A COST EFFICIENT SCALABLE STRATEGY FOR ESTIMATION OF ADVERTISING ROI

Max Kilger and Ellen Romer, Experian Marketing Services

Introduction

One of the metrics of advertising that continues to be an elusive but "front and center focus" for advertising agencies, advertisers, content providers and market researchers is the return on investment (ROI) for the large amounts of financial capital dedicated to reaching and messaging customers and prospects.

Typically, ROI has often been difficult to measure given the large number of complex issues involved, including but not limited to the often lengthy path-to-purchase, struggles to disentangle the effects of advertising through multiple, often competing media channels as well as the ability to directly attribute the effects of advertising to actual purchases of products and services.

One of the traditional ways in which advertising ROI has been measured is the utilization of tracking studies for specific advertising campaigns and brands. Typically these studies measure characteristics such as unaided and aided brand awareness, specific advertising recall and resultant consumer behaviors such as purchase of the advertised product or service across pre-, post- and concurrent time periods relevant to the specific advertising campaign under evaluation[1]. Often information about competitive products and services is gathered at the same time, along with a set of key demographics and attitudinal statements.

Usually these tracking studies suffer from several key disadvantages. The first involves the financial cost of a tracking study, which is often significant and therefore must be conformed to budgetary constraints within the commissioning organization. Secondly, tracking studies collect primary data and thus often for budgetary reasons collect data on a modest number of measures necessary for an elementary evaluation of the effectiveness of advertising efforts. Thirdly, tracking studies often have limited data collection time frames during which time data is collected, analyzed and reported. Subsequent advertising campaigns that require ROI tracking generally require additional tracking studies, which in turn adds to the mounting expenditures for ROI measurement.

If there were a less expensive alternative to tracking studies for measuring advertising ROI then the various parties involved would all benefit. One of the alternatives that we considered was the adaptation of a large, syndicated consumer study to provide a basic measure of ROI for advertising campaigns.

It should be noted that the application of an existent, syndicated data collection instrument precludes the collection of some of the ROI-associated measures found in tracking studies such as brand awareness and advertising recall. However, the advantage is that the utilization of a syndicated study by its very nature implies that there would be significant financial savings involved both because the cost of the syndicated study is generally spread across many different subscribers to the study.

Further, particularly when considering tracking multiple advertising campaigns over an extended period of time, syndicated studies are ongoing concerns that collect data over an extended period of time, which would allow multiple campaigns to be evaluated for the same subscription price. Finally, utilizing a syndicated study for advertising ROI purposes also provides clients with a exponentially larger number of variables to utilize in gaining a better understanding of the relationship of the campaign to consumer purchases.

Unfortunately there is no such thing as a "free lunch". There are significant methodological, statistical and client-facing challenges involved in attempting to utilize a large scale syndicated study for the purposes of advertising ROI. This paper documents a number of these challenges and issues along with describing how each one of them was met. The efforts necessary to undertake this initiative were not insignificant and the decision to accept this challenge was not lightly taken. However, as we will see in subsequent sections, the obstacles are not insurmountable and the result of the initiative can be labeled a reasonable success.

1 For a basic description of advertising tracking studies see Feldwick, 1998.

Description of the Large Syndicated Study

The large syndicated data set that was selected for adaptation to a time series format was Experian Simmons' National Consumer Study (NCS). The NCS is a national probability sample of approximately 25,000 respondents in the contiguous United States who do not live in groups quarters and who speak either English or Spanish. There is a significant Hispanic oversample in the study design, as well as an income stratification schema and geographic oversampling in certain large metropolitan areas. The study itself has over 60,000 consumer variables that cover a variety of topics including but not limited to media usage, product category and brand consumption and a large battery of attitudinal questions. The NCS is currently reported in quarterly releases.

Necessary Conditions for Adaptation

There were a number of conditions necessary for the successful adaptation of the National Consumer Study into an ROI-oriented tool. Without being able to satisfy each of these conditions it would not be possible to adapt this traditional syndicated study into a time series data set appropriate for trending and ROI analyses.

Continuous Measurement

While traditionally some syndicated studies have field periods that are scheduled at certain time periods throughout the year, it is important in the case of converting syndicated studies to a time series structure that data is collected continuously throughout the entire year, rather than in discrete, isolated field periods. This avoids having data collection "gaps" that might appear in the middle of advertising campaigns and otherwise disrupt trending or evaluation efforts. This also ensures that traditional holiday periods such as Christmas, Fourth of July and other holiday periods are equally represented in the data. Fortunately, the National Consumer Study went to a continuous measurement sample design in 2007 and so every day of the year is covered during data collection.

Equivalency of In-Field and In-Pool Samples

Traditionally, sample designs and operational procedures for large syndicated studies rely upon strategies such as using sample replicates to ensure that respondents are randomly selected and contacted throughout the field period. Sample replicates ensure that the characteristics of the sample in the field are statistically equivalent to those still awaiting assignment in the sample pool. The practice of making sure that each sample replicate is exhausted before opening a new sample replicated helps ensure that every element in the sample frame has a non-zero chance of being selected into the sample.

The equivalency of in-field and in-pool sample pools becomes even more important when constructing time series related data sets because rather than just experiencing the effects of potential non-coverage of some sample elements, there is the opportunity for *respondent selection by time period interactions* to occur. For example, if sample replicates are not utilized and regularly exhausted, there may be some characteristic of the respondent or the selection process that is associated with the respondent currently residing in the in-field sample or the in-pool sample waiting to be assigned. If this respondent characteristic or process element is also correlated with a specific season, holiday period or real world event then the sample for any given time period may not be equivalent to those at other time periods. The equivalency of sample in the field with that still remaining in the sample pool, whether maintained by sample replicates or some other mechanism, is an important component for the adaptation of syndicated studies for use in time series ROI evaluations.

