

THE COOPER UNION  
ALBERT NERKEN SCHOOL OF ENGINEERING

Inferring the causal impact of  
Super Bowl marketing campaigns  
using a Bayesian structural time series model

by  
Neema Aggarwal

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Engineering

May 2017

Professor Sam Keene, Advisor

THE COOPER UNION FOR THE  
ADVANCEMENT OF SCIENCE AND ART

ALBERT NERKEN SCHOOL OF ENGINEERING

This thesis was prepared under the direction of the Candidate's Thesis Advisor and has received approval. It was submitted to the Dean of the School of Engineering and the full Faculty, and was approved as partial fulfillment of the requirements for the degree of Master of Engineering.

---

Dean, School of Engineering                      Date

---

Prof. Sam Keene, Thesis Advisor              Date

## Acknowledgments

I am extremely grateful to the Cooper Union for providing an incredible education and environment for research, to the faculty and students for motivating and inspiring me, and my friends and family for their continuous encouragement. I'd especially like to acknowledge and thank Morgan Macklin for her detailed comments on my drafts, Brenda So for the LaTex help, and above all Dr. Sam Keene, my advisor, for his unwavering support, guidance, and time.

## Abstract

The causal impact of a treatment can be defined as the difference between the observed value of a chosen response variable and the unobserved value that would have been obtained had the treatment not taken place. An area where this plays a key role is in the field of marketing, where the main objective is to incur some form of positive effect on a chosen business metric in order to increase revenue in the short or long term. Inferring the causal impact of marketing events is a critical yet imperfect science in the business world. This thesis investigates the application of Bayesian structural time series (BSTS) models to isolate the impact of marketing campaigns launched during the 2017 Super Bowl. The model combines multiple control markets and prior knowledge on trends to produce a synthetic counterfactual of the desired response metric, had the Super Bowl ad never occurred. The difference between the counterfactual and the true observed values is taken as the causal impact. Being Bayesian in nature, the final posterior density depends only on the actual observations, while accounting for all other states and parameters. The objectives of this thesis include (a) demonstrating the application of BSTS in the marketing setting, (b) investigating the effects of parameter choices and covariate inclusion, (c) comparing the BSTS model to a traditional difference-in-differences algorithm, and (d) illustrating the usefulness of the BSTS model in gaining business insights regarding Super Bowl marketing.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background</b>	<b>3</b>
2.1	Super Bowl Marketing . . . . .	3
2.1.1	Advertising strategies and the shift to digital platforms . . . . .	4
2.1.1.1	Display advertising . . . . .	5
2.1.1.2	Search engine optimization and paid search . . . . .	5
2.1.1.3	Deals, discounts, and sweepstakes . . . . .	6
2.1.1.4	Social media marketing . . . . .	6
2.2	Measuring returns . . . . .	7
2.2.1	Return on marketing investment . . . . .	8
2.2.1.1	Short term ROMI . . . . .	8
2.2.1.2	Long term ROMI . . . . .	9
2.2.2	Marketing Mix Modeling . . . . .	9
2.2.3	Brand lift . . . . .	10
2.2.4	Emergence of new metrics . . . . .	11
2.3	Analysis techniques for inferring causality . . . . .	13
2.3.1	Randomized experiments . . . . .	13
2.3.2	Causal inference without randomized experiments . . . . .	14
2.3.2.1	Difference in differences analysis . . . . .	15
2.3.3	Bayesian structural time series analysis . . . . .	16
<b>3</b>	<b>Application of BSTS to Super Bowl ad campaigns</b>	<b>22</b>
3.1	Statement of Problem . . . . .	22
3.2	Data sources . . . . .	25
3.2.1	Data treatment and normalization . . . . .	28
3.3	BSTS analysis . . . . .	28
3.3.1	Implementation in R . . . . .	29
3.3.1.1	Model assumptions . . . . .	29
3.3.1.2	Utilizing and adjusting the model . . . . .	29
3.3.2	Comparison to difference-in-differences method . . . . .	32
<b>4</b>	<b>Results</b>	<b>33</b>
4.1	BSTS model testing . . . . .	33
4.1.1	Varying the post period window . . . . .	34
4.1.2	Controls . . . . .	35
4.2	Super Bowl marketing impact . . . . .	37
4.3	Comparison to difference-in-differences method . . . . .	42
<b>5</b>	<b>Conclusions and Future Work</b>	<b>43</b>
5.1	Conclusions . . . . .	43
5.2	Future Work . . . . .	44

