**Upper funnel ad effectiveness and seasonality in consumer durable goods**

Vivian Qin, Koen Pauwels

Amazon Ads

**ABSTRACT**

Brands usually invest in a portfolio of ad products for brand consideration and conversion. To gauge performance, brands often use ad attributed metrics to compare Return on Advertising Spend (ROAS) across different ad channels. However, we see two major shortcomings with this approach. First, it relies on the 14-day last-click attribution window. Though it is common industry practice, the resulting ROAS does not only favor lower funnel ads, such as paid search, but could change with different attribution models. Second, attributed performance metrics usually don’t consider seasonality, and organic consumer demand changes over the course of the year, which could bias the estimates of the effectiveness of ads. This could result in mistakenly attributing high organic demand to campaign performance. We address these issues with a Seasonal Autoregressive Integrated Moving Average with X/Exogenous Variables (SARIMAX) model, accounting for seasonality and brands’ past performance. This analysis compares ad efficacy on total retail metrics, regardless of attribution methods. We apply this method to Amazon Ads, for fifteen brands in U.S. consumer durables. While the lower funnel ad product (Sponsored Products) sees a lot more current usage, we find that higher funnel ad products such as Sponsored Brands, DSP Display, and Streaming TV Video all have higher efficacy. We recommend brands evaluate their ad product performance at regular cadence to avoid under-utilizing high-performing ad products. We also encourage future research to replicate this methodology in different verticals and locales for generalizability.

**Keywords:** digital advertising, marketing analytics, time series model, Amazon Ads, consumer durables, brands, ROAS

**INTRODUCTION**

Brands typically reference Return on Advertising Spend (ROAS) based on ad-attributed metrics to evaluate ad product’s effectiveness, and tend to do one advertising action at a time (Olsson, 2022)1. However, brands rarely invest in only one ad product, which calls for a holistic view of advertising performance rather than each ad channel evaluated in silo2. Furthermore, many consumer brands have demand seasonality. Not accounting for it not only renders performances at different points in time incomparable, but also likely erroneously attributes organic demand fluctuations to campaign effectiveness.

Based on specific advertiser requests to do so, this analysis leveraged data from fifteen brands’ requests to compare ad product effectiveness on Amazon.com. Data for each brand ranges from 1 year to 1.5 years at the daily level from Jan 2021 to June 2022. All brands are in consumer durables (labeled ‘Hardlines’ by Amazon), and all except one are from U.S. Hardlines’ four root browse nodes3: Electronics, Tools & Home Improvement, Automotive, and Home and Kitchen. For each brand, we use daily data on Sponsored Products (SP), Sponsored Brands (SB), DSP Display and Streaming TV Video (STV) spending and on-Amazon sales to estimate a Seasonal Autoregressive Integrated Moving Average with X/Exogenous Variables (SARIMAX) model, accounting for seasonality and brands’ past performance.

This analysis has three key findings. First, all brands have at least one ad product other than SP that significantly improves performance metrics such as detailed page view or retail revenue. This is important to brand managers as besides SP, other ad products have low adoption rate, contrasting to their high performance uncovered in this analysis. Second, while DSP Display and STV significantly increase sales in a model that accounts for all invested channels, the indexed effectiveness is higher for SB and SD. Third, SB and STV effectiveness has smaller variances than SD and DSP Display, suggesting the latter two should be tested on a case-by-case basis.

Based on those findings, we recommend the followings. First, brands should adopt ad products beyond SP especially if they want to grow brand consideration, and better still, those ad products improve conversion as well. Relatedly, we encourage brands to be more open-minded for upper-funnel and middle-funnel ad products and adopt a test-and-learn approach for each brand. Second, we encourage brands to measure ad products’ effectiveness appropriately beyond last-click heuristic view, especially when a holistic brand-level media planning is needed. Last but not least, for product categories with high seasonality, current methodology accounting for this not only informs brands’ marketing strategies but also enables more efficient ad impact estimates. So, we suggest choosing methodology appropriately for the brand’s specific product category.

**RESEARCH METHODOLOGY**

Given our objectives, we have four specific requirements for the research methodology.

First, the method should explain brand performance by all relevant ad products at the same time, empirically estimating their impact instead of using heuristic attribution such as last click, first click, or equal importance.

Second, the method should provide insights for brand media planning with budget allocation across ad products at the strategic level. It should not require tracking individual customers or building audiences, which are part of more tactical campaign level optimization. The current scope of analysis is brand media strategy by channel, not campaign by audience types.

Third, the method should go beyond relating performance to same-period ad spending. It is well known that current performance values are strongly driven by past values, and that failing to account for this may lead to spurious and over-estimated effects of KPI drivers4. Controlling for brands’ past performance, the estimated impact is much less likely to be driven by past performance but more likely to be attributed to ad investment. As a result, ad efficacy estimates become more precise.

Fourth and finally, the method should identify and control for both time-related (i.e., seasonality) and random fluctuations (i.e., noise), to provide more efficient estimates and enable right-sizing the impact of ad investment.

