Dynamic Effects of Ad Impressions on Commercial Actions in Display Advertising

Joel Barajas*
UC, Santa Cruz
Santa Cruz CA, USA
jbarajas@soe.ucsc.edu

Ram Akella
UC, Santa Cruz
Santa Cruz CA, USA
akella@soe.ucsc.edu

Marius Holtan
AOL Research
Palo Alto CA, USA
marius.holtan@teamaol.com

Jaimie Kwon
AOL Research
Palo Alto CA, USA
jaimie.kwon@teamaol.com

Aaron Flores
AOL Research
Palo Alto CA, USA
aaron.flores@teamaol.com

Victor Andrei
AOL Research
Palo Alto CA, USA
victor.andrei@teamaol.com

ABSTRACT

In this paper, we develop a time series approach, based on Dynamic Linear Models (DLM), to estimate the impact of ad impressions on the daily number of commercial actions when no user tracking is possible. The proposed method uses aggregate data, and hence it is simple to implement without expensive infrastructure. Specifically, we model the impact of daily number of ad impressions in daily number of commercial actions. We incorporate persistence of campaign effects on actions assuming a decay factor. We relax the assumption of a linear impact of ads on actions using the log-transformation. We also account for outliers with long-tailed distributions fitted and estimated automatically without a pre-defined threshold. This is applied to observational data post-campaign and does not require an experimental set-up. We apply the method to data from one commercial ad network on 2,885 campaigns for 1,251 products during six months, to calibrate and perform model selection. We set up a randomized experiment for two campaigns where user tracking is feasible. We find that the output of the proposed method is consistent with the results of A/B testing with similar confidence intervals.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*; G.3 [**Mathematics of Computing**]: Probability and Statistics—*Time Series Analysis*

General Terms

Algorithms, Economics, Management, Measurement

Keywords

Attribution, DLM, Marketing, Online Display Advertising

*Main contact.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.

1. INTRODUCTION

Display advertising represents a significant percentage of the online advertising market and it is comparable to that of sponsored search [7]. This often triggers online users to search for information about commercial products [1]. Therefore, evaluating the effectiveness of display advertising campaigns in all possible respects is increasingly important to allocate a given budget effectively. Recent research on campaign evaluation has focused primarily on two approaches: running randomized experiments (A/B testing), and bias correction based on user features for observational data. Lewis *et al.* corrects over-estimation issues due to user activity bias using A/B testing [8]. However, this approach demands the instrumentation and expensive use of unexposed users. Chan *et al.* propose to correct the bias in the user selection for ad exposure in observational studies [4]. This approach relies heavily on user features which are often incomplete and their collection is biased toward active users. Dekimpe and Hanssens demonstrate the presence of long term and transient impact of campaigns on actions [5]. They suggest the concept of *persistence* as the true impact. Since A/B testing do not consider the time lag between ad exposure and conversions, the persistence impact is not properly measured. A key requirement for the methods discussed is the use of reliable tracking cookies. In practice, a significant number of web users either reject tracking cookies outright or frequently delete them. We have identified approximately 17% of users not tracked via cookies. Thus, developing a time series approach to incorporate persistence, and to relate commercial actions to ad impressions shown to users with unreliable cookies is very valuable.

We propose an approach to estimate the effects of display marketing campaigns that differs from recent literature in three respects. We focus on commercial actions rather than surrogates such as search terms or clicks. We consider the context where user tracking is not available. The proposed method is simple to implement based on aggregated data. We develop a time-series approach to model the impact of ad impressions on actions. In the absence of user tracking, we account for confounding effects by a base time series model. We rely on the prediction power of this model to capture the effects attributed to other factors. In prior work, we developed this base model incorporating weekly seasonality [2]. We incorporate a decay factor to model impression effects

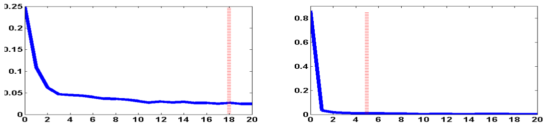


Figure 1: Time distribution of latest impression shown to a converting user. Dotted line shows 95% cumulative probability. x-axis is in days and y-axis is the probability mass.

on actions that automatically provides different lags. Our approach process outliers and account for non-linear impact of impressions using a logarithmic model.

This paper is organized as follows. In section 2, we present the proposed model. We also define the measures to evaluate campaigns. In section 3, we provide results for several variants of the method under different assumptions. We compare our results with A/B testing for two campaigns. In section 4 we discuss the main findings and future work.

2. METHODOLOGY

2.1 Notation and Definitions

Let T be the total number of days we observe commercial actions and impressions, and N the number of advertising campaigns running during the observed time period. The indices: t denotes discrete time in days, c denotes a particular campaign, and k refers to the number of forecast look ahead steps. Let $X_t^{(c)}$ be the number of ad impressions at time t for campaign c, and Y_t be the number of actions at time t for a given product. $Y_{1:t}$ represents the vector of $[Y_1, \dots, Y_t]$.

