

TV Impact on Online Searches

Ying Liu, Yonathan Schwarzkopf, Jim Koehler

Google Inc.

August 31, 2017

Abstract

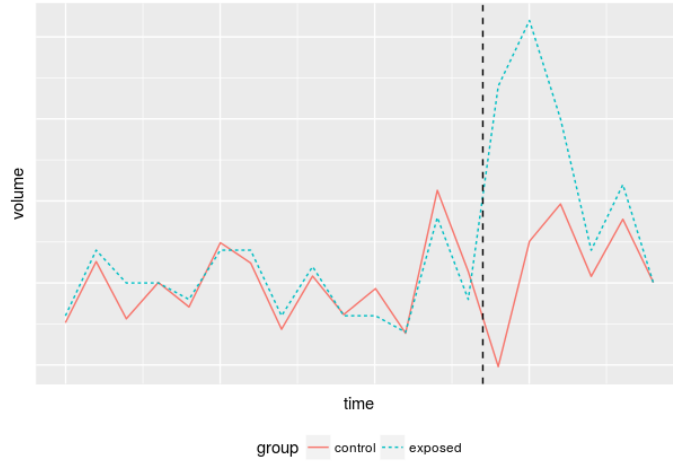
We study the impact of TV advertising on viewers’ online search behaviors. In particular, we develop methods to estimate how many incremental searches can be causally attributed to TV ad spots, based on Bayesian Structural Time Series (BSTS) models. Simulation studies show that the TV-induced incremental search volumes can be accurately estimated in most cases. Our work provides a way of comparing incremental searches from TV v.s. web video ads. We demonstrate our methods with a case study, and present some insights from the analysis.

1 Introduction

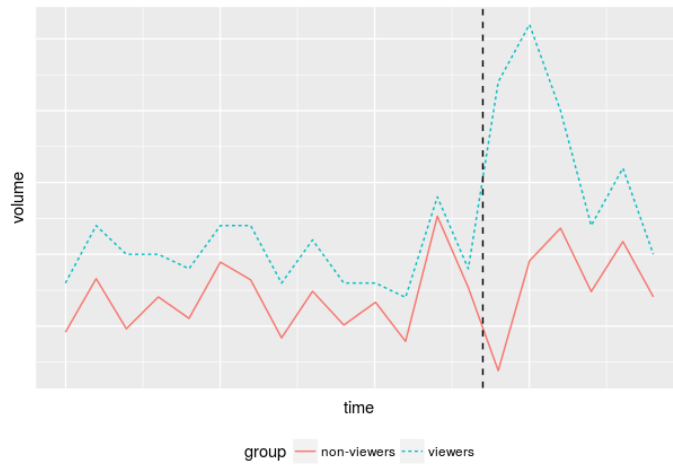
TV advertising is one of the most common forms of a marketing campaign. Much research has been conducted to measure the effectiveness of TV campaigns [3, 8]. The most direct metric is the viewership of TV ads (such as impressions and reach), which provides information on the size of the campaign audience. Survey based approaches have been widely used to measure brand metrics such as purchase intent [2]. However, such surveys are difficult to conduct, and suffer from problems such as non-responses. Ken et al. looked at queries that matched movie titles and were able to correlate search series with mentions of the movie titles in ads or TV content [6]. In this study we propose studying viewers’ interest in the brand (or related topics) as advertised in the TV ads, using their online searches that are related to the TV ads (such as brand, product or ad creative related queries). In particular, we look at minute-by-minute aggregated search volume time series during a TV campaign, and develop methods to measure the incremental searches that can be causally attributed to the ad spots of the campaign.

The proposed method in this paper follows the Rubin Causal Model framework [11]. In an ideal world, randomized experiments would be practical for measuring the causal effect of TV ads [7]. We would randomly split the audience into two groups, and expose one to a campaign and use the other as a control. Data from one ad spot in such an experiment may look like Figure 1(a), where the dashed vertical line indicates the ad spot, and the blue (dashed) and the red curves show the search volume time series from exposed and control groups, respectively. Then the difference between the two curves in a *post period* after the ad is the estimated incremental searches that are caused by this spot, given the same sizes of the exposed and control groups. However, current TV advertisement is not personalized, so this random assignment is not feasible.

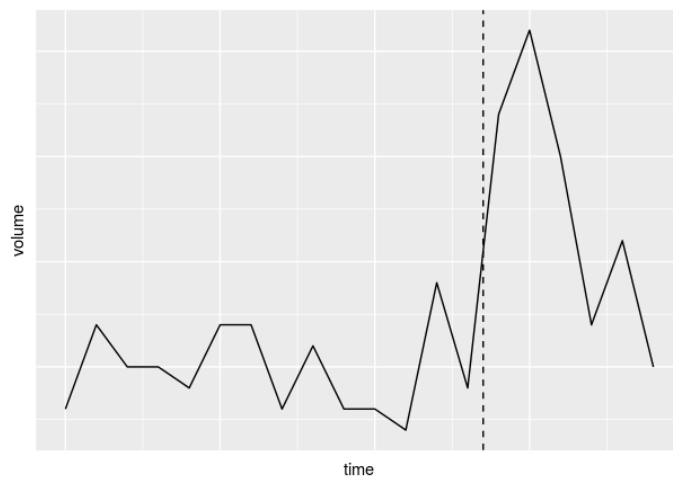
If we could track the ad viewership for a group of TV viewers, we might get data like Figure 1(b), where the blue (dashed) curve is the search volume time series from the ad viewers, while the red curve is from the non-viewers. With such data, we could estimate the campaign effect, but taking care to reduce bias—it is very likely that the normal search behaviors of the two groups of audiences



(a) Potential data from a randomized experiment



(b) Searches from ad viewers and non-viewers



(c) Aggregated search volume time series from all the online population

Figure 1: Scenarios for measuring incremental searches due to TV ads

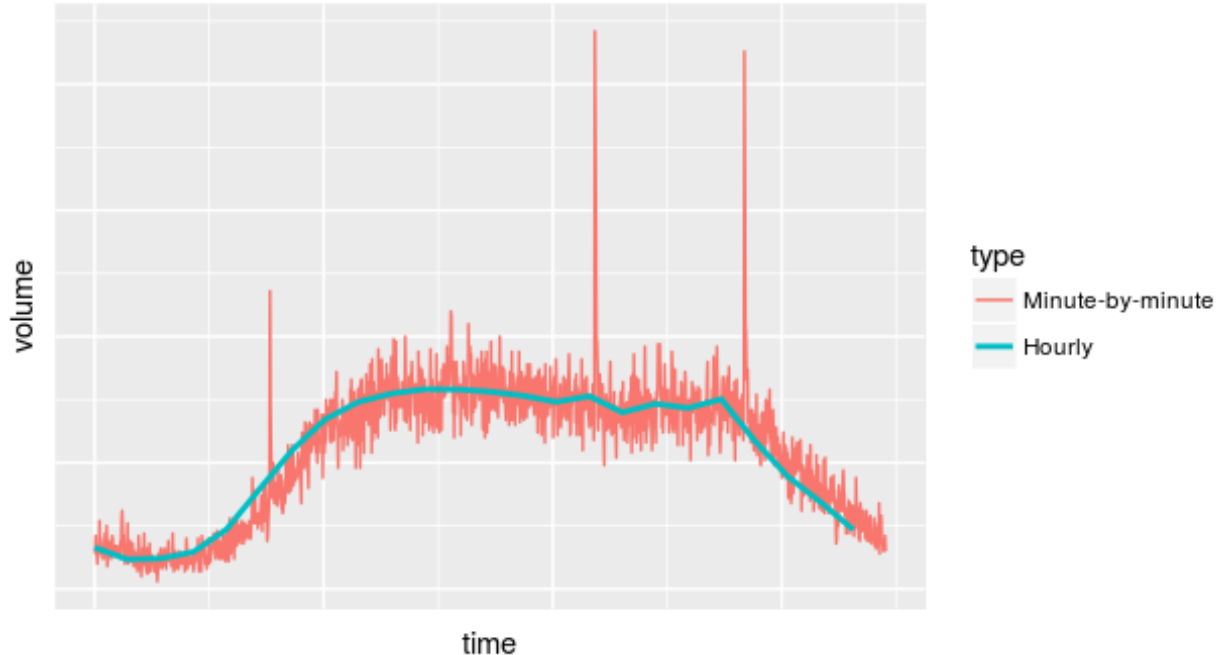


Figure 2: Search time series at different aggregation levels

are intrinsically different (as demonstrated by the systematic difference in the “pre-period” before the ad airs), resulting in a self-selection bias. Causal methods such as propensity scores [10] are commonly used to remove such bias, provided that additional covariates are available. However, individualized ad viewership data is still needed for such modeling, which can be very difficult to get.

In this paper we propose analyzing aggregated search time series from the total online population, irrespective of ad exposure. The data might look like that in Figure 1(c). Two characteristics of the data are worth mentioning. First, the measurable ad impact on searches is usually transient (Figure 1(c)). Second, the finer the granularity of the time series, the more observable the search spikes (if any) are, as shown in Figure 2. Thus we use minute-by-minute search volume time series in this research.