Based on these requirements, we propose the SARIMAX (Seasonal Autoregressive Integrated Moving Average with X/Exogenous variables) model. Designed specifically for time series data, the SARIMAX will

1. explicitly models Seasonality (S);
2. explicitly accounts for past performance in its Autoregressive (AR) component;
3. empirically tests if dependent variables are non-stationary, i.e., integrated (I);
4. uncovers patterns from noise in the Moving Average (MA) component;
5. efficiently estimates the effect of each ad products in the (X) component.

Below is an example representation of the model formation of SARIMAX (p,d,q) (P,D,Q, s), where (p,d,q) are the non-seasonal parts’ autoregressive, differencing and moving average order, (P,D,Q) are the order for the seasonality counterpart, and s is the periodicity of the seasons, i.e., the interval a recurring patterns occurs:

Where

is the non-seasonal autoregressive lag polynomial

is the seasonal autoregressive lag polynomial

is the time series, differenced d times, and seasonality differenced D times

is the trend polynomial (including the intercept)

is the non-seasonal moving average lag polynomial

is the seasonal moving average lag polynomial

X is the contemporaneous exogenous ad spend variable

is the error term

The exact orders of (p,d,q) (P,D,Q,s) are done through Python Library [pmdarima](https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.ARIMA.html)’s auto\_arima procedure, where the parameter combination with the lowest Akaike Information Criterion (AIC) is selected as the best model. We execute this procedure for two performance variables: (1) Detail Page Views (DPV), representing consumer consideration of the brand on Amazon5 and (2) retail revenue, representing conversion. The two performances are estimated separately because ads could have different effects on each KPI. Consistent with past academic research6, we extract the coefficients from best-fitting model and interpret all as relative to SP as the benchmark.

Variable inclusion and lag order selection follow the established tests for time series models6. First, we tested the stationarity of all variables with Augmented Dickey-Fuller (ADF) test. All variables were classified as stationary and thus no further differencing is needed. Second, before each model fitting, we checked multicollinearity by using Variance Inflation Factor (VIF) and took out ad product variables with a VIF above 10. Third, we selected the lag order preferred by both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Fourth, we verify that the residuals are approximately normally distributed with Q-Q plot visuals and show no substantial serial correlation via the Ljung-Box test.

Beyond the coefficient estimates, we also derive from the SARIMAX model the importance of seasonality, which is the magnitude of KPIs that repeat at regular cadence. This importance is expressed as a percentage (%) of the performance driven by seasonality.

**RESULTS**

The SARIMAX model adequately captures the data generating process for each studied brand, and the marketing coefficients are all significant at the 90% (p < 0.10) level. Brand level results are available upon request, subject to Amazon Approval process.

Our first substantive finding, shown in Figure 1, is that the upper funnel ads (SB, SD, DSP Display Video and STV) have higher effectiveness than SP on Detail Page Views (DPV). This means they significantly improve brand consideration relative to SP.

Figure 1. Indexed Ad Efficacy on Detail Page Views Relative to Benchmark\*

*\* Bar height indicates average value across all brands, while the line indicates standard deviation. SP is used as benchmark of 1.*

The findings are similar for revenue, as shown in Figure 2. Thus, increased investment in these ad products is associated not only with increased consideration, but also with increased conversion.

Figure 2. Indexed Ad Efficacy on Revenue Relative to Benchmark\*

*\* Bar height indicates average value across all brands, while the line indicates standard deviation. SP is used as benchmark of 1.*

Interesting differences arise by root browse node. Figure 3 shows that DSP Display (Video) and SB (SBV) have the highest efficacy on Detail Page Views for Electronics, whereas SD has the highest efficacy for Tools and Home Improvement. However, for revenue (Figure 4), both SB (SBV) and SD have the highest efficacy in Electronics, while DSP Display performs the best for Home and Kitchen.

Figure 3. Indexed Ad Efficacy on DPV by Browse Node\*

*\* Bar height indicates average value across all brands, with SP as benchmark equal to 1.*

Figure 4. Indexed Ad Efficacy on Revenue by Browse Node

*\* Bar height indicates average value across all brands, with SP as benchmark equal to 1.*

We note the contrast between consideration versus conversion goals for SD and DSP Display (Video), consistent with the upper funnel nature of these ad products. Specifically, in Tools and Home Improvement, SD performance is about 8x higher than SP for consideration, but only twice as high for revenue. In Electronics, DSP Display (Video) performance is 4x that of SP, but it is at par with SP as a revenue generator. In contrast, DSP Display (Video) compares favorably to SP in revenue impact for Tools and Home Improvement and for Home and Kitchen. We speculate that video is especially effective for helping consumers choose among such experience products8, where they can access dynamic demonstration of the focal products. This offers a much more engaging experience than search ads with only textual information. As search ads (e.g., SP) appear on search results and close to conversion, it is not surprising the disparity in efficacies are smaller in revenue than DPV. Thus, the effectiveness of ad products depends on both the advertising goal and the nature of the product category.

