are held constant is practically non-existent. And even if the marketer allows for such "leakage", setting up and conducting these types of tests is extremely time consuming and can be a detriment to the business.

TRACKING ERRORS

The well-known limitations of cookies, pixels, and other forms of following an individual across multiple devices means that the treatment and control groups are corrupted. It is highly likely that the control group contains a significant number of individuals that have "seen" a given ad, just on a different device or in a non-trackable context.

THOUSANDS OF ADS

Digital advertising has exploded the number of individual ads that an advertiser may have running at any given time, and this challenge has created an incredibly complex minefield for the marketer attempting to run a proper controlled test. To get a sense of the practical complexity and the impractical reality of managing such a test, consider the following simple example:

- 1,000 ads, mixed among many different channels and ad networks
- Measured over a 30 day period
- On average an individual is exposed to 5 ads

In order to appropriately apply the controlled testing paradigm, two groups of individuals need to be identified where 4 of the 5 ads viewed are exactly the same and the only difference is whether or not they viewed the 5th ad. If the sequence of the views is considered, there are 10^{15} possible paths! That's far beyond the population of the entire planet. Throw out ad sequences and the possibilities only decrease to about 8×10^{12} . Now imagine you have 10,000 or 100,000 ads in the field (remember, each keyword in paid search is an ad). There is simply no way that treatment and control groups of any meaningful size can be identified, even if some paths are far more popular than others.

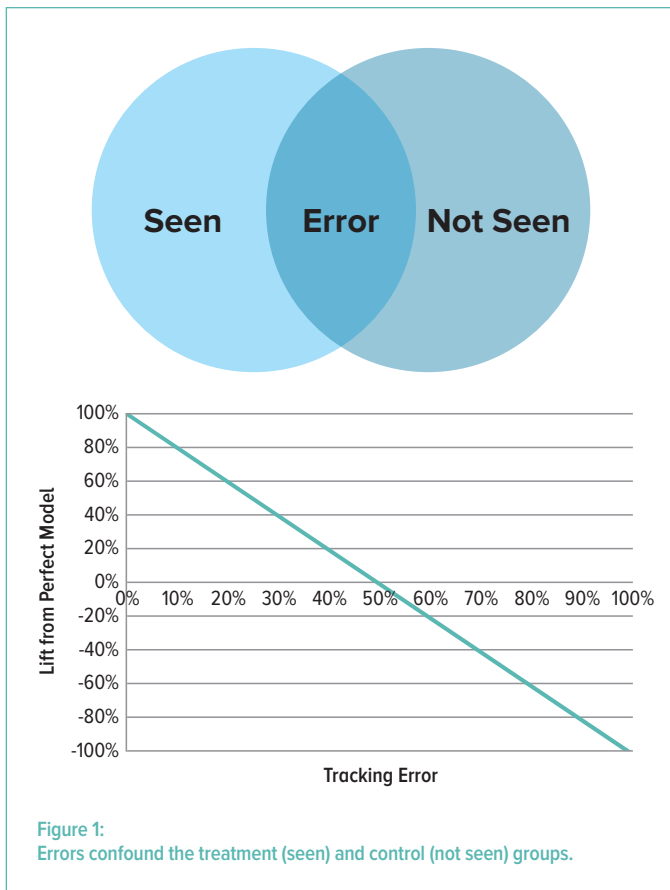
In practice what typically occurs with attempts to use path analysis to simulate a controlled test is two simplifications:

- Only consider clicks (engagements)
- Aggregate data to the marketing channel level

Consider the previous example. Instead of 1,000 ads there are probably less than 10 advertising channels (e.g. paid search, display, social). And by considering only clicks, the average path length over a 30 day period in most cases across a wide variety of industries drops to less than 2. This reduces the complexity down to fewer than 100 possible paths. As long as ad views, and therefore all of offline advertising or even upper funnel digital channels with low click rates (mobile, display, video) have no value and there is no desire to understand value below a channel level, this is a valid simplification. In fact, this is what many attribution vendors do in order to provide a valid model.