Our approach produces incremental search (“search lift”) estimates for each TV ad spot. As you might imagine, when spots are very dense, the impact from close-by spots may overlap. We propose grouping dense spots in such circumstances (see Section 2).

The paper is organized as follows: in Section 2 we describe our methods based on BSTS, taking special care in scenarios where the TV ad schedule is dense; in Section 3 we conduct systematic simulation studies that demonstrate the performance of the proposed methods; in Section 4 we show the results from a case study; in Section 5 we compare search lift from TV v.s. web video ads, using results from both the case study; Section 6 concludes with discussion.

2 Methods

We align the TV ad schedule and the aggregated search data to estimate the incremental searches that can be attributed to TV ad spots. The basic idea is to predict the baseline search volume in absence of TV ads, or the “counterfactual”, and then take the difference between the observed

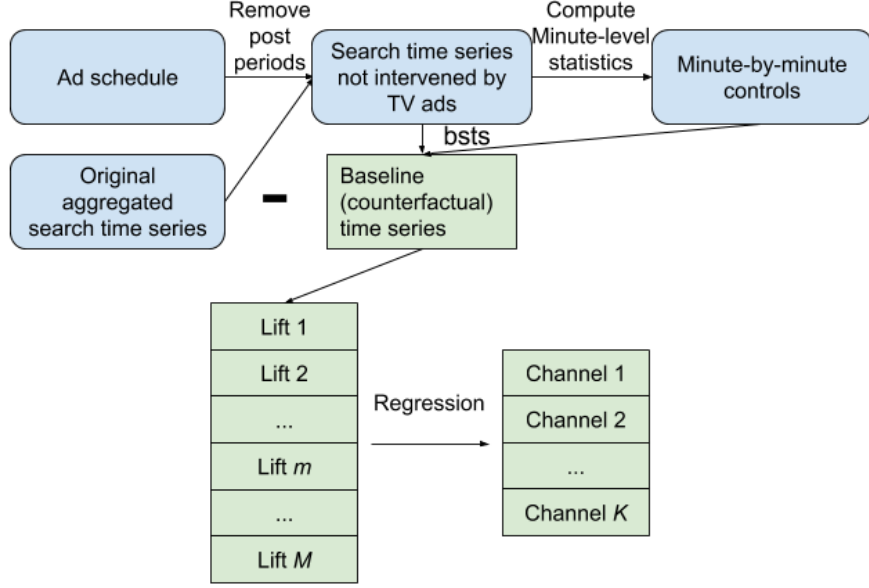


Figure 3: Diagram of the TV search lift methodology (where blue boxes represent data and green boxes are output from our models), details of which will be discussed throughout Section 2

volume and the counterfactual as the estimate of the incremental searches. We do this by removing a post period after each ad spot from the original search volume time series, fitting the remaining curve, and imputing the missing data. We also require confidence (or credible) interval estimates so that we can make statistical significance claims. These considerations bring us to the Bayesian Structural Time Series (BSTS) model [12] that has the ability to fit time series with missing data and generate credible intervals.

BSTs models are state space models for time series estimated using Bayesian methods. They have two components: a time series component and a regression component. An observation y_t at time t (in our case minute-by-minute search volume time series) is linked to the state space through the *observation equation*

$$y_t \sim \text{Poisson}(\exp(\mu_t + \mathbf{x}_t^T \boldsymbol{\beta})), \quad (1)$$

where \mathbf{x}_t is a d -dimensional control time series at time t (in our case minute-by-minute daily search volume time series to capture any daily patterns in searches detailed in Section 2.1) and $\boldsymbol{\beta}$ is a vector of regression coefficients. The state space is a random walk defined by the *state equation*

$$\mu_t = \mu_{t-1} + \epsilon_t \quad (2)$$

where ϵ_t 's are i.i.d. normally distributed random variables.

The control series are selected using a spike-and-slab prior [4, 5, 9, 12] and used in the regression component. Specifically, the “spike” is a point mass at zero that shrinks a subset of coefficients to zeros, and the “slab” is a weakly informative distribution on the complementary set of nonzero coefficients. Thus each draw of posterior coefficients has a subset of zeros (hence achieving the goal of variable selection), and the final model is an average of all the models each with a set of posterior coefficients.

Readers can refer to [12, 1] for the model and estimation details. Figure 3 shows a diagram of the TV search lift estimation methodology, which we discuss in detail in the following sections.

2.1 Fitting a BSTS model using non-contaminated data

To estimate the baseline search volume that would occur in the absence of the TV ad campaign, we fit the BSTS models to data that excludes a post period after each TV ad spot, within which search volumes may be affected by TV. Our research shows that TV’s impact on search volume is transient in time in most cases, usually less than 10 minutes. Thus we set the post period to 8 minutes by default, and extend it when necessary (we discuss this in Section 2.3). We set the values of the search time series within the post periods to missing when fitting the BSTS models.

We calculate the following minute-level time series across different days using the above remaining data:

- Average volume
- Median volume
- First and third quartiles of the volume

Take average volume for example. For each minute it is the average search volume of this minute across days that do not have this minute in a post period. Figure 4 shows an example with three days. First we remove post periods, then for each minute we take an average of the volume across three or two days (depending on whether that minute is within a post period). These minute-by-minute summary statistics provide information on the typical daily search patterns. So we use them as the candidate controls for the regression in BSTS, and these controls are selected by the spike-and-slab prior depending on their correlation to the original non-contaminated time series.

We fit a local-level BSTS model with the Poisson observation equation (1) on the original series with post-ad periods removed, plus the above four daily time series as controls in the regression component. This results in a baseline (counterfactual) time series with the missing values in the post periods imputed. Then the lift for a time period (e.g. a post period) is estimated as the difference between the observed number of searches during the post period and the sum of the counterfactual over that period. Figure 5 demonstrates the estimated counterfactuals from BSTS with non-contaminated data, with shaded areas representing post periods.

2.2 Ad groups

The case where multiple ads occur quickly bears special attention. We define two or more ads with separations smaller than the length of a post period to be an ad spot group. We use counterfactuals between the time of the first spot in the group to the end of the post period of the last spot in the group to calculate the total lift for that group. Let t_1 be the time of the first ad spot within an ad group m , and t_k be the time of the last spot within the group. We calculate one lift number l_m for the period $[t_1, t_k + post_period]$.

The lift estimate above should be prorated to each ad spot within the group. We fit a linear regression model to the spot and spot-group level lifts, using TV channel and time of day (and optionally creative IDs) as predictors:

$$l_m = \alpha_0 + \sum_{\text{spot } j \text{ in group } m} (\alpha_{ch_j} \text{Imp}_j + \gamma_{dp_j}) + \epsilon_m, \tag{3}$$

where ch_j , dp_j and Imp_j are the TV channel, time of day (hour of day) and the ad impression of spot j in group m , thus α_{ch_j} is the response rate for ch_j , and γ_{dp_j} is the “baseline” lift for time of day dp_j . Figure 6 illustrates a spot group with four airings, with the shaded area the post period for this group.