## List of Figures

1	Average cost of a 30 second commercial during the Super Bowl . . . . .	3
2	Ads aired during the 2015 Super Bowl . . . . .	4
3	Illustration of difference in differences definition . . . . .	15
4	Graphical representation of model . . . . .	21
5	Pre-Super Bowl Google search trends . . . . .	31
6	Avocados from Mexico causal impact output . . . . .	34
7	Bud Light controls as weighted by the spike-and-slab algorithm . . . . .	36
8	Audi control weighting versus Buick . . . . .	36
9	Skittles control weighting versus Snickers . . . . .	36
10	Causal impact results with threshhold on control market inclusion probability . . . . .	37
11	Results of BSTS model with Google search data . . . . .	38
12	Pre-marketing Bud Light vs. 84 Lumber Google search history . . . . .	39
13	Post-marketing Bud Light vs. 84 Lumber Google search history . . . . .	40
14	Causal impact on Google searches vs. Twitter mentions . . . . .	41
15	Comparison of BSTS method versus DD method . . . . .	42

# 1 Introduction

Marketing has long been one of the most vital and perplexing elements of making a business successful. When done properly, television advertisements combined with digital marketing can bring in new customers, influence buying decisions, and define a brand's reputation. When not done properly, even the most perfect products can become obsolete, quietly overshadowed by louder voices. In 2015 the total US spend for TV advertising was over \$60B, from companies setting aside 5-15% of their revenue for marketing [1]. Naturally, the question arises of how to measure the value gained to justify spending such copious amounts of money. This raises one fundamental question: how to infer the causal impact of a designed campaign on a chosen economic metric.

Over the years, econometricians have developed concepts such as return on marketing investment (ROMI) [2] and techniques such as difference-in-differences [3] to try and quantify the impact of each marketing dollar spent. Both methods have their benefits and limitations (to be discussed in section 2.2). In the last few years, econometricians and data scientists have looked to advanced machine learning techniques paired with the growing quantities of available data as a potential approach. In 2015, Broderson et al. proposed Bayesian structural time series (BSTS) models as a powerful tool for inferring causal impact of marketing campaigns [4].

The suggested BSTS model combines multiple control markets and prior knowledge on product trends to produce a counterfactual post-ad campaign time series for the response variable in question assuming the campaign never took place. The difference between the observed response variable time series and this synthetic counterfactual is then defined as the estimated causal impact of the marketing.

The objective of this thesis is to demonstrate the applicability of the BSTS model by applying it to several marketing campaigns. The aim is to fit the model to

multiple novel sets of data from varying sources, tune the various parameters, and compare the results to other non-Bayesian methods. Additionally the thesis will address the valuable business insights gained through the results of the model.

One of the most important times for brand marketing is during football season, TV's most sought after and expensive slot for advertisers [5]. For the 2016-2017 season, a 30 second spot during NBC's Sunday night football cost an average of \$650,000 [5]. For the purposes of this thesis, the specific marketing event chosen was Super Bowl Sunday. In 2017, an advertising spot during the big game cost \$5M+ for 30 seconds of airtime [6]. Inferring the causal impact of an ad campaign of this caliber is a crucial business need.

The rest of this thesis is organized as follows. Section 2 provides background content on typical Super Bowl marketing strategies, methods of measuring returns. This section will introduce analysis techniques for inferring causality, including details on the BSTS model. Section 3 discusses the application of the model to Super Bowl marketing. Section 4 presents the technical results of applying and tuning the BSTS model, as well as the marketing results of the various product campaigns tested. Finally, section 5 details the corporate insights gleaned from the exercise, the future work to potentially expand the Super Bowl analysis or further test the BSTS model, and the additional possible applications for the model.

