Attribution of Conversion Events to Multi-Channel Media

Brendan Kitts, Liang Wei, Dyng Au, Amanda Powter, Brian Burdick

Lucid Commerce
159 S. Jackson St, Suite 200, Seattle, WA 98104

***Abstract*—This paper presents a practical method for measuring the impact of multiple marketing events on sales, including marketing events that are not traditionally trackable. The technique infers which of several competing media events are likely to have caused a given conversion. We test the method using hold-out sets, and also a live media experiment in which we test whether the method can accurately predict television-generated web conversions.**

Keywords-credit assignment;attribution;advertising

I. Introduction

Television is a highly successful medium. The average American spent almost 37 hours a week watching television in 2009 – this is over twice time spent online (Leaders, 2010). Yet at the same time television presents formidable challenges in measuring and optimizing television advertising. Customers almost always view ads on TV and convert through other channels including web and retail stores. This is a fundamental problem. Kokernak (2010) suggests that “until we can develop cross-platform metrics, additional new business models for television will be nearly impossible to establish.”

This problem is shown in a real-life example in Figure 1. The top graph shows phone responses due to a television commercial (where people used the telephone number shown in the advertisement), and the bottom shows web responses for the same product. Almost all of the web transactions are occurring due to television advertising, yet there is no tracking technology available that can associate the web conversions to the television broadcast. Without visibility into these effects marketer would have enormous difficulty managing their television advertising.

This paper discusses a model for inferring attribution between a set of competing media events. We test the method using hold-out sets, and also a live media experiment in which we test whether the method can accurately predict television-generated web conversions. The method has been implemented in a television media tracking system and provides for fully automated analysis of media impact.

II. Previous Work

A variety of studies have begun to look into the problem of multi-channel conversion tracking, and specifically have called into question web-based conversion tracking numbers. Brooks et. al. (2009) noted that 71% of conversions in clients of the Atlas system were from navigational queries. Rimm-Kaufman (2007) noted that 50% of clicks may be on brandname keywords. This is suggestive that the customers already knows about the product and so had essentially been acquired through a different marketing event or offer. Chandler-Pepelnjak (2009) also noted that assigning credit to the last click ignores all other channels that may be bringing about the conversion.

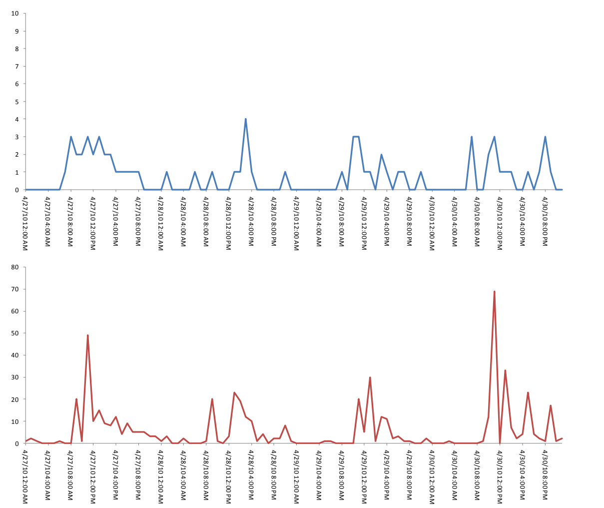

Figure 1. Phone orders (top) versus web orders (bottom) for project 10031. x-axis is hour.

III. Problem Definition

Let s_i be the ith spot or media event instance. A spot instance is an advertisement, such as on 10/10/2009 there was an airing at 8:30pm on FOX for a particular advertisement. Each media event has the following attributes (a) date, (b) product that was advertised, (c) media cost amount, (d) geography or market (eg. Boston vs Philadelphia), (e) advertising creative, which is the actual video that was shown. Let r_j be a response or purchase event which has a variety of attributes including (a) date, (b) customer, (c) address (eg. Pleasant Street, Philadelphia) (d) phone number, (e) credit card (f) email (g) product that was purchased, (h) sale amount (i) other information, eg. transcripts, and so on. The attribution problem is, for each response r_j, to associate it with the correct spot s_i that generated it.

