Using Generalized Additive Models in Marketing Mix Modeling
Patralekha Bhattacharya, Thinkalytics

ABSTRACT
Marketing professionals have always struggled to accurately measure the carry over effects of their marketing efforts. While it is widely accepted that campaigns impact sales over several time periods, adding multiple lagged terms in an attempt to model them, greatly increases the collinearity among variables, and increases the variance of the estimated coefficients. To compound the problem, the non-linear nature of the relationship often leads to incorrect model specifications and may result in irrelevant lagged terms showing up as significant in the model. We propose a generalized additive model (using PROC GAM) to estimate the lagged effects of advertising efforts in simulated models. The GAM procedure is better able to handle multi-collinearity than other procedures. Besides, the flexibility of the functional form allowed leads to more accurate estimates of the duration of the lag. In this paper, we present the approach to building marketing mix models using PROC GAM and then use simulations to demonstrate its effectiveness over other approaches.

INTRODUCTION
It is well established that advertising not only has immediate short term effects on the sales of a product but influences sales in the longer term as well. Plenty of evidence exists of the long term effects of advertising. One of the most famous pieces of evidence was a study conducted by IRI in 1992 (Lodish and Lubetkin, 1992). They selected 44 advertising tests that generated returns in the first year and received no advertising support in the following 2 years. They found that the average incremental sales effect of these tests in the 1st, 2nd and 3rd years were 22%, 14% and 7% respectively. Therefore according to this study, the total long term impact (in the 2nd and 3rd years) was as large as the immediate incremental effect in the 1st year. Broadbent (2000) performed a meta-analysis of 113 case studies in eight countries. He found evidence of long term effects of advertising sometimes lasting more than a year later. Dyson (2008) indicates that adding together the long and short term impacts of advertising, the total advertising effect is increased by a factor of 3 or 4. Lodish et al. (1995) looked at 55 tests in each of which a test group was exposed to heavier advertising than a matched control group for a one year period after which the advertising was stopped. They showed that on average, the initial one year sales impact of successful TV advertising campaigns is approximately doubled when the sales impact over the next 2 years are added.

The effects of advertising that are felt in future periods can be due to (1) Delayed response effects: a delay between the actual advertising expenditure and the sales of the product or (2) Holdover effects: customers who try and like the product after the first purchase may continue to buy it in future periods. Delayed response effects in turn may result from (1) airing delay: the ad may air a few weeks after the expenditure (2) viewing delay: the customer may actually see the ad a few weeks after it first appears and (3) customer inertia: even after the ad is viewed it may be several weeks before the customer actually makes a trip to the store to purchase the product. These delayed response effects are an important reason for advertising wearin which typically occurs at the start of a campaign. Advertising wearin is the phenomenon due to which advertising responsiveness does not kick off until a few weeks have passed after the start of the campaign. It occurs because repetition of the campaign in subsequent periods enables more people to see the ad, talk about it, think about it and respond to it. There may be lead effects of a campaign as well. Consumers may sometimes react in anticipation of a marketing stimulus. This is especially true for price promotions. If consumers know about a future price promotion, they may trim down their purchases in the current period. On the other hand, an expected price increase may encourage consumer hoarding and we might observe an increase in sales before the price increase and a dip in sales right after.

Marketing mix modeling practitioners have always struggled to accurately measure these delayed effects of advertising. Since sales in any period is influenced by marketing expenditure of previous periods, accurately measuring and predicting sales will require correctly estimating these lagged effects. Besides, accurately estimating the delayed response effects of advertising, helps the practitioner plan his marketing strategy in a way that is most effective in generating sales.
REVIEW OF LITERATURE

Estimation of advertising lags has received a great deal of attention in the marketing literature and many papers have been published on the subject. One way of modeling delayed effects of advertising is to introduce lagged advertising terms as independent variables in the model. Therefore Sales in period \( t \) depends on advertising in period \( t \) as well as in previous periods as shown in equation (1)

\[
S_t = \alpha + \beta_1 a_t + \beta_2 a_{t-1} + \ldots + \beta_s a_{t-s+1}
\]  

(1)

where \( S_t \) represents Sales and \( a_t \) represents advertising in period \( t \). \( \beta_1 \) through \( \beta_s \) are the lag coefficients.