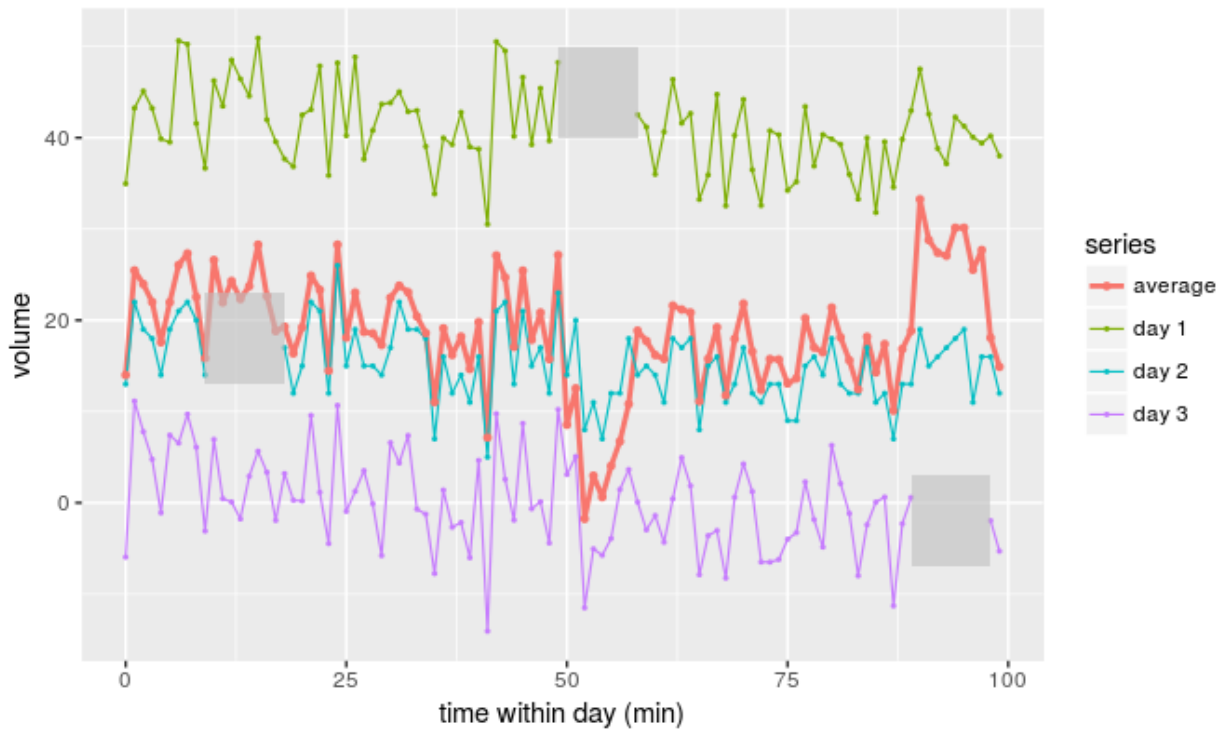


Figure 4: Generating control series

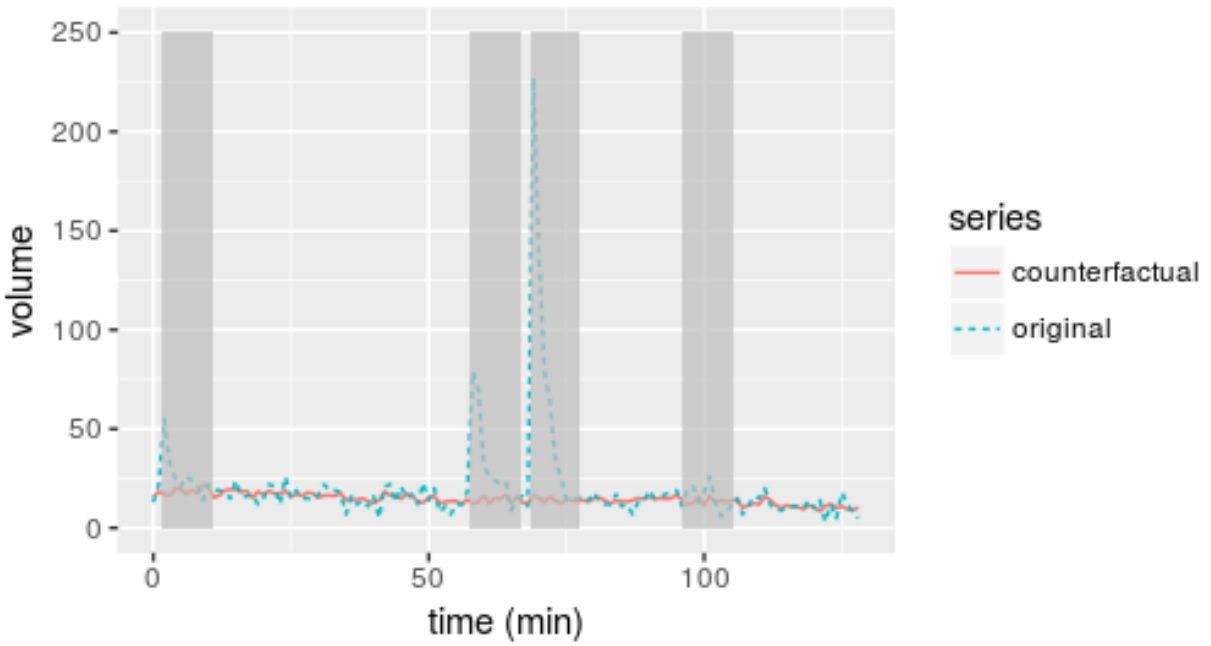


Figure 5: Original and counterfactual search volume time series, where shaded areas represent post periods

Time	Channel	Impressions
20:00	CHIL	60K
20:06	CBS	8000K
20:10	BET	170K
20:16	HGTV	70K

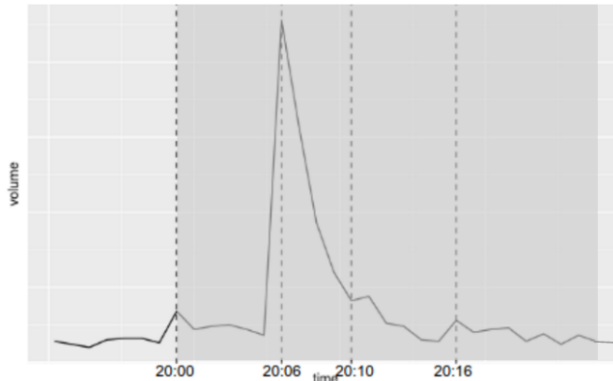


Figure 6: TV schedule and search volume time series for a spot group with four airings

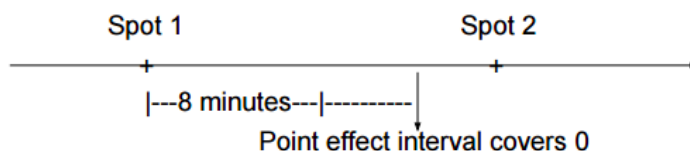


Figure 7: Adaptive post periods

We then use the predicted lift from the above regression to allocate the total search lift into each of the spots in a group. Specifically, take a group with two spots with total lift estimate l_m as an example. Suppose the two spots are in dp_1 and dp_2 , and aired on ch_1 and ch_2 , respectively. We allocate the total lift to ad spot 1 as (similarly for spot 2)

$$l_m \times \frac{\alpha_0 + \alpha_{ch_1} \text{Imp}_1 + \gamma_{dp_1}}{(\alpha_0 + \alpha_{ch_1} \text{Imp}_1 + \gamma_{dp_1}) + (\alpha_0 + \alpha_{ch_2} \text{Imp}_2 + \gamma_{dp_2})}. \quad (4)$$

2.3 Adaptive post periods

The impact from TV ad spots could last more than 8 minutes (the default length of the post periods). We look for evidence of longer effects, and adjust the post periods accordingly, from the current default value (8 minutes) to the point at which the impact is no longer significant (Figure 7), except truncated at the following ad spot (when we extend periods we do not form spot groups) and with a maximum of four hours.¹ In practice this rarely results in periods longer than 8 minutes.

2.4 Aggregating spot-level lift estimates

The spot-level lift estimates can be aggregated by different data slices, such as TV channels. For example, let $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_N$ be lift estimates for each of the total N ad spots, aired on channels ch_1, ch_2, \dots, ch_N , respectively. For spots that are part of groups, these are the result of the allocation process in equation (4). Then the estimated search lift for a specific channel ch is

$$\sum_{\{i|ch_i \in ch\}} \hat{\lambda}_i. \quad (5)$$

¹This is consistent with Google's YouTube search lift measurement methodology.

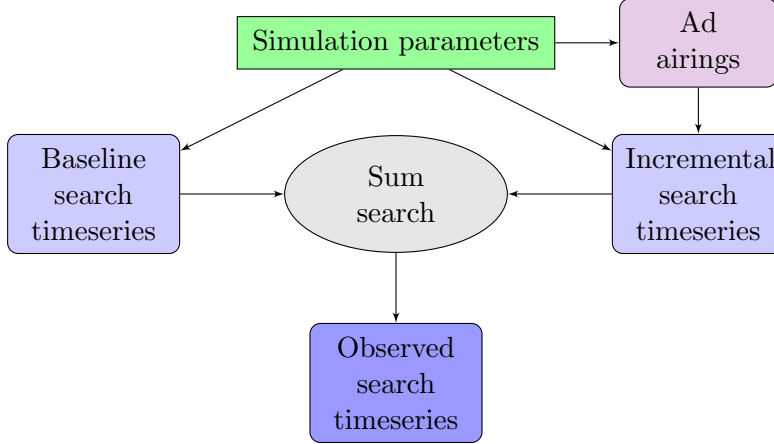


Figure 8: Diagram of the synthetic data generator

2.5 Interval estimates

We use the BSTS-generated posterior counterfactual time series for all the interval estimation. Specifically, we generate the spot-level lift as described above using each of the B (default 100) posterior samples, and aggregate by channels, etc. Then we get the posterior intervals at both spot and channel levels from the computed posterior lift samples.

3 Simulation studies

In this section we evaluate our proposed methodology with synthetic data and systematic simulation studies. We first describe our synthetic data generator, then show some results of the model performance.

3.1 Synthetic data generator

The data generation involves simulating three components: the ad airings data, the baseline search counts and the incremental search counts, as shown in Figure 8. The incremental search series depends on the generated ad airings data.

We use the following parameters:

- \bar{I} : The average TV impressions per spot;
- R_{spots} : The average number of TV spots per hour;
- p_{search} : The baseline search rate, i.e., the probability that a given search is on the keyword we are interested in out of all google.com search queries;
- r : Relative lift during post period;
- τ : A time constant that determines the effect duration of TV impact on search.

See Appendix for details about the parameters and the simulation.