Finally, the importance of seasonalityalsodiffers by root browse node, with the highest from Tools & Home Improvement, and the lowest from Home & Kitchen. Figure 5 shows that, on average, Tools & Home Improvement has the highest seasonality at 10%, whereas Home & Kitchen has the lowest at 6%. Similar results appear for Detail Page Views.

Figure 5. Percentage of Seasonality in Revenue Variance Explanation by Browse Node

A higher percentage of seasonality highlights the importance of accounting for seasonality when evaluating ad effectiveness. Furthermore, the presences of seasonality also suggest that brands can tailor their marketing strategy to the season. For example, brands could utilize key seasonal shopping moments by doubling down on ad investment, and continue with an ever-green in low seasons to maintain momentum9.

**CONCLUSION**

With a SARIMAX model, we are able to test the relative effectiveness of ads based on the advertising portfolio leveraged by brands. Although Amazon Ads offer a variety of ad products across search, DSP display, video, and Streaming TV Video, internal analysis shows that 8 out of 10 CPG and Softlines brands invest only in Sponsored Products. For fifteen brands included in the current study, we consistently see the significant impact coming not only from Sponsored Products, but also from other Sponsored ads (i.e., Sponsored Brands, Sponsored Brand Video and Sponsored Display), DSP Display, and STV. The relative effectiveness of each depends on both advertising goal (consideration vs conversion) and product category. Furthermore, the importance of seasonality ranges from 6% to 10% across product categories, underscoring the importance of accounting for seasonality when measuring ad effectiveness.

This research has several limitations that suggest areas for future research. First, the 15 brands were studied because of requests from both Amazon Ads media managers and respective brand managers. Each brand has at least one year of ad spend and KPI at daily grain so that the seasonality effect could be measured. Future research would ideally include a random selection of brands with a larger sample across more product categories. Second, research is needed in non-Amazon settings and in other countries and categories. We encourage future replications primarily for two reasons: first, both commonality and differences in the patterns could inform us how to serve consumers in a specific product category better with more tailored ad recommendation; second, academic literature also shows shoppers’ reactions to ads can change depending on macro factors such as cultural settings10, ad platforms11, and individual context such as device12. Finally, future research should explicitly test that marketing effectiveness differs across seasons, so that peak and maintenance budgets can be optimized.

In sum, we learned that upper funnel marketing effectiveness can be higher than lower funnel marketing effectiveness for 15 brands across US consumer durable categories on Amazon.com. We recommend brands to explore these opportunities and adopt a test-and-learn approach for verticals/locales not covered in the current study.

references

1. Olssen, J. (2022) “What is ROAS?”, *Match2one*, May 17th, <https://www.match2one.com/blog/what-is-roas/>
2. Reimer, K., Rutz, O. J., & Pauwels, K. (2014). How online consumer segments differ in long-term marketing effectiveness. *Journal of Interactive Marketing*, *28*(4), 271-284.
3. Root browse nodes are the highest level of Amazon’s hierarchy of nodes: https://webservices.amazon.com/paapi5/documentation/use-cases/organization-of-items-on-amazon/browse-nodes.html
4. Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, *2*(2), 111-120.
5. Amazon (2022), Increasing Consideration in the Shopping Journey on Amazon, <https://advertising.amazon.com/library/guides/increasing-consideration-in-the-shopping-journey-on-amazon>
6. Li, J., Konuş, U., Pauwels, K., & Langerak, F. (2015). The hare and the tortoise: do earlier adopters of online channels purchase more?. *Journal of Retailing*, *91*(2), 289-308.
7. Pauwels, K. (2018), Modeling Dynamic Relations among Marketing Actions and Performance Metrics, [*Foundations and Trends in Marketing*](https://econpapers.repec.org/article/nowfntmkt/), 11 (4), 215-301
8. Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy, 78*(2), 311–329.
9. Hanssens, D. M., Wang, F., & Zhang, X. P. (2016). Performance growth and opportunistic marketing spending. *International Journal of Research in Marketing*, *33*(4), 711-724.
10. Lepkowska-White, E., Brashear, T. G., & Weinberger, M. G. (2003). A test of ad appeal effectiveness in Poland and The United States-the interplay of appeal, product, and culture. *Journal of Advertising*, *32*(3), 57-66.
11. Bayer, E., Srinivasan, S., Riedl, E. J., & Skiera, B. (2020). The impact of online display advertising and paid search advertising relative to offline advertising on firm performance and firm value. *International Journal of Research in Marketing*, *37*(4), 789-804.
12. Stewart, K., Kammer-Kerwick, M., Auchter, A., Koh, H. E., Dunn, M. E., & Cunningham, I. (2019). Examining digital video advertising (DVA) effectiveness: The role of product category, product involvement, and device. *European Journal of Marketing*, *53*(11), 2451-2479.