Equation (1) indicates that advertising has an effect up to \( s-1 \) periods in the future. However estimation of equation (1) can become difficult for the several reasons. First of all it is difficult to decide how many lagged terms to include in equation (1). Without prior knowledge about how long the effects of advertising last, we cannot choose a value of \( s \). Secondly with \( s-1 \) lagged terms the number of parameters turns out to be \( s + 1 \). For large values of \( s \), this can require estimation of a large number of parameters, which may cause problems because of loss of degrees of freedom. Besides, these lagged advertising variables may be correlated with the original variable thus adding to the collinearity in the model.

This has led researchers to look into other methods of modeling the carryover effects of advertising. One of the methods has been to postulate relationships between the different lag parameters in order to reduce the number of parameters in the model. One of the most popular lag structures assumes that the advertising effect is geometrically decreasing over time. This is known as the geometric model or the Koyck model. Other lag structures used in the literature has been the Pascal lag structure where the lags follow a negative binomial distribution and Almon’s method of using a Polynomial lag structure where each lag is a polynomial function of time.

While much of the literature has been devoted to various ways of handling the lags, there have been very few papers that have looked at how an incorrect model specification can affect the estimation of advertising carryover effects. Since real world data is unlikely to adhere to a specific functional form, using any particular function to model the data may lead to misspecification problems. Besides, most of the time media variables do not show much variation within the data. Years of experience enable marketing managers to more or less optimize their media levels and we only observe advertising levels that vary within a very small range. Therefore, for that kind of advertising data, several models can be a good fit. And one can use a number of different functions to model that data. In this paper we use simulation methods to look at the role of model specification in the estimation of advertising lags. We find that when the true model is different from the model that is specified, incorrect advertising lags will often show up as significant in the model. In all of our simulation exercises we assume that the advertising affects Sales in the current period as well as in 3 future periods. Therefore a good model should pick out lags 0 through 3 as significant in the model. We find that two kinds of errors can occur when an incorrect model specification is used. The first type of error is one where some of the lags that are actually significant show up as insignificant in the model. Therefore, one or more of the lags 0 through 3 are left out by the model. The second type of error occurs when far out irrelevant lags are picked out as significant by the model. We propose that using a generalized additive model instead to estimate the underlying relationship will help to obtain better estimates of the number of significant advertising lags. We find that the generalized additive model is also able to determine with a great degree of accuracy the approximate number of periods for which advertising has an effect. Most of the time the GAM procedure will pick out lags 0 to 3 as significant in the simulation exercises. Sometimes GAM also picks out lag 4 as significant but far out lags are never picked out as significant.

FUNCTIONAL FORMS USED IN THE PAPER

In this section we will briefly digress from the main topic in order to discuss the functional forms that we have used in the paper. Most marketing practitioners agree that the relationship between Sales and advertising is nonlinear. In fact, it is well known that while Sales increase with advertising at low levels, at higher levels of advertising the marginal increase in sales begins to decline as advertising increases. In other words, each incremental unit of advertising causes a progressively lesser increase in Sales. This is known as diminishing returns to scale of advertising and results in a concave shape of the Sales response curve to advertising. In order to model diminishing returns, the usual approach is to transform the advertising variable to a nonlinear scale such as a logarithmic transformation. If Sales is modeled as a logarithmic function of advertising as shown in equation (A), the corresponding model is called the semi-logarithmic model and equal increases in advertising leads to progressively lower increases in Sales. Therefore the Sales response curve to advertising has the desired concave shape as shown in figure 1.

\[
S_t = \alpha + \beta \log(a_t) + u_t, \quad \alpha>0, \beta>0
\]

(A)
As we can see from the graph, the semi-logarithmic model always shows diminishing returns to scale as advertising increases. This model however may be deficient for extreme values of advertising. For instance, the model implies that Sales goes to infinity as advertising goes to infinity which is not realistic since we expect Sales to reach a finite asymptote for high levels of advertising, due to saturation effects. Also equation (A) suggests that Sales is negative for low values of advertising. This problem is sometimes taken care of by adding 1 to the advertising variable so that log(a_t+1) equals 0 when a_t = 0.

Apart from the logarithmic transformation, other transformations can also result in concave sales response curves. For instance a square root transformation is represented by equation (B) and results in the graph shown in figure 2.

\[ S_t = \alpha + \beta \sqrt{a_t} + u_t \quad \alpha>0, \beta>0 \]  \hspace{1cm} (B)

This model too, implies that Sales goes to infinity as advertising goes to infinity and therefore is deficient for high values of advertising.