We define a latent state $\theta_\mathbf{t}$ to be a stochastic process describing the behavior of the series. Let Y_t be normally distributed given $\theta_\mathbf{t}$, and $\theta_\mathbf{t}$ be normally distributed given $\theta_\mathbf{t-1}$:

$$\begin{aligned} Y_t &= F_t'\theta_t + \nu_t & \nu_t &\sim N(0, V_t) \\ \theta_t &= G_t\theta_{t-1} + w_t & w_t &\sim N(0, W_t) \end{aligned} \tag{1}$$

G_t is the evolution matrix, F_t is the observation matrix, ν_t is the observational noise with variance V_t, and w_t is the state evolution with covariance matrix W_t. This is denoted as a Dynamic Linear Model (DLM) [12]. By choosing G_t and F_t, we can model different types of behavior of the series.

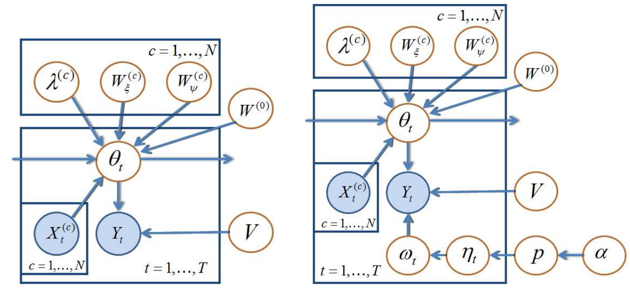
2.2 Modeling Campaigns

To model the effect of impressions on the actions, we assume a persistence component as suggested in previous work [5]. To verify this effect, we estimate the distribution of the number of days before the last impression is delivered to a user who performs a commercial action. Fig 1 shows this distribution for two campaigns.

Let $\xi_t^{(c)}$ be the *effect* of ad impressions on the number actions at time t for campaign c. $\lambda^{(c)}$ represents the constant rate of decay of this effect. $\psi_t^{(c)}$ is the dynamic impact per impression. For N campaigns we have:

$$\begin{gathered} Y_t = \sum_{c=1}^{N} \xi_t^{(c)} + \nu_t \\ \xi_t^{(c)} = \lambda^{(c)}\xi_{t-1}^{(c)} + \psi_t^{(c)} X_t^{(c)} + w_t^{(\xi,c)}, \quad \psi_t^{(c)} = \psi_{t-1}^{(c)} + w_t^{(\psi,c)} \end{gathered} \tag{2}$$

By defining $\xi_t^{(c)}$, we automatically fit the number of previous days with impact on Y_t through $\lambda^{(c)}$ unlike modeling a fixed lag in autoregressive models [5]. We incorporate independent $\lambda^{(c)}$ and $\psi_t^{(c)}$ for each campaign. Then, we write this model as a DLM, based on the definition of Eq 1, as follows:



(a) (b)

Figure 2: (a) Graphical model for multiple campaigns and a base model, $M^{(0:N)}$. (b) Model for multiple campaigns with outliers processing $M\omega^{(0:M)}$.

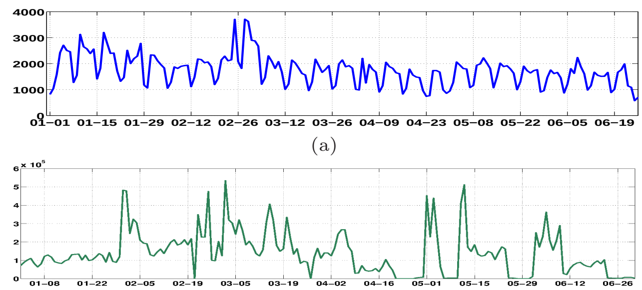


Figure 3: (a) Commercial actions and (b) number of impressions through time. X-axis is time in dates.

$$\begin{gathered} \theta_t'^{(c)} = [\xi_t^{(c)}, \psi_t^{(c)}] \qquad w_t'^{(c)} = [w_t^{(\xi,c)} + X_t^{(c)} w_t^{(\psi,c)}, w_t^{(\psi,c)}] \\ W_t^{(c)} = \begin{bmatrix} W_\xi^{(c)} + \left(X_t^{(c)}\right)^2 W_\psi^{(c)} & X_t^{(c)} W_\psi^{(c)} \\ X_t^{(c)} W_\psi^{(c)} & W_\psi^{(c)} \end{bmatrix} \\ G_t^{(c)} = \begin{bmatrix} \lambda^{(c)} & X_t^{(c)} \\ 0 & 1 \end{bmatrix} \qquad F'^{(1:N)} = [(1,0)^{(1)}, \cdots, (1,0)^{(N)}] \\ \theta_t'^{(1:N)} = [\theta_t'^{(1)}, \cdots, \theta_t'^{(N)}] \qquad G_t^{(1:N)} = \text{diag}[G_t^{(1)}, \cdots, G_t^{(N)}] \\ w_t'^{(1:N)} = [w_t'^{(1)}, \cdots, w_t'^{(N)}] \qquad W_t^{(1:N)} = \text{diag}[W_t^{(1)}, \cdots, W_t^{(N)}] \end{gathered} \tag{3}$$