TRACKING ERRORS

The fact that technologies like cookies and pixels cannot track all of the people all of the time, especially across multiple devices, is well known [1,2,3]. What is less commonly appreciated is the effect this has on attempts to use path analysis and the controlled testing paradigm to determine incremental ad value. Only having a sample of the population trackable is perfectly fine for the personalization use case, but it is deadly for determining ad value. Take a look at the Venn diagram and associated graph below.



The problem with tracking errors is that the same individuals are now in both groups. Even worse, there are very likely to be cases where the ad clicks are in one group and the conversion value is in the other. A common example occurs when the individual views the ad on a mobile device and might even click-through to get more information. They are now part of the treatment group. However later that day the same

person logs onto their laptop computer, navigates directly to the website and completes a purchase. From a tracking perspective the ID associated with the laptop is a different individual. Now the ID that “saw” the ad did not convert and the ID that “did not see” the ad converted. Every time this happens the measured incremental lift of the ad is decreased two-fold. Once by the treatment ID not converting and a second time by the control ID converting. If 50% of a treatment group is also in the control, the measured incremental lift will be zero as shown in the figure, regardless of the true effectiveness of the ad. If the overlap exceeds 50% one can actually measure negative lift! E.g. conclude that the addition of an ad has reduced the propensity to purchase, when in fact it has increased it.

Coupled with the thousands of ads discussed in the previous section, it does not take a particularly large error rate to wipe out any measured incremental lift. The common approach to tackling the tracking error issue is to add in additional data integration efforts in an attempt to “stitch” IDs together across devices and even to offline behaviors by converting each ID to a unique name & address and then integrating with the order information. While these efforts definitely bring the tracking error rate down, it can be prohibitively expensive and it is impossible to completely eliminate the tracking errors. Finally, solutions from large publishers such as Facebook and Google only solve this issue within their own walled gardens and only represent a fraction of any marketer’s efforts. As such, they do not solve this problem for the marketer in any comprehensive way.

CLICK CENTRIC

The final point about using digital tracking to simulate the controlled testing paradigm for marketing measurement is that it is very “click-centric”, meaning it is heavily biased towards media that directly generate digital engagements. This bias exists because trackability is far higher when an individual clicks through to the advertiser’s website. Offline media does not get included in such approaches since there are no clicks. Offline value such as in-store and call centers are also at a major disadvantage even when an ID stitching add-on is included. Anecdotal evidence indicates that the match-rate for online to offline IDs is less than 10%. This means that attempts to measure digital ads’ influence on non-digital activities only captures a small set of activity- certainly not enough

to make informed marketing decisions. Even upper funnel digital media such as display banner ads are penalized since the click-through-rates are inherently lower than other digital media.

Impressions are a much better analog than clicks to the size of audience reached by a digital ad (aka media “reach”). The poorly understood issue with impressions is that many of the popular digital marketing platforms like Google and Facebook do not allow third-party tracking of impressions, only clicks. This means that the IDs of individuals who saw but did not click-through an ad are not available to the advertiser, only the publisher. Plus because the controlled testing paradigm depends on dividing individuals into treatment and control groups, the whole issue of ad “viewability” also becomes a problem when trying to consider impressions. What if the ad was displayed at the bottom of the screen below the fold and the user doesn’t scroll down? Does it count? As a result, while there is a lot of discussion about “view-through” conversions (IDs that viewed an ad and subsequently purchased through a different channel), impressions are rarely if ever used in a controlled testing context for determining ad value. View-through conversions do come up in the context of “fractional attribution”, which is discussed in the next section.