Assignment of Calendar Date to Completion of Survey

Sometimes it is the case in syndicated surveys that the actual data of the completion of the survey is not known. It is important for the purposes of tying the data collected to a calendar date that this information or a reasonable estimate is available to assign to the data collected in the survey instrument. In the case of the National Consumer Study this is even more challenging because the size of the survey – it's an approximately 130 page survey booklet – means that it is likely that they respondent fills out the survey instrument over some period of time.

Given this challenge, it was necessary to find the best estimate of the date that should be assigned to the data collected for a survey booklet in the NCS. A study was undertaken to enumerate the lifecycle of a National Consumer Study booklet from receipt of the booklet by the respondent to receipt of the completed booklet by the mail center. This study revealed, taking into account average mail delivery schedules for the mailing and return of the booklet, that the mean time to completion was 8.4 days and that 86% of the respondents filled out the booklet within 14 days. Given that our selection of basic unit of time for the time series data set was one week, we set the date stamp for a particular booklet to be 2 weeks prior to the day it was received at our processing facilities. This was felt to provide the best estimate of the week in which the booklet was filled out.

Sample Size

It turns out that the determination of the basic unit of time to be used in the time series adaptation relied heavily upon the average size of the sample that would be present in that time period. Many of the decisions, challenges and strategies involved in the adaptation of a syndicated study to a time series architecture require serious considerations of the consequences of the size of the sample in the basic fundamental time unit.

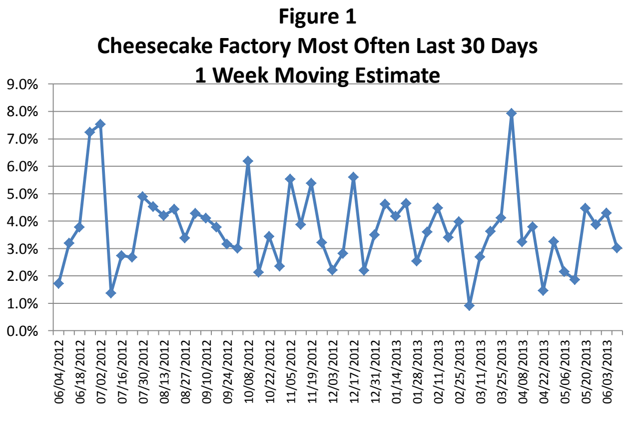
In fact, it is often the case that sample size is cited as the main objection why data sets like syndicated data studies cannot be adapted to time series. The reduction in sample size per time unit that occurs when an aggregated data set collected over a period of time and analyzed as a whole is suddenly "cut up" and distributed across a discrete set of time periods often discourages even the bravest researcher from attempting this objective. For example, a sample size of 2,000 respondents spread out over 365 days equals about 5.5 respondents per day or a little over 38 respondents per week.

The obvious concern here was the size of the sampling error encountered in dealing with these sample sizes. The sampling error present in these samples was simple too large to be of practical use. The National Consumer Study maintains an intab sample of approximately 25,000 respondents per year. The end result of discussions was that one calendar week was chosen as the basic fundamental time unit after 30 days was rejected as not granular enough for ROI tracking purposes and one day was rejected for insufficient sample size. Each week of the National Consumer Study then averages out to have about 480 respondents. Even this decision to use one week as the basic time unit presented significant statistical and user challenges that will be discussed in the following section.

Estimation Challenges

The Development of a Useful Type of Estimate

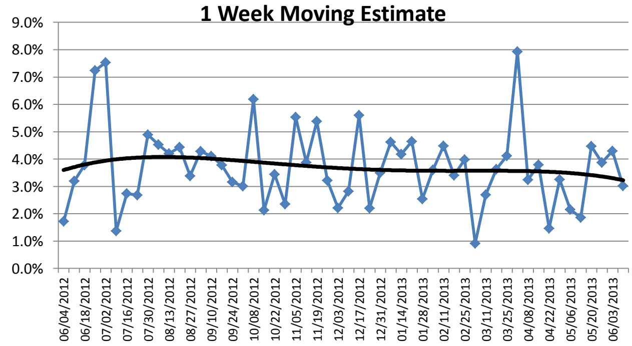
One of the key challenges that emerged from the previous section was the issue of sample size, its corresponding sampling error and the relevant difficulties that may result from adapting a syndicated national study to a time series architecture. For example, the maximum standard error for a sample proportion with a sample of size 480 is approximately 4.5%, which at a 50% incidence rate is high but tolerable. More problematic however, is when incidence rates are small. A typical sampling error for p=.1 and n=480 is 2.7% - while less than the maximum error previously computed - is now a very large percentage of the estimate. Given that the incidence of many of the consumer variables in many syndicated studies such as the National Consumer Study can be modest, this presented some immediate problems. Figure 1 below illustrates a typical trend for a weekly estimate over a short period of time and visually confirms some of the immediate issues with sampling error.



Note that there is a non-trivial amount of apparent variation in the estimate from week to week. We know that some of this is noise due to sampling error and that some is due to true market movement during the week. The issue of course is that it is difficult to determine what portion of the variation is due to what source.

There are several ways to improve the interpretation picture here, depending upon the objective of the analyst. One of the more obvious strategies is to draw a trend line through the series of data points. If the objective is to show a trend over time then this is a helpful maneuver. Figure 2 below shows the same chart with a simple linear trend line constructed through the data series. The trend line provides the researcher with a data visualization that is more interpretable and meaningful.

