Anti-Ad Blocking Strategy: Measuring its True Impact

Atanu R. Sinha
Adobe Research
atr@adobe.com

Meghanath Macha
Carnegie Mellon University
mmacha@cmu.edu

Pranav Maneriker
Adobe Research
pmanerik@adobe.com

Sopan Khosla
Adobe Research
skhosla@adobe.com

Avani Samdariya
IIT Kanpur
savani@iitk.ac.in

Navjot Singh
IIT Bombay
navisngh11@gmail.com

ABSTRACT

The increasing use of ad blocking software poses a major threat for publishers in loss of online ad revenue, and for advertisers in the loss of audience. Major publishers have adopted various anti-ad blocking strategies such as denial of access to website content and asking users to subscribe to paid ad-free versions. However, publishers are unsure about the true impact of these strategies [2, 3]. We posit that the real problem lies in the measurement of effectiveness because the existing methods compare metrics after implementation of such strategies with that of metrics just before implementation, making them error prone due to sampling bias. The errors arise due to differences in group compositions across before and after periods, as well as differences in time-period selection for the before measurement. We propose a novel algorithmic method which modifies the difference-in-differences approach to address the sampling bias due to differences in time-period selection. Unlike difference-in-differences, we choose the time-period for comparison in an endogenous manner, as well as, exploit differences in ad blocking tendencies among visitors' arriving on the publisher's site to allow cluster specific choice of the control time-period. We evaluate the method on both synthetic data (which we make available) and proprietary real data from an online publisher and find good support.

CCS CONCEPTS

•Computing methodologies →Model development and analysis; •Mathematics of computing →Regression analysis;

KEYWORDS

Anti-Ad blocking, Difference in Differences, Endogenous Control Group Selection, User Modelling

ACM Reference format:
Atanu R. Sinha, Meghanath Macha, Pranav Maneriker, Sopan Khosla, Avani Samdariya, and Navjot Singh. 2017. Anti-Ad Blocking Strategy: Measuring its True Impact. In *Proceedings of ADKDD'17, Halifax, NS, Canada, August 14, 2017,* 6 pages.
DOI: 10.1145/3124749.3124756

ADKDD'17, Halifax, NS, Canada

DOI: 10.1145/3124749.3124756

1 INTRODUCTION

The use of ad blocking software grew by 41% globally from 2014 to 2015 [1]. In the US alone, ad revenue loss was estimated to reach $20.3 B in 2016 [1]. While digital publishers find this trend outright costly, it hurts newspapers as well since "A quarter of advertising revenue comes from digital" [14]. As for advertisers, this loss of audience to which they communicate messages is a major setback. The increasing threat posed by visitors' adoption of ad blocking software [1–4], has resulted in major publishers adopting various anti-ad blocking strategies. These include denial of access to website content, or, to subscribe to paid ad-free versions, or, to opt for subscription with few ads [3, 4]. Claims of effectiveness [2] and ineffectiveness [3] of anti-ad blocking strategies abound. Either way, quantifying the true effectiveness of mitigation strategies is a problem of great interest to publishers and advertisers alike. To date, scant attention has been paid to ad blocking by the research community (exceptions being [16–18]).

To appreciate the problem at hand it is instructive to view the data we obtained from the publisher. The data comprise ad blockers subjected to anti-ad blocking strategy, which is the treatment group. However, the data do not have a natural control group; that is, it lacks ad blockers not subjected to anti-ad blocking because the publisher implemented anti-ad blocking for all ad blockers. While the prudence of site-wide implementation can be questioned, in practice that has been the norm for many major publishers [2–4]. The absence of a natural control group in the data precludes consideration of existing statistical tools such as difference-in-differences (DiD), or, the usual treatment vs. control group comparisons. It is in the context of addressing the nascent ad blocking problem with this type of data at hand, we propose a practicable method which makes the following contributions. One, the method is ex post; that is, true effects can be uncovered from past, observational data without running new, costly experiments. Two, our novel approach allows for the endogenous selection of the control time-period from the available data. This extends the DiD method in a new direction. Three, it allows endogenous selection of clusters of visitors along with the selection of the time-period within the DiD framework, thereby recognizing heterogeneity in ad blocking tendencies among visitors. Hence, we obtain a measure of true effectiveness of the treatment by clusters of site visitors who vary in ad blocking tendencies. Four, we show that to quantify the effect of anti-ad blocking strategy the negative binomial regression model accounts for over-dispersion in count metrics, better than commonly used models such as linear and Poisson regressions.