In general we can solve this problem by estimating the conditional probabilities and then selecting the most probable spot s_j : max $\Pr(s_j \mid r_i)$ where $r_{id} >= s_{jd}$ and $w < r_{id} - s_{jd} < W$ and $r_{ip} = s_{jp}$. We have added some constraints to help to ensure that the problem can be solved in a computationally efficient manner and which are also standard methodology for click counting (IAB, 2005; IAB, 2009). 1. The spot event must occur earlier in time than the response event. 2. The sale must occur at least w after the airing, and within a certain timeframe *W*, such as 30 days. 3. The response event should be for the same product *p* that was advertised.

IV. MODEL

In order to help create an accurate attribution, it may be possible to add some cases for which the attribution is known. This can be done by inserting a special Toll Free Phone Number (TFN), URL, or coupon offer into the advertisement. If a user accesses this URL, phone number, or redeems the coupon offer, then they can be linked to the television broadcast. Typically only a small number of viewers use the linking keys, and the remainder must be inferred. We can therefore treat this as a supervised machine learning problem in which the problem is to train a model to predict, based on demographic distance, spot viewership, time difference between spot and response, the probability that this response-spot pairing is correct. A wide variety of models can be used for this problem including logistic regression and neural networks. A logistic regression is shown below

$$\Pr(r_i \mid s_j) = \frac{1}{1 + e^{-\sum_k R_k r_{i,k} - \sum_l S_l s_{j,l} - \sum_m + C_m c_{i,j,m} - A}}$$

where $r_{i,k}$ are features of the response, $s_{j,k}$ are features of the spot, and $c_{j,k}$ are features of the response-spot pair, and R_k, S_l, C_m, A are estimated using maximum likelihood or another parameter estimation technique. If the probability is too low then attribute to “unknown spot” that we will discuss later.

A. Drag Orders

Several modifications are needed to ensure that the prediction problem is practical. Toll-Free Numbers are often unique for a television station, however different numbers tend not to be given for separate airings on the same station. Therefore, if an advertisement displays on a station like FOX once at noon and again at 3pm, and a direct response occurs at 3:30pm, the same Toll Free Number is being used for each spot. It is likely that the response is due to the airing at 3pm. However there is a small chance that the response is due to the airing at noon. This phenomenon is referred to in the television industry as “drag orders”. In order to solve this problem we developed a technique that we referred to as the “Lone Spot” method. Drag orders can only exist if multiple airings occur in a short period of time. We therefore identified cases where only a single Spot aired within *W* hour period such as 24 hours. This allows us to create probabilities for response after airing.

B. Partial Attribution

Throughout this paper we will generally be looking for the most probable media event that caused a response. However it may be useful to also know what set of media events are responsible for the conversion, and their relative influence, which is sometimes referred to as “Partial conversions” (Biggs and Hollis, 1997; Chandler-Pepelnjak, 2009. Partial conversions are supported in the present model by predicting conditional probabilities for each of many events and selects the maximum. The probability is added to the schema beside each spot,response,probability pair to show the relative weighting for each of the media events in bringing about the response.

V. VARIABLES

A variety of variables are available to predict the probability of the response being conditionally dependent upon the spot.

A. Geography

Television advertising airings sometimes are localized by geography, and other times are national in nature. When the airings are localized by geography, it is possible to de-weight the probability of a response if it occurs outside of the geographic area.

Some examples of responses that are out-of-geo are shown in Fig. 2. This shows a broadcast in Philadelphia and Los Angeles. Phone responses are occurring in Philadelphia, and LA – but also other scattered locations around the country. Those other responses in the country are due to cell phone and shipping address location differences. In the example below the attribution model actually only credits web responses that are occurring in their local broadcast region because the probability of an out-of-geo response is relatively low. Fig. 2 also shows a national broadcast. This creates phone and web responses which span the entire country.