Figure 2
Cheesecake Factory Most Often Last 30 Days
1 Week Moving Estimate



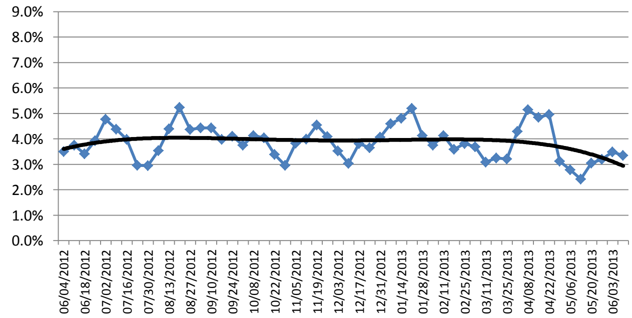
A second strategy that we deployed in this project was to develop an estimate that would provide users with a more useful and interpretable visual representation of the data in addition to the use of a trend line. Following along the lines of the traditional idea of a simple moving average we constructed a similar estimate that we label a "moving estimate", so as not to confuse it with simple moving average as a term of statistical art.

Both simple moving average and moving estimate rely upon the mechanism of pooling data across time periods and averaging them in some manner in order to provide an estimate for a particular point in time. One of the consequences of this averaging process is that random errors around the data points tend to cancel each other out, thus resulting in a smoothing effect on the data series.

A moving estimate can be distinguished from simple moving average in that the raw data points in each moving estimate are separately projected to a population apart from other moving estimates. This means that each data point represented in the graph will have a different set of weights depending upon the time period encompassing the estimate as well as the position of that estimate within the time frame. More about this weighting schema will be discussed in the next section.

Figure 3 below illustrates the effects of on the same data presented in the previous figure utilizing a 4 week moving estimate process on the data points within the graph. This estimate pools data points from 4 consecutive weeks of data. The most obvious effect is that like a simple moving average, variance present in the original data points is attenuated. This results in a much less chaotic and more interpretable visual representation of the data. The trend line present in the chart also adds to the user's ability to comprehend the meaning of the data.

Figure 3
Cheesecake Factory Most Often Last 30 Days
4 Week Moving Estimate



If one enlarges the time period for the moving average then the result is even more smoothing of the data. This is useful in the case where the data series is even more chaotic than expected. The approach utilized by the development team was to offer - in addition to 1 week estimates - 4 week, 8 week and 12 week moving estimates so that the appropriate amount of data smoothing would be available for a wide range of data series.

Of course there is no such thing as a "free lunch" and so there are some costs to deploying the moving estimate schema that are similar to those incurred with the use of simple moving averages, including the fact that non-random sources of variation – such as true market movement – also get attenuated. These will be discussed in the second part of the next section on weighting.

Weighting Issues

There are two quite different types of weighting issues that need to be discussed in this section. The first deals with weighting or projecting the estimates in the data series from their initial raw data counts to estimates for the U.S. adult population. The second discussion revolves around some of the issues involving the effects of moving average-like estimators on data series and how weighting has been utilized to compensate for those issues. Finally, some logistics about the weighting process itself are discussed.

Weighting for Purposes of Projecting to a Population

Each of the raw data points present in the time series data set represents an unweighted estimate of the variable in question. Taking the 1 week estimate as the simplest case, there are approximately 450 cases each week that need to be weighted and projected to national estimates. Every week of data that appears in the time series adaptation of the syndicated study needs to have a weight associated with it. Given the small sample size, rather large weights are necessary to project the raw data points into an estimate for total adults in the U.S. Thus, in addition to the larger native sampling error attributable to smaller raw sample sizes, additional variance is injected into the estimate during the processes of post-stratification weighting. Given this situation, it is not surprising to see significant variation in week to week estimates like those found in the 1 week estimate charts in Figures 1 and 2.

As was previously mentioned, one of the ways in which some of this variation can be shed is through processes like simple moving averages or in our case moving estimates. For example, in Figure 3 above, each of the data points in the chart consists of the data from that specific week as well as all the data from the previous three weeks, which results in an estimate with an average sample size of about 1,920 respondents. The reduction in variation seen in the estimates from week to week in Figure 3 are due in part to the larger sample size that results from pooling the data as well as the cancellation of some of the variation due to the random error component that would be found in the four separate estimates as one would expect from a simple moving average-like process. Also note that the larger sample allows smaller post-stratification weights to be attached to each observation, thus reducing the increase in variance due to weighting of small cells among the various demographic classes utilized.

One of the complications of constructing multiple week moving estimates is that multiple weights must be constructed for each set of weekly data. That is, the raw data for a single week during the calendar year will be counted in creating a number of different estimates, depending upon the position of the week within the time period selected and the nature of the length of the moving estimate.

In the case of a 4 week moving estimate, a single week's worth of data may reside at time t_1, t_2, t_3 or t_4 within a 4 week time period. For example, the 4 week moving estimate for 6/25/2012 in Figure 3 consists of data from 6/4/2012, 6/11/2012, 6/18/2012 and of course 6/25/2012. In this case the data for the week of 6/25 resides in the t_4 position.

The 4 week moving estimate for the following week of 7/2/2012 consists of data from 6/11/2012, 6/18/2012, 6/25/2012 and the aforementioned 7/2/2012. In this case the weekly data for 6/25 resides in the t_3 position. Repeating this logic for the two weeks subsequent to 7/2/2012, weekly data, the data collected during the week of 6/25 is a component of four different 4 week moving estimates, depending upon its position within the temporal window (time t_1, t_2, t_3 or t_4). Thus for any of the 4 week moving estimates within the 6/25 to 7/16 time frame, the weekly data from 6/25 is utilized a total of four times - once for each of the four estimates that are generated within the time frame.