We define the above DLM model as $M^{(1:N)}$ where:

$$M^{(1:N)} = \text{DLM}\left(F^{(1:N)}, G_t^{(1:N)}, V^{(1:N)}, W_t^{(1:N)}, \theta_t^{(1:N)}, w_t^{(1:N)}\right) \tag{4}$$

This model incorporates the effects of all N campaigns on the number of actions. However, there is no base model to describe the number of actions when no impression is displayed. This is highly relevant to separate the actions attributed to external factors. Let $M^{(0)}$ be a base model and $M^{(0:N)}$ be the full model. Thus, we define:

$$\begin{gathered} \widetilde{F}' = [F'^{(0)}, F'^{(1:N)}] \qquad \widetilde{G}_t = \text{diag}[G^{(0)}, G_t^{(1:N)}] \\ \widetilde{w}_t' = [w_t'^{(0)}, w_t'^{(1:N)}] \qquad \widetilde{W}_t = \text{diag}[W^{(0)}, W_t^{(1:N)}] \\ \widetilde{\theta}_t' = [\theta_t'^{(0)}, \theta_t'^{(1:N)}] \qquad M^{(0:N)} = \text{DLM}\left(\widetilde{F}, \widetilde{G}_t, \widetilde{V}, \widetilde{W}_t, \widetilde{\theta}_t, \widetilde{w}_t\right) \end{gathered} \tag{5}$$

A key advantage of the model $M^{(0:M)}$ is that it allows us to incorporate *any* DLM to attribute the time series of the actions. Thus, we can include any assumption about the dynamics of the actions and model the remaining variability by the impressions. Fig 2(a) shows the graphical model for the model $M^{(0:N)}$. We test two base models: a random walk and a seasonal weekly model. The random walk model [10] is defined as: $M_b^{(0)} = \text{DLM}(1, 1, V_b, W_b, \theta_{b,t}, w_{b,t})$. Fig 3 depicts a seasonal component on the number of actions that is synchronized with the day of the week [2]. To model this seasonality, we use the Fourier representation of DLMs [12] and denote it as $M_s^{(0)}$.

Algorithm 1 Generative Model to Handle Outliers

Draw $p|\alpha \sim Dir(\alpha)$
for $t \leftarrow 1$ to T **do**
  Draw $\eta_t|p \sim Mult(1, p)$
  Draw $\omega_t|\eta_t \sim \Gamma(\frac{\eta_t}{2}, \frac{\eta_t}{2})$
  Set $V_t^* = \omega_t^{-1}V$
end for

Algorithm 2 Gibbs Sampling Algorithm

Define $D_{1:T} = \left\{Y_{1:T}, X_{1:T}^{(1:N)}\right\}$
Initial guess $\Phi^0 = \left\{\lambda^{(1:N)}, W_\psi^{(1:N)}, W_\xi^{(1:N)}, W^{(0)}, \widetilde{V}\right\}^0$,
Initial guess $\Omega^0 = \{\omega_{1:T}, \eta_{1:T}, p\}^0$
for $s \leftarrow 1$ to $N_0 + N_s$ **do**
  Draw $\theta_{1:T}^s \sim p\left(\theta_{1:T}|\Phi^{s-1}, \Omega^{s-1}, D_{1:T}\right)$ using FFBS
  Draw $\Phi^s \sim p\left(\Phi|\theta_{1:T}^s, \Omega^{s-1}, D_{1:T}\right)$ using Eqs from Appx B.
  Draw $\Omega^s \sim p\left(\Omega|\theta_{1:T}^s, \Phi^s D_{1:T}\right)$ using Eqs from Appx C.
end for

2.3 Log-Transformation and Outlier Handling

To relax the assumption of a linear impact of ad impressions on actions, we use the log transformation for both variables. We consider the following model for one campaign:

$$\begin{aligned} Z_t &= \log(Y_t), & X_t^* &= \log(X_t) \\ Z_t &= \theta_t^{(0)} + \psi_t X_t^* + \widetilde{\nu}_t & Y_t &= \exp\left\{\theta_t^{(0)} + \widetilde{\nu}_t\right\} X_t^{\psi_t} \end{aligned} \tag{6}$$

If $\psi_t < 1$, the effect of ad impressions on actions decreases as more of them are shown. As depicted in Fig 5, the daily number of impressions is more dynamically changing than the number of commercial actions. By assuming this model, we smooth these changes of the time series of impressions. Since this is a monotonic function $f(Y_t)$, the cumulative distribution for Y_t is the same as the cumulative distribution of the inverse transform of the Z_t, $cdf(Y_t) = cdf(f(Y_t))$. Thus, for model fitting we transform $X_{1:T}, Y_{1:T}$ to the new space $X_{1:T}^*, Z_{1:T}$ for all campaigns and model $Z_{1:T}$ with $M^{(0:N)}$.