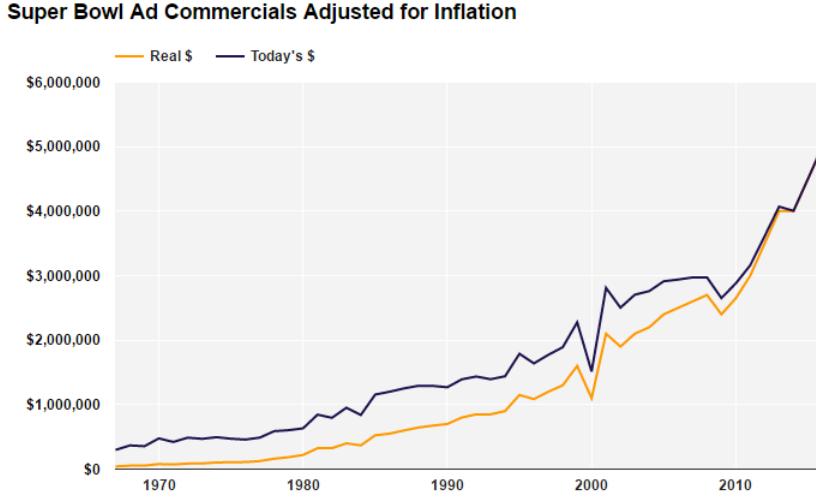


Figure 1: Average cost of a 30 second commercial during the Super Bowl adjusted for inflation

## 2 Background

### 2.1 Super Bowl Marketing

The Super Bowl, the annual championship of the National Football League, the highest level of professional football in the world, has generated record levels of viewership since its very start. From the first victory by the Green Bay Packers in 1967 to the New England Patriots in 2017, 51 years later, the sport continues to captivate millions. While the number of viewers has increased more than 3 fold from 50 million to 170 million [7] over that time, the larger growth has been in the marketing dollars spent by companies looking to capitalize on the event. The average cost of a 30 second commercial during the Super Bowl started at \$40,000 about 50 years ago and has since risen to \$5 million dollars (see Figure 1 [6]. Total annual ad spend in the last 20 years alone has drastically risen from close to \$100 million to upwards of \$400 million. According to Bloomberg Business [8], more than one third of the 2015 Super Bowl broadcast was commercials (see Figure 2). The Super Bowl can be as much career defining for the players, as for the advertisers staking millions in hopes of being memorable. Companies go all out with one goal

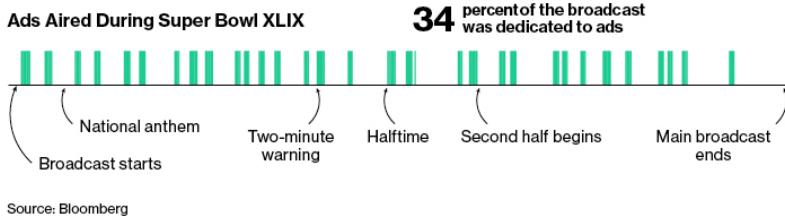


Figure 2: *Ads aired during the 2015 Super Bowl*

in mind: that their products and commercials be remembered and talked about just as much as the team that scored the winning touchdown.

### 2.1.1 Advertising strategies and the shift to digital platforms

For decades the main and only form of Super Bowl advertising was through television commercials. While it's true that the average price tag for a 30 second commercial has risen over 75% in a decade [6], there has also been a significant shift towards digital marketing platforms. The constant technological changes our society faces today has put more screens at our fingertips than ever before, creating a plethora of advertising opportunities. When TV was the only advertising avenue available, marketers could rely on the big reveal ads. However, as consumers become continuously more exposed to digital advertising every moment they browse the internet, watch videos on YouTube, or utilize apps on smartphones, advertisers are forced to develop new strategies defining what content is viewed and when. According to Salesforce research, 73% of Super Bowl fans said they plan to use at least two devices during the big game [9], further increasing the importance of a cross-channel digital engagement strategy. In order to fully monetize and gain exposure for TV commercials, it has become necessary to branch out and explore other means of staying relevant. All forms of digital marketing, combined with traditional TV marketing, serve to create buzz around the brand in question. That buzz can be in the form of word-of-mouth chatter regarding the product, or, more

commonly these days, in the form of "digital buzz", i.e., online chatter.

### **2.1.1.1 Display advertising**

Display advertising is a common form of online advertising in which images, videos, or audio clips are utilized in banner ads, pre-roll ads, companion banners, rich media, and more. Companies that properly utilize this type of digital media optimize a combination of pre-Super Bowl clips to build anticipation, usually paired with a post-Super Bowl campaign taking advantage of ads on sites like YouTube to remain linked to any Super Bowl related searches. This way, as people go online to rewatch the epic pass interception in the second quarter or the half-time performer messing up, they are also faced with display advertising reminding them of the TV commercial for product x.