Continuing to use this example, the four 4 week moving estimates available in the time frame 6/25 to 7/16 must each be individually weighted and projected to the U.S. population. Because each of these four estimates contains a somewhat different set of data, four independent weighting runs must be made – one for each estimate. This means that the data collected during the week of 6/25 for example, will have four different weights linked to it – one for each of the four estimates that are during the time frame that are associated with its position within the time frame - position t_1, t_2, t_3 or t_4.

Following this logic out for all of the possible moving estimates, for any arbitrary week of data collected in the time series data set, there will be 1 weight for the 1 week moving estimate, 4 weights for the associated 4 week moving estimate, 8 weights for the 8 week moving estimate and 12 weights for the 12 week moving estimate. The result is 25 total weights for any single weekly set of data points. The weight that is used depends upon the length of the moving estimate it is being incorporated into (1, 4, 8 or 12 week) as well as the position - t_1, t_2, t_3 or t_4 - within the moving estimate time frame.

Issues in Utilizing Moving Average-like Processes

As was mentioned briefly in the previous section, there is no "free lunch" when it comes to the application of moving average-like algorithms. The statistical literature outlines a number of issues in utilizing moving average types of algorithms[2]. Most commonly, a number of these issues are addressed by unequal weighting of data elements. That is, in simple moving averages each of the data points - regardless of when they occurred in time – are equally weighted. One consequence of this is that more recent data has the same level of influence on the estimate as less recent data. Another consequence of equal weighting in simple moving average processes is that it tends to obscure cyclical patterns such as seasonality within the time frame of the moving average estimate.

One way to emphasize more recent data is to apply simple exponential smoothing to the moving average process where the weights for each time period exponentially decrease from the most recent time period to the oldest time period. This places more emphasis on more current data and less on less recent data points. This strategy is usually deployed in cases where the data series is stationary.

A more advanced variation of exponential smoothing is the popular Holt-Winter method of exponential smoothing. In this case three different formulas are applied that deal with the level of the data, the trend component and seasonality components present in the data series (see Shumway and Stoffer, 2010 for background on the Holt-Winter approach). This method can be used where the data series is not stationary and there is evidence of cyclical or seasonal components.

The current algorithms utilized in developing moving estimates within the time series data utilize a straight-forward simple unweighted averaging presently. Extensive testing of numerous data series within the National Consumer Study found that the short time period nature of the moving estimates being used resulted in few seasonal or cyclical patterns that might be adversely affected by not utilizing a more sophisticated strategy. It was also found that trends in the data, while attenuated to some degree, were for the most part preserved in these short time periods for which estimates were being made. It is anticipated that some form of the Holt-Winter method of exponential smoothing will be deployed in the next major software release of the time series DataStream product after extensive testing to evaluate differences between the moving estimates generated by the original and proposed Holt-Winter algorithms.

Logistical Considerations for the Weighting Processes

One of the changes in logistical structures that comes with the adaptation of a syndicated study into a time series architecture is that the weighting cycle, which often can take a number of days or weeks to perform and evaluate for a syndicated study, has only a day or so for processing in a time series data set that is updated weekly. In addition to this constraint, there are a much larger number of weights that must be calculated for each weekly raw data sample - 25 weights in all per week as was seen in the previous discussion. Weighting processes are often assembled, executed and evaluated by hand and so this becomes more problematic in our case.

We designed an automated, cascading weighting process to deal with the weighting logistics present in the adaptation. There are three potentially cascaded weighting stages that are utilized in the weighting process. Each stage has its own separate weighting model. The primary model contains most of the weighting variables that are utilized in the syndicated product. The secondary model contains a subset of the primary model variables and the tertiary or "last chance" weighting model contains an even small subset of weighting variables found in the syndicated study.

The process begins with a traditional raking weighting strategy using the variables present in the primary weighting model. Once the data has been weighted, a set of 10 "canary" variables – variables known to be correlated to many other variables and sensitive to weighting changes – are examined and compared to previous recent historical data for those variables. Each variable gets a vote in allowing the weighting model to stand. A positive vote is generated if the current week's variable is within certain bounds of the recent historical figure for that variable. It takes 7 or more positive votes for a model to pass – there are no penalties if a variable does not pass the evaluation.

If the primary weighting model receives 7 or more votes, then the weights are retained and processing continues. If the primary stage fails to gather the necessary votes, then the secondary weighting model is applied and the voting process starts over again. If the secondary weighing model gains 7 or more positive votes, then those weights generated are retained. If the secondary weighting model fails, then the fallback is the tertiary or "weighting model of last resort" which contains the most parsimonious weighting schema. This weighting schema is then applied and the positive votes counted. If the tertiary model can obtain 7 or more positive votes then the weights are retained. If the tertiary model fails, then the automated process stops, a message is generated and human intervention is required.

[2] For an introduction to moving averages and time series analyses in general see Shumway and Stoffer, 2010.

Interpretation and Client Education Challenges

One of the not so obvious revelations to come out of the adaptation of a syndicated study into a time series architecture and eventual product were some of the initial difficulties in interpretation of time series data, especially in its raw visual form. It was also a challenge to encourage users to focus upon looking at data within a time series perspective and it took both significant client service and client education to assist users in understanding data from this perspective.

One of the key origins of these difficulties is the effects that larger sample variances due to smaller sample sizes will have on the visual representation of estimates throughout time. The peaks and valleys that are seen in the visual representation of the data series contain variation due to both true market movement as well as non-trivial levels of sampling error and other error sources. The resultant display is often visually distracting and hinders easy interpretation of the data.