We demonstrate that true effectiveness is related to cluster characteristics that underlie visitors' ad blocking tendencies. The real data from the publisher is limited in scope since obtaining this kind of sensitive data over a longer duration is very difficult, constraining our ability to do a direct evaluation. Thus, we use an indirect approach for evaluation, which is described in Section 4. Moreover, we implement the method on a synthetic dataset which is made available to researchers[1].

The rest of the paper is organized as follows. Section 2 shows the contribution with respect to the literature. Section 3 describes our methodology for calculating the effectiveness of a strategy and carry out visitor segmentation. In Section 4, we evaluate our technique on two datasets. We conclude with section 5.

2 RELATED WORK

We are unaware of any academic work that specifically addresses anti-ad blocking. We mention recent works adjacent to this area. In [18] the authors address how users interact with ads by monitoring HTTP requests of users on the network. Two other works [16, 17] study the prevalence of anti-ad blocking among the most visited websites on the internet (using their Alexa rankings [5]). None of these offers a method to measure true effectiveness of anti-ad blocking strategies. Our approach examines data based on site-wide implementation of anti-ad blocking strategy, as commonly observed, and does not ask publishers to run costly, new controlled experiments. The approach uncovers ex post from the available observed data, the true impact of strategies on desirable outcomes such as pageviews, number of visitors, etc. On the methodological side, the novelty lies in extending the DiD method [23], which becomes necessary when obtaining data on a control group is not possible.

For concreteness, consider that a publisher has implemented an anti-ad blocking strategy, labeled the 'treatment' condition. Let t_0 be the time of its implementation. The publisher's goal is to find effectiveness of the treatment on outcome metrics. Conventional wisdom and current practice suggest two types of comparison: (i) compare the outcome metrics in the after period $(t_0, t_+]$ with that in the before (baseline) period $[t, t_0)$; and / or (ii) compare the outcome metrics in the after period $(t_0, t_+]$ with that of the most recent past period of equivalent duration (baseline) period $(t_0 - t, t_+ - t]$ with $t_0 - t > 0$. Call the period after the treatment as the 'treatment period', and the baseline period as the 'control period'. Note that in the control period for both the above comparisons, the metrics are generated by visitors not subjected to the treatment of anti-ad blocking. We posit that both approaches are prone to errors in the context of traffic coming to a website. First, the comparisons are biased because the ceteris paribus condition does not hold across the treatment and control periods, in the internet news cycle of fast moving stories. Second, site visitors' reactions to an intervention such as anti-ad blocking are likely to be heterogeneous. Moreover, different types of visitor may come to the website in different frequencies and periodicities. A comparison across two exogenously set fixed time periods fails to recognize these nuances and that failure gives rise to sampling biases. A DiD approach overcomes one kind of sampling bias arising out of the differences between the treatment population and the control population. However, DiD assumes an exogenously determined control time-period and a natural control population. Our approach overcomes these sampling biases by modifying the DiD approach.

In the problem we address, finding a competing publisher's data (which could be the equivalent of the control population in DiD) is impossible for any publisher, besides other legal issues involved in sharing that kind of data. Also noteworthy is that non-ad blockers cannot be used as a control population because they fundamentally differ from ad blockers with respect to the problem being researched. The confinement to one publisher's dataset prompted us to innovate an approach by statistically choosing from a prior time-period a control group of ad blockers on whom anti-ad blocking was not applied. That control group has characteristics matching those of the treatment group, which is ad-blockers subjected to the anti-ad blocking strategy. For clustering, we borrow from the literature on integration of behavioral targeting into contextual marketing actions [1, 12]. For feature extraction, we find [6, 10, 21] helpful as they demonstrate different internal session features and genre sequences [15] to constitute the web behavior of a user. Also, we leverage the list of features from clickstream data provided in [7]. In the next section we provide technical details.