Although geography is a very important variable, it can also be erroneous due to data integrity errors. Some broadcasts are improperly labeled local, or may be improperly coded with strange values. When the Geography of the broadcast and response both match exactly, the probability of the spot being the correct for the response – in our data set - is high, but only 48%. Therefore this is a useful variable, but on its own is not definitive. For a national broadcast, a response occurring somewhere in the country only has a 4% chance of being due to the national broadcast. When there is no match at all, there is a 1% chance of the spot being correct.

B. Population

The total number of viewers reached by a television broadcast is also predictive as to the probability that a given response should be assigned to that spot. For example, if we have a very small television spot on the “Do It Yourself” channel, and another spot on “Fox”, then it is more likely that the response is due to “Fox” because of its greater audience. Viewer population for a TV airing can be estimated by a variety of methods, including Nielsen panel data, which indicates what percent of its panel was watching a particular program.

C. Demographics

Each responding customer has certain attributes including their gender, age, and so on. Let us define each of these variables as $r_{i,Dj}$ where D_j is the jth demographic variable for customer response *i*. This can be used for determining whether the customer is "like" the airing which we know about. We can also create an average profile for customers that we have linked to each television station program. For every television station program S_i we can define the $S_{i,Dj}$ as the jth demographic of the television station program S_i. The disparity Δ between the responding customer and the television station program profile can be calculated as below.

$$\Delta(r_i, s_k) = \frac{1}{J}\sum_j Z(r_{i,D_j}, s_{k,D_j})$$

$$Z(r_{i,D_j}, s_{k,D_j}) = \frac{\delta(r_{i,D_j}, s_{k,D_j}) - \frac{1}{MN}\sum_{m,n}\delta(r_{m,D_j}, s_{n,D_j})}{\sqrt{\frac{1}{MN}\sum_{i,k}\left(\delta(r_{m,D_j}, s_{n,D_j}) - \frac{1}{MN}\sum_{m,n}\delta(r_{m,D_j}, s_{n,D_j})\right)^2}}$$

$$\delta(r_{i,D_j}, s_{k,D_j}) = \left| r_{i,D_j} - s_{k,D_j} \right|$$

D. Time

Time is one of the most important variables for determining which spot might have caused a particular response. We employ the "lone spot" method to observe the time between responses and spots, whilst ensuring that no drag orders or other issues are possible. Based on this filtered set of spot and response pairs, we can calculate the probability distribution of how quickly after an airing consumers respond, $\Pr(r|r_d\text{-}s_d=T)$. The shape of this spot-response probability of response given time curve is shown in Fig. 5. In a 30 minute infomercial responses occur up until the 30th minute, and then decrease dramatically helping to isolate in time the television effect.

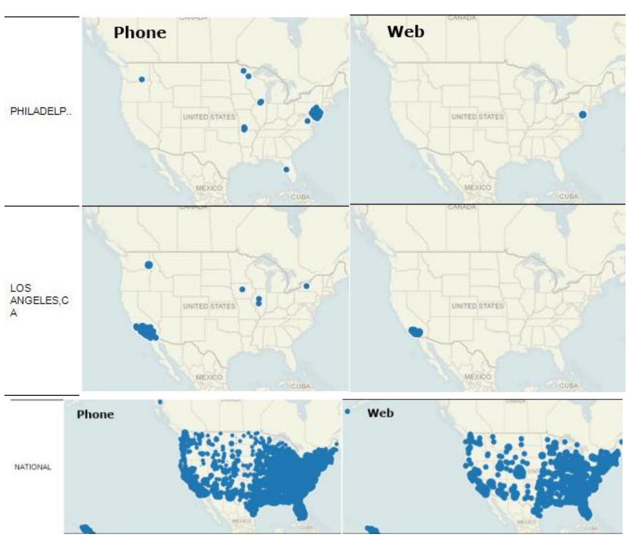

Figure 2. (Top) Phone orders due to a television broadcast are on the left, and web orders that have been attributed by the algorithm on the right. The television advertising is a local broadcast centered in Philadelphia (top) and Los Angeles (middle). The attribution model credits web responses that are in the broadcast geography in a strict fashion. Some of the telephone responses originate from out-of-broadcast area, which is possibly some inaccuracy introduced due to shipping addresses and phone numbers. (Bottom) shows another US map with phone responses from a national broadcast. The attribution algorithm attributes web responses from across the country.