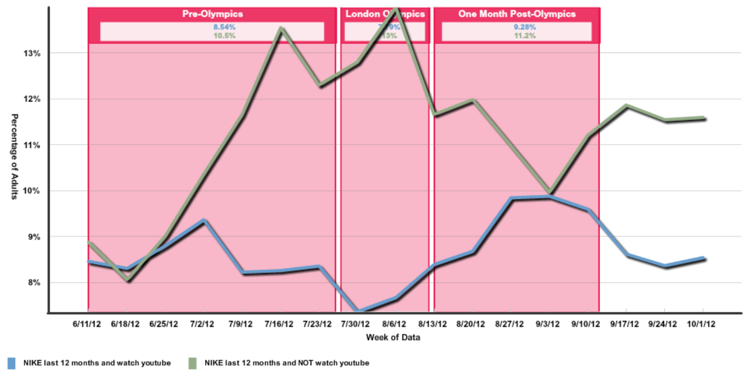
Additionally, one of the initial struggles for potential users was drawing their focus away from single data points – single peaks and valleys - and to instead concentrate on trends over time. It seemed almost second nature for a number of early project testers to want to focus on and interpret individual data points. Given the noisy nature of the time series data, this was a tendency that we wanted to extinguish in users.

Trend lines and the smoothing effects of averaging processes such as the ones used in the moving estimate metric are ways in which we worked to focus the user’s attention away from single data point initiated peaks and valleys and towards longer term trend indicators. These strategies are useful in reducing difficulties in interpretation and avoiding misattribution of the data. However, even these tactics are not without challenges to the end user or analyst. For example, when choosing trend lines, one has the option of linear or polynomial based trend lines. Within polynomial trend line deployment there are choices in terms of the polynomial function to be used in forming the trend line. The selection of the correct form of trend line is situationally dependent and so it was difficult in many cases to form strict rules about their use.

There were similar questions posed by users when it came to choosing the best moving estimate metric for the data – either a 1, 4, 8 or 12 week moving estimate. Determining the most useful version of the metric – especially for novice users – sometimes depended upon trial and error among the versions along with a visual inspection of each trial. It took some experience with the time series data in this initiative before users were able to select the best choice on the first try. Failing to choose the best moving estimate would result either in a data series that was still characterized by excessive variation in the chart or would result in a data series that too severely smoothed out the variation in the data series, resulting in excessive attenuation of what was likely true market movement or evidence of the effects of real world events on the variable in question. In the end, it turned out that the mode for the choice of best estimate was the 4 week moving estimate. It most often modestly attenuated likely random error while for the most part preserving likely market movement within the data series.

Finally, the addition of the ability to partition the data series into discrete segments also was of significant assistance in interpreting the data. The natural comparison of consumer behaviors for advertising return on investment often involves dividing data into pre-, post and campaign time windows. Thus data present in each of these three relevant segments can be summarized to assist in evaluating metrics such as ROI. In Figure 4 below, you can see three discrete time segments before,

Figure 4
Nike Jogger Olympic YouTube Video



during and after the 2012 London Olympics. In addition to being able to evaluate the data series lines themselves across the three segments, the average incident rates for each of the three time segments for both data series can be seen above the series, assisting in the assessment of the data.

Figure 4 also illustrates a key point in ROI evaluation - the presence of the olive colored data series which represents a baseline during the time periods under observation. Without a baseline or control group, upward or downward trends that coincide with advertising campaigns could be due to confounding factors other than the campaign itself, such as advertising or sales campaigns of competitors or even real world events.

This specific case represents the attempt to ascertain if the release of a Nike advertising video on YouTube – in this instance 12-year-old overweight jogger Nathan Sorrell – had an effect on Nike sales. The olive line represents the baseline of individuals who purchased Nike athletic shoes during the last 12 months but did not watch YouTube in the last 30 days. The blue line represents individuals who purchased Nike athletic shoes during the last 12 months and did watch YouTube in the last 30 days. Nathan's video was released on YouTube on or about July 31, 2012. You can see that shortly after the release of the video on YouTube, those individuals who reported purchasing Nike athletic shoes during the last 12 months began to climb, while the baseline of those who purchased Nikes but did not watch YouTube in the last 30 days peaks during the first week in August and then declines during the rest of August. Indeed, as the effects of television advertising for Nike begin to fade after the Olympics, the persistence of the YouTube link appears to be paying off. This suggests that the Nike YouTube campaign may have had an effect on people who had the opportunity to see the video and it paying dividends as more and more people visit the video on the YouTube website.

This example brings up other methodological issues that also need to be addressed in the adaptation of a large syndicated study into a time series architecture. The first is the fact that unlike tracking studies, the measures that are utilized in the ROI evaluation are often already fixed in place and form long before the advertising campaign is deployed. That is, unlike a tracking study that might ask if the respondent purchased Nike shoes in the last 30 days, the researcher must utilize the measure and time frame that is already in place in the syndicated study. Because the measure asks about purchase behavior over the last 12 months, any recent purchase behavior will be contaminated to some extent by purchase behavior that occurred during the previous 12 months. Thus, any market movement that occurs during the time window under observation – in this case just prior to just post Olympics – will also be influenced by prior behaviors in the consumer marketplace. However, ceteris paribus, if the Nike video on YouTube had an influence on Nike sales, we should see some indicator of that – which we do.

Also note that we are using the reported behavior of using YouTube as a proxy for having the opportunity to see the Nike video. Unlike a tracking study we cannot ask respondents whether or not they watched the Nike video on YouTube during the time period under observation. Thus there are undoubtedly people who went to the YouTube website but did not watch the Nike video. These are some of the tradeoffs that are made when utilizing a syndicated study as an ROI evaluation tool rather than committing to a custom tracking study. The benefits of utilizing an adaptation of a syndicated research study to cost effectively conduct repeated advertising campaign evaluations and evaluate them on a large number of subgroups and descriptors has to be weighed against some of the methodological issues mentioned above. Evaluating advertising campaigns often means utilizing "data in the wild" and under this kind of environment there will always be some tradeoffs in the process. In the end, the consensus of the research and development team was that the cost/benefits ratio of adapting a syndicated study into a time series architecture was well worth the tradeoffs involved.