### **2.1.1.2 Search engine optimization and paid search**

Another key digital marketing strategy is search engine optimization (SEO). SEO is the process of optimizing a website to increase its visibility in a search engine's unpaid results. This involves various methods of increasing the relevance between the website and keywords, optimizing content, increasing the links to other sites or pages, and more. In the context of Super Bowl advertising, top tier marketing departments will create Super Bowl-focused landing pages for their products that funnel in traffic by ranking highly for Super Bowl related search queries. A key aspect of SEO is mobile search, which as of May 2015 surpassed desktop search [10]. Keeping this in mind, it becomes important for these Super Bowl-focused landing pages to also be well optimized for mobile devices.

Paid search marketing allows companies to advertise in the sponsored listing section of a search engine or website's results by paying the website each time the ad is clicked upon by a user or even displayed. Displayed ads are charged on a cost-per-impression (CPM) basis while clicked ads are slightly more expensive

on a pay-per-click (PPC) basis. Both are critical methods for websites to make money and companies to make an impression on potential customers. The growing importance with paid search in recent years also falls on extending it to mobile devices, much like SEO.

#### **2.1.1.3 Deals, discounts, and sweepstakes**

With each growing market, advertisers look for ways to monetize on the Super Bowl. As food delivery becomes more popular, both restaurants and delivery services alike will offer deals and discounts. As online shopping takes off, the Super Bowl suddenly becomes a fantastic time to purchase a TV or other electronics. Sweepstakes start popping up to win everything from NFL tickets to straight cash. These additional advertising opportunities typically differ from TV commercials in that they have a longer window ahead of Super Bowl Sunday. But much like TV, the main factor making them successful, is the newest form of marketing - social media.

#### **2.1.1.4 Social media marketing**

Facebook, Twitter, Instagram, Snapchat - the latest marketing channels being utilized around Super Bowl time. Super Bowl campaigns are no longer complete without the accompanying hashtag or tweet. In a society where every major brand or company has a Twitter and Facebook page, it becomes critical to further supplement TV advertising with social media posts. In order to maximize the full investment made for a 30 second commercial, marketing teams may spend as much as an additional 25% on promoting the commercial itself through memorable social media posts to ensure the ad is spoken about, retweeted, Snapchatted, etc... [11]

One of the most useful aspects of social media marketing is the incorporation of real-time user engagement. Retweeting or answering tweets on Twitter, reposting Instagram photos with a given hashtag, product giveaways, live user contests se-

lecting winners, are all examples of ways campaigns continue to boost their brand during major events like the Super Bowl.

With the increased usage of digital media marketing methods and the continued increase in cost of TV commercials, the Super Bowl remains one of the most critical advertising campaigns of the year for dozens of brands. The most creative companies find ways of combining multiple marketing avenues, incorporating real-time user input, and running campaigns over the longest windows possible before, during, and after Super Bowl Sunday to ensure each dollar spent has the greatest return possible.

It is important to note that increased buzz or engagement on social media does not necessarily have to tie to social media marketing. All the aforementioned marketing techniques, including television marketing, can have an impact on social media buzz. For example, 30% of 2017 Super Bowl TV commercials contained a reference to a Twitter hashtag [12] - thus creating user engagement on Twitter. Social media buzz can be a metric for measuring returns of any form of marketing, digital or otherwise. Salesforce research shows that people who engage with Twitter while watching a TV commercial for a product are 88% more likely to purchase the product [9], further proving that the importance of the metric. Section 2.2 will provide further details on metrics and methods of measuring returns.

## **2.2 Measuring returns**

If millions of dollars are being put on the line for half a minute of TV air time, the returns must be justifiable. One of the most challenging aspects of marketing is the fact that while it is commonly recognized as imperative, the exact value gained through is nontrivial to measure. How does one tell if sales of their beer brand increased due to the particular TV commercial that was run or because the main competitor increased their prices around the same time? Or maybe the increase was

drive by the famous actress who drank the beer in her last movie? Or maybe the additional sales are simply due to the seasonality and being close to spring break when college students are stocking up? Could it be that if half the money had been spent on a lower quality commercial, sales would have still gone up the same amount? What if a TV commercial was not aired at all and the brand relied solely on the fact that the competitor increased prices? These are the kinds of questions marketing teams try to answer as they justify their upcoming fiscal year budget projections to the company CEO.

### **2.2.1 Return on marketing investment**

Return on marketing investment (ROMI) is fundamentally the increase in profits attributed to marketing spend divided by the marketing spend risked [2]. Unlike traditional return on investment calculations which are based on money that is put into more tangible assets like inventory or facilities, ROMI measures marketing funds that are 'risked' in the current period. The formula for ROMI is simple:

$$\frac{\text{Incremental Revenue Attributable to Marketing (\$)} * \text{Contribution Margin (\%)} - \text{Marketing Spend (\$)}}{\text{Marketing Spend (\$)}} \quad (1)$$

As long as the ROMI is positive, the marketing spend is justifiable. The formula itself is simple enough, but the difficult part is understanding what the incremental revenue attributable to marketing really is. First, it is important to make the distinction between the two types of ROMI metric.