The reciprocal model leads to a graph like the one shown in figure 3 and also demonstrates diminishing returns to scale with increase in advertising. Equation (C) represents the reciprocal model. Sales is assumed to be a linear function of the reciprocal of x. This model implies that Sales reach a finite asymptote as x goes to infinity which is more realistic than the previous models. However, the model does not work well for low levels of advertising, since Sales become negative.

\[ S_t = \alpha + \beta / a_t + u_t \quad \alpha>0, \beta<0 \]  \hspace{1cm} (C)
Rarely in the real world does the relationship between Sales and advertising follow a specific mathematical functional form. Moreover, predictor variables usually do not show much variation in the sample. Therefore we may only observe values of advertising within the range PQ in each of the above figures. Sometimes with only small variation in the sample, several models can be a good fit for the data. We show that when a specification that differs from the true model is used to fit the data, incorrect lags may show up as significant in the model. In these cases, it may be better to use a generalized additive model to fit the data.

For our simulation exercise we use these 3 different types of models one at a time to simulate the ideal relationship between advertising and Sales. For each ideal model, the other 2 alternative models as well as the generalized additive model are fitted to the data to determine whether the correct lags show up as significant terms in the model.

SIMULATION METHOD

We use a dataset that contains media impressions for magazines. The data are simulated to be as close to real world data as possible. Actual magazine accumulation curves were obtained from the MEMRI website and those were used to create a variable with magazine impressions. For the sake of simplification, throughout this paper we will assume that Sales are affected by only one media variable. While in reality several media variables as well as other factors affect Sales, we propose that if incorrect model specification is a problem even in this simple case, it is likely to lead to further inaccuracies in the more complicated scenario. We assume that carryover effects exist so that the media variable influences sales not only in the period in which it is aired but also in future periods. Therefore Sales in any period is determined by the value of the media variable in that period as well as lagged values of the media variable. In this paper we assume that there are 3 significant lags so that in the true model, Sales in the current period is affected by the media variable in the current and 3 preceding periods. The way we conduct our experiment is as follows. We postulate the true relationship between Sales and the media variable by specifying the model and the values of the parameters. Next we use Monte Carlo simulation methods to try to fit several different models (including the true model) to estimate the relationship between Sales and the current and lagged media variables and see which lags come up as significant in the model. For example, suppose we assume that the true relationship between Sales and the media variable (M) can be represented by a semi-log model as follows:

\[ S_t = \alpha + \beta_1 \log(M_t) + \beta_2 \log(M_{t-1}) + \beta_3 \log(M_{t-2}) + \beta_4 \log(M_{t-3}) \]  

(2)

Using the current and lagged values of the media variable that we have in our dataset, and a randomly chosen set of parameters (\( \beta_1, \beta_2, \beta_3, \beta_4 \)), we calculate the value of Sales using equation (2). This is assumed to be the true relationship between Sales and the media variable, M.

Taking this as the true model, we will next try to simulate a bunch of data sets each with different random scatter. In order to do this we first create a new variable stream (called \( \delta \), say) the values of which are chosen from the standard normal distribution with replacement. We next add this new variable, \( \delta \) to our dependent variable, Sales to create a new Sales variable (New_S_{it}).
In most cases, fitting the dataset with the true model results in the best fit and only the correct lagged terms are significant in the model. When trying to fit alternative model specifications however, incorrect lags may show up as significant in the model. Of all the alternative model specifications, we show that the generalized additive model specification works the best in obtaining the correct lags as suggested by equation (2).

Next we use Monte Carlo simulations to determine other plausible values that the dependent variable can take assuming that the true relationship is (2). These are the values of Sales that may be observed in practice when the true sales stream is $S_t$ in equation (2). To obtain these possible values for the Sales stream, we proceed as follows. To each ideal point we add random scatter drawn from a Gaussian distribution with a mean of 0 and SD equal to the value of $S(yx)$ reported from the linear regression of our experimental data. This gives us the probable values that the Sales stream can take when the true values is given by equation (2). We repeat this step 50 times to obtain 50 different data sets each containing different Sales streams. With each dataset and each new Sales stream we try to fit the simulated data with linear regression using different model specifications including the true model specification. For instance since the true model specification is semi-log the simulated data set is fitted using a semi-log model, a reciprocal model, a square root model as well as the generalized additive model.

In the above example, we used the semi-logarithmic model to obtain the ideal data set and then tried to fit other types of models to the simulated data that we derived from this ideal data. We repeat this exercise outlined in the previous paragraph for other types of models as well. More specifically, apart from the semi-log model, the above simulations are also performed using the reciprocal model and the square root model as the ideal models. Therefore, in the second phase of the experiment we use the reciprocal model as the true relationship between Sales and advertising and then try to derive a bunch of simulated data sets from this ideal data set. Next these simulated data sets are fitted with the reciprocal model, the semi-log model, a square root model as well as the generalized additive model. In the third phase of the experiment, we assume that the true relationship between Sales and advertising is represented by the square root model. All of the above steps are then repeated assuming that the square root model is the true model.