3.2 Model performance under synthetic data

We use the synthetic data generator to generate 14 days of ad schedule on 10 TV channels, and the corresponding minute-by-minute search volume time series, with the following default parameter settings:

- $\bar{I} = 10^5$
- We set p_{search} such that there are ~ 50 baseline searches per minute on average
- $r = 10$
- $\tau = 8$

We generate data with different numbers of spots per hour, $R_{\text{spots}} = 0.5, 1, 1.5, \dots, 5$, which covers most of the cases in reality (Figure 14 in Appendix). For each value of R_{spots} , we simulate 100 data sets, each containing 14 days of data on 10 channels. Then we apply our proposed method to the 100 simulated data sets, and compute the root mean squared error (RMSE) and empirical coverage probabilities at the 80% nominal level at both spot and channel levels (see Equation 5). Specifically, the RMSE is defined as

$$\text{RMSE} = \sqrt{(1/N) \sum_{i=1}^N (\hat{\lambda}_i - \lambda_i)^2} \quad (6)$$

where the meaning of N , $\hat{\lambda}_i$ and λ_i depends on the case. Channel-level RMSE is a single number that combines information from all 100 datasets; $\hat{\lambda}_i$ and λ_i are the estimated and (simulated) true lift per spot for a specific TV channel from replicate i , and $N = 100$. Spot-level RMSE estimates are produced for each simulated dataset, $\hat{\lambda}_i$ and λ_i are the estimated and (simulated) true lift for spot i , and N is the total number of spots (all channels) in the dataset.

For a single spot that is not part of a group, $\hat{\lambda}_i - \lambda_i = \text{number of baseline searches} - \text{estimated baseline}$. Usually the estimated baseline is an accurate estimate of the expected baseline, so the RMSE is driven by variability in the actual number of baseline searches in the post period.

Figure 9 shows the spot lift RMSE and coverage probability under different spot densities. The boxplots represent RMSE or empirical coverage calculated from the 100 simulated datasets under different mean spots per hour. In most cases the estimated search lift is off by a median of ~ 25 searches. The coverage probabilities by the posterior intervals are close to the nominal level (80%), with a slight decay when the spots are denser (in which cases our methods depend on the regression to allocate search lift to nearby spots).

Figure 10 shows the TV-channel lift RMSE and coverage probability under different spot densities for each of the 10 TV channels. Specifically, each bar represents the RMSE (or empirical coverage) under a specific spot density for a specific TV channel calculated from the 100 simulated datasets with that density. Aggregating by channel results in even smaller RMSEs that are around 10 searches in most cases. The empirical coverage probabilities also decay as the spots get denser. There are two possible reasons for this decrease. First, since we remove a post period from the original search volume time series after each spot, the more the spots, the less training data is left for fitting the BSTS model. Second, as the spots get denser more and more airings are in groups (see Figure 15 in Appendix) and our model depends on the regression in Equation (3) to allocate the overall lift to each spot in a group.

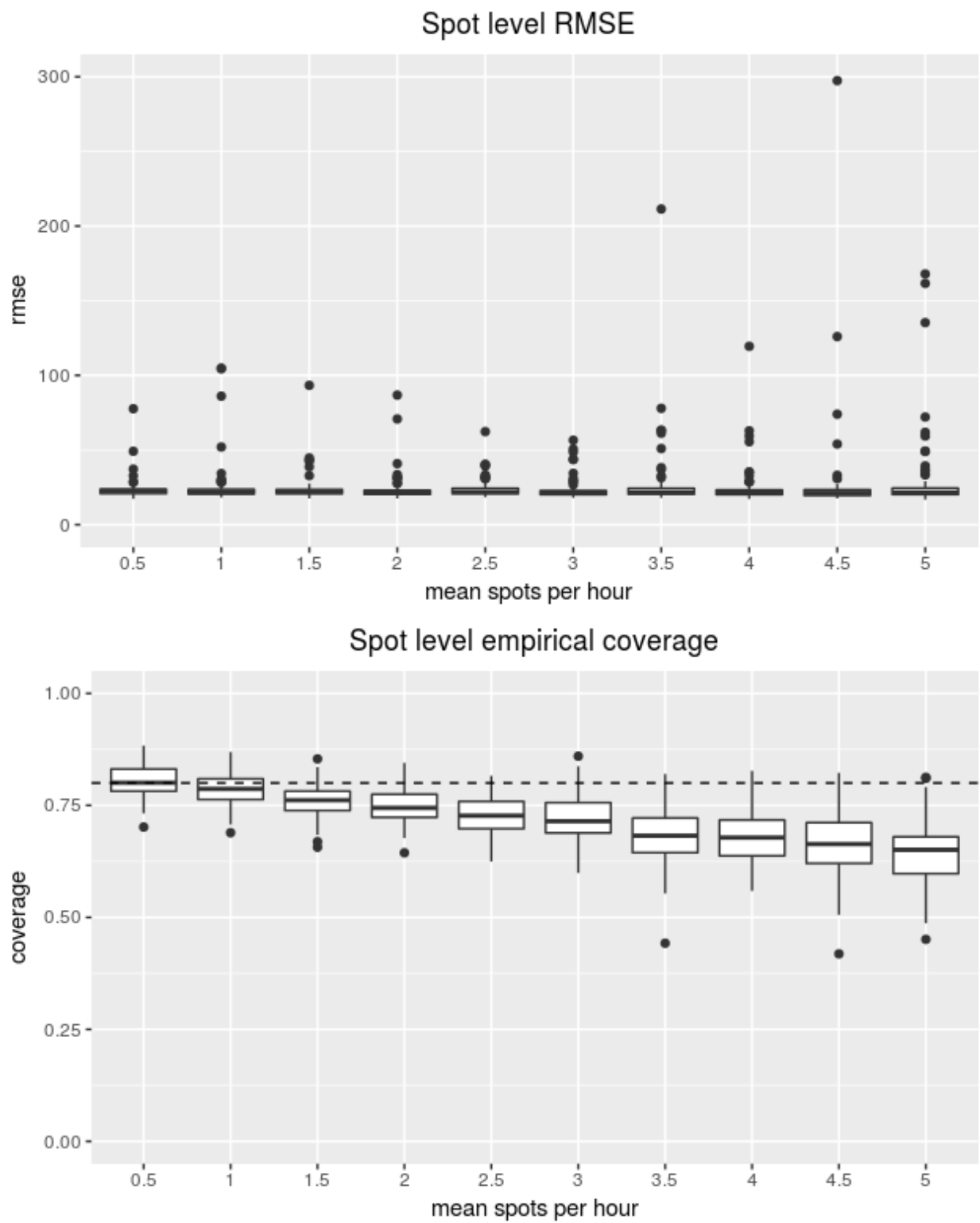


Figure 9: Spot level RMSE and coverage probability under different ad airing densities