3 METHODOLOGY

3.1 Examining Effectiveness

Let Z = treatment, or, the anti-ad blocking strategy implementation. In the data only ad blockers are subjected to the treatment Z; that is, non ad-blockers are not subjected to Z. The period immediately after the implementation is referred as the treatment period. We use the word "group" to reflect the group of ad blockers who visit the site in the specified time period. We use treatment group (control group) for ad blockers visiting the site in the treatment period (control period).

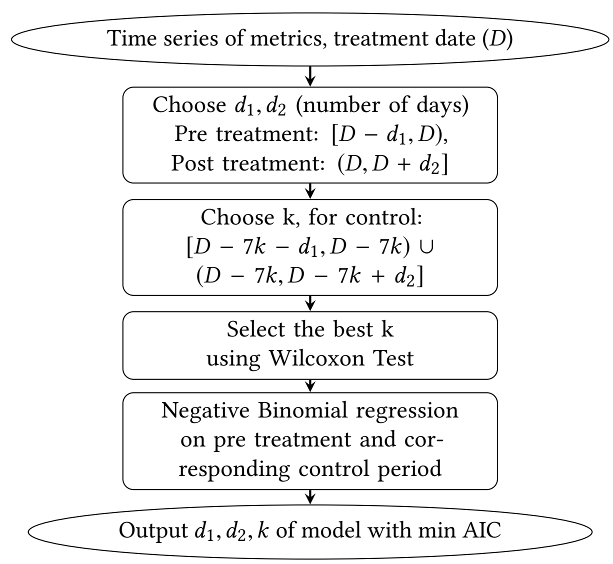

Figure 1: Summary of Effectiveness algorithm

[1] http://bit.ly/SyntheticDataAdblock

3.1.1 Control Group Selection. The first problem is to identify the control group which provides an appropriate baseline so that the effectiveness of treatment Z can be examined for an outcome metric. For the control group, our goal is to observe the behavior of ad blockers in a counter factual setting; that is, not subjected to Z. Hence, we need the behavior of ad blockers as if the anti-ad blocking strategy were not implemented, during the same time period as that of anti-ad blocking. Since all ad blockers are exposed to the treatment [18], for the choice of control group, we look for similarity of behavior of ad blockers in the time periods prior to the treatment. We emphasize that non-ad blockers cannot be used as a control population because they fundamentally differ from ad blockers with respect to the problem being researched.

While in the DiD estimation the treatment and control groups are concurrent, we modify this estimation by finding control groups in time periods prior to the treatment. Our model performs automatic control group selection from among many candidate control groups. These groups may vary by time periods over which to be observed, making the time period endogenously determined. The control group and treatment group contain the same series of days (e.g Sunday - Saturday). To ensure this, we choose time periods in multiples of seven days going backward from the treatment (Figure 1). The seven days reflect the natural weekly cycle. Readers' arrival to the site varies across weekdays and weekends. We select the best matching control group by measuring the web-behavioral similarity of the ad blockers in the control groups with those of the treatment group. This is done by first determining the variables that are relevant for matching control group and the treatment group. These variables include: browser, first touch marketing channel, page-name, referrer, types of article read, operating system, etc. Then we compare the distribution of visitors on those variables. The Wilcoxon Test is used to choose the best match among the candidate control groups [22]. The control group selection is conditioned upon the choice of clusters of visitors. In Section 3.2 we elaborate on how the clustering is achieved in our method.

3.1.2 Modified DiD estimation. Having identified the control group we propose estimating the effect using the modification necessitated by recognising that the control group belongs to a retrospective time period and not to the concurrent time period as the treatment group. We need to choose two suitable durations of time: immediately before and immediately after the intervention of anti-ad blocking strategy. Then we apply durations of same length to the control group. To control for effects of time of day and weekend dummy variables are used.

The regression thus becomes:

$$\begin{aligned} y =& \beta_0 + \beta_1 * timeperiod + \beta_2 * grouptype + \\ & \beta_3 * hour_{dummy1} + \beta_4 * hour_{dummy2} + \\ & \beta_5 * weekend + \beta_6 * timeperiod * grouptype + error \end{aligned} \tag{1}$$

where:

$$\text{timeperiod} = \begin{cases} 1, & \text{observation in post-treatment period} \\ 0, & \text{observation in pre-treatment period} \end{cases}$$

$$\text{grouptype} = \begin{cases} 1, & \text{observation in treatment group} \\ 0, & \text{observation in control group} \end{cases}$$