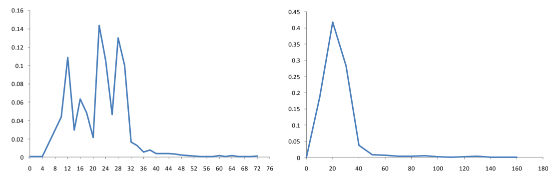

Figure 3. (Left) x-axis is minutes rounded to 2-minute intervals. y-axis is probability / proportion of responses that occur this number of minutes after the airing. The spikes correspond to "calls to action" in the television advertisement. (Right) x-axis is minutes rounded to 10 minute intervals showing how responses collapse quickly after the airing.

VI. Unknown Signal Identification

The television spots that we know about may not be a complete list of the marketing events that are running in the world. There may be other marketing events, including other television broadcasts, direct mail, and so on. There may also be organic web activity as people convert online. The method for identifying unknown spots is if the predicted highest probability of a spot being responsible for a response is lower than a threshold, then we assume that another unknown spot may be causing this response.

$$\begin{aligned} Attribution(r_i) &= s_j : \max \Pr(r_i \mid s_j), if \quad \exists s_j : \Pr(r_i \mid s_j) > \theta \\ &= UNKNOWN, otherwise \end{aligned}$$

It is possible to gather a little more information on the unknown spot. After identifying an unknown spot, we have information about the kind of spot it was. Amongst other things we know (1) The geography of the responder, and so the spot, (2) The time of the response, and so approximately the time of the unknown spot (3) The demographics of the response, and so the spot. Using this information we can measure a discrepancy between known spots that we have in our data. If this distance is larger than a threshold called the vigilance threshold V, we can spawn a new unknown source S_n which has centroid equal to the incoming response that has been measured as being different from other known and unknown spots in the code book (Duda and Hart, 1973)

$$If \Delta(r_i, s_k) > V \quad then \quad S^*_{n,D_j} = r_{i,D_j} \forall j$$

This new spot S^*_n is literally a newly identified unknown source. Unknown sources are identified in an unsupervised manner using this algorithm, and so model the structure of the input data. Although unsupervised methods cannot be guaranteed to appropriately segment unidentified sources, they can regardless provide valuable insights into the presence of unknown sources, and this in turn can be reviewed by advertisers to determine if anything unusual is happening on the campaign. After being initialized, the spot is optimized based on new responses that match to it so that it more closely represents these unidentified sources.

$$S^{*'}_{n,D_j} = \alpha(r_{i,D_j} - S^*_{n,D_j})$$

I. RESULTS

A. *Web attribution rates versus Web-TV relatedness*

Fig. 5 shows web timeseries versus phone timeseries for four projects. In some projects the web timeseries are clearly correlated with phone timeseries, which is suggestive that the web responses are being driven by the television advertising almost exclusively. 10010 and 10020, in particular, have phone and web timeseries with strong correlation coefficients (0.939 and 0.944 respectively).

We noticed some interesting behavior from the web attribution algorithm on these projects (Table V summarizes these results). Web sales are attributed at a very high rate by the algorithm for these projects (95% and 90% respectively).

In contrast, a project with a weak correlation coefficient (project 10023 at 0.5427) only shows an attribution rate of 10%

The web attribution percent therefore appears to be correlated with the R statistic for the phone versus web timeseries – which itself is a measure of the degree of relatedness between television and web activity. This is suggestive that the web attribution algorithm – by looking at the underlying spots and which is most probable for a response - is properly estimating the degree of relationship between television and web and at least making good decisions in aggregate around the overall quantity of web sales that should be attributed.