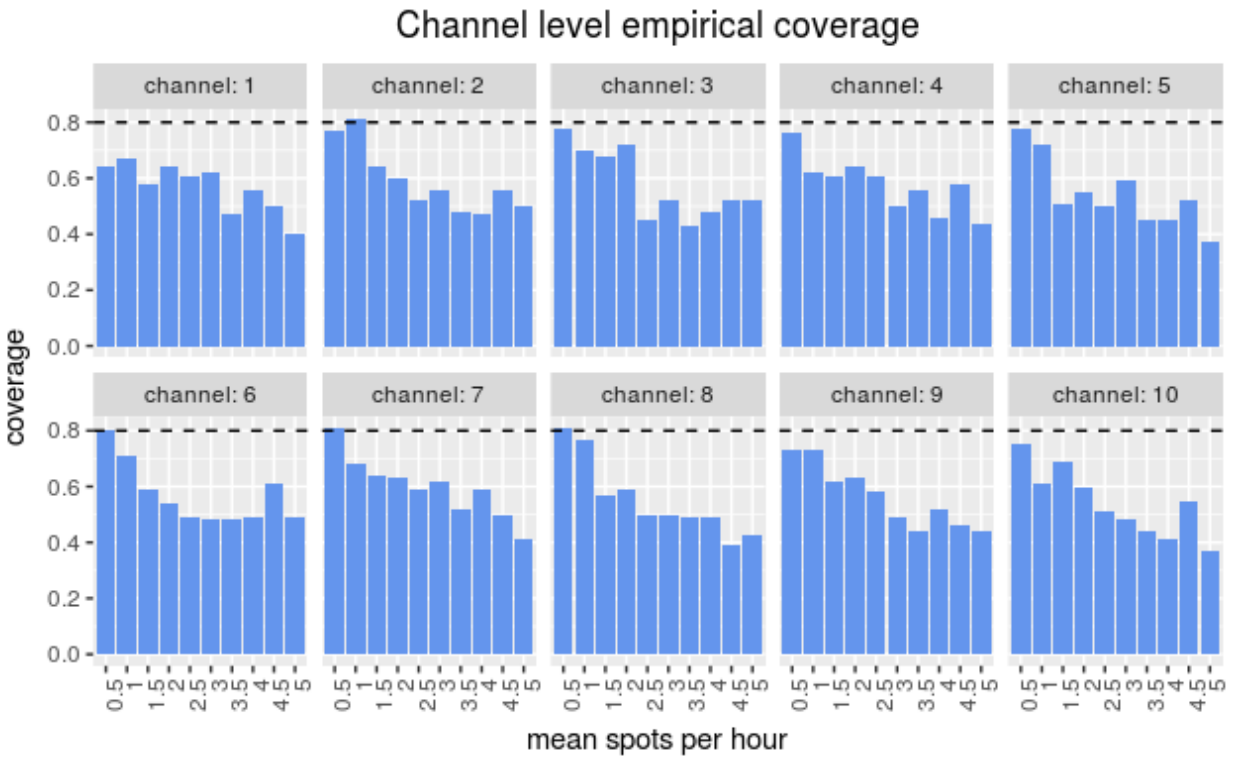
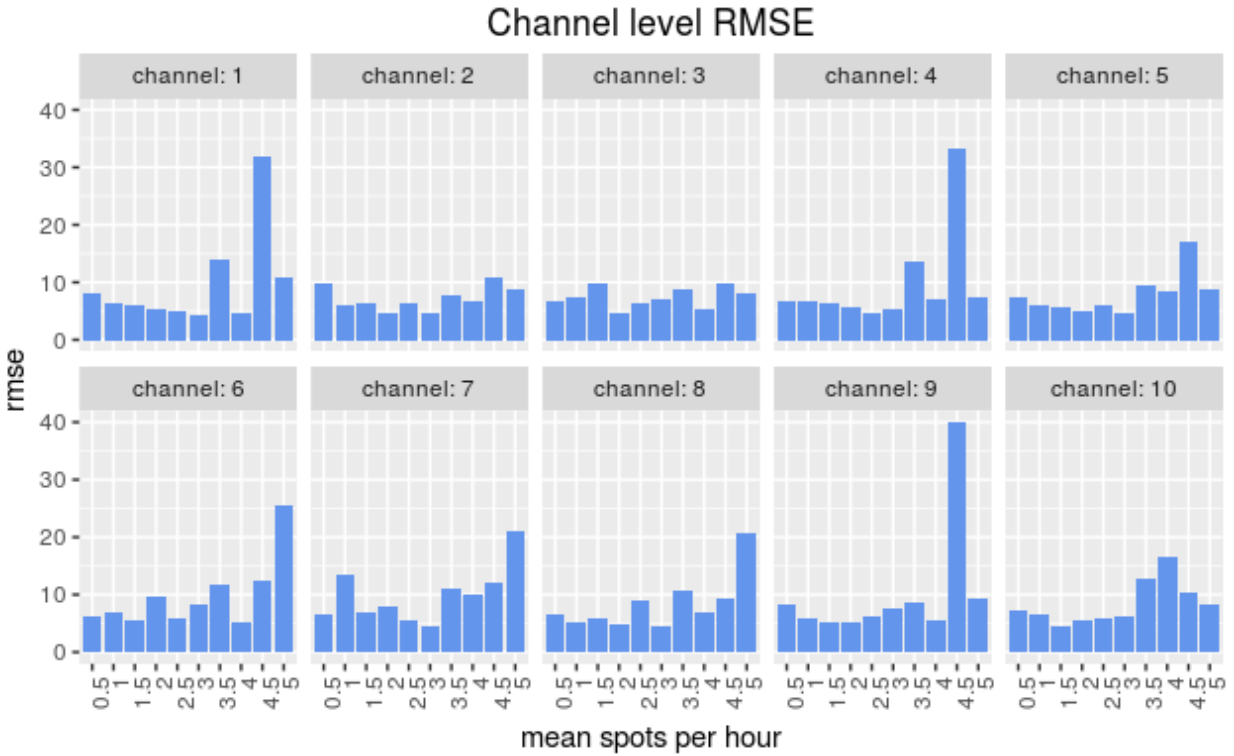


Figure 10: Channel level RMSE and coverage probability under different ad airing densities

We also apply the proposed methods to only the simulated baseline search time series to investigate the type I error. In all cases the empirical type I error is close to the nominal level (20%, data not shown).

We then conduct sensitivity analyses on the relative lift (r) and the length of the post periods (τ). We set relative lift to 1, 5, and 10. Both the RMSE and the empirical coverage probability are stable under all the three settings. We set the post period to 4, 8, and 12 minutes. The RMSE and empirical coverage probability are similar under 4 and 8 minutes, but there is slightly higher RMSE and lower coverage probability at 12 minutes. This is understandable since in this case the signal to noise ratio decays drastically at the end of the post periods given the exponential decaying shape of the simulated incremental search volume.

4 A case study on an Android campaign

In this section we show the results from one specific study for Android, and demonstrate how to use the proposed approach to analyze TV campaigns and generate insights that can be helpful for optimization against incremental searches on different dimensions such as time of day.

Google ran a campaign from March 1st to 28th in 2016 on TV to advertise its Android system. This campaign had 85 million impressions on 11 networks. We apply the proposed methodology to measure the incremental searches due to the TV ad airings in the TV campaign. We analyse two keywords that are related to the TV ad: “Android” and “Cool”. Aggregated search time series for the above two keywords is pulled from Googles search logs. We use Rentrak², a third party data provider, combined with a Google internal TV ad tracker, to get the TV ad airing data.

The first plot in Figure 11 shows the incremental searches per spot (ISPS) for different times of day, where “Prime” is from 19:00 to 23:00. The numbers are indexed so that the average is 10. On weekdays primetime spots have more incremental searches per spot, likely because those spots have more TV impressions. The highest ISPS happens during weekend non-prime time. A closer investigation reveals that it is largely driven by a few spots on ESPN and Freeform, for example, an ESPN spot at 11:52 on March 5 during College Gameday.

One can also do such comparisons by taking into account TV impressions or cost. The middle plot in Figure 11 shows the indexed incremental searches per impression (ISPI) for different time of day. The weekday pattern actually reverses, where non-prime time spots have a larger ISPI than primetime spots. Weekend non-prime spots still have the largest ISPI. The bottom plot in Figure 11 shows the indexed incremental searches per cost (ISPC) where the pattern is similar to that of ISPI. These results highlight the cost effectiveness of weekend non-prime time spots in this campaign in terms of TV induced searches, while weekend primetime has the lowest ISPS, ISPI and ISPC.

5 Comparison between TV and YouTube incremental searches

In this section we compare incremental searches from TV and YouTube, based on results from the Android case study in Section 4. Note that the incremental searches combine the effects of the difference in media and ads (for example, YouTube viewers already have a device at hand so it may be easier to do a search). First we explain why we think such cross-media comparison is reasonable. There are two issues: 1) Are we measuring the same metric across TV and YouTube? 2) Are both estimates accurate enough to make a comparison meaningful?

²www.renrak.com.