Although equation (1) uses a linear model, the approach captures any generalized linear model with a suitable function of y on the left hand side. The model allows variation in site visitations within a day as well as across weekdays and weekends through the self-evident dummy control variables. This recognizes the empirical evidence that the KPIs of a website depends on the hour of day as well as weekend/weekday. We find that the estimates of these dummy variables are significant confirming our hypothesis. The estimate for our effect is shown as follows. The first equation below finds the difference within treatment group between post and pre-treatment time periods. The second equation below finds the difference within control group between post and pre-treatment time periods. The third equation finds the difference between the first and second equations to arrive at the desired effect.

$$\begin{aligned} & (y_{treatment,post}) - (y_{treatment,pre}) \\ & = (\beta_0 + \beta_1 + \beta_2 + \beta_6) - (\beta_0 + \beta_2) = (\beta_1 + \beta_6) \end{aligned} \tag{2}$$

$$\begin{aligned} & \left(y_{control,post}\right) - \left(y_{control,pre}\right) \\ & = (\beta_0 + \beta_1) - (\beta_0) = (\beta_1) \end{aligned} \tag{3}$$

The true effectiveness is thus (2) − (3) = β_6,

The baseline, consistent with common industry practice, is given by $\left[(y_{treatment,post}) - (y_{treatment,pre})\right] = (\beta_1 + \beta_6)$

Depending upon signs of the parameters, estimates of the true effect can be less than or greater than the baseline.

3.1.3 Model for the dependent variable. Generally, the KPIs of interest to a publisher are count metrics such as page views, which are observed to be over-dispersed (variance larger than mean). Poisson distribution works well for count variables, but does not account for over-dispersion. Negative binomial regression captures over-dispersion (captures the variability in a quadratic fashion [9, 11]) and we empirically find that the negative binomial performs better in terms of the fit statistics than the Poisson regression model (using AIC, BIC, and Normal Q-Q plots). See results in Table 1. The negative binomial regression model is written as [9]:

$$f(y_i|x_i) = \frac{\Gamma(y_i + \theta)}{\Gamma(\theta) y_i!} \left(\frac{\theta}{\theta + \mu_i}\right)^{\theta} \left(\frac{\mu_i}{\theta + \mu_i}\right)^{y_i} \tag{4}$$

$$\begin{aligned} E(y_i) = \mu_i \quad \text{and} \quad Var(y_i) &= \mu_i[1 + \frac{1}{\theta}\mu_i] \\ &= E(y_i)[1 + \frac{1}{\theta}E(y_i)] > E(y_i) \end{aligned} \tag{5}$$

As shown, the above accounts for over-dispersion. The above model is the Gamma mixture of Poisson rate, a.k.a., Gamma-Poisson model. It can be obtained from the Poisson model by including unobserved heterogeneity [9, 11] in the following manner:

$$E(y_i \vee \varepsilon_i) = \mu_i \varepsilon_i \tag{6}$$

where ε_i is the error term distributed as $\Gamma\left(1, \frac{1}{\theta}\right)$.

3.2 Heterogeneity among visitors

Since ad blocking tendencies potentially reflect differences in behavioral characteristics, we incorporate heterogeneity as follows. Our premise is that the publisher has access to hit level data collected through its site. The hit level data are pre-processed in order to filter crawlers and empty rows and later converted into a sessionized, user level data. Since our model focuses on web users, mobile users are omitted. That said, the approach easily extends to mobile users. The steps in our method are described below:

3.2.1 Feature Engineering. Apart from out-of-the-box fields provided in the data, we engineer the following features to capture both behavioral as well as static features of a user's web profile:

- **Tags:** Publisher websites contain a basic set of tags associated with articles published. In our data, the publisher defines 10 different tags; e.g., Culture, Technology. To capture tag based information for empty fields, meta-content like URL, content title, page title, are parsed and mapped with the existing tags. Extraction of content (tag) related information leads to creation of features that allow the model to capture the reading interests of a user. Some features are listed below:
  - **Average time spent on a tag:** For each browsing session of a user we find the amount of time (in seconds) spent by a user on a tag, normalized by the number of times she visited a page with that tag.
  - **Bi-grams and tri-grams:** To capture the user's interest in reading articles in sequence, we extracted bi-grams and tri-grams from user sessions.
- **Number of hits per visit:** To capture user's interaction with the webpage, we capture number of hits recorded per session.
- **Browser/OS family:** In order to capture variations due to browser/OS preferences, we heuristically group the browsers and OS into families. The browser families comprise Explorer, Chrome, Firefox, Safari, Opera, Edge, Others and the OS families are Linux, Windows and Mac-OS.
- **Browser/OS versions:** The version of the particular browser or OS is also incorporated into the model. The version is an instrument for the technological savviness of the user under the premise that more savvy users update to the latest version. It helps the model uncover relation between ad-blocking tendencies and technological savviness.

