No Measurement is Better than Bad Measurement

Steven Gordon, Ph.D.

Digital advertising allows businesses and organizations to quantify the impact of their advertising expenditures on sales and other positive outcomes. The assumptions inherent in statistical estimates of such effects, however, are often unrealistic. As a result, overly optimistic estimates of the returns to online advertising frequently lead to flawed decision making regarding the allocation of advertising dollars.

Online advertising expenditures have seen rapid growth in recent years. Digital ad spending in the U.S. alone is forecasted to reach $83 billion in 2017, surpassing that of television for the first time.[1]

The appeal of online advertising lies in its ability to answer many elusive questions regarding consumer behavior. In contrast with traditional mediums of advertising, such as radio or television, digital ads can be precisely targeted to narrow segments of potential customers, and–equally important–their use and performance can be measured in a variety of ways.

Today, a host of digital metrics such as click through rates (CTRs), cost per click rates (CPCs), cost per lead rates (CPLs), cost per acquisition rates (CPAs), and cost per engagement rates (CPEs) assist in answering one key question: If I spend $1 more on ads, how many more dollars in sales will I get?

The online marketing and analytics industry attempts to answer this question, but a simple understanding of incentives suggests that these companies may not be well motivated (and consequently, not well equipped) to provide would-be clients with a scientifically rigorous and objective measurement of the value of their services.

In 2011, the *Interactive Advertising Bureau* (IAB) commissioned a research study that found controlled experiments to be the only reliable method of measuring the impact of advertising.[2] The report also noted that these types of analyses were also the least common in the industry. This finding highlights a critical weakness of digital advertising service providers and analysts: The problem is not the lack of measurement; the problem is *poor* measurement.

> A 2011 report by the *Interactive Advertising Bureau* found controlled experiments to be the only methodology capable of providing reliable estimates of the impact of advertising, yet the report also found these types of analyses to be the least common in the industry.

The Selection Effect

Consider a hypothetical example. Suppose your business decides to subscribe to an online advertising service where 1.5 cents is charged to your account every time a user clicks on your ad. Imagine that the true causal impact of the ad campaign, unknown to you, is such that 1 out of 6,000 people are convinced to buy your product–a realistic assumption according to existing scholarly research.[3] If your product sells for $30, the true return on investment (ROI) for this service would therefore be $0.

Discovering the true impact of this ad campaign on your business's sales, however, is not easy. Can you really assume that the 6,000 users who clicked on your ad are truly representative–i.e., a random sample–of the entire population of all users? The reality is that whether through intentional targeting of your ads to users more likely to purchase your product (which is how most online ad campaigns are designed to operate) or by the fact that people clicking on your ad are probably more likely to be searching for your product anyway, users clicking on your ads will most likely have a higher likelihood of purchasing your product than the complete pool of potential customers.

Let's assume then that, on average, 10 percent of users who click on your ads end up buying your product (a conservative number given the amount of targeting that online advertising campaigns generally aim for). This means that, out of the 6,000 people who clicked on your ad in this case, odds are that 600 of them will buy your product for reasons other than the impact of the ad, while just 1 user will buy simply because of the ad. In this case the "selection effect"–that is, the effect of just being the type of user who clicks on your ad–will be *600 times larger* than the actual impact of the ad itself on sales. Thus, in measuring the impact of the ads on your sales, it will be all too easy to mistake the selection effect for the actual causal effect.

Suppose also that the probability of a random user purchasing your product is just over 1.5 percent. By comparing the users exposed to your ads to a random pool of users not exposed to your ads, you would have observed a difference of 8.5 percent between the exposed and unexposed groups of users. Not taking into consideration this selection bias in your estimate of advertising would have resulted in an estimate that would have seemed too good to be true: At a cost of $3,000 for the ad campaign described above, your estimate would have led you to calculate an ROI of roughly 600 percent.

This hypothetical (but also realistic) example highlights the reality that uncovering the causal impact of online advertising using data is a delicate undertaking. Unfortunately, the marketing and analytics industry does not approach this task with the nuanced understanding necessary to obtain meaningful results.

In a 2008 *Harvard Business Review* article, the president of comScore–a leading media measurement and analytics company–stated that measuring the impact of online advertising requires only a simple comparison of the purchasing activity of the exposed versus unexposed users.[4] This approach assumes that the purchasing behavior of the people who did not click on the ad represents an accurate picture of what the people who did click on the ad would have purchased had they not clicked. Clearly, this is a big assumption and an erroneous one one at that.

> Statistically estimating the causal impact of online advertising requires a sophisticated understanding of the assumptions inherent in any methodological approach. Unfortunately, the marketing and analytics industry is naive to the methodological nuance necessary to obtain meaningful results.

In a 2014 analysis of online advertising, several economists from *Yahoo! Research* found that a similarly naive approach to analyzing the impact of online ads would have yielded an estimated impact of online advertising that was nearly three times larger and in the opposite direction than their carefully calculated experimental estimate.[3]

In the example above, simply comparing the purchasing totals of people who clicked on the ads with the purchasing totals of people who did not click on the ads would have led to an estimate naturally "biased" upwards–by a factor of nearly 600. This makes the careful choice of statistical modeling assumptions vital if data are to be used at all to make decisions about advertising expenditures.

A Better Way

The key to improving the measurement of the effectiveness of advertising on sales is straightforward: Estimate what would have happened in the absence of advertising, and compare that with what did happen. However, this seemingly simple estimation task represents a daunting challenge.

Imagine you possess a time machine that allows you to turn back the clock and repeat history. In this case, a statistical analysis of the impact of advertising would be straightforward: Increase spending on ads and record your sales, then hop in your time machine, turn back the clock, and record sales again–this time leaving your advertising budget unchanged. The difference would be the true causal impact of advertising on sales since you would have perfect knowledge about what would have happened in the absence of advertising.

Without a time machine, however, any statistical analysis must rely on a guess about what would have happened had there been no increase in advertising. This guess is the "counterfactual," and any measurement of the impact of advertising on sales will only be as good as the counterfactual is realistic.

The challenge of creating a reliable counterfactual is not primarily a problem that requires vast computing power, complicated algorithms, or "big data." Rather, it is a challenge that can only be overcome with a combination of creativity, intuition, and sophisticated statistical methods that rely on careful reasoning.

Statistical models have a variety of ways of solving this problem. Randomized experiments are generally considered to be the best way of generating a counterfactual: Users or locations are randomly selected into "treatment" and "control" groups, and differences in sales are calculated while taking into consideration the effect of other relevant variables, such as demographics and seasonality. However, businesses and organizations without the resources to undertake these often expensive experimental tests can use other methods to still get a reasonably good counterfactual. Forecasting techniques, synthetic control methods, and instrumental variables models are all methods that even small businesses and organizations with limited resources can utilize to ensure that their advertising dollars are actually generating sales, increased membership, or whatever positive outcomes are desired.

> Instead, the challenge must be met with a combination of creativity, intuition, and sophisticated statistical methods that rely on careful reasoning.

The real world reality of data distorting tendencies, such as the selection effect described at the beginning of this paper, puts a premium on the value of logically sound, empirical techniques in estimating the impact of advertising. What is needed is a systematic way of drawing (or not drawing) data-driven conclusions while maximizing the amount of information in the data.

A good statistical analysis of the above scenario would have, at worst, shown a small or negligible estimate of the impact of advertising on sales with insufficient statistical precision to give us confidence in the results. The virtue of accurate analyses, however, is that even in the worst case of finding not enough usable information in the data, such an approach would still lead to a rather useful insight that conforms to common sense: Given the data we have, we cannot say if there is any impact of advertising on sales.

It would be far better in this case to rely on simple intuition than a poorly designed statistical analysis.

Faulty Statistics Corrupt Common Sense

Returning to the example from the beginning of this paper, an ROI of 600 percent would have warranted a huge increase in advertising spending which you are (rightfully) not willing to make–your intuition as the leader of your organization would probably have told you that this ROI could not be true, and yet it might also have helped to "move the needle" on your decision making regarding advertising expenditures. Consequently, you might have invested a bit more in advertising than you would have in the absence of such results.

> What is needed is a systematic way of maximizing the amount of information extracted from the data, and every statistical technique should be subject to the "sniff-test."

While careful statistical modeling is capable of complementing and informing the intuition of business leaders and decision makers, no measurement is better than *bad* measurement. In the example described above, your natural intuition would have been closer to the truth, but introducing flawed assumptions into your decision making process might have caused you to second guess your gut and make an erroneous decision.

Clearly, what is needed is for the leaders of businesses and organizations to recognize the virtues of careful statistical analyses and know that intuition or common sense is not precluded by the use of sophisticated modeling techniques. Every statistical technique should be subject to the "sniff-test": Do the results make sense? If not, a return to the drawing board is warranted.

References

[1] Lauren Johnson. *U.S. Digital Advertising Will Make $83 Billion This Year, Says EMarketer*. `http://www.adweek.com/digital/u-s-digital-advertising-will-make-83-billion-this-year-says-emarketer/`. 2017.

[2] Paul J Lavrakas. "An evaluation of methods used to assess the effectiveness of advertising on the internet". In: *Interactive Advertising Bureau Research Papers* (2010).

[3] Randall A Lewis and David H Reiley. "Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo!" In: *Quantitative Marketing and Economics* 12.3 (2014), pp. 235–266.

[4] Magid Abraham. "The Off-Line Impact of Online Ads". In: *Harvard Business Review* 86(4):28 (2008).