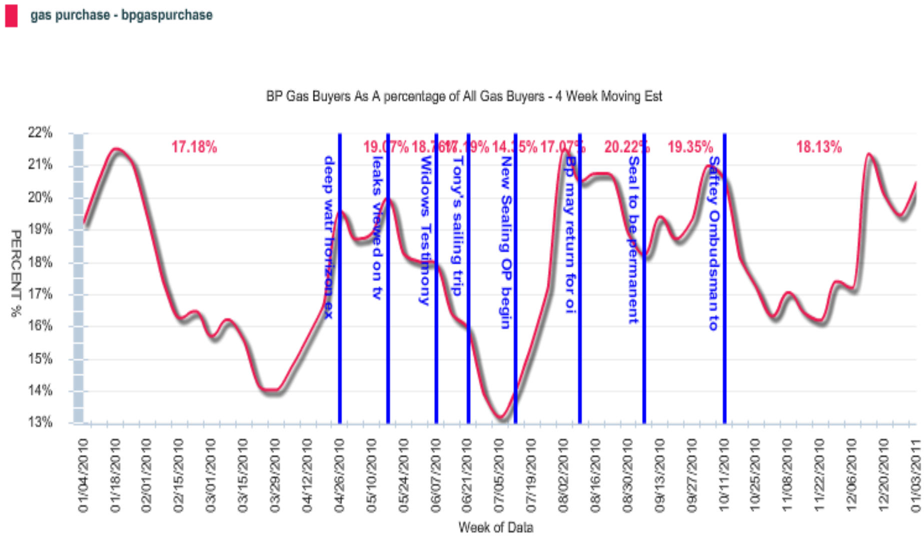
Validation Efforts and Some ROI Examples

During the early days of the adaptation of the National Consumer Study to a time series format, there were efforts made to develop some validation evidence that the adaptation worked, as well as some early collaborative efforts among client "friendlies" to attempt to provide further evidence of the ROI utility of the initiative. One of the first areas that we focused upon was the ability of the newly adapted time series data to reflect real world events. Unlike the example of advertising ROI, where mentioned above there were was some ambiguity about exposure to advertising campaigns, exposure to national level real world events was something that was easier to assume that all respondents were experiencing.

There were two efforts undertaken to determine if the adapted time series data would reflect real world events in the U.S. The first was an effort to link attitudes towards consumer confidence to national events such as changes in government policy or large scale economic events such as major banking system failures and major financial firm closures during the course of the 2008 economic crisis (Kilger and Palit, 2011). The second effort involved tracking consumer responses to the April 2010 British Petroleum oil spill.

One of the questions on the minds of marketers, especially those representing the oil and gas industry and in particular British Petroleum, was the extent to which the 2010 BP oil spill negatively affected consumer sales of their products. Palit and Kilger (2011) outline in their paper an investigation of how events that occurred during the timeline of that disaster

Figure 5
Consumer Reaction to BP Oil Spill



likely affected the consumer sales of British Petroleum gasoline. Figure 5 reveals that here were eight significant events during the crisis that were mapped onto the week by week BP share of sales of gasoline reported being purchased by consumers[3].

On April 20, 2010 an explosion on the drilling rig Deepwater Horizon killed 11 people and injured a number of other oil platform workers. The even worse news was the fact that crude oil was now spilling into the Gulf of Mexico and might be doing extensive environmental damage. As can be seen in Figure 5 there seems to be only a small decline in BP gasoline purchases by U.S. adults as little is known about the crisis as of yet. After the first week in May, comprehensive television coverage of the true extent of the spill and the environmental damage it is causing is given extensive coverage on the nightly network news programs. Consumers react a bit more by purchasing less BP gasoline.

Television coverage continues to play a part in the drama as a couple of weeks later network news programs interview some of the widows of the crew members that were killed in the initial explosion. There is nothing like widows on television to put a dent in things, as can be seen as the share of BP gasoline in the marketplace continues to decline. On or about June 19th, network television news reports that BP chief Tony Hayward has been relaxing and sailing his personal yacht while the crisis continues to worsen. According to the graph, the share of BP gasoline purchased by the American public continues to decrease.

On or about July 11, British Petroleum begins new, more vigorous efforts to cap or seal the leaking oil well. Efforts begin to pay off and are reported in the news and BP's share of the U.S. gasoline market begins to climb again. It reaches a peak on or around August 2, 2012 when the "static kill" procedure successfully completely caps the leaking oil well. On August 6th, BP announces that they might return to the Deepwater Horizon well and reopen it for resumption of oil drilling. Consumers react and the share of BP gasoline sales in the U.S. declines once again.

During the first week in September, 2012 British Petroleum announces it will permanently shut the Deepwater Horizon well and as Figure 5 reveals, BP's share of U.S. gasoline sales again begin to climb. This climb continues until the day that the ombudsman for British Petroleum announces that they are discontinuing accepting claims for damage and compensation linked to the Gulf oil spill, where upon BP gasoline sales again take a decline, signaling the American consumers displeasure at British Petroleum's decision to seemingly abandon those people affected in the Gulf region.

Given the methodological challenges previously outlined for adapting a large syndicated consume study to a time series format, the development team felt that the BP oil spill analysis was a good indicator of several key validation points. The first validation point was that the converted time series data appeared to be sensitive enough to pick up shifts in the consumer marketplace. This was in spite of some the previously discussed issues involving sample size, larger variation due to weighting processes and smoothing effects that might threaten to smooth out any marketplace originating movement.