3.2.2 Feature Buckets. The features are grouped into the following buckets to capture different aspects of web profile.

- **Loyalty:** Loyalty of a user towards the website encapsulates the features such as visit number, total time spent, read five pages.
- **Reading Interest:** Features pertaining to the reading behavior of a user include frequency of visit on culture related pages, average time spent on tech based articles, etc.
- **Technology:** Technological aspect of a user's profile includes information about user's Browser/OS version and family, cookie, user-agent, JavaScript version etc.
- **Geo-Segmentation:** A user's geographic location comprises country, region, city (Tier1, Tier2 etc.), language, etc.

3.2.3 Unsupervised Clustering. The above-mentioned feature buckets are used as input to Unsupervised clustering. We used K-means clustering [13] algorithm with Euclidean distance. Column values are scaled to give equal importance to each of them (since feature importance may vary for different publishers). The stability of clusters is assessed using silhouette scores [19]. The normalized silhouette scores presented in Figure 2, justify our selection of six clusters. We note that K means and silhouette score generate appropriate clusters for the size of our dataset. However, for very large datasets we can use approximate methods such as Mini-batch clustering [20] along with Calinski-Harabasz criterion [8] (which scales linearly with data).

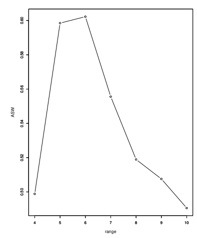

Figure 2: Silhouette scores for different number of clusters

4 EVALUATION

The evaluation is performed on two separate datasets. The first dataset is synthetic, but mirrors the statistical distributional characteristics of the second, real life dataset, which is proprietary. The synthetic data allows us to validate the proof of concept. Then we apply the method to the real life data to further validate the model and present more insights relevant for the publisher. For reasons of confidentiality about the publisher we cannot disclose the exact date of implementation of its anti-ad blocking strategy.

4.1 Evaluation - Synthetic Data

We validate the negative binomial model for our over-dispersed count data. The comparison shown is against a linear model. Separately, a comparison with the Poisson model also supports our thesis and is available upon request. Since the synthetic data are also made available for the community the results can be reproduced for any of these models.

In Table 1 the models presented under Method 1 are a formalization of the first conventional comparison, namely, compare the outcome metrics in the period $(t_0, t_+]$ with that of the baseline period $[t_-, t_0)$. Further we use control variables to account for time of day and weekend effect, as applicable. Method 2 refers to the comparison of the outcome metrics in the after period $(t_0, t_+]$ for the treatment group with that of the past endogenously selected control group for a period of equivalent duration. Note that the it is wrong to compare the overall goodness of fit measures AIC / BIC across the methods because the baseline data for each method are different from that of the other method. The goodness of fit comparisons are meaningful only within a method and across the models.

4.2 Evaluation - Real Life Data

4.2.1 Model Free Evidence. To give readers a feel for the data we start with model free evidence. Consider the metric PageViews (defined as the number of pages viewed). Control_1 is our endogenously chosen baseline. Control_2 is an industry practice in which the control group is an equivalent duration as in the treatment

Table 1: Regression Model Comparison for Synthetic Data

Page Views

	Method1		Method2		Method3	
	linear	neg. bin.	linear	neg. bin.	linear	neg. bin.
Intercept	828.19**	6.77**	1058.70**	7.09**	724.74**	6.60**
timeperiod	26.05	-0.03	NA	NA	176.46**	0.30**
grouptype	NA	NA	-110.24*	-0.29**	14.58	-0.01
dummy_1	-550.93**	-1.27**	-719.02**	-1.33**	-479.87**	-1.12**
dummy_2	-33.98	-0.07	-148.54*	-0.22*	84.81**	0.10*
weekend	-359.23**	-0.83**	NA	NA	-269.68**	-0.54**
t:g[2]	NA	NA	NA	NA	-124.82*	-0.23**
AIC	3382.04	3178.97	2086.59	2005.50	6642.84	6385.93
BIC	3402.93	3204.01	2101.43	2026.37	6676.23	6419.33