B. *Signal separation*

We also observed an interesting phenomenon in which web sales the attribution algorithm attributed. Fig. 6 shows project 10031's original web timeseries which can be decomposed into an attributed web timeseries (predicted to be due to TV), and an unattributed web timeseries (due to unknown or organic web activity). The unattributed web timeseries is shown in Fig. 7 with the aid of Fourier analysis. This shows that the unattributed timeseries is strongly periodic. There is a spectral spike at a period of 7 days, as well as a strong auto-correlation with 7 day period. The traces of this 7-day period are missing entirely from the attributed web timeseries. Thus the web algorithm seems to have cleanly separated transient television-related web transactions, from web transactions that could be organic in nature. This kind of signal separation would be extremely useful for the web marketer, since it allows them to observe their web sales without being "contaminated" by the arbitrary effects of television.

In order to validate the web attribution rates, we planned a live television advertising experiment. Using a particular television advertisement, some Direct Marketing Association (DMA) geographies received television advertising, and others did not. The objective of the experiment was to measure the lift in web sales due to the presence of television, holding as many other factors the same as possible.

The technique is extremely useful because any cross-medium effects on retail or web sales can be quantified by comparing the treatment DMA to control DMA. Using this technique, a "ground truth" web attribution during the period of the experiment can be directly computed based on the lift observed in the experimental group.

We selected 10 DMAs for the test and paired with control DMAs that had similar demographics. Demographic similarity was measured as L1 distance between aggregated census demographics for zip-codes in the targeted area. We selected the controls as the top N DMAs that were closest to the demographics of the experimental DMA.

As an additional control we also tracked the performance of all of the other DMAs in the United States. This helped to show whether any seasonal effects might have been occurring in the control groups.

We purchased short-form (60 and 120 second) television advertisements from May 17-31 2010 in each of the 10 local DMAs. The results on web channel for the customer is shown in Fig. 9. We have shown the results as a 7 day moving average because the retailer experienced highest sales on Saturday, and also because the web channel had experienced minimal sales prior to the test.

Prior to the test all test groups were normalized to their "baseline sales", which was the average of their web sales over a 3 week period. Just before the start of the media test, in fact the experimental group dropped to 0 sales – clearly there was very little activity in the web channel.

During the May 17 – 24 period, web sales increased by approximately 6x. Starting May 24 we started up a national television campaign which impacted the control markets. The control markets can be seen to lift by around 2.5-3x. However, the experimental market with local television increased even further. These lift results were statistically significant ($p<0.01$ Wilcoxon test).

Because we know that all incremental sales in the experimental market were due to television (since that is the only factor that is different), we can attribute those incremental sales to television. Also assuming that prior to the test, the baseline web sales (equal to 1.0 lift) were due to non-television-related factors – which means that we are supposing that there was no television-related sales prior to the test. This assumption may in fact be incorrect, in which case the attribution % estimate calculated below will actually be a *lower bound* on the correct attribution rate.

Given observed lift in the Correct attribution rate can be calculated as follows:

$$\textit{Correct Attribution \%} = L / (1+L)$$
$$\text{where } L = (\textit{Sales-in-exp} / \textit{Sales-in-control})$$

Given the lift of web sales by 6x, we can calculate the attribution rate as 6 / (1+6) = 85%. The subsequent lift due to the national program was 16, so 94% attribution.

The behavior of the machine learning algorithm before and during the market test is shown in Fig. 9. After the first week of local television broadcasting, web attribution rate had increased to 30%. When national broadcasting ran in the second week, attribution rate increased to 80%. The attribution rates calculated algorithmically are similar, although lower than the actual attribution rates. One of the reasons for the discrepancy is many of the media markets were improperly coded, and so were failing to match when a response originated in one of the targeted DMAs. We are working on methods for standardizing market naming automatically to address this problem.

II. Conclusion

Tracking cross-channel effects due to television in an automated manner is a central problem of television advertising. Without quantification of television effects, marketers may misallocate budgets, sometimes with disastrous results. One anecdote from a diet company was that in 2009 they shut down television advertising due to the economic downturn and because the conversions couldn't be tracked. At the time it was an easy program to cut. After just 6 months all web conversions (a completely different channel which had been extremely successful and which they had spent a great deal of money on creative site design) disappeared. They had to re-activate their television budget to stay in business.

The method that we have presented uses tell-tale signs from the responder to identify their most likely media event of origin including their proximity in time, geography, and demographics.