FRACTIONAL ATTRIBUTION

Fractional attribution has arisen as a new technique offered by many vendors to avoid the issues described above with cleanly identifying treatment and control groups. That said, fractional attribution still maintains the tracking or path analysis mindset, approach and drawbacks. The main idea behind fractional attribution is to ignore any issues with setting up a proper controlled test and simply assign value based on the presence of an ad in a path. In other words, completely ignore the concept of incremental lift. For example, if there is a click on a display ad prior to a click that leads to a purchase, then the display ad gets a “fraction” of the credit. The most common phrase applied to fractional attribution is “it isn’t perfect but it is better than nothing”. In most cases this turns out not to be true due to the lack of consideration of incrementality. Consider this simple example: An advertiser is making \$1M in topline revenue per week through a Website. The advertiser then

launches an aggressive new display campaign that manages to get the ad in front of 50% of the IDs that purchase from the site- all during the next week. However, in that next week, the revenue is still \$1M. According to the fractional attribution paradigm, the new display ad deserves some credit for the revenue simply because it is present in 50% of the paths to purchase, even though there is a 0% increase in actual revenue! Ultimately fractional attribution rewards putting ads in front of likely buyers, regardless of whether there is any actual incremental lift.

The notion of view-through conversions most often comes up in the context of fractional attribution as a way of countering the argument that digital tracking is very click-centric. The use of view-throughs in an analysis exacerbates the tracking error and incrementality issues discussed above. Since digital impressions are inherently less trackable than clicks due to limitations imposed by many publishers, the tracking error associated with paths including ad impressions will be much larger, pushing the measured lift down. However because incrementality is completely ignored in a fractional attribution context, the measured lift is at the same time pushed up. This leads to a situation where the view-through conversions are either a gross over-estimate or a gross under-estimate of ad value. In general, the higher percentage of mobile impressions involved the more likely the view-through numbers will under-estimate value due to tracking errors, and the higher percentage of desktop that is reached the more likely the view-throughs will over-estimate value due to lack of consideration of incrementality.

The final big risk with fractional attribution is the assignment of the fractional weights themselves. In most cases they are arbitrarily assigned. An example of this arbitrary weighting can be seen in the notorious “u-shape” assignment that splits the value evenly between the first and last click, assigning 0% to any clicks in-between. When arbitrary assignment of weights are added to a Lack of consideration of incrementality and issues with view-through conversions, the result is a value assignment paradigm that really doesn’t deserve to be called a statistical model. While it is definitely different than “nothing”, it is certainly not better than nothing if any meaningful measure of statistical accuracy or validity is applied.

ECONOMETRIC MODELS

The major issues discussed above with the controlled testing and fractional attribution paradigms naturally lead to the other well established method for determining incremental ad value - econometric modeling. Econometric modeling is a venerable technique backed by academic departments [4] and publications [5, 6], and has been used effectively for decades. The main concept behind econometric models for marketing is to follow the ads over time, not consumers. As a simple example, imagine a retailer that has one physical store, no ads, and makes \$100k/week of top-line revenue. Then the retailer puts a billboard up on the highway and starts making \$110k/week. Four weeks later the billboard comes down and the revenue drops back down to \$100k/week. In its simplest form, the econometric model would assign the billboard an incremental value of \$10k/week. The model has no notion of individuals or paths, just that the revenue went up along with the billboard and went back down when the billboard was removed.

Traditionally econometric models, often referred to as “media mix” or “marketing mix” models when used for advertising [7], have been associated with the following benefits:

- ◆ Seamlessly handle online and offline media and value
- ◆ Designed to consider additional non-media effects such as seasonality, economic conditions, competition, and pricing
- ◆ Automatically handle ongoing campaigns and a changing media mixes