#### **2.2.1.1 Short term ROMI**

Short term ROMI is what is most commonly measured by companies when determining how many marketing dollars should be spent. This looks purely at highly quantifiable values such as revenue or market share and their fluctuation due to marketing. The output here is typically something like: every dollar spent on a par-

ticular form of marketing results in an additional \$x dollars in profit or an increase of x% market share.

### **2.2.1.2 Long term ROMI**

Long term ROMI on the other hand focuses on more intangible benefits of marketing such as brand perception or awareness. This is a more complex and sophisticated metric factoring both marketing analytics and business/consumer insights. When done correctly, it can be used to determine a company's marketing strategy, channels, and messaging to accomplish a variety of goals such as brand turnaround or customer demographic improvement or reduction in churn rate (the annual percentage of customers lost).

Long term ROMI will utilize many of the same methods as short term ROMI, but further supplement with models like customer lifetime value (CLV). CLV models the long term value of a single incremental customer acquisition (especially important in subscription based business models like a magazine or gym membership). The most common approach for both short and long term ROMI is through the application of Marketing Mix Modeling.

### **2.2.2 Marketing Mix Modeling**

Marketing Mix Modeling is a statistical regression approach which uses historic data such as product sales to correlate impact of specific marketing campaigns [13]. Using linear regression, the relationship between marketing events and a response variable, like sales, is mapped. By using the equation to define the exact effectiveness of a campaign, or the sales volume generated by that particular activity, companies can then use the ROMI formula to look at how much return their marketing investment produced. This insight taken across a variety of campaigns, regions, products, times, etc can then be utilized to make strategic and tactical decisions to optimize future marketing spend. In order to set up a marketing mix model, the

dependent and independent variables have to be selected carefully, balancing automated programs working through large data sets and experienced econometricians selectively adjusting the inputs. Once the variable sets are created, it takes multiple iterations before an appropriate model is created. The goal of the analysis is to inform what incremental gain in sales can be obtained from an increase in the marketing element by one unit. This one unit could be a dollar spent, it could be an extra 5% discount on a sale, or one extra spot during the Super Bowl game day. This method is invaluable in understanding the effectiveness of various types of marketing, especially given the shift to new digital advertising forms in recent years. Once the model is complete, a marketing team can run scenarios with different allocations of marketing spend on different types of advertising to fully optimize their budget.

The marketing mix modeling method, used by most large consumer packaged goods companies like P&G, Kraft, and Coca-Cola, does have one major limitation. The regression modeling used to form a relationship between the marketing activity and the product sales relies heavily on a large quantity of historic sales data for that product. This makes MMM highly ineffective in managing marketing investments for new products without much historic data.

### **2.2.3 Brand lift**

Another important metric to measure besides just sales, is brand lift. Especially in our increasingly digital world where every opinion is voiced through social media, topics like brand lift and brand perception become more indicative of how well a company or product is doing. Brand lift is defined as an increase in interaction with a brand as a result of an advertising campaign [14]. The importance of brand lift lies in increasing awareness of the product or brand and improving the perception. The purpose of brand lift is to think more long term and look past straightforward sales data. Brand lift metrics include the following components:

- Awareness of the brand or product
- Attitude or opinion on the quality, value, and appeal of the brand or product
- Recall or the ability to remember the brand or product
- Favorability or the likelihood of recommending the brand or product
- Intent or the likelihood of purchasing the brand or product

One possible method of measuring these components is through primary market research. Online surveys and polls are a great way of gathering information on brand awareness and favorability. Many online tools such as Google Consumer Surveys allow your one or two simple questions to be shown as a precursor to accessing content as an alternative to subscriptions. Through social media platforms like Facebook or Twitter, marketing teams can set up polls leveraging the vast online population willing to publicly post their opinions. These polls or surveys can be set up to run right before or right after a marketing campaign, providing a snapshot view of brand lift at any given moment; but with limited insight into the effect over time. Additionally the results are often skewed as it is most likely for customers with strong opinions to respond to a survey. This is most useful not in determining how much money to invest in a marketing campaign, but rather how to tailor the wording and messaging to better improve specific brand lift metrics that may be lagging.