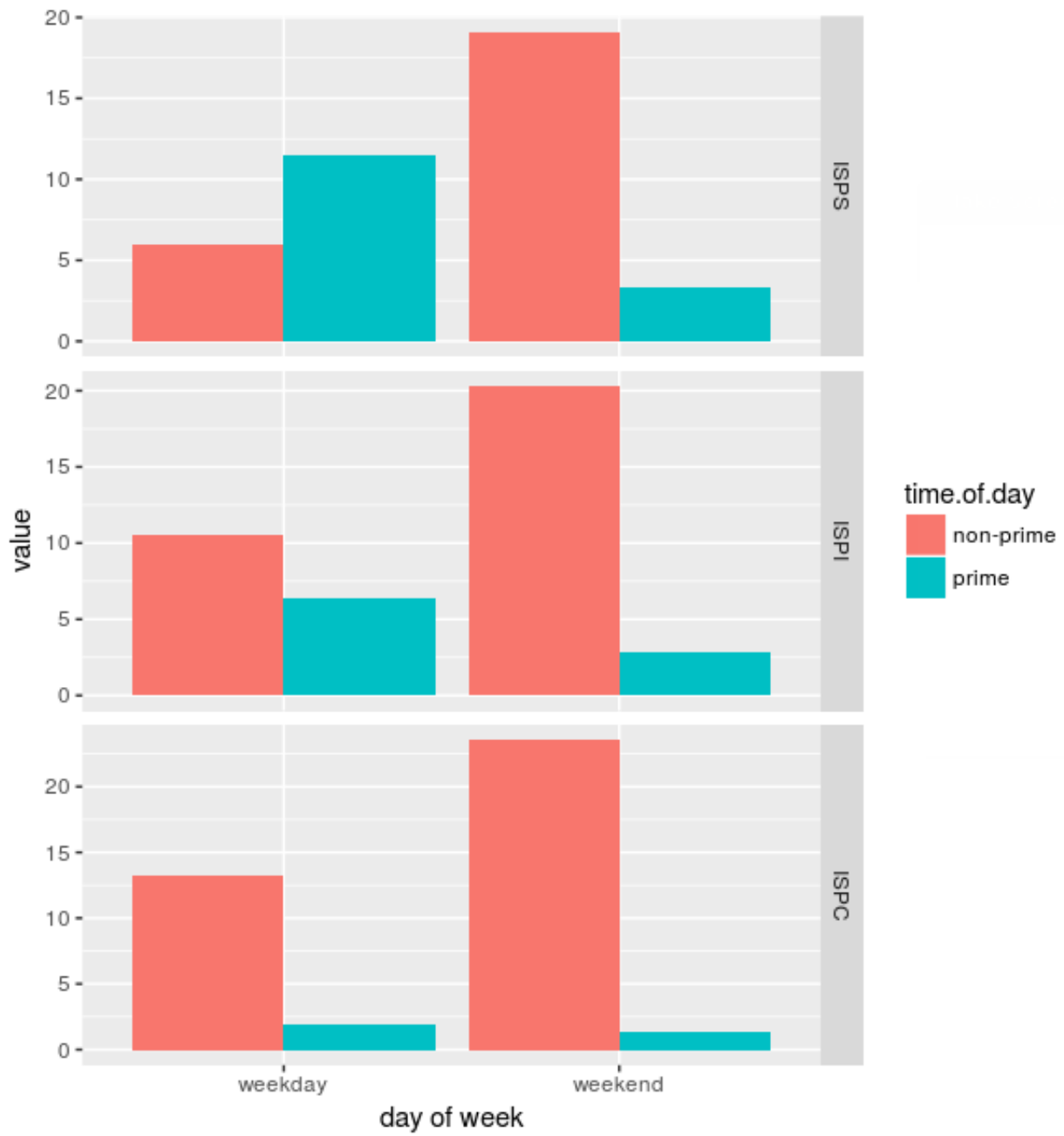


Figure 11: ISPS, ISPI and ISPC for different time of day in the Android campaign. The metrics are indexed so that the average is 10 in each panel

Regarding the first issue, both TV and YouTube search lift try to measure the same metric, i.e., total incremental searches due to ads. Furthermore, both incremental searches should be from the “treated” audience (i.e., “treatment effect on the treated”), where in the cases of both TV and YouTube, it is the “campaign audience”, i.e., ad viewers. In our TV methodology we actually analyse the aggregated search volume from all the google.com online population, with the assumption that users not exposed to TV ads are not affected by the ad campaign. We use an adaptive post window up to 4 hours (the same as Google YouTube Search Lift) so are not able to capture the potential very-long term effect.³ However, our data shows that in most cases most of the incremental searches happen within minutes of the TV spots. Bearing the above in mind, both TV and YouTube search lift try to measure the same metric, and if there is no measurement error, it is reasonable to do the cross-platform comparison (with appropriate normalization such as “incremental searches per impression” like we use in this paper since the total audience sizes could be different).

The second level of comparability is whether the measured metrics produced by potentially different methodologies for different platforms (TV and YouTube) are comparable. Despite different techniques used (whether the proposed BSTS based method for TV or A/B experiments for YouTube), both of them produce unbiased estimators of the true metric for them to be comparable. If they also produce intervals of the right coverage, these intervals can also be used for comparison. In the case of TV, we have run systematic simulations to make sure that the incremental search estimates are accurate (Section 3), and our simulation assumptions are supported by data analysis. In the case of YouTube, we use randomized experiments to avoid systematic bias. Since both methodologies measure the same metric and are accurate, the results they produce can be compared to each other.

Google ran a YouTube campaign to advertise Android during the same time period as the TV campaign discussed in Section 4. The campaign had 13 million impressions. We measure the incremental searches from the YouTube ads using Google’s YouTube Search Lift product. This uses randomized experiments, and the same keywords as for the TV search lift study. Figure 12 shows ISPC and ISPI for top TV channels and YouTube for the Android case study we discussed in Section 4. We also include the TV average (labeled as TV-ALL-CHANNELS) in the figure. We apply a 25% discount to the TV cost rate card when calculating TV ISPC, to approximate actual discounted rates. All the metrics in the figure are indexed against that of YouTube. The boxplots represent the 80% intervals generated by our model. YT ranked #5 and #3 among different TV networks in terms of ISPC and ISPI, respectively. Interestingly, Freeform had both the biggest ISPC and the biggest ISPI; it targets teenagers and young adults, who are perhaps more receptive to Android campaigns.

6 Discussion

In this paper we propose an approach to measuring the impact of TV ads on online search behavior. Our method models the baseline search volume time series (or counterfactual), by using data not contaminated by TV ad spots. One advantage of this approach is that we do not need to model the search spikes, which would require additional assumptions on the shape of the spikes.

We evaluate the performance of the proposed methods with systematic simulation studies. Results show that the incremental search point estimates are unbiased, and the interval coverage

³Also this could be confounded with a power issue as shown in our sensitivity analysis that the signal to noise ratio is likely to decay drastically at the end of the post periods so it is less likely that the actual post windows last for 4 hours.

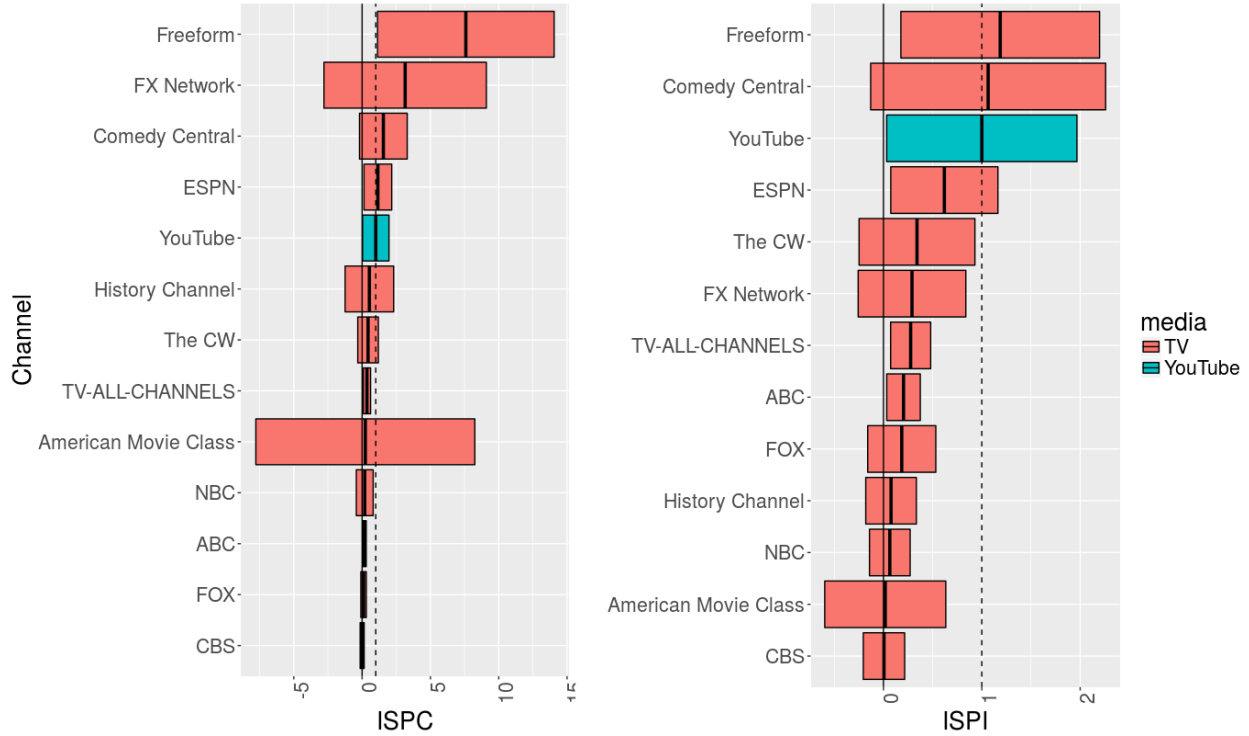


Figure 12: Incremental searches per cost and per impression (relative to YouTube)

is close to the nominal level.

We then showcase the proposed approach with a case study for an Android campaign with results broken down by time of day and day of week. These results can help TV advertisers optimize their TV campaign. Such optimization can be done on different dimensions in addition to time of day, such as TV creative, and program genre. We also compare the TV ad effectiveness to that of YouTube by using the incremental searches per impression.

In this paper we measure the effectiveness of TV campaigns based on search, which is an integral part of the consumer experience. However, an extra question is whether incremental search is reasonable as a surrogate for other metrics (e.g., brand metrics). Our data shows that YouTube audience generally searches more than TV audience (data not shown). Thus the comparison of incremental searches from TV and YouTube reflects the combined effect of both the ads and the medium. If advertisers care about viewers' online search behavior due to their TV campaigns, the approach in this paper provides a useful way of measuring incremental searches in a causal way. However, online searches might not be as important for some advertisers. If that is the case, one might want to consider other effectiveness metrics, for example, survey based metrics as mentioned in Introduction. We hope that this paper will generate more research on TV effectiveness measurement.

Acknowledgement

The authors would like to thank Tony Fagan, Elissa Lee and Lu Zhang for their encouragement and support, Georg Goerg, Tim Hesterberg and Tony Fagan for their thorough review of the manuscript and constructive comments.