In most cases, fitting the dataset with the true model results in the best fit and only the correct lagged terms are found to be significant in the model. When trying to fit alternative model specifications however, incorrect lags show up as significant in the model. Of all the alternative model specifications, we show that the generalized additive model specification works the best in obtaining the correct lags.

Notice that to obtain the ideal relationship between Sales and media variables (and their lagged values) as shown in equation (1) we need to come up with values for the parameters $\beta_1, \beta_2, \beta_3, \beta_4$. This parameter combination is chosen randomly (with certain restrictions) in order to make sure that the choice of parameters does not influence any of the results. In fact, for each model type, 100 different parameter combinations are used to obtain the dependent variable and create 100 ideal data sets. Therefore for each model type, the method of simulating 50 datasets outlined in the previous paragraph was repeated for each of the 100 different parameter combinations.

Therefore, a total of 15000 model simulations were run with 50 simulations for each of 3 model types and 100 parameter combinations.

The GAM procedure was invoked using the following code:

```plaintext
proc gam data = test;
model Y = spline(M) spline(lag1M) spline(lag2M) spline(lag3M) spline(lag4M) spline(lag5M) spline(lag6M) spline(lag7M) spline(lag8M) spline(lag9M) spline(lag10M) spline(lag11M) spline(lag12M) spline(lag13M) spline(lag14M) spline(lag15M) spline(lag16M) spline(lag17M) spline(lag18M) spline(lag19M) spline(lag20M) / dist = normal;
ods output ANODEV = Anodev_out;
run;
```

where lag1M represents the variable that is obtained by taking the the ith lag of M.

**RESULTS**

The tables in this section illustrate the results obtained from the simulation exercises. Tables 1a – 1d shows the results when the true model is semi-logarithmic. Recall that we have assumed that the current media stream as well as the 3 lagged terms are significant in the ideal model.

Table 1a shows a typical result for *one of the parameter combinations* when the true model is semi-logarithmic. If the fitted model is also semi-logarithmic, then almost 100% of the simulated models show the correct lags as sig-
significant. However, if the fitted model is the square root model or reciprocal, some irrelevant lag terms always show up as significant in the model. Note that for this particular parameter combination, when the square root model is used, lags 0 to 3 show up as significant but so does lag 10. When a reciprocal model is used to fit the same data, lag 3 is never picked up by the model. However, when a generalized additive model is used to fit this data, lags 0 – 3 always show up as significant. In this case, lag 4 also shows up as significant in 40% of the model simulations.

![Table 1a](image)

Table 1b summarizes the results for all the parameter combinations when the true model is semi-log and the fitted model is the square root model. Recall that the simulation exercise is repeated for 100 different parameter combinations. The left most column in Table 1b shows the percentage of models for which the corresponding lag shows up as significant. Therefore for each of the 100 parameter combinations, 80% to 100% of the simulated models pick up lag 0 as being significant. The same conclusion is true for lag 1. Lag 2 (as well as lag 3) is picked up as being significant in at least 80% of the model simulations for each of 99 parameter combinations. For one parameter combination however, less than 20% of the models show lag 2 (and lag 3) as being significant. In general when the square root model is used, the relevant lags (lags 0 through 3) show up as significant in the model most of the time. Now if we look at lags 8 -11 we see that for each of 90 parameter combinations, at least 80% of the model simulations pick up one of those far out lags as significant. In fact for 74 of those parameter combinations, at least 90% of the models pick out one of the lags 8 through 11 as significant. Therefore, when the square root model is used the relevant lags almost always show up as significant in the model but so does one of the far out lags from lag 8 to lag 11.

![Table 1b](image)

Table 1c summarizes the results when the true model is semi-log and the fitted model is the reciprocal model.

![Table 1c](image)

In this case, as we can see from the above table, the model sometimes fails to pick up one of the relevant lags, lags 0 – 3. For 22 of the parameter combinations, lag 0 shows up as significant in less than 20% of the model simulations. In fact, for 19 parameter combinations, lag 0 never shows up as significant in any of the model simulations. Similarly for 29, 21 and 28 parameter combinations, lags 1, 2 and 3 never show up as significant in any of the model simulations. In fact, at least one of the relevant lags, lag 0 - lag 3 never gets picked up as significant in
80 of the 100 parameter combinations. As we can see from the results, the reciprocal model does not usually select the far out lags but often fails to pick up the relevant lags in the model.