Although we have discussed an application in tracking television effects, the method is general purpose and can be used for any marketing event. Tracking is a problem that affects numerous marketing channels including direct mail, print advertising, as well as television, and linking keys are typically used in those mediums as well to track small numbers of responders. The method should be extendible to these other mediums.

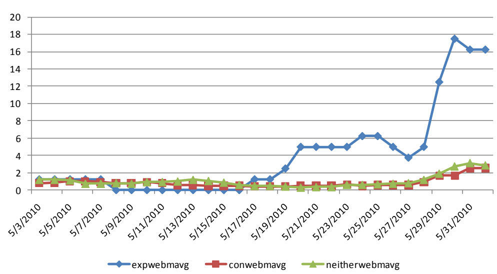

Figure 4. Web orders in targeted DMAs compared to control DMAs which did not have television advertising. In the experimentat markets web sales increased by around 16x. We can use this behavior to calculate a correct attribution rate during the experiment – clearly most of the web sales during the experimental period were due to

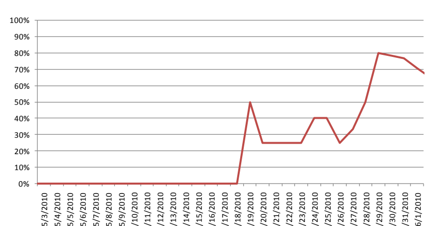

Figure 5. Web attribution rate from algorithm in experimental DMAs while the market test was underway.

References

[1] Brooks, N. (2009), Paying for Navigational Search, Atlas Digital Media Insights, http://www.atlassolutions.com/uploadedFiles/Atlas/Atlas_Institute/Published_Content/dmi-NavigationalSearch.pdf

[2] Chandler-Pepelnjak, J. and Song, Young-Bean. (2003) Optimal Frequency—The Impact of Frequency on Conversion Rates, Atlas Digital Insights http://www.atlassolutions.com/uploadedFiles/Atlas/Atlas_Institute/Published_Content/OptFrequency.pdf

[3] Chandler-Pepelnjak, J. (2009), Measuring ROI beyond the Last Ad, Atlas Digital Marketing Insights, http://www.atlassolutions.com/uploadedFiles/Atlas/Atlas_Institute/Published_Content/dmi-NavigationalSearch.pdf

[4] Kokernak, M. (2010), What's Television's Next Business Model? Media Post Daily News, Wednesday, March 17, 2010 http://www.mediapost.com/publications/?fa=Articles.showArticle&art_aid=124424

[5] IAB (2005), IAB Impression Measurement Guidelines, http://www.iab.net/media/file/US_meas_guidelines.pdf

[6] IAB (2009), IAB Click Measurement Guidelines, http://www.iab.net/media/file/click-measurement-guidelines2009.pdf

[7] Leaders, (2010), In Praise of Television: The great survivor, The Economist, April 2010.

[8] Microsoft adCenter DOM (2010), Microsoft adCenter Help, referenced 2010-02-26 https://adcenterhelp.microsoft.com/Help.aspx?market=en-US&project=adCenter_live_Std&querytype=topic&query=MOONSHOT_CONC_ClickMethod.htm

[9] Rimm-Kaufman, A. (2007), PPC And Your Good Name: Sales From Brand Searches Aren't Incremental May 27, 2007 http://searchengineland.com/ppc-and-your-good-name-sales-from-brand-searches-arent-incremental-10825

TABLE I. Attribution Rates for 4 Projects

Project Key	Date Start	Date End	Phone - Web R	Web Attrib %
10010	7/29/2009	5/5/2009	0.9399	95%
10020	6/22/2009	3/30/2009	0.9448	90%
10031	5/9/2010	2/8/2010	0.8880	34%
10023	5/20/2010	4/2/2010	0.5427	10%