References

- [1] Kay H. Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, 9:247–274, 2015.
- [2] A. Djambaska, I. Petrovska, and E. Bundalevska. Is Humor Advertising Always Effective? Parameters for Effective Use of Humor in Advertising. *Journal of Management Research*, 8(1), 2016.
- [3] M. Elmore. An Exploration of Advertising Effectiveness Methodologies: Comparing Recall to Opportunity to See. *Insight Express*, 2012.
- [4] Edward I. George and Robert E. McCulloch. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.
- [5] Edward I. George and Robert E. Mcculloch. Approaches for Bayesian variable selection. In *Statistica Sinica*, pages 339–374, 1997.
- [6] K. Harrenstien, Miner E., Patel A., and Varian H. AdWorth for TV movie ads. *Technical report*, May 2006.
- [7] J. Liaukonyte, T. Teixeira, and K. Wilbur. Television advertising and online shopping. *Marketing Science*, 34(3):311–330, 2015.
- [8] L. M. Lodish, M. Abraham, S. Kalmenson, J. Livelsberger, B. Lubetkin, B. Richardson, and M. E. Stevens. How TV Advertising Works: A Meta-Analysis of 389 Real World Split Cable TV Advertising Experiments. *Journal of Marketing Research*, 32(2):125–139, 1995.
- [9] Nicholas G. Polson and Steven L. Scott. Data augmentation for support vector machines. *Bayesian Analysis*, 6(1):1–23, 2011.
- [10] Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55, 1983.
- [11] Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, October 1974.
- [12] Steven L. Scott and Hal R. Varian. Predicting the present with bayesian structural time series. *IJMNO*, 5(1/2):4–23, 2014.

A Appendix

In this appendix we describe in detail our simulation assumptions and settings. We redefine parameters used in generating the synthetic data as the same in the main text for completeness.

A.1 Ad airing data

Ad airing data consists of data from multiple TV networks. When creating TV ad spot data, our goal is to create ad airings for multiple TV channels where we can control the density of the spots. This allows us to investigate cases where ads on different channels run close to one another. We define the following parameters.

- N_{ch} : The number of TV channels we want to simulate data for.
- N_{days} : The number of simulated days.
- R_{spots} : The average number of spots per hour.
- \bar{I} : The average impression count per spot.

The spot data is generated independently per channel as follows:

1. Each channel has a mean impression count drawn from an exponential distribution, i.e., $I_{ch} \sim \text{Exp}(\bar{I})$.
2. Per channel, we generate spots with a time difference between them drawn from a uniform distribution such that the rate of spots per hour R_S , is maintained. The time between spots is drawn from $\Delta_t \sim U(0, 2/R_S)$.
3. The number of impressions for a spot k on channel ch is drawn from an exponential distribution (rounded up) $I_k \sim \lceil \text{Exp}(I_{ch}) \rceil$.

Here (and for the following subsections) we explicitly list the assumptions we make.

- We assume that ad airings are independent;
- We assume that there is no diurnal pattern for impressions.

A.2 Search data

Search data consists of baseline searches, that are not related to TV ads and searches that are due to the ads. We first simulate the baseline searches and then given the set of TV ad spots, we generate the incremental searches, which can be negative because viewers might no longer need to search because they get the information from the TV ads. The “observed” search time series is a sum of both.

A.2.1 Baseline search data

We define the following search rate parameter.

- p_{search} : The search rate, i.e. the probability that a given search is for the keyword we are interested in.

We follow three steps for generating the baseline search data:

1. We extract 12 weeks of minute-by-minute search volumes from google.com US search data (on all the queries).
2. We take the average for each minute of a day, denoted by \bar{N}_t .
3. For a given time t , we simulate the baseline search count at that time N_t via

$$N_t \sim \text{Poisson}(p_{\text{search}} \bar{N}_t). \quad (7)$$

We make the following assumptions in generating the baseline search volume.

- Diurnal patterns follow general patterns for search, with no adjustment for day of week;
- We assume that search counts follow a Poisson distribution where in reality they can be overdispersed with a variance that has a diurnal pattern as well. We leave it as a Poisson for simplicity.

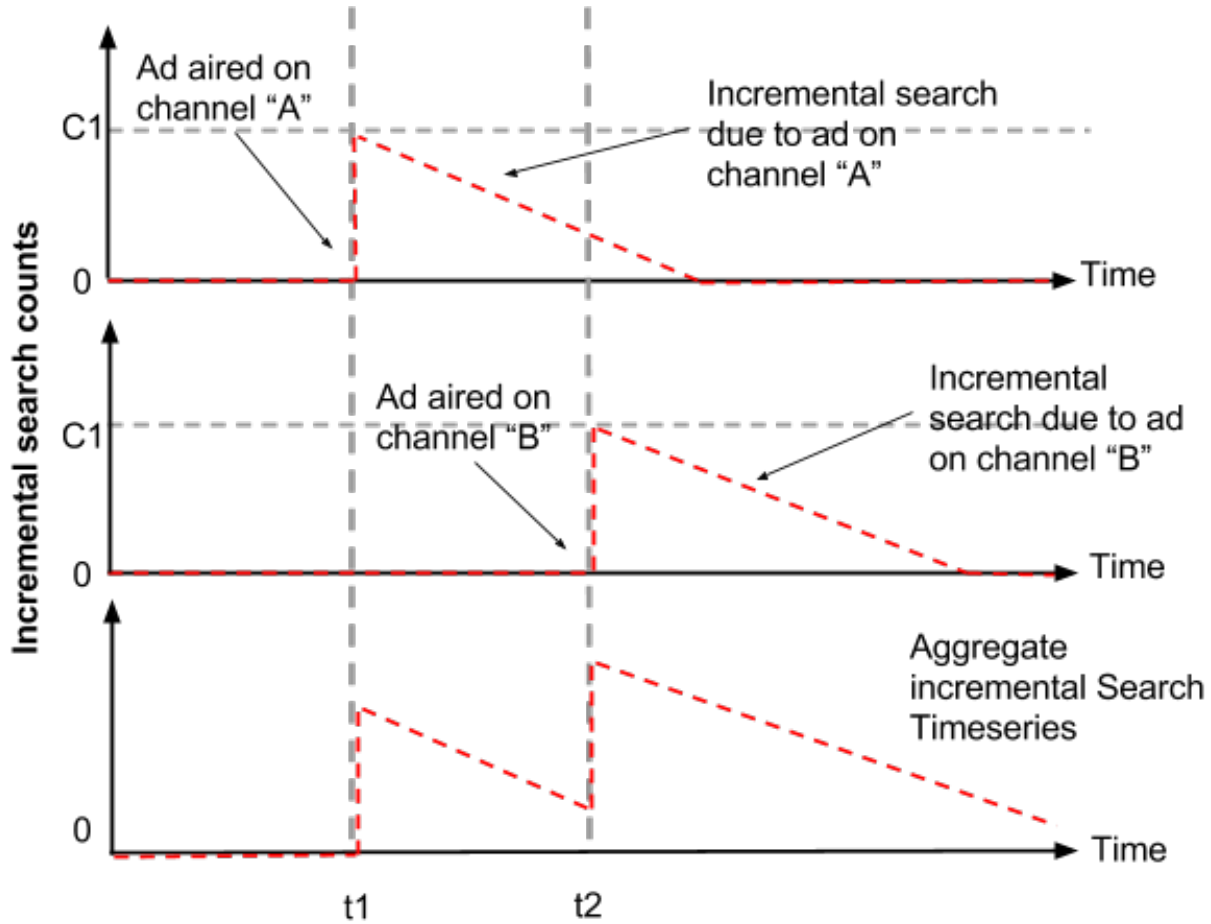


Figure 13: Generating incremental searches

A.2.2 Attributed (Incremental) search data

The incremental search counts are searches that occurred as a result of users seeing a TV ad and are incremental to any searches that a user would have done had they not seen the ad. The way we approach the problem is by generating the incremental search volume for each ad spot independently, which could mean ad airings on different channels occurring at the same time, and then combining all the incremental searches into a single aggregated incremental search time series (Figure 13).

We assume an exponential decay for the incremental search time series in the post period, i.e., the effect of the ad airing on search volume decays exponentially with time⁴. More specifically, the incremental search counts are drawn from a Poisson distribution with an exponential decaying mean. The number of incremental search counts for ad airing k , at time t is simulated by

$$N_{k,t} \sim \text{sign}(C_k) \times \text{Poisson}(1(t \geq t_k) | C_k | e^{-(t-t_k)/\tau}), \quad (8)$$

where

- t_k is the time of ad airing k

⁴In our simulation studies we actually defined other types of the incremental effects, and the model achieved similar performance (results not shown).

- C_k is the strength of the response for ad view k in terms of number of searches per minute. This determines the effect magnitude.
- $1(\cdot)$ is the indicator function (1 if the condition is true, else 0)
- τ is a time constant that determines the effect duration

The the magnitude of the effect C_k depends on the number of impressions in a spot, and is given by

$$C_k = r_{\text{ch}} p_{\text{search}} I_k, \tag{9}$$

where

- r_{ch} is the relative lift per channel, drawn from a normal distribution on a channel level such that $r_{\text{ch}} \sim rN(1, 0.5)$, where r is a predefined relative lift.