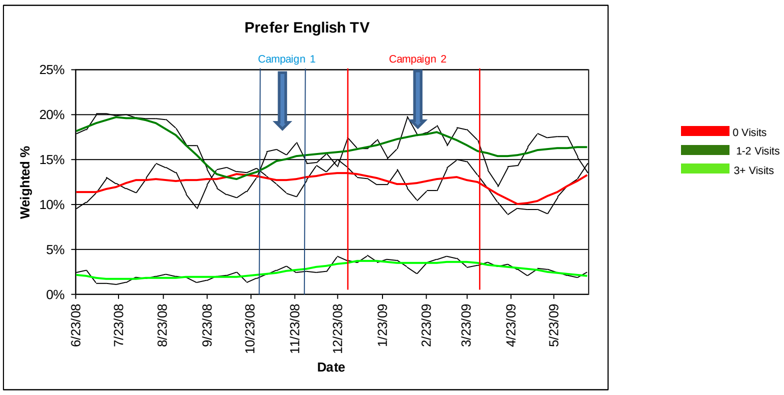
[3] Note here that a baseline is not necessary for this comparison, given the assumption that all U.S. adults were exposed to this national level event.

The second key validation point was that despite the heterogeneity of the time frames involved in booklet date completion estimation, the fact that survey measure time frames were larger than desired and other timing factors, the fact remained that the time series data appeared to do a quite respectable job of capturing the effects of real world events on a temporally accurate basis.

Another early validation effort involved the cooperation of a Spanish language media client and access to advertising schedules for a specific advertising campaign involving a national restaurant chain that was advertising with them at the time. We examined the number of visits to this restaurant chain over a period of time that included a pre-campaign period, two different advertising campaigns executed within a short period of time of each other and post-campaign time periods.

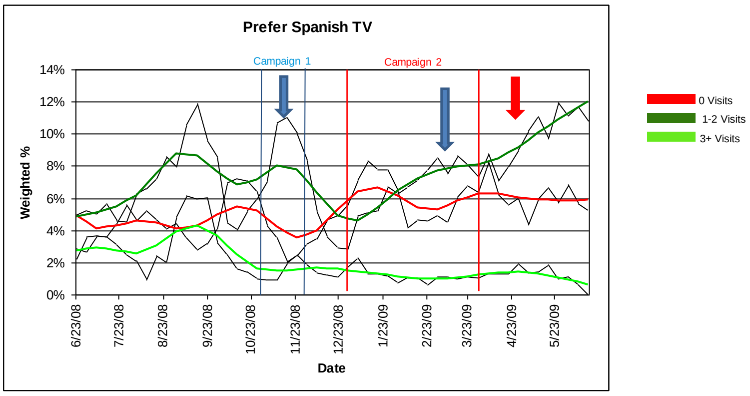
Figure 6 plots the data series for volumetric categories of visits to the restaurant within the last 30 days. Zero visits meant that the respondent considers themselves a patron of the restaurant but had not visited it within the last 30 days. The characteristics of the group in the chart are that they are Hispanic adults who visit the restaurant and who mostly/only prefer to watch English language television. This chart will serve as a reasonable baseline for our comparison.

Figure 6
Hispanic Adults Who Prefer Watch Mostly/Only English Language TV



As can be seen from the figure, there are two advertising campaigns on this Spanish language media provider. The first one is of rather short duration, followed by a short fallow period and then a second, longer advertising campaign is executed. The colored lines are polynomial trend lines for each of the three data series. If you examine the trend lines you will notice that there are some modest gains for the 1-2 visit and 3+ visit data series which may be attributed to the “mostly” part of the nature of the variable. This is to be expected because the target group in Figure 6 prefer to watch television in English while the advertisements appeared on a Spanish language television network.

Figure 7
Hispanic Adults Who Prefer Watch Mostly/Only Spanish Language TV



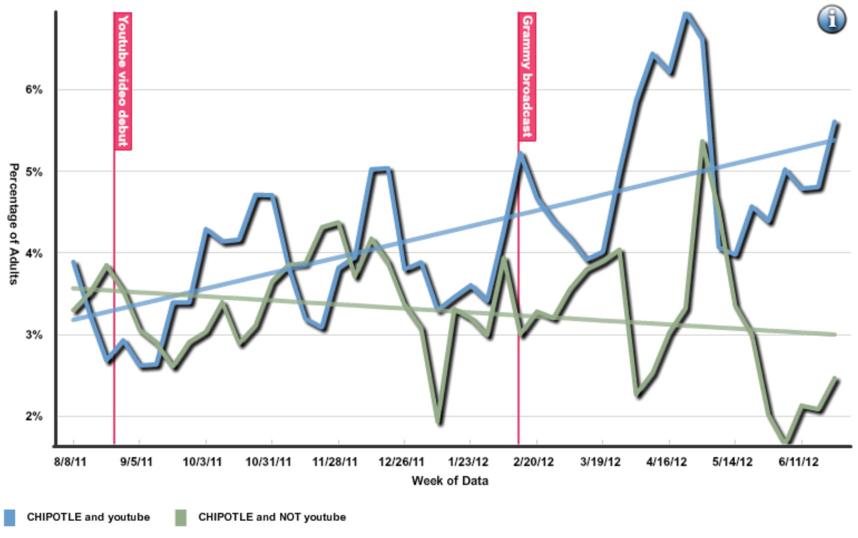
Now compare this to the trend and data series lines in Figure 7 where we have adult Hispanics who prefer to mostly/only watch Spanish language television. In campaign 1, it can be seen that there are large increases for the percentage of visits for both 1-2 visits as well as a significant reduction in the number of 0 visits reported. The effects are even more dramatic if you reference the data series rather than the polynomial trend line, which tends to over-smooth the data series due to the short duration of the campaign. As for campaign 2, there seems to be a steady increase in the 1-2 visit category, with the 0 visit graph oscillates modestly with the overall effect across campaign 2 ending up flat and the 3+ visit series decreasing a bit. When comparing Figure 7 with the baselines in Figure 6, this evidence suggests that both advertising campaigns 1 and 2 had positive effects on the incidence of visitation for the national restaurant chain.