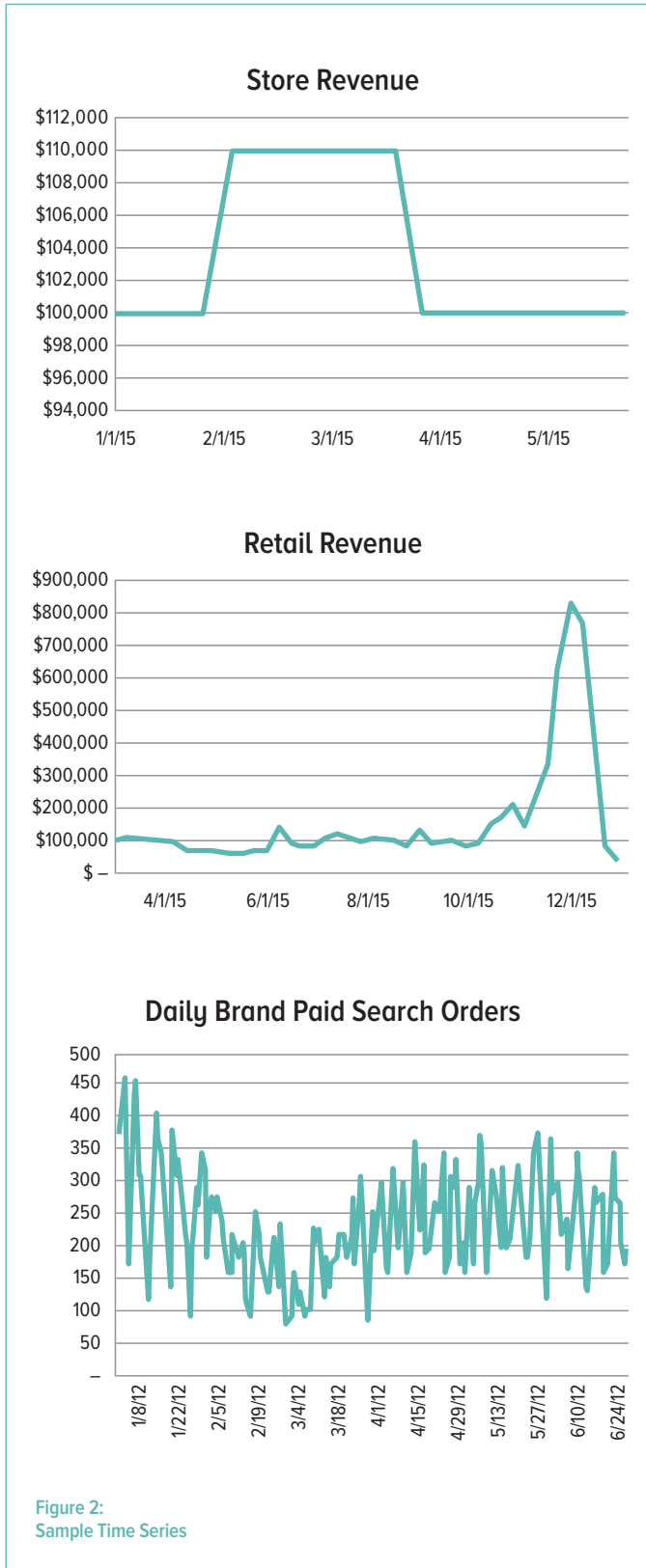
However there are also a few issues that have been associated with traditional econometric approaches:

- ◆ **High-Level** — because the traditional use case has been offline media, the econometric models typically provide measures at a channel-week level. While this is perfectly appropriate for traditional media such as TV, Radio and Print, the models are not nearly granular enough for digital media. This is a significant issue in that there is frequently a very large range of performance among the campaigns and ads within any given digital channel.

- ◆ **Slow** — econometric models are typically applied by consultants over the course of a multi-month engagement, and are time consuming to update- especially if the marketer wants to measure new or different KPIs. Further, because of the time to hand-craft such models, they are typically very expensive.
- **Black-box** — when compared to a controlled test, models always appear to be more complicated and harder to understand. This leads to questions about the validity of the results.

It turns out the the granularity and speed issues aren’t really inherent in the econometric paradigm. They are really just a function of how the technique has been traditionally applied. The black-box or validity concern is a legitimate one that exists for all modeling approaches, not just econometric modeling, but is addressable.