#### **2.2.4 Emergence of new metrics**

As businesses and customers undergo the digital transformation of the consumer purchasing journey from advertising to product research to actual purchasing through to product reviewing, marketing teams need to recognize the emergence of new data and metrics. Supplementing traditional key performance indicators (KPIs) such as

sales data, by adopting other quantitative indicators of a brand's success is now crucial. A few examples of this include:

- **Social media followers or likes:** How many Twitter followers or Facebook likes a brand has is becoming increasingly important as a direct indicator of brand awareness, perception, and likelihood of purchase
- **Share of search:** Quickly becoming a leading indicator of market share, share of search shows how many consumers are searching for a specific brand or product relative to its competitors. Given the 75% increase in e-commerce over the past 5 years [15], searching for a brand online is key in both the product research and product purchase phases of the consumer journey
- **Online searches:** While share of search looks to understand how a brand is stacking up against its competition, purely looking at absolute increases in search volume is telling of how many potential customers have your brand top of mind
- **New visits to a website:** Measuring the change in percentage of new visits to a brand's website, perhaps as a result of clicking on a specific marketing ad, is useful for considering new customer acquisition, important for a brand's sustainability
- **Social buzz:** Similar to social media followers, this metric allows for social response monitoring and any changes in brand sentiment, especially directly following an advertising campaign [16]

As data on these new metrics becomes more readily available, combining it with a strong time series modeling method to measure true causality of marketing activity is the key next step. This paper explores a specific application of Bayesian state space models on such data to explore a number of marketing strategy questions.

## 2.3 Analysis techniques for inferring causality

As previously discussed, there are many difficulties in analyzing the causal impact of marketing events. Dozens of variables are at play at any given moment on top of shifting consumer behavior and seasonal variations, continually complicating the process of isolating the impact of a single discrete advertising event. Additionally from an analytical standpoint, the impact of TV campaigns cannot be tracked at the individual consumer level, the goal is to see how the entire target population acts to affect the response metric.

### 2.3.1 Randomized experiments

While randomized experiments would be the most reliable method of conducting marketing analytics, they are not always possible given the situation. Websites and email based campaigns can often be set up to run randomized control experiments, however TV commercials are less likely to allow for the same level of manipulation. They are often paired with digital marketing to further the reach and impact, thus blurring results across channels. Therefore with pure randomized experiments being unlikely, time series analysis at the market level becomes the ideal method of evaluation.

A/B testing, a type of randomized experiment related to causal inference, is another tool used to survey the effectiveness of various marketing treatments [17]. A/B testing is a form of multi-variant hypothesis testing in which a subject's response to variable A is tested against the response to variable B, typically done to see which variable, or marketing treatment, is most effective. This technique is best used in highly controllable situations like an email advertisement being sent out or website design. An example would be sending one email with a certain subject line (variable A) to a randomly selected population of potential customers, and the same email with a different subject line (variable B) to a different randomly se-

lected population. By measuring the click rate - percent of subjects that proceeded to open the email and click on the website's link - marketers may see if one variable was more effective than the other. As mentioned above, like all randomized experiments, this works best in situations where the reach and effect of the marketing can be contained and appropriately manipulated.

### **2.3.2 Casual inference without randomized experiments**

The fundamental question marketing teams aim to answer is whether treatment T (the ad campaign, commercial, email, product release, etc...) caused the observed outcome Y (an increase in product sales, the acquisition of new customers, a lift in brand awareness, an increase in Google searches, etc...). This can be done by looking at the counterfactual: if treatment T had not occurred, what would the outcome look like? The difference between *that* outcome and the observed outcome Y is the causal effect of treatment T [18]. Without randomized experiments, the key to this is the creation of a control [19]. Analyzing interventions through time series modeling requires two basic steps [4]:

- 1. Identifying control markets** that follow the general trend of the test market where the treatment took place - but were not affected by the treatment in question. This could take the form of different geographies, different but similar brands or products, some proxy variables that measures activity in the industry as a whole, etc... This control helps account for seasonal trends and other external variables affecting the broader picture (e.g., natural disasters, political movements, health studies, consumer behavior shifts, etc...). Time series matching based on data prior to the treatment finds the best set of controls.
- 2. Comparing the observed outcome Y to the outcome modeled** based on the relationship between the test and control markets.