Significance codes: $* < 0.05$, $** < 0.01$

Time Spent Per Visit

Method1		Method2		Method3	
linear	neg. bin.	linear	neg. bin.	linear	neg. bin.
60515.43**	11.04**	73653.99**	11.23**	57826.54**	10.98**
-5827.46**	-0.09*	NA	NA	14491.36**	0.26**
NA	NA	-16840.44**	-0.30**	1748.02	0.00
-26873.14**	-0.75**	-30890.23**	-0.65**	-28182.26**	-0.72**
15262.84**	0.21**	12903.21**	0.18**	14203.51**	0.20**
-29235.29**	-0.73**	NA	NA	-23179.06**	-0.53**
NA	NA	NA	NA	-18588.47**	-0.30**
5244.71	5194.48	3102.66	3117.91	10432.35	10395.97
5265.59	5219.53	3117.51	3138.78	10465.74	10553.54

Significance codes: $* < 0.05$, $** < 0.01$

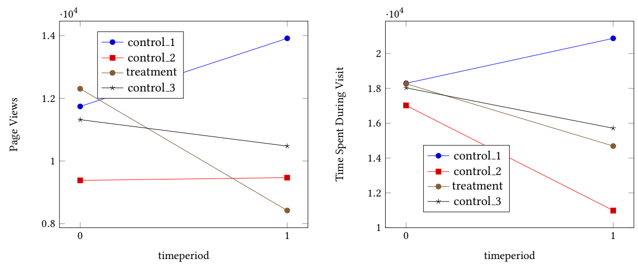

Figure 3: Real Life Data - Aggregate Statistics for Adblockers

Table 2: Real Life Data - Aggregate level model for Visitors

Visitors	Method3	
	linear	neg. bin.
Intercept	379.69**	5.92**
timeperiod	129.16**	0.43**
Grouptype	46.68**	0.13**
dummy_1	-267.12**	-1.16**
dummy_2	13.81	0.03
weekend	-123.11**	-0.44**
t:g	-174.82**	-0.55**
AIC	5896.92	5730.64
BIC	5930.31	5764.03

Significance codes: $* < 0.05$, $** < 0.01$

group, but selected from the most recent preceding time. Control_3 represents another practice whereby multiple equivalent duration are selected going back in time and then averaged. From Figure (3) we find that in the treatment period, both conventional comparisons - within treatment group, across time period; and within timeperiod=1 across treatment and control_2 as well as control_3 - show a decrease in PageViews; although the decrease is modest. Our endogenous choice of control group (control_1) baseline confirms that; however, the extent of the decrease is much greater due to anti-ad blocking as seen by within timeperiod=1 across treatment and control_1. Hence, the finding should concern the publisher much more than if it were to use the conventional comparison. Now considering another metric, TimeSpentDuringVisit, the comparison within treatment group across time shows a decrease, and so does for timeperiod=1 across treatment and control_3. However, the comparison within timeperiod=1 across treatment and control_2 groups shows an increase. Which is the valid effect? Our endogenous choice of control based comparison - within timeperiod=1 and across treatment and control_1 groups - confirms the decreased effect and shows it is substantial. In timeperiod=1, control_3 shows that both metrics decrease in the treatment group, a finding which is consistent with that of the endogenously chosen control_1. In summary, the conventional comparison method may give ambiguous answers in some cases (e.g., TimeSpent), and in other cases underestimate the effect size. Our method provides answer that is consistent across metrics, confirms the effect is a decrease, and furthermore shows that the effect size is more substantial than what the conventional method shows. Hence, the true magnitude of this adverse effect should concern the publisher.