Before diving into the potential issues it helps to understand what the models are actually predicting. For determining incremental ad value, an econometric model predicts the value of a conversion point over units of time. The ads are inputs into the model (aka independent variables). The value is the model target (aka outcome or dependent variable). If the behavior of the ad is correlated with changes in value it is assigned a portion of that value by the model. In the previous example the conversion point is the retail store and the unit of time is a week. Conversion points can be broken down by location, product hierarchy, digital source, or a wide variety of other methods. As examples, physical stores, call centers, a particular paid search keyword, a natural search keyword, or a direct mail source code are all conversion points. The things they all have in common is their primary purpose is to facilitate a transaction and they can be associated with the “last touch” prior to purchase. In contrast, the primary purpose of watching a TV show is not to shop, and a brand-oriented TV ad cannot currently be the point of purchase. In general, value comes from an individual visiting a store, a website, an authorized agent, or calling a call center. The value associated with any of these conversion points can be tracked over time. The figures below show the revenue by week for the retail store example, the revenue per week for an actual store, and the orders per day for an actual paid search keyword.



The purpose of the econometric model is to determine the most likely causes of the changes in value from one time period to the next. In this simple example there is only one potential cause (the billboard) and it correlates perfectly with the change in value. For real world scenarios, there are thousands of potential causes of the changes, and not all of them are advertising. Some of the changes might even be random noise not associated with any particular cause, while others may be due to economic conditions, seasonality, day of week, the weather, physical location- all having nothing to do with advertising. Econometric models, when properly built and executed, account for these factors and influences alongside measuring the incremental lift generated by marketing.

GRANULARITY

With respect to granularity, the ultimate goal is to estimate how each individual ad contributes to value, not just an entire marketing channel. When considering one out of thousands of ads and it’s influence on overall value it is highly unlikely that traditional media mix models will be able to detect a pattern. This is one of many reasons why traditional media mix has rolled media up to the channel level for modeling. In reality however, when the value is broken down into individual and more granular conversion points, detectable patterns start to emerge. For example if the ad of interest is a paid social post about motorcycle policies for a consumer insurance company, measuring the marketing influence of this ad across all home, auto, life, and motorcycle policies is unlikely to succeed. If just the motorcycle policies are considered, the probability of measurement success is much higher.

To achieve this level of granular measurement, many hundreds or thousands of models need to be built. Training hundreds or thousands of granular econometric models with thousands of potential inputs is clearly much more difficult than training a handful of high-level models with a few dozen inputs, but this is precisely the type of task that modern “big data” infrastructures and machine learning algorithms are designed to handle. There is nothing that prevents thousands of econometric models being trained and deployed at a very granular level using modern techniques.

SPEED

The slow speed issue associated with econometric modeling is very much tied to the use of custom services engagements and manual efforts that are common for strategic media planning projects. If the goal of the project is a full strategic media plan for the next four to eight quarters, the speed of the analysis is not on the critical path. There is no compelling reason to switch over to a modern big data machine learning infrastructure if the results aren't needed for 90 days. However just like the granularity issue, once the proper software infrastructure is in place, speed becomes a non-issue. With proper software, once data feeds are setup and configured, the time it takes to train and score thousands of models is only limited by the available computing resources. Such automation is the key to solving the granularity challenge and to staying current with marketing questions and market conditions that evolve over time.

VALIDITY

There is no question that given the choice between a proper controlled test and a statistical model, most people prefer the controlled test. Ironically, the dawn of the Internet era, once hailed as the most trackable and testable medium ever, has made controlled testing for media value virtually impossible. To be clear, there are many opportunities for controlled tests in areas like website design or click-through-rates for different offers or content designs. For ad value, the question has become not “should a model be used”, but “what modeling technique should be used”. When it comes to judging the validity of a particular set of models, there are four common techniques.

- ◆ **Explainability** — Do the modeled relationships “make sense” based on domain knowledge.
- ◆ **Backtest** — for an event in the past with a known root cause, do the models correctly identify the cause.
- ◆ **Hold-out Accuracy** — given media spend over a new time-period not used for training, how closely do the predicted values match the actual values.
- ◆ **In-Market Test** — Make a change based on the results and validate that the change in value is as expected.

Each of the validation techniques have their pros and cons as shown in the table below. In general, best practice is to use explainability coupled with one other technique.