Table 1d shows the results when a generalized additive model is fitted to the data.

<table>
<thead>
<tr>
<th>% Simulations</th>
<th>lag0</th>
<th>lag1</th>
<th>lag2</th>
<th>lag3</th>
<th>lag4</th>
<th>lag5</th>
<th>lag6</th>
<th>lag7</th>
<th>lag8</th>
<th>dum9</th>
<th>dum10</th>
<th>dum11</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%-20%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>20%-40%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>40%-60%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>60%-80%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>80%-100%</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>90%-100%</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>100%</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>95</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The generalized additive model almost always picks out the relevant lags (lags 0 – 3). 100% of the model simulations pick out lags 0 through 2 for each of the 100 parameter combinations. For 95 of the parameter combinations, lag3 is also picked out as significant in 100% of the model simulations. In this sense the generalized additive model performs better than the reciprocal model which often leaves out some of the relevant lags. As we see from the above table, the GAM will occasionally pick out lag4 as significant in the model. This does not happen very often and in only 17 of the 100 parameter combinations does lag4 show up as significant in at least 80% of the model simulations. The Generalized additive model also never picks out the far out lags (lags 8 – 11) and therefore performs better than the Square Root model. So GAM does a little better than the other types of models in the sense that the relevant lags are never left out and far out lags are never included in the model. The flexibility of the functional form allowed by the GAM procedure helps to pick out the correct lags in the model.

The results are very similar when the true model is reciprocal. In this case, when the fitted model is square root or semi-log, the relevant lags are always picked out by the model. However both the square root model and the semi-log model also frequently pick out some far out irrelevant lags (usually lag 10) as significant. As before, we also try to fit a generalized additive model to the data. Like the semi-log and the square root model, the generalized additive model also performs very well when it comes to selecting the relevant lags, lags 0 to 3. Unlike the other two models however, the GAM does not pick out any of the far out lags. However, it frequently also picks out lag 4 as significant. In that sense, GAM performs somewhat better than the other two models since it never picks out a far out lag and instead picks a lag that is close to the relevant lag terms in the model.

Next we repeat the same exercise when the true model is assumed to be square root. Therefore, the relationship between Sales and advertising is assumed to follow a square root function. When a semi-log model is fitted to the data, it picks up the relevant lags most of the time and sometimes also selects lag 4 as significant in the model. Therefore when the true model is square root, the semi-log model really does an excellent job of fitting the data. When a reciprocal model is fitted to the data, we find that for most of the parameter combinations, at least one of the relevant lags 0 to 3 is missing from the final model. Therefore the reciprocal model once again leaves the relevant lags out of the model. When we try to fit a generalized additive model to the same data, we find that GAM performs slightly better than the semi-log model in picking up the relevant lags. However, GAM also selects some of the irrelevant lags (lag4 ) more frequently than the semi-log model. Therefore in that sense, GAM performs slightly better than the semi-log model in selecting the relevant lags but slightly worse when choosing the irrelevant lags.

Overall, if we take into account the different model specifications, we find that in most cases GAM works better and almost always picks out the significant lags in the model and never chooses the far out lags. Therefore, in the real world when we do not know the true model specification, using a GAM analysis would help to get one as close to the true model as possible.

CONCLUSION

In this paper, we look at carryover effects of advertising. Marketers have long known that advertising has delayed effects. However, accurately measuring the long term effects of advertising campaigns has always been a challenge faced by practitioners in the industry. In this paper, we show that model specification may play an important role in determining which lags show up as significant in the model. We propose that using a generalized additive model (proc GAM in SAS) instead of a specific functional form may help to more accurately identify the significant lags in the model. Since the true relationship between Sales and advertising, rarely follows a precise functional form, using an explicit function to model the relationship may lead to incorrect estimation of the lagged effects of
advertising. A Generalized Additive Model allows greater flexibility of the functional form and helps to get more accurate results. As a side note, we would like to point out that in this paper we have used a very simplistic model to show the accuracy of GAM vis a vis other functional forms. More research is needed to investigate how well GAM performs when we use a more complicated model.

REFERENCES

Dyson, Cutting adspend in a recession delays recovery, WARC Online, March 2008
L Lodish & B Lubetkin, General truths? Nine key findings from IRI test data, Admap, February 1992

ACKNOWLEDGMENTS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Patralekha Bhattacharya
Thinkalytics
Email: pbhattacharya@thinkalytics.com
Web: www.thinkalytics.com

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *