Causal impact of digital display ads on advertiser performance

Abstract: Brands are searching for innovative ways to reach customers online. Sponsored Display (SD) by Amazon Ads is a new way to do so, and allows customer reaching strategy by category, product and audience. However, advertisers are uncertain how much SD improves their performance over different time horizons. This paper studies more than 40,000 brands with two different methods: a diffusion-regression state-space time-series analysis that predicts response counterfactuals during a 20-weeks period post SD adoption of audience reaching strategy, and a non-parametric Gaussian Process algorithm that generates counterfactuals using Bayesian multitask learning to draw causal inferences in shorter time frames (i.e., one-month post SD adoption of category and product customer reaching strategy). The performance variables include impressions, page views, sales, new-to-brand consumers and Return on Advertising Spend. The results are consistent and quantify how much adding SD to the ad mix increases performance.

Keywords: display advertising, machine learning, causal inference

Track: Digital Marketing and Social Media

1. Objectives and Overview of the Research

Digital marketplaces have continued to grow over recent years, with the Amazon marketplace bringing together over 1.9 million independent suppliers with over 100 million Amazon Prime customers in addition to non-Prime customers (Rangaswamy et al. 2020). Likewise, digital advertising effectiveness and its causal attribution have seen strong research attention. However, advertisers are regularly offered new ways to reach audiences online and are uncertain how much such digital ads can improve their performance measures over different time horizons.

Display ads are online ads that combine copy and visual elements with a call-to-action message that links to a landing page. Consumers typically see display ads along the top or sides of a website—or sometimes, in the middle of the content they are reading. Recently, Amazon.com introduced Sponsored Display (SD) based on three potential audience strategies: (1) the category the consumer is browsing in, (2) the specific product, and (3) the audience remarketing to customers who browsed the product in the past. These three strategies are not mutually exclusive and they can be combined to engage with customers who are exploring products within a category, evaluating a specific product, or re-engage those customers who have browsed specific products without making a purchase to leverage missed sales opportunities. Although these SD strategies show promise, Digital advertisers are uncertain to what extent they drive performance variables such as impressions, Buy Box detailed page views (i.e., visits to the product pages), sales, and new-to-brand consumers (i.e., consumer who have not bought the brand in the previous year on Amazon.com). Furthermore, beyond SD's intrinsic value, advertisers are interested in learning know how they can use SD to complement their existing portfolios of other Amazon Advertising product such as Sponsored Products (SP), i.e., ads showing individual products to Amazon shoppers in related shopping results and product pages, and Sponsored Brands (SB), ads showcasing the advertiser's brand to Amazon shoppers in related shopping results and product pages (Amazon Learning Console 2021). While SP is widely seen as a bottom-funnel tactic and SB as a mid-funnel tactic, SD is considered an upper-funnel tactic, which may increase the customer base and sales over the longer run, but it may decrease efficiency, typically measured as Return on Advertising Spend (Robb 2021).

To measure the performance impact of the different SD strategies over long- and short-term-time horizons, we used two different, complementary methodologies. First, we estimate a diffusion-regression state-space model that predicts the counterfactual response that would have occurred over a 20-week look forward horizon had the advertiser not adopted audience reaching strategy SD. Second, we employ Causal Multitask Gaussian Process (CMGP) to evaluate the impact of SD adoption among advertisers using category and product SD strategies over a 1-month time horizon after enabling SD. This method, proposed by Alaa and Van der Schaar (2018), generates counterfactuals using a Bayesian multi-task learning approach.

The results from these two approaches are consistent. The counterfactual time series analysis shows that advertisers activating SD with audience reaching strategy for the first-time increased sales by +14% on average within the first 20 weeks after adopting SD as compared to not enabling SD. On the other hand, CMGP shows that brands that began using SD category reaching strategy for the first time saw, on average, positive impacts across different metrics during the next month after adoption as compared to advertisers that didn't: +33.9% more impressions, a +3.6% increase in Buy Box Detailed Page Views (DPV) and a +2.6% increase in New-To-Brand (NTB) customers. Similarly, Brands that created an SD product customer reaching strategy campaign for the first time saw, on average, increases of +28.8% in sales, +12.4% in DPV, +3.2% in NTB customers' awareness, and +4.2% in NTB customers' consideration the following month, compared to brands that did not.

2. Methodology

2.1 Diffusion-regression state space model

In contrast to difference-in-differences schemes (Lechner 2011), state-space models allow inferences about the temporal evolutions of attributable impact, and flexibly accommodate multiple sources of variation, including the time-varying influence of contemporaneous covariates, local and linear trends, and seasonality components. In addition, these models can adopt a fully Bayesian nature by incorporating empirical priors on the model parameters which adds extra flexibility and robustness to the analysis. In this context, for the first part of our analysis we applied a Bayesian Structural Time Series Model (Brodersen et al. 2015). We selected 284 advertisers that satisfied the following conditions within a 50-week timeframe: (1) advertisers were active SP and/or SB for at least 30 weeks prior to SD Audience Reaching Strategy activation, (2) in the last 20 weeks of the analyzed time period, the only advertising-specific action they took was launching an SD Audience Reaching Strategy campaign, and (3) advertisers should be similar in term of business size and ad-campaign activity. Finally, we selected advertisers with ad-support sales higher than 5-th percentile in SD to prevent skewing our results towards those advertisers that are still in a test-and-learn phase for these ads.

For the selected advertisers, we calculated the impact on sales during the 20 weeks following their SD-Audience activation by predicting how their sales would have evolved if the SD-Audience activation had not occurred. We trained our model the first 30 weeks (pre-SD adoption period) to predict the counterfactual response for the following 20 weeks (post-SD adoption), using 10 additional covariates including sales, units, glance views, etc. at a vertical aggregation level and compares the same covariates for 600 advertisers that never activated SD within the same timeframe. Finally, we measured the lift in sales by subtracting the observed sales (real value) from the counterfactual sales (predicted sales).

2.2 Causal Multi-Task Gaussian Process Machine Learning Model

Second, to measure a shorter-term causal impact (i.e., 1-month post-activation) on advertisers who adopted SD product customer reaching and SD category customer reaching strategy for the first time we selected 43,720 advertisers in the US marketplace. Our analysis is based on a method proposed by Alaa and Van der Schaar (2018) called Causal Multi-task Gaussian Processes (CMGP) which estimates Conditional Average Treatment Effects (CATE), and has competitive performance on various metrics (e.g., RMSE, Coverage) as compared to existing methodologies in causal inference such as Causal Forests (Wager and Athey, 2018) and Propensity Score Matching (Rosenbaum and Rubin, 1983), when applied to observational data. This algorithm, builds upon the idea of Gaussian Processes (GP) in the context of Multitask Learning for the estimation of Individual Treatment Effects (ITE) (Bonilla, E. V., et. al. 2008), and according to its properties and results provided by its authors it is a suitable alternative for impact estimation study.

This method consists in a non-parametric Bayesian approach that uses a multi-task GP with a Linear Model of Coregionalization kernel (LMC) (Álvarez et. al. 2016). It learns simultaneously from both treated and untreated population response functions which reduces the impact of selection bias in the final estimates. These are obtained through a risk-based empirical Bayes method that jointly minimizes the empirical error in observed outcomes and the variance in unobserved counterfactual outcomes. Similarly, as several observational causal inference methods, this approach assumes overlap between treatment and control population and strong ignorability.

The authors show that is possible to connect the regularized empirical risk minimization of the PEHE (Precision in Estimating Heterogeneous Effects) with Bayesian inference using GPs. The estimated CATE function (denoted by $\hat{T}(x)$, where $x \in R^d$ is a vector that describes advertiser characteristics – 50+ attributes) can be interpreted as a posterior mean given a GP prior with a covariance kernel K_θ, thus $T(.) \sim GP(0, K_\theta)$. Here, $\theta \in R^q$ corresponds to the vector of hyper-parameters that parametrizes the kernel and defines the functional behavior of the CATE function. This way of approaching the CATE estimation has the following advantages: (1) Enables the estimation of individualized uncertainty in $\hat{T}(x)$, (2) Provides a natural proxy for the empirical PEHE minimization required to estimate the CATE via Bayesian risk, and (3) it incorporates the posterior uncertainty in counterfactual outcomes without explicit propensity modeling. This last benefit is especially relevant in the context of advertising as it adds more robustness in the results, especially in those cases where for some products/segments the availability of sample sizes may be limited. These types of scenarios add complexity in estimating either propensity scores, or the nuisance functions required by other causal inference methods (e.g., Causal Forests, Propensity Score Matching) for selection bias mitigation. Note that although the estimator produces treatment effect estimates at the individual level, it is possible to roll these up and obtain the population Average Treatment Effect (ATE).

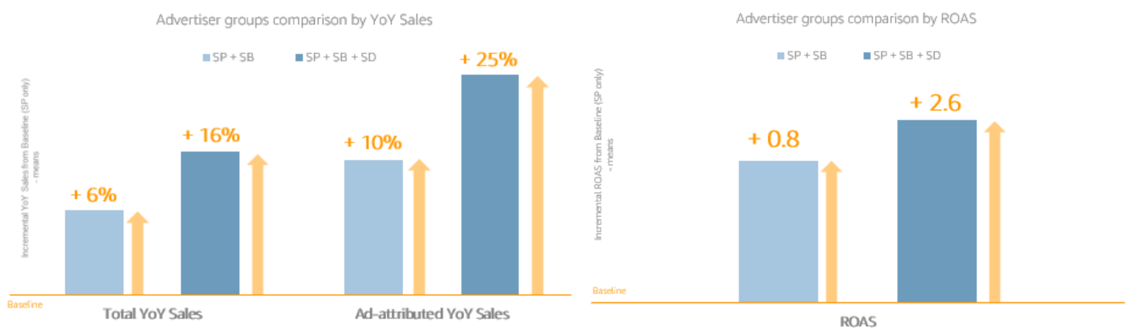
Although flexible and adaptive for different CATE functions without the need of a-priori specification, some of the short-comings of this method are: (1) Computational cost. Since the estimation requires the inversion of a covariance matrix spanned by the kernel $\boldsymbol{K_\theta}$, the LCM structure increases the $O(n^3)$ complexity of the GP by a factor of 8, which reduces computational efficiency for large sample settings; (2) Since it is based on Gaussian Processes, the method is sensitive to the Kernel specification. For our analysis, we used Radial Basis Function (RBF) kernel with Automatic Relevance Determination (ARD).

3. Major Results

In our analysis we used data between April 2020 and June 2021. As to model-free evidence, ***Figure 4*** below compares the Year-over-Year (YoY) growth of brands combining SD with SP and SB vs. those using SP + SB (which we considered the baseline). The left panel shows: (a) Total Sales, and (b) Total Ad-attributed Sales Year-over-Year (YoY) growth. The right panel shows the

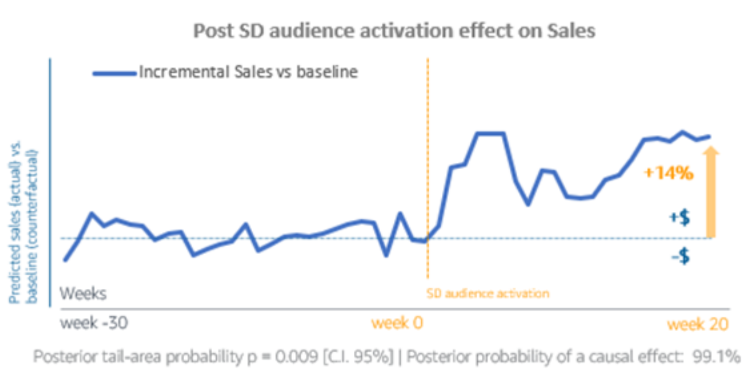
Return on Advertising Spending (ROAS), i.e., the $ sales revenue earned for each $1 spend on advertising. Brands combining SP + SB + SD generated +16% more in total sales and +25% more in ad-attributed sales, respectively, vs. brands using SP only, and they achieved +2.6 higher ROAS. Thus, adding Sponsored Display audience reaching strategy (an upper-funnel tactic) to the existing lower- and mid-funnel tactics, is associated with higher sales growth (effectiveness) without decreasing efficiency (as measured by ROAS), answering a key advertiser question.

Figure 4: Higher performance for advertisers who added Sponsored Display (Model-free)



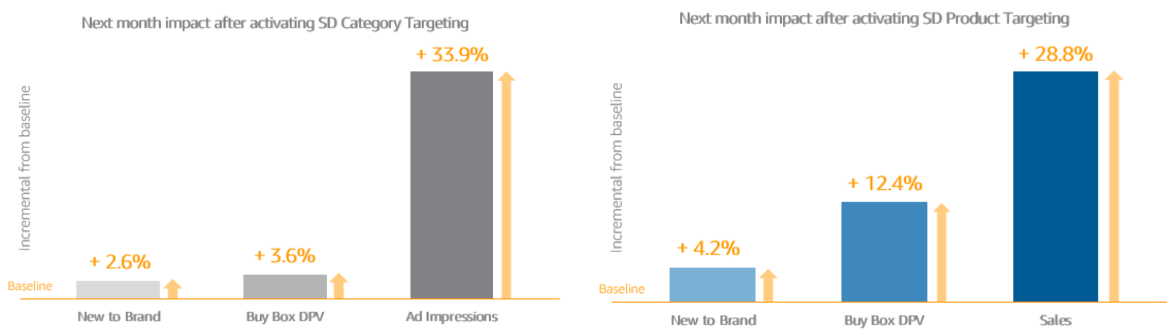
From a medium-term impact perspective, in our *counterfactual causal analysis*, the 284 advertisers that adopted SD with *audience reaching strategy* saw, on average, a +14% increase in total sales in the following 20 weeks, compared to their estimated sales without SD adoption. Figure 5 shows on the left panel that the model works well predicting the average baseline of actual sales (i.e., weeks -30 to 0 in the horizontal dotted line) and on the right side the impact of adding SD to SP+SB during the 20 weeks post SD adoption. The posterior probability of a causal effect is 99.1%. We performed validation separately with posterior tail area probability of p=0.009.

Figure 5: Incremental sales from adopting Sponsored Display Audience vs. counterfactual baseline



Finally, using our CMGP method we estimated the impact of adding SD over the next month post SD adoption. Brands that began using *category customer reaching* strategy within SD strategies for the first time saw, on average, +33.9% more impressions, a +3.6% increase in Detailed Page Views and a +2.6% increase in New-To-Brand customers the following month compared to those that didn't. Similarly, brands that created an SD *product customer reaching* campaign for the first time saw, on average, a +28.8% sales increase, +4.2% DPV increase and a +2.6% NTB increase the next month compared to those that didn't. ***Figure 6*** summarizes these results We measured the statistical significance with a 5% significance level of these estimates using a bootstrapping procedure.

Figure 6: Two-stage Gaussian Process estimates of impact of SD audience reaching strategy.



4. Implications: Incorporate Sponsored Display strategies in digital advertising

Using a multi-method approach, we conclude that brands that incorporated Sponsored Display experienced increases in total sales ranging from +10% to +29%, as well as increases in impressions, Detailed Page Views, New-to-Brand customers, ad-attributed sales, and Return on Advertising Spend (ROAS), compared to brands that only use Sponsored Products, or Sponsored Products and Sponsored Brands on Amazon.com. Based on these results, we recommend that advertisers incorporate Sponsored Display to their media plans. We also recommend that brands consider using multiple SD tactics, such as audience, category and product customer reaching strategy. Future research is needed to explicitly compare audience with category and product customer reaching strategy, and to quantify how combinations of these approaches are best for advertisers under different conditions, such as category and brand characteristics (e.g., consumer involvement and brand strength).

References

Alaa, A. M. and van der Schaar, M. (2018) "Bayesian nonparametric causal inference: Information rates and learning algorithms." *IEEE Journal of Selected Topics in Signal Processing*, 12(5),1031–1046.

Amazon Learning Console (2021), https://advertising.amazon.com/library/courses/reach-shoppers-with-sponsored-display/?ref_=a20m_us_rfy_lbr_crs_1 .

Brodersen, K. H., F. Galluser, J. Koehler, N. Remy and S.L. Scott (2015), "Inferring Causal Impact using Bayesian Structural Time-Series Models." *The Annals of Applied Statistics*, 9 (1), Institute of Mathematical Statistics, 247–274, http://www.jstor.org/stable/24522418.

Bonilla, E. V., Chai, K. M. A., & Williams, C. K. I. (2008). "Multi-task Gaussian Process Prediction". In *Advances in Neural Information Processing Systems 20 (pp. 153-160)*. NIPS Foundation. http://books.nips.cc/nips20.html

Rosenbaum PR and Rubin DB. 1983. "The central role of the propensity score in observational studies for causal effects". *Biometrika, 70(1); 41-55*.

Álvarez, Mauricio A., Lorenzo Rosasco and Neil Lawrence. "Kernels for Vector-Valued Functions: a Review." Found. Trends Mach. Learn. 4 (2012): 195-266.

Jennifer L. Hill (2011), "Bayesian Nonparametric Modeling for Causal Inference", *Journal of Computational and Graphical Statistics*, 20:1, 217-240, DOI: 10.1198/jcgs.2010.08162

Lechner, M. (2011), "The Estimation of Causal Effects by Difference-in-Difference Methods", *Foundations and Trends® in Econometrics*: 4 (3), 165-224.

Rangaswamy, A., Moch, N., Felten, C., van Bruggen, G., Wieringa, J. E., & Wirtz, J. (2020). The role of marketing in digital business platforms. *Journal of Interactive Marketing*, *51*, 72-90.

Robb, K. (2021), Amazon Sponsored Display Ads: A Full-Funnel Approach, *Teikametrics*, https://www.teikametrics.com/blog/amazon-sponsored-display-ads-a-full-funnel-approach/

Wager, S. and Athey, S. (2018), "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association*, 113 (523),1228–1242.