So one remaining question is, what is happening with the 1-2 visit data series in Figure 7? It continues to climb long after the advertising campaign was over and it may be stretching it to suggest that these are residual effects from advertising campaign 2. The research team puzzled over this for some time and made inquiries to the media company and confirmed that there was not an additional advertising campaign present for the post-campaign 2 period. After some additional media detective work, it was discovered that the national restaurant chain had made a deal with the producers of the television program where the advertisements had appeared to "write-in" the restaurant as the setting for a number of show segments and these shows were broadcast immediately after campaign 2. This suggests that the placement of the restaurant as a setting for program segments for a number of show episodes during this time period had a positive effect on at least some categories of visitors to the national restaurant chain.

Finally, we present to the reader an example of the effectiveness of online video advertising and its synergy with the more traditional advertising medium of television. On August 25, 2011 the restaurant chain Chipolte launched a 2 minute animated Youtube video entitled "Back to the Start" that featured Willie Nelson covering a Coldplay song. The video quickly gained an audience and was so popular that Chipotle ran the same video as a television spot during the Grammy broadcast February 12, 2012. Did the video elevate sales at Chipotle? According to the Wall Street Journal (Vranica, 2012), revenues at Chiptole increased 23% during the first half of 2012 where the campaign was in effect.

Do we see any corresponding increase in the incidence of visiting Chipolte during the time window of August, 2011 to June 2012? In Figure 8 below we can see the 4-week moving estimates for visiting Chiptole in the last 30 days and also viewing YouTube in the last 30 days. A short time after the August 25, 2011 premiere of "Back to the Start" it can be seen that YouTube viewers who also ate at Chipotle begins to climb while non-Youtube viewers dropped for awhile and then began a gentle climb as well. As time progresses and the Chiptole video gains more cumulative YouTube viewers, the trend for eating at Chipotle continues to rise while the non-Youtube viewers who also ate at Chipotle begins to slack off towards the end of the year and starts to decline more steadily thereafter. What is interesting is right around the time of the debut of the video as an advertisement on television during the Grammys, the trend for eating at Chipotle and also being a YouTube viewer begins to increase at a higher rate.

Figure 8
Visiting YouTube and Eating At Chipolte



This may suggest that there could a synergy between viewing the video on YouTube and perhaps also seeing the same video on television. Previous research has suggested that viewing advertisements through multiple platforms may increase the effectiveness of those advertisements[4] and we may be seeing some evidence of that here. At any rate, the linear trend lines show a pretty clear contrast between the two subpopulations.

Now we should note here that there are several issues related to the adaptation of a syndicated consumer study to a time series format that need to be taken under advisement. The first of course is that individuals using YouTube may not have actually seen the "Back to the Start" advertisement. Being a YouTube user during the time frame only suggests the opportunity to have seen the video. It should also be pointed out that we did not plot the incidence of those individuals who watched the Grammys but rather marked the air date and observed the subsequent data series for both YouTube and non-YouTube users, assuming that the Grammy episode contributed to the rise in incidence. These are some of the tradeoffs that have to be made when determining whether the more rigorous and specific indicators that come from a custom tracking study are worth the time and expense when compared with utilizing time series data from a large syndicated survey adaptation.

Summary

The purpose of this paper was to demonstrate that there are potential alternatives to utilizing custom tracking studies for the purpose of evaluating advertising return on investment. More specifically, we outlined some of the non-trivial methodological issues involving things like discordant time frames, the ability to sometimes provide only circumstantial evidence of exposure to specific advertising, larger than normal variance due to smaller sample sizes and attenuation effects due to smoothing processes that are encountered in the adaptation of a large syndicated study to a time series format. The evidence presented in this paper and the experience of the research team with this approach suggests that in many situations the advantages of continuous or consecutive coverage of the effects of advertising across multiple time periods, the ability to extensively examine large number of competitive products and services simultaneously, the opportunities to examine advertising effects by an almost unlimited number of descriptors and the savings that can be achieved from a time series adaptation of a large consumer survey are likely worth the tradeoffs against custom tracking studies highlighted in this paper.

References

2011 Assael, H. From Silos to synergy: A fifty year review of cross-media research shows synergy hs yet to achieve its full potential. Journal of Advertising Research, vol. 51, #1, pp. 1-17.

1998 Feldwick, P. Tracking studies. Pp. 234-243 in How Advertising Works, J. Jones, editor. Sage Publications: Thousand Oaks, CA.

2011 Kilger, M. and C. Palit, Consumer confidence and consumer purchase behavior during the economic crisis of 2008. Paper delivered at the American Association of Public Opinion Research, May, 2011.

2011 Palit, C. and M. Kilger, An analysis of the relationship between public sentiment and corporate performance during a crisis as illustrated by the British Petroleum Gulf Coast crisis. Presentation delivered at the American Association of Public Opinion Research, May, 2011.

2010 Shumway, R. and Stoffer, D. Time Series Analysis and Its Applications, 3rd edition. Springer: New York.

2012 Vranic, Suzanne, Wall Street Journal December 23, 2012, retrieved 08/10/2013 from http://online.wsj.com/article/SB10001424127887324731304578189323042824276.html

[4] For an excellent review of this topic see Assael, 2011.