Given that commercial action data is collected by advertisers, there is no control over this process by the Ad network [9]. Then, we often observe *drastic* changes in the daily number of actions that can increase the estimation variance. One approach to handle outliers is to weight the observations based on the variance modeled for each output Y_t [12]. For this analysis, we use a simplification of the model presented in [10]. We switch from using the Normal distribution for Y_t to using a t-distribution with a set of degrees of freedom to choose from. We use the Normal-Gamma mixture to represent this t-distribution.

Fig 2(b) shows the graphical model for all campaigns with outliers handling. Main changes are introduced at the observation level with a hierarchical model to handle individual number of degrees of freedom for each Y_t. Algorithm 1 shows the generative model for the observational variance when we process the outliers. We now have independent variances for each observation. Given on $\omega_{1:T}$, we have a DLM model $M^{(0:M)}$ with $V_t^* = \omega_t^{-1}V$ for $t = 1, \ldots, T$. The possible η_t values are predefined a priori.

2.4 Inferring the Model Parameters

We assume all parameters to be random variables and follow a Bayesian approach. Observations are the daily number of actions and impressions for each campaign. Algorithm 2 defines the Gibbs sampling approach to fit the model. We sample $\theta_{1:T}|\Phi, \Omega, D_{1:T}$ based on Forward Filtering Backward Sampling (FFBS) method explained in Appendix A. To find the percentage of actions attributed to a campaign, we constrain the campaign effects to be positive $\psi_t, \xi_t > 0$ and the base model trend [2]. We constrain these components when they are sampled following the method from [11].

Table 1: Model variants used for experimentation.

Model	Process Outliers	Aggregate Camp	Log Transform	Positively Constrained
Mω	X			
MωAgg	X	X		
Mω+	X			X
Mω+Agg	X	X		X
Mω log	X		X	
Mω logAgg	X	X	X	
Mω log+	X		X	X
Mω log+Agg	X	X	X	X

To evaluate the model fitting, we use the median and 90% credible intervals. We also evaluate model prediction based on one step ahead forecasting, $Y_t^{k=1}|D_{t-1}$, estimated from the samples we generate in FFBS at the filtering stage.

$$\begin{aligned} &\hat{Y}_t|M^{(0:N)} \approx \text{Median}(F'\theta_t^s), \quad \hat{Y}_t|M_{\log}^{(0:N)} \approx \text{Median}(\exp(F'\theta_t^s)), \\ &\hat{\omega}_t|_\omega M^{(0:N)} \approx \text{Median}(\omega_t^s), \quad \hat{Y}_t^{k=1}|M^{(0:N)} \approx \text{Median}(Y_t^{k=1,s}) \end{aligned} \tag{7}$$

2.5 Campaign Evaluation

We present two campaign evaluation approaches. We interpret the model and find dynamic campaign performance, and we use the variability attributed to a campaign as a whole. To interpret the model for campaign evaluation, we find the proportion of actions attributed to a campaign:

$$\begin{aligned} &Y_t^{(c)s}|M^{(0:N)} = F'\theta_t^s - F'^{\neg c}\theta_t^{\neg cs}, \quad \pi_t^{(c)s} = Y_t^{(c)s}/Y_t, \\ &Y_t^{(c)s}|M\log^{(0:N)} = \exp(F'\theta_t^s) - \exp(F'^{\neg c}\theta_t^{\neg cs}) \end{aligned} \tag{8}$$

where $\theta^{\neg cs}$ are the state attribution samples from the full model except campaign c. This measure provides the expected difference attributed to campaign c, with respect to other ones and the confounding effects. Daily samples of the campaign effects $\pi_t^{(c)s}$ are estimated by this metric.

We also estimate the variability attributed to a given campaign, $R^2(c)$, by fitting the model without campaign c and finding the difference with the full model. This measures the improvement in model fitting by campaign impressions. We measure this variability with respect to: the data variance $var(Y)$, the base model $M^{(0)}$, and full model without campaign c, $M^{(0:N\neg c)}$.

$$\begin{aligned} R^2(c|\text{var}(Y_t)) &= \frac{\text{MSE}\left(M^{(0:N\neg c)}\right) - \text{MSE}\left(M^{(0:N)}\right)}{\text{var}(Y_t)} \\ R^2(c|M^{(0)}) &= \frac{\text{MSE}\left(M^{(0:N\neg c)}\right) - \text{MSE}\left(M^{(0:N)}\right)}{\text{MSE}\left(M^{(0)}\right)} \\ R^2(c|M^{(0:N\neg c)}) &= 1 - \frac{\text{MSE}\left(M^{(0:N)}\right)}{\text{MSE}\left(M^{(0:N\neg c)}\right)} \end{aligned} \tag{9}$$

where MSE stands for mean squared error. For these measures, we do *not* process outliers since this process *weights* the observations based on how they deviate from the others. As the model is run multiple times to estimate these outliers are estimated differently.