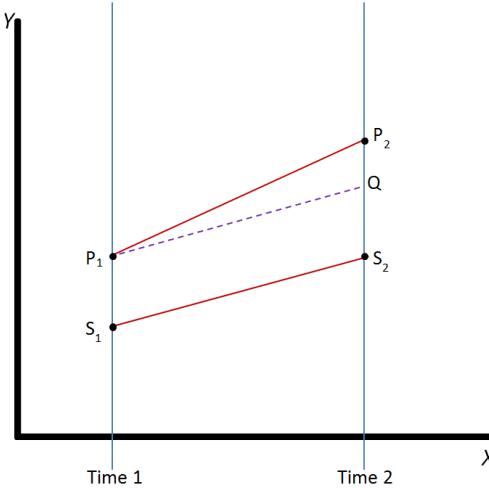


Figure 3: Illustration of difference in differences definition

The most basic approach for time series matching (step 1) is to use Euclidean distance to find controls that best align with the test data [4, 20]. However this approach often over simplifies to throw out potential controls for the slightest deviations between markets. Ideally the control market and test market would be consistently matching, but in reality this is rare. This simplistic approach eliminates viable control market candidates with temporary shifts in relation to the test data. In a later section this paper will discuss more robust methods of selecting controls. For the second step listed above, the traditional inference approach is the difference in differences technique.

#### 2.3.2.1 Difference in differences analysis

The difference in difference technique is typically a static regression model that calculates the difference between *the average change over time in the test variable* from pre to post treatment, and *the average change over time in the control variable* [3]. As seen in Figure 3, the difference in differences formula is  $(P_2 - S_2) - (P_1 - S_1)$ , where P receives the treatment at some point between Time 1 and Time 2.

While a difference in differences model does aim to mitigate a number of external biases through the use of a control market, it typically also requires a few key

assumptions. The first states that the observations be independent and identically distributed (i.i.d.). The second key requirement is the parallel trend assumption which states that the difference between the test and control markets remain constant. Both these assumptions rarely hold true for time series data. Naturally during any given time there will be a number of variables affecting the control market unrelated to the test market undergoing its treatment, thus violating the parallel trend assumption. One of the main limitations of the DD model, is that it does not provide any insight into the change in effect over time. All data points prior to the treatment are averaged to one number, and all data points post treatment are average to another. The model looks only at the difference between the two and outputs a singular causal impact value.

To summarize, the traditional difference in differences approach is limiting in three critical ways:

1. It is a static model with i.i.d data and regression coefficients that do not change with time
2. It only considers two time points, one before and one after the treatment without accounting for the change in effect over time
3. There are significant restrictions on the construction of the control

### 2.3.3 Bayesian structural time series analysis

The alternate approach utilized for this analysis revolves around fitting a Bayesian structural time series model with multiple markets feeding into the control data set to produce a counterfactual post-treatment time series for the response variable assuming the treatment never took place. The difference between the observed response variable time series and this synthetic counterfactual is then taken as the estimated causal impact of the treatment.

A useful machine learning technique for feature selection, time series forecasting, and inferring causal impact, is the Bayesian structural time series model. It is a powerful alternate approach to difference-in-differences for the purposes of marketing analytics. One of the key advantages of BSTS is that it makes it possible to infer evolution of impact over time [4]. Additionally, the model allows for the incorporation of Bayesian empirical priors, building in what is known through observations to prevent over-fitting. BSTS allows for a greater level of flexibility when accounting for sources of variation such as seasonality, or larger trends. Thus the limiting factors of difference-in-differences mentioned above are mitigated.

A structural time series model is a state space model for time series data. State space models represent  $N^{\text{th}}$  order differential equations as  $N$  first order differential equations. The model distinguishes between an observation equation specifying how a system state translates into measurements and a state equation that describes how the state vector evolves through time [4]. These two equations can be defined as follows:

$$y_t = Z_t^T \alpha_t + \epsilon_t \quad (2)$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad (3)$$

The first equation above is the observation equation defining the relationship between observed data  $y_t$ , which is a scalar, and a latent  $d$ -dimensional state vector  $\alpha_t$ .  $Z_t$  is an output vector also of dimension  $d$  and  $\epsilon_t$  is a scalar observation error normally distributed with mean 0 and standard deviation of  $\sigma_t^2$ .

The second equation is the state equation with state vector  $\alpha_t$ ,  $d \times d$  transition matrix  $T_t$ ,  $d \times q$  control matrix  $R_t$ , and  $q$ -dimensional system error,  $\eta_t$ , normally distributed with mean 0 and standard deviation  $Q_t$ .