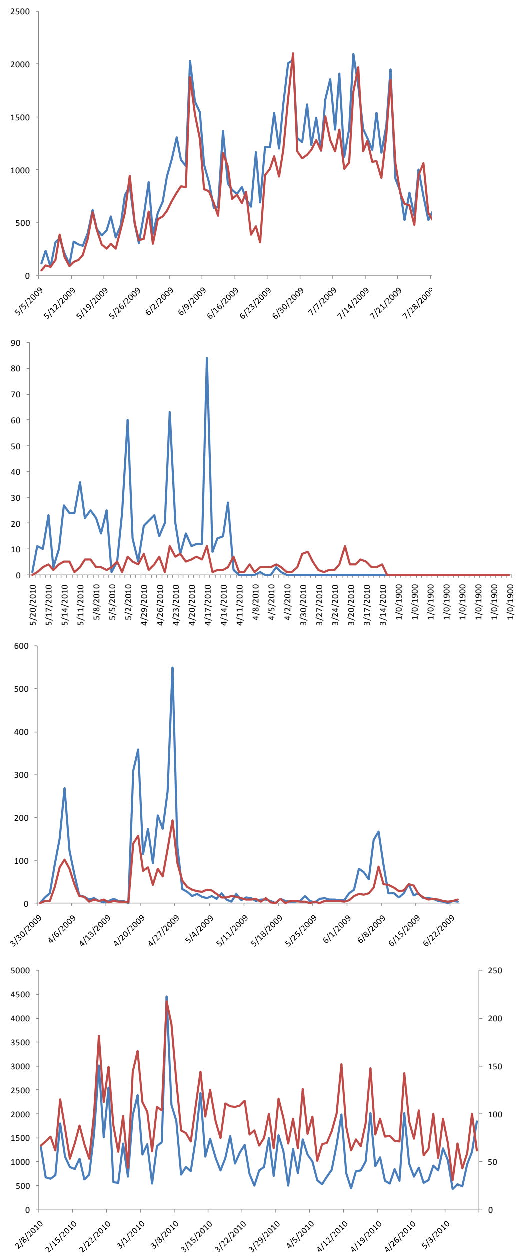

Figure 6. Phone and web timeseries for Projects 10010, 10023, 10020, and 10031 in order top to bottom. Correlation coefficients are 0.939, 0.54, 0.944, 0.888. Web attribution rates from the algorithm are 0.95, 0.10, 0.90, 0.34.

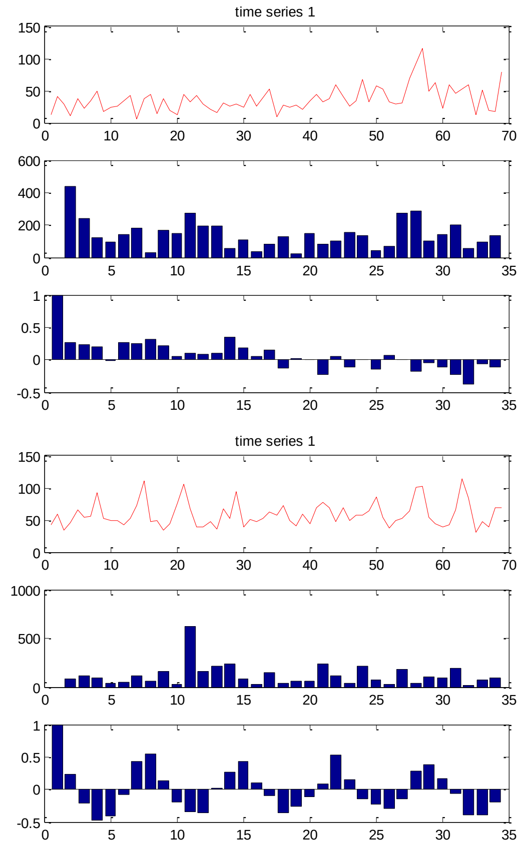

Figure 7. Project 10031 web timeseries broken nto attributed and un-attributed web sales and analyzed using Fourier analysis and Auto-correlation. Top subplot is the timeseries, middle is the fourier periodogram, bottom is an auto-correlation plot. (Top) Attributed timeseries does not show any 7 day pattern. (Bottom) Un-attributed web sales show a strong 7 day period.