3. VALIDATION AND RESULTS

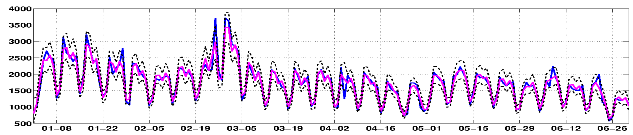
To evaluate model fitting, we measure the mean relative squared error (MRSE). Given model M, we have:

$$\begin{aligned} &\text{MRSE}^f(M) = \frac{1}{T}\sum_t \hat{\omega}_t|M\left(\frac{Y_t - \hat{Y}_t|M}{Y_t}\right)^2, \quad Y_t > 0 \\ &\text{MRSE}^f_{\omega=1}(M) = \frac{1}{T}\sum_t \left(\frac{Y_t - \hat{Y}_t|M}{Y_t}\right)^2 \end{aligned} \tag{10}$$

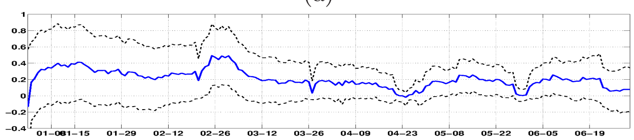
We provide two metrics as our model processes outliers. If some outliers produce large errors for $MRSE_{\omega=1}$, we might

Table 2: Model evaluation results averaged over products with 95% confidence intervals. Estimates scaled by 10^{-2}.

	Random Walk Base model $M_b^{(0)}$, MRSE				Weekly Seasonal Base model $M_s^{(0)}$, MRSE			
Model	Fitted	Forecast	Fitted $\omega_t = 1$	Forecast $\omega_t = 1$	Fitted	Forecast	Fitted $\omega_t = 1$	Forecast $\omega_t = 1$
Mω	7.91 ± 1.85	61.77 ± 7.13	14.87 ± 2.40	72.13 ± 7.58	8.26 ± 2.13	61.13 ± 7.34	12.65 ± 2.11	70.75 ± 7.67
MωAgg	9.79 ± 2.73	57.65 ± 8.75	16.63 ± 3.60	68.21 ± 9.31	8.06 ± 2.00	58.56 ± 7.05	13.40 ± 2.47	71.47 ± 7.87
Mω+	15.78 ± 3.15	59.99 ± 8.06	21.79 ± 3.56	65.97 ± 7.89	13.69 ± 3.23	62.11 ± 7.85	16.34 ± 3.10	68.55 ± 8.11
Mω+Agg	14.86 ± 2.74	53.41 ± 6.53	21.44 ± 3.15	62.03 ± 6.55	10.32 ± 2.15	59.39 ± 7.06	14.18 ± 2.47	66.14 ± 7.33
Mω log	1.33 ± 0.32	**13.25±2.21**	5.49 ± 1.02	20.00 ± 2.57	**0.72±0.14**	15.51 ± 2.81	**3.92±0.92**	21.20 ± 3.12
Mω logAgg	1.48 ± 0.32	**11.76±2.42**	5.75 ± 1.03	18.52 ± 2.76	**0.64±0.13**	12.45 ± 2.20	**4.43±1.05**	19.14 ± 2.83
Mω log+	1.87 ± 0.37	13.87 ± 2.77	5.70 ± 1.15	19.39 ± 3.08	1.48 ± 0.27	15.07 ± 2.44	4.22 ± 0.95	19.39 ± 2.79
Mω log+Agg	1.84 ± 0.42	12.70 ± 2.87	7.33 ± 1.52	21.10 ± 3.61	1.20 ± 0.22	15.31 ± 2.43	**3.24±0.74**	19.16 ± 2.81



(a)



(b)

Figure 4: (a) Model fitting and (b) proportion of actions attributed to campaign in Fig 3. X-axis is time in dates.

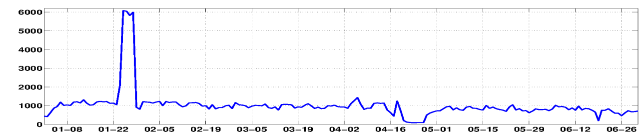
under-estimate the model fitting. To evaluate model prediction, we use one-step-ahead forecast estimates $\hat{Y}_t^{k=1}$ to calculate $MRSE$, denoted as $MRSE^{k=1}$.