Validation Technique	Advantage	Disadvantage
Explainability	<ul style="list-style-type: none"> ◆ Simple 	<ul style="list-style-type: none"> ◆ Qualitative, not quantitative ◆ Domain knowledge is not always correct
Backtest	<ul style="list-style-type: none"> ◆ Uses data already gathered and modeled ◆ Non-media events can be used as well as media events 	<ul style="list-style-type: none"> ◆ Data may not contain an appropriate event with known causes
Hold-Out Accuracy	<ul style="list-style-type: none"> ◆ Quantitative assessment of predictive accuracy 	<ul style="list-style-type: none"> ◆ Requires gathering additional data ◆ Predictive accuracy does not necessarily equate to business value
In-Market Test	<ul style="list-style-type: none"> ◆ Gold standard for validity. Did it actually work in the field? 	<ul style="list-style-type: none"> ◆ Test has to be big enough to measure ◆ Adds significant time ◆ No single test is guaranteed to be successful

HYBRID APPROACH

More recently several vendors have advocated a hybrid approach where digital tracking is used for the granular online to online influences and an econometric approach is used to add in high-level effects from offline media and offline value. Some vendors have even gone so far to acquire other technology vendors to put together this hybrid offering. This so-called “Top-Down/ Bottom-Up”

hybrid approach is attractive in theory because it claims to provide the best of both worlds.

The reality is merging in digital tracking with econometric modeling does absolutely nothing to address the underlying issues discussed above with controlled testing or fractional attribution. All of the problems with tracking errors and incremental lift still exist. In fact any attempt to merge the paradigms tends to make the problems worse. If an econometric model uses data with unknown and potentially large error rates like view-through conversions as inputs, the models become invalid. The golden modeling rule of “garbage in - garbage out” still applies.

In addition to multiplying the time, cost, and complexity of the solution by at least an order of magnitude, the final draw-back to the hybrid approach is inconsistency of the results. Because of the fundamental differences in the datasets and the portions of the ecosystem that are covered by each technique, the granular “bottom-up” numbers from online tracking never match the high-level “top-down” results from econometric modeling. This makes the validation task nearly impossible.

CONCLUSION

Given the additional complexity that digital marketing has added to the world of advertising, understanding the incremental value of individual marketing techniques can only be achieved through modeling. Controlled testing to determine incremental value, while still practical for large offline media campaigns, has become prohibitively expensive and time consuming when digital advertising is added to the mix. Fractional attribution as a replacement for controlled testing does not hold up to any sort of scrutiny unless the goal of the exercise is to produce a random number larger than zero. While not a panacea due to the added task of model validation, a granular econometric modeling approach using a modern big data architecture and machine learning techniques is a practical and cost effective solution for determining the incremental value of marketing.

LEARN MORE

To learn how OptiMine Insight can help you maximize your cross-channel marketing goals, contact us at www.OptiMine.com or info@OptiMine.com.

References

1. Cookies Are So Yesterday; Cross-Device Tracking Is In -- Some Tips. Michael Whitener, 2015-01-27. <https://iapp.org/news/a/cookies-are-so-yesterday-cross-device-tracking-is-in-some-tips/>
2. The cookie is dead. Here's how Facebook, Google, and Apple are tracking you now. Richard Byrne Reilly, 2014-10-06. <http://venturebeat.com/2014/10/06/the-cookie-is-dead-heres-how-facebook-google-and-apple-are-tracking-you-now/>
3. The Impact of Cookie Deletion on the Accuracy of Site-Server and Ad-Server Metrics: An Empirical comScore Study. Dr. Magid Abraham, Cameron Meierhoefer, Andrew Lipsman, 2017-06.
4. Do Display Ads Influence Search? Attribution and Dynamics in Online Advertising. Pavel Kireyev, Koen Pauwels, Sunil Gupta, 2013-02-28. <http://hbswk.hbs.edu/item/do-display-ads-influence-search-attribution-and-dynamics-in-online-advertising>
5. Big Data: New Tricks for Econometrics. Hal R. Varian, 2013-06.
6. Freakonomics: A Rogue Economist Explores the Hidden Side of Everything. Stephen J. Dubner, Steven Levitt, 2005-04-12.
7. Marketing Mix Modeling: What Marketing Professionals Need to Know. Nielsen Holdings N.V., 2014.