4.2.2 Model based evidence. After showing that the proposed model improves upon existing comparisons in measuring true effectiveness, we demonstrate that our approach to recognize heterogeneity among visitors is valuable. Recall that we posited cluster characteristics may reflect ad blocking tendencies which impacts measure of true effectiveness. For space constraints, we consider only a single outcome metric, Visitors (number of unique visitors), a first-among-equal metric for publishers. The aggregate analysis shows that the Negative Binomial outperforms the linear model (AIC / BIC: 5730 / 5764 and 5897 / 5930, respectively). The real life data validate the Negative Binomial model's use, consistent with the synthetic data.

Going forward, for the sake of brevity, the results presented will be for the Negative Binomial model only. We find that [Table 2] there is a decrease in visitors ($-0.12 = 0.43 - 0.55$) in $(t_0, t_+]$ relative to the baseline period $[t_-, t_0)$ and a decrease ($-0.42 = 0.13 - 0.55$) compared to the outcome metrics in the algorithmically selected baseline period $(t_0 - t, t_+ - t]$. However, the measure of true effectiveness, as depicted in the interaction term is -0.55. That is, the magnitude of the decrease in visitors, as a consequence of the anti-ad blocking strategy, is actually more pronounced than what either of the naive comparisons suggests.

We now analyze at the level of clusters. This is important since aggregate level analysis can cover up for cluster level differences. As confirmation, the results [Table 3] for cluster 4 show a non-significant true effectiveness ($\beta = 0.11, p = 0.23$). The other five clusters depict significantly negative true effectiveness which varies

[2]timeperiod:grouptype

Table 3: Real Life Data - Cluster level model for Visitors

Visitors	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6
Intercept	5.09**	4.80**	5.21**	3.10**	2.62**	3.61**
timeperiod	0.46**	0.49**	0.44**	0.07	0.50**	0.27**
grouptype	0.25**	0.17**	0.19**	-0.20**	-0.16**	0.02
dummy_1	-1.22**	-1.18**	-1.21**	-0.81**	-1.17**	-1.04**
dummy_2	-0.05	0.05	-0.05	0.24**	0.05	0.21**
weekend	-0.43**	-0.45**	-0.40**	-0.44**	-0.31**	-0.56**
t:g	-0.59**	-0.66**	-0.56**	0.11	-0.54**	-0.40**

Significance codes: $* < 0.05$, $** < 0.01$

Table 4: Cluster Details

	Visitors (%) of total	Visits (%) of total	End of Article Reached (%)	Viewed 5 Pages (%)
Cluster1	32.44	32.41	8.45	0.35
Cluster2	22.51	22.24	90.22	0.04
Cluster3	34.69	34.63	11.47	0
Cluster4	2.18	2.15	92.72	29.17
Cluster5	1.94	1.90	56.52	0.56
Cluster6	6.24	6.66	26.57	1.56

from -0.39 to -0.66. These are different from the aggregate measure of -0.55.

4.3 Cluster properties - Real Life Data

In the absence of data to directly evaluate our method, we resort to an indirect evaluation. We relate cluster properties to the true effectiveness estimates. Table 4 shows that Cluster 4 stands out from others in its engagement with the publishers, as measured by the metric Viewed 5 Pages (29.17%). It is expected that this cluster will show the least effect from anti-ad blocking. The Table 3 finds the true effectiveness parameter for this cluster to be non-significant ($\beta = 0.11, p = 0.23$), corroborating that this cluster is least impacted by the treatment Z. The next highest on engagement Viewed 5 pages (1.56%) is cluster 6, which has a significant negative effect ($\beta = -0.40, p < 0.01$), but which is an appreciably smaller effect in magnitude from those of the effectiveness parameter estimates for the Clusters 1, 2, 3 and 5. While we do not have the ability to do an actual evaluation on the publisher's data, these provide indirect evidence of the validity of our approach: that endogenous time-period selection along with cluster level analyses matter and our clustering provides meaningful results.

5 CONCLUSION

In this application, we propose a method to quantify the true effect of an anti-ad blocking strategy implemented by a publisher. We show the method produces more accurate estimates of the true effectiveness of strategy implementation. For mitigation of ad blocking, the technology proposed can become useful for publishers because proper measurement is a necessity for managing a phenomenon. We find that the true effectiveness is on average substantially more negative than what the conventional measurements show. Importantly, we present a new method that identifies heterogeneity among visitors and then for each segment of visitors endogenously selects the proper control group going back in time. The data show how segment heterogeneity manifests in differences in effectiveness of anti-ad blocking actions of publishers. While we do not have an answer to prevent ad blocking, we provide a tool that can show how to find segments that are likely to respond least negatively to anti-ad blocking. Provided appropriate data are available (e.g., from controlled testing) it should be possible to perform a direct evaluation of our proposed approach, a goal we have for the future.