We analyze all the transactions for 2,885 campaigns associated with 1,251 products during six months, from January 1^{st} to June 31^{st}, 2011. We measure campaign performances for each product independently. We use $N_0 = 1000$ samples for burn-in and $N_s = 4000$ for the posterior distribution.

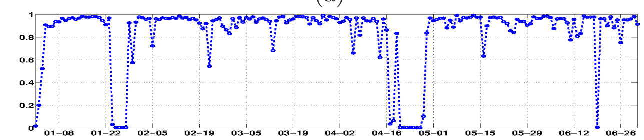
3.1 Experimental Results

We show qualitative results to analyze the model performance and the outlier handling. We test different variants, displayed in table 1. MAgg is the model where impressions from all campaigns are aggregated. Our goal is two fold: 1)Modeling the impact of the set of campaigns as a whole. 2)Hierarchically disaggregating campaigns to evaluate the impact in detailed. Finally, we evaluate the fitting improvement provided by campaigns using R^2.

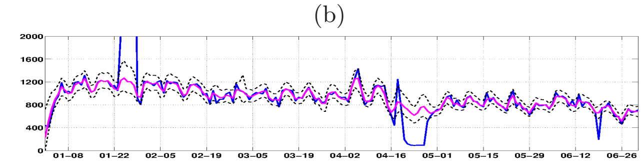
Fig 4 shows the model fitting and proportion of actions attributed to the campaign shown in Fig 3 using $M_s\omega$log. The seasonal base model seems to be a good choice given the weekly periodicities in the action series. Fig 5 shows the actions and impressions with the model fitting for a product with outliers Y_t is likely to be an outlier when $\hat{\omega}_t$ is low. If sudden changes are observed in the daily actions, the outlier handling *weights* these changes automatically. Table 2 shows the fitting and prediction results. Performance estimates based on $MRSE$ instead of $MRSE_{\omega=1}$ are better because they diminish the effect of the outliers on the measure. We observe better performance when we include the log transformation consistently. We observe that weekly seasonal base model results show significantly better fitting than the random walk. The predictive power of both models is equivalent. The fitting performance of the constrained models is lower than without this constraint. This is expected as the imposed constraint might not be optimal for the $MRSE$. Table 3 shows the campaign performance.



(a)



(b)



(c)

Figure 5: (a) Commercial actions. (b) Median weights fitted for outliers. (c) Model fitting with 90% credible intervals. X-axis is the time in dates.

Table 3: Averaged campaign evaluation results $\bar{\pi}^{(c)}$. Distribution of campaign effect significance (%).

Model	% attributed	Campaign Significance		
		(+)	(-)	(±)
Random Walk Base model				
Mω	14.07 ± 1.36	23.13	0.71	76.09
MωAgg	19.07 ± 2.63	40.96	2.01	57.03
Mω log	21.31 ± 1.63	18.65	0.58	**80.71**
Mω logAgg	24.75 ± 2.40	34.44	1.31	**64.25**
Weekly Seasonal Base model				
Mω	10.39 ± 1.23	19.66	1.31	78.96
MωAgg	16.10 ± 1.91	34.33	1.56	64.11
Mω log	19.84 ± 1.64	14.83	0.60	**83.98**
Mω logAgg	21.09 ± 2.28	25.24	0.36	**73.18**

Table 4: Attributed variability results.

	RandWalk $M_b^{(0)}$		WeekSea $M_s^{(0)}$	
Measure	Mean	Std Dev	Mean	Std Dev
$R^2(c\mid\mathrm{var}(Y_t))$	**0.1241**	0.2704	**0.0667**	0.1750
$R^2(c\mid M^{(0)})$	0.2804	0.3827	0.3002	0.3701
$R^2(c\mid M^{(0:N\neg c)})$	0.4967	0.4114	0.4703	0.3729

Models with best fitting results show the largest percentage of campaigns without statistically significant effect illustrating the challenge of this signal in evaluation. Table 4 depicts the mean and variance over all the campaigns for variability attribution. Here, attribution with respect to the data variability is lower for the weekly seasonal model because actions are attributed to the day of the week and to campaigns.

3.2 Comparison with A/B Testing

We compare our results with A/B testing for two independent campaigns during 27 days. We count the users who are first exposed and later perform an action (exposed and ac-

Table 5: A/B testing comparison with the attribution given by Mωlog and Mωlog+ for the RandWalk model.