This means that we assume each exposed user performs an extra search with probability $r_{\text{ch}} p_{\text{search}}$, i.e. r_{ch} above or below the average search rate. The resulting total number of searches are then truncated to be non-negative. In addition we make the following assumption.

- Incremental search is not correlated between different ad viewings.

Note that our methodology (see Methods) actually does not depend on most of the assumptions we make in generating the synthetic data.

A.3 Mean spots per hour from real studies

Figure 14 shows the histogram of the average number of TV spots per hour from 2000+ Google Brand Lift studies. Most studies have fewer than 10 spots per hour. There are campaigns with ~ 60 spots per hour because some advertisers constantly run TV ads on different TV networks.

A.4 Percentage of overlapping spots under different values of mean spots per hour

Figure 15 shows the percentage of overlapping spots under different values of mean spots per hour in the simulated data. When there are an average of 5 spots per hour more than 70% of the spots overlap.

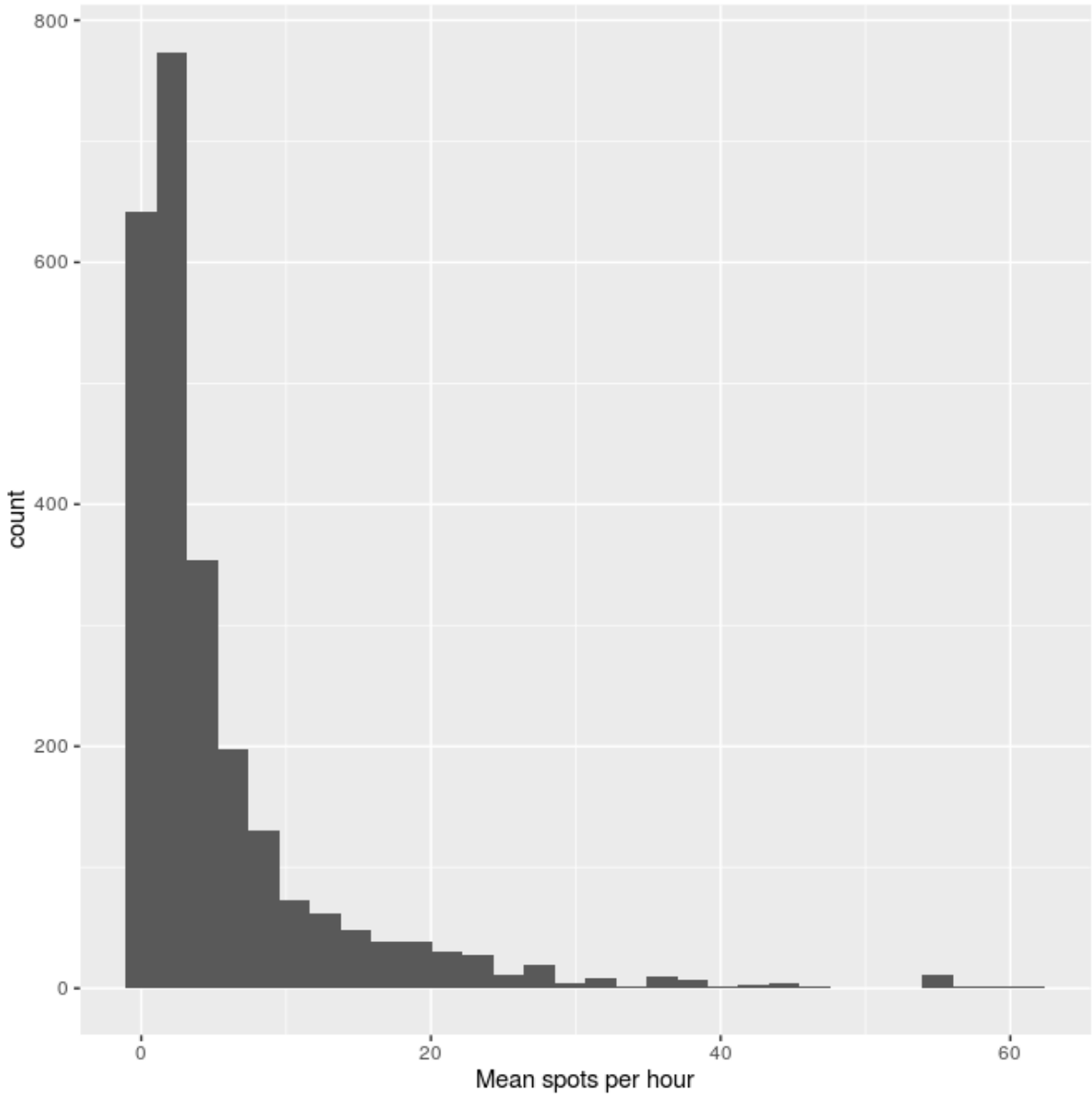


Figure 14: Histogram of mean spots per hour

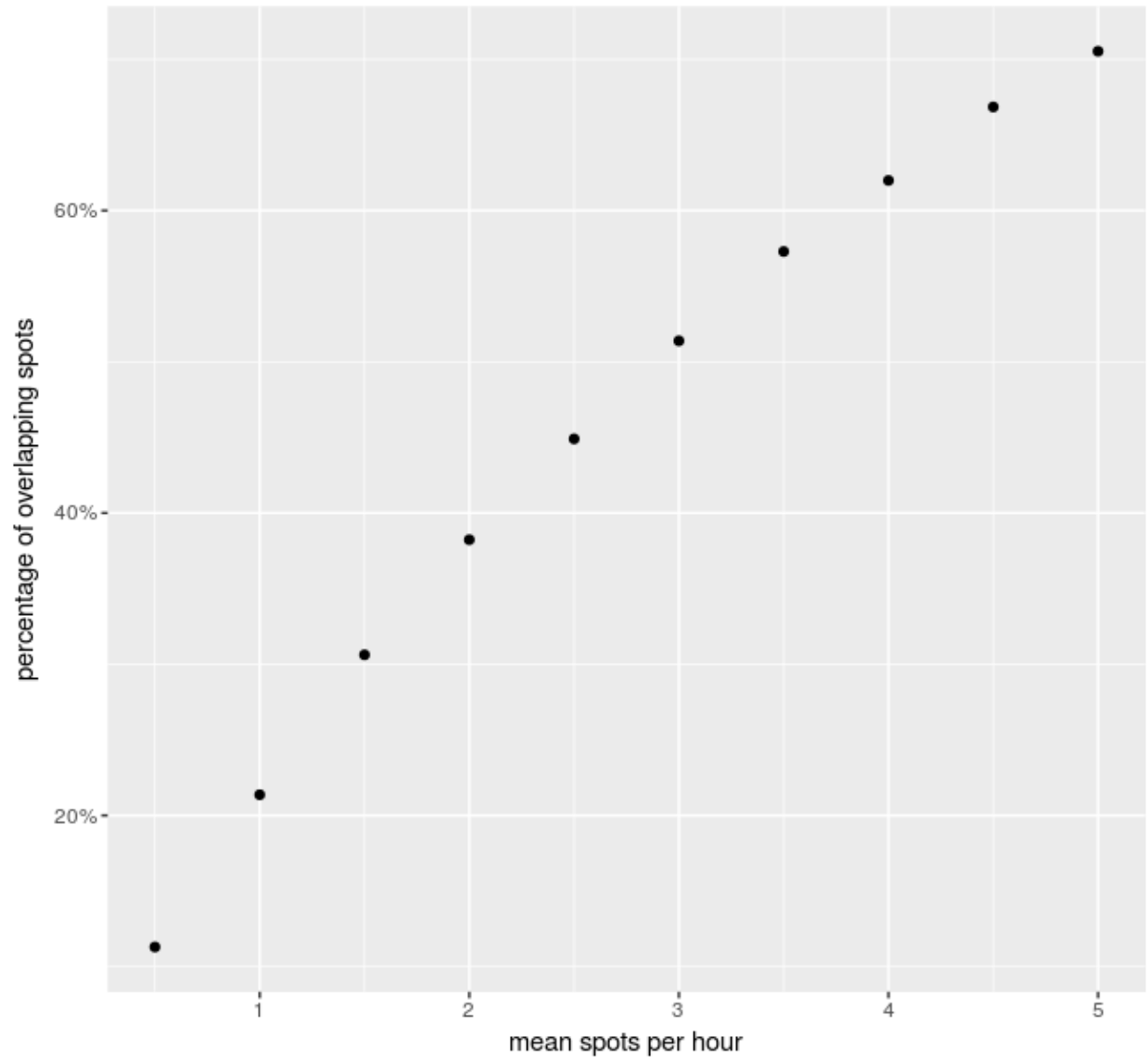


Figure 15: Percentage of overlapping spots under different values of mean spots per hour