REFERENCES

[1] 2015. *The cost of ad blocking.* Technical Report. PageFair and Adobe.

[2] 2016. Axel Springer says it is winning fight against adblockers. http://www.ft.com/cms/s/0/0e58d68a-8305-11e5-9dc0-186bb1146746.html. (2016). Online; accessed 14 July 2016.

[3] 2016. Sites that block adblockers seem to be suffering. https://thestack.com/world/2016/04/21/sites-that-block-adblockers-seem-to-be-suffering/. (2016). Online; accessed 14 July 2016.

[4] 2016. Ways Publishers have reacted to ad blockers. https://searchenginewatch.com/2016/03/15/seven-ways-publishers-have-reacted-to-ad-blockers/. (2016). Online; accessed 14 July 2016.

[5] 2017. Alexa Top Sites. https://alexa.com/topsites. (Jan. 2017). Online; accessed 29 January 2017.

[6] M. Abramson and D. W. Aha. 2013. User Authentication from Web Browsing Behavior. In *Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference (FLAIRS '13)*. 268–273.

[7] Eugene Agichtein, Eric Brill, and Susan Dumais. 2006. Improving Web Search Ranking by Incorporating User Behavior Information. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*. 19–26.

[8] Tadeusz Caliński and Jerzy Harabasz. 1974. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* 3, 1 (1974), 1–27.

[9] William H Greene. 1994. Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. (1994).

[10] Jian Hu, Hua-Jun Zeng, Hua Li, Cheng Niu, and Zheng Chen. 2007. Demographic Prediction Based on User's Browsing Behavior. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*. 151–160.

[11] Zvi Griliches Jerry Hausman, Bronwyn H. Hall. 1984. Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica* 52, 4 (1984), 909–938. http://www.jstor.org/stable/1911191

[12] Ting Li, Ning Liu, Jun Yan, Gang Wang, Fengshan Bai, and Zheng Chen. 2009. A Markov chain model for integrating behavioral targeting into contextual advertising. In *Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising*. ACM, 1–9.

[13] James MacQueen and others. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1. Oakland, CA, USA., 281–297.

[14] G. Stocking Mitchell, A and K.E. Masta. 2016. Long-form reading shows signs of life in our Mobile News World. (5 May 2016).

[15] Alan L. Montgomery, Shibo Li, Kannan Srinivasan, and John C. Liechty. 2004. Modeling Online Browsing and Path Analysis Using Clickstream Data. *Marketing Science* 23, 4 (2004), 579–595.

[16] Muhammad Haris Mughees, Zhiyun Qian, Zubair Shafiq, Karishma Dash, and Pan Hui. 2016. A First Look at Ad-block Detection: A New Arms Race on the Web. *arXiv preprint arXiv:1605.05841* (2016).

[17] Rishab Nithyanand, Sheharbano Khattak, Mobin Javed, Narseo Vallina-Rodriguez, Marjan Falahrastegar, Julia E Powles, Emiliano De Cristofaro, Hamed Haddadi, and Steven J Murdoch. 2016. Ad-Blocking and Counter Blocking: A Slice of the Arms Race. *arXiv preprint arXiv:1605.05077* (2016).

[18] Enric Pujol, Oliver Hohlfeld, and Anja Feldmann. 2015. Annoyed Users: Ads and Ad-Block Usage in the Wild. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*. ACM, 93–106.

[19] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[20] David Sculley. 2010. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*. ACM, 1177–1178.

[21] Kosuke Takano and Kin Fun Li. 2009. An Adaptive Personalized Recommender Based on Web-Browsing Behavior Learning. In *Proceedings of the 2009 International Conference on Advanced Information Networking and Applications Workshops (WAINA '09)*. 654–660.

[22] Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics bulletin* 1, 6 (1945), 80–83.

[23] Jeffrey Wooldridge and Guido Imbens. 2007. What's new in econometrics? Difference-in-differences estimation. *Lecture Notes* (2007).