Method	Campaign 1			Campaign 2		
	Low	Med	High	Low	Med	High
A/B	0.009	0.199	0.458	-0.034	0.115	0.312
Mωlog	0.076	0.289	0.421	-0.049	0.347	0.809
Mωlog+	0.133	0.272	0.623	0.094	0.180	0.519

tor) or not (exposed and non-actor). For the control group, we select the users who are targeted first but are not selected for exposure randomly. We find if the user performs an action (actor and non-exposed) or not (non-actor and non-exposed). We use standard Beta conjugate prior distribution to find the probability of action by sampling. We estimate the change in action probability respect to the control group (lift). For our method, we find the mean increase of actions attributed to the campaign respect to those attributed to other components. Table 5 shows the comparison. Even when a controlled experiment is run, sparsity of actions is an issue. We have one positive significant campaign at the 90% confidence level and one leaning towards positive effect for both A/B testing and Mωlog. For positively constrained contributions, we have similar estimates but with larger confidence interval. The median estimates for the two variants tested fall in the confidence interval of A/B testing, except Mωlog for campaign 2. Similarly, the median estimates for A/B testing fall in the confidence intervals of the proposed method.

4. DISCUSSION AND FUTURE WORK

We have presented a time series based approach to measure the effects of online display marketing campaigns when user tracking is not available. We have modeled the effect of ad impressions on commercial actions through a DLM, and provided daily effect estimates. We have incorporated persistence of campaign effects, through a decay factor, and accounted for outliers automatically without any threshold. We have presented several different campaign evaluation measures. These are intended to provide a spectrum of choices at the time of evaluating a campaign. We have found that a model in the log-scale is more effective to describe the relationship between ad impressions and commercial actions. Results indicate that a seasonal base model will give less attribution to campaigns. We observed several campaigns where the effect is not statistically significant. This is consistent with A/B testing results. We consider this challenge as future work for commercial actions. Combining A/B testing and the dynamic analysis of campaign effects is a direction for improvement when user tracking is available. Our ultimate goal is to provide daily significant estimates of the campaign effects on commercial actions.

5. ACKNOWLEDGEMENTS

This work is partially funded by CONACYT UC-MEXUS grant 194880, CITRIS and AOL Faculty Award.

6. REFERENCES

[1] Atlas. Engagement mapping: A new measurement standard is emerging for advertisers, 2008. White Paper.

[2] J. Barajas, R. Akella, M. Holtan, J. Kwon, and B. Null. Measuring the effectiveness of display advertising: a time series approach. In *WWW*, pages 7–8, 2011.

[3] C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

[4] D. Chan, R. Ge, O. Gershony, T. Hesterberg, and D. Lambert. Evaluating online ad campaigns in a pipeline: causal models at scale. In *ACM SIGKDD*, pages 7–16. ACM, 2010.

[5] M. G. Dekimpe and D. M. Hanssens. Sustained spending and persistent response: A new look at long-term marketing profitability. *Journal of Marketing Research*, 36:397–412, 1999.

[6] L. Devroye. *Non-Uniform Random Variate Generation.* Springer-Verlag, 1986.

[7] Google. CPA perfomance trends on the Google display network, 2010. White Paper.

[8] R. A. Lewis, J. M. Rao, and D. H. Reiley. Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising. In *WWW2011*, pages 157–166. ACM, 2011.

[9] M. Mahdian and K. Tomak. Pay-per-action model for on-line advertising. *Int. J. Electron. Commerce*, 13:113–128, 2008.

[10] G. Petris, S. Petrone, and P. Campagnoli. *Dynamic Linear Models with R.* use R! Springer-Verlag, 2009.

[11] G. Rodriguez-Yam, R. Davis, and L. Scharf. Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. Technical report, 2004.

[12] M. West and J. Harrison. *Bayesian forecasting and dynamic models (2nd ed.).* Springer-Verlag, 1997.

APPENDIX

A. FFBS ALGORITHM

FFBS is a method to sample the hidden states θ_t given the parameters in a DLM [10, 3]. Here, samples are generated backwards after filtering given the generated states:

1. Estimate $p(\theta_t|\Phi, D_{1:t}) = N(m_t, C_t)$ for $t = 1, \dots, T$ using Kalman Filtering equations (Forward Filtering)
2. Draw $\theta_T|D_{1:T} \sim N(m_T, C_T)$
3. For $t = T-1, \dots, 1$ draw $\theta_t|\theta_{t+1}, D_{1:T} \sim N(h_t, H_t)$ (Backward Sampling)

$$\begin{aligned} h_t &= m_t + C_t G'_{t+1} R^{-1}_{t+1}(\theta_{t+1} - a_{t+1}) \\ H_t &= C_t - C_t G'_{t+1} R^{-1}_{t+1} G_{t+1} C_t \\ R_{t+1} &= G_{t+1} C_t G'_{t+1} + W_t \qquad a_{t+1} = G_{t+1} m_t \end{aligned} \tag{11}$$

B. SAMPLING DISTRIBUTIONS FOR Φ

$IG(\alpha, \beta)$ represents Inverse Gamma distribution with shape α and rate β. $TN(m, C, a, b)$ refers to the Normal distribution truncated at $[a, b]$. T_0^c and T_f^c represent the start and end times of campaign c. $T^c = T_f^c - T_0^c + 1$. $[\alpha_{v0}, \alpha_{w0}^{(0)}, (\alpha_{\xi 0}, \alpha_{\psi 0})^{(1:N)}] = 0.5$; $[\beta_{v0}, \beta_{w0}^{(0)}, (\beta_{\xi 0}, \beta_{\psi 0})^{(1:N)}] = 10^{-6}$.

$$\begin{aligned} &\widetilde{V} \sim IG(\alpha_v, \beta_v), && \alpha_v = \alpha_{v0} + \tfrac{T-1}{2}, \\ &SS_y = \textstyle\sum_{t=1}^{T} \omega_t (Y_t - \widetilde{F}'\theta_t)^2, && \beta_v = \beta_{v0} + \tfrac{1}{2} SS_y \end{aligned} \tag{12}$$

$$\begin{aligned} W^{(0)} \sim IG(\alpha_w^0, \beta_w^0), \quad \alpha_w^0 &= \alpha_{w0}^0 + \tfrac{T-2}{2}, \\ \beta_w^0 &= \beta_{v0}^0 + \tfrac{1}{2} \textstyle\sum_{t=1}^{T-1} (\theta_{t+1}^{(0)} - G^{(0)} \theta_t^{(0)})^2 \end{aligned} \tag{13}$$

$$\begin{aligned} &W_\xi^{(c)} \sim IG(\alpha_\xi^{(c)}, \beta_\xi^{(c)}), && \alpha_\xi^{(c)} = \alpha_{\xi 0}^{(c)} + \tfrac{T^c-2}{2}, \\ &\hat{\xi}_{t+1}^{(c)} = \lambda^{(c)} \xi_t^{(c)} + \psi_{t+1}^{(c)} X_{t+1}^{(c)}, && \beta_\xi^{(c)} = \beta_{\xi 0}^{(c)} + \tfrac{1}{2} \textstyle\sum_{t=T_0^c}^{T_f^c-1} (\xi_{t+1}^{(c)} - \hat{\xi}_{t+1}^{(c)})^2 \end{aligned} \tag{14}$$

$$\begin{aligned} W_\psi^{(c)} \sim IG(\alpha_\psi^{(c)}, \beta_\psi^{(c)}), \quad \alpha_\psi^{(c)} &= \alpha_{\psi 0}^{(c)} + \tfrac{T^c-2}{2}, \\ \beta_\psi^{(c)} &= \beta_{\psi 0}^{(c)} + \tfrac{1}{2} \textstyle\sum_{t=T_0^c}^{T_f^c-1} (\psi_{t+1}^{(c)} - \psi_t^{(c)})^2 \end{aligned} \tag{15}$$

$$\begin{aligned} &\lambda^{(c)} \sim TN\left(m^{(c)}, C^{(c)}, 0, 0.88\right) \\ &m^{(c)} = \frac{\sum_{t=T_0^c}^{T_f^c-1} (\xi_{t+1}^{(c)} - \psi_{t+1}^{(c)} X_{t+1}^{(c)}) \xi_t^{(c)}}{\sum_{t=T_0^c}^{T_f^c-1} (\xi_t^{(c)})^2 + 1}, \quad C^{(c)} = \frac{W_\xi^{(c)}}{\sum_{t=T_0^c}^{T_f^c-1} (\xi_t^{(c)})^2 + 1} \end{aligned} \tag{16}$$

$\lambda \in \{0, 0.88\}$ is equal to $\{0, 5.44\}$ days for 50% effect decay.

C. SAMPLING DISTRIBUTIONS FOR Ω

We sample $\Omega|\theta_{1:T}, \Phi, D_{1:T}$ using Algorithm 1. $\Gamma(\alpha, \beta)$ is the Gamma distribution with shape α and rate β. We use a Dirichlet prior, α=1, for p.

$$\omega_t \sim \Gamma(\alpha_\omega, \beta_\omega), \quad \alpha_\omega = \tfrac{\eta_t + 1}{2}, \quad \beta_\omega = \tfrac{1}{2} \widetilde{V}^{-1} (Y_t - \widetilde{F}'\theta_t)^2 \tag{17}$$

$$\eta_t \sim p(\eta_t = i), \quad p(\eta_t = i) \propto \Gamma(\omega_t | \tfrac{i}{2}, \tfrac{i}{2}) p_i \tag{18}$$

$$p \sim Dir(\alpha + N_y), \quad N_y = [N_{y1}, \dots, N_{yL}]', \quad N_{yi} = \textstyle\sum_{t=1}^{T} (\eta_t = i) \tag{19}$$

We set η_t to be $\{1, 2, \dots, 10, 20, \dots, 50\}$ with cardinality, L. We sample η_t from Eq 18 using inverse transform sampling [6].