A large number of autoregressive integrated moving average (ARIMA) models, like those often used for short term forecasting, can be written in the form of

the two above equations, making the flexibility of structural time series models an important benefit. By assuming the errors of various state-component models to be independent, we can concatenate a number of commonly used trend or seasonality models to form the state vector  $\alpha_t$ . State components are assumed to evolve according to independent Gaussian random walks [4]. More details regarding the formation of the state vector for this thesis will be provided in Section 3.3.

By combining the trend and seasonality state-components, with regression state-components from our multiple untreated control markets, it is possible to construct a merged control off of which the synthetic counterfactual prediction can be obtained. The observations from the untreated control markets help account for any other short term trends or variance that is not captured by the generalized trend and seasonality models.

The first state-component model utilized in the approach is the local linear trend, as defined by the following equations:

$$\mu_{t+1} = \mu_t + \delta_t + \eta_{\mu,t} \quad (4)$$

$$\delta_{t+1} = \delta_t + \eta_{\delta,t} \quad (5)$$

In the equations,  $\mu_t$  represents the value of the trend at time  $t$ , while  $\delta_t$  represents the expected change from  $\mu_t$  to  $\mu_{t+1}$ . The two system errors,  $\eta$ , are normally distributed about mean 0. This model is good for quick adaptation to local variation for short term predictions, which is useful when considering shorter term marketing campaigns.

Another state-component model incorporated is seasonality. The most common seasonality model is as follows:

$$\gamma_{t+1} = - \sum_{s=0}^{S-2} \gamma_{t-s} + \eta_{\gamma,t} \quad (6)$$

Here  $S$  is the number of seasons and  $\gamma_t$  is their joint contribution to the observed data  $y_t$ . Since  $s$  goes from 0 to  $S - 2$ , the most recent  $S - 1$  seasonal effects are accounted for. This model can be applied to any form of "seasonality" whether that be the four seasons, or  $S = 7$  to represent a weekly cycle or even  $S = 24$  for a daily cycle if looking at a short time period. Both the  $\mu_t$  and the  $\gamma_t$  vectors become subsets of the  $\alpha_t$  state vector.

The final step is to add the components of state representing the untreated control markets. As previously discussed, these control markets are critical for the creation of an accurate counterfactual since they account for additional variance from unobservable causes beyond the aforementioned trend and seasonality models. The basic way to incorporate the contemporaneous covariates is through linear regression, with coefficients that can be static or time-varying. For static coefficients, the state space form sets  $Z_t = \beta^T \mathbf{x}_t$  and  $\alpha_t = 1$ , where  $\mathbf{x}_t$  denotes the control markets. Since the treatment is fully Bayesian, a large number of covariates can be included with the confidence that a spike-and-slab prior (described in detail next) will incorporate posterior uncertainty regarding which control markets should be included and how heavily they should be weighted to reduce overfitting.

The spike-and-slab method is a Bayesian regression technique useful for picking variables out of large sets of possible predictors. It allows for a number of potential predictors to be narrowed down to the most important ones based on more than just Euclidian distance as previously discussed. The model starts with a vector the length of the number of possible predictors. The values in this vector are either 0 or 1 depending on whether the predictor is included in the regression. Assuming no prior knowledge on which predictors are preferred over others, the model defaults to a Bernoulli prior distribution. A normal prior is then applied to the model coefficients for included predictors. A subset of the vector is taken for the variables for which the value is 1. A subset is also taken from the prior distribution

of the corresponding regression coefficients. A posterior probability distribution is then calculated for both the variables to be included and the coefficients. The process is repeated thousands of times using Markov chain Monte Carlo for posterior inference calculations. From this, a posterior distribution of the variables to be included, the regression coefficient values, and the corresponding prediction can then be obtained. To recap, the steps of the algorithm are as follows:

1. Start with a vector the length of the controls
2. Assign values of 0 or 1 based on a Bernoulli prior distribution indicating whether the control is included in the regression
3. Assign coefficients to all the controls based on a Gaussian prior distribution
4. Take a draw of control markets that were assigned a 1, and the corresponding coefficients
5. Calculate a posterior probability distribution for both (the controls to be included and the coefficients)
6. Repeat the process, revising the inclusion probabilities and the coefficients until steady state is reached
7. Repeat the entire process a thousand times using Markov chain Monte Carlo

The name of the model refers to the shape of the two prior distributions: a spike for the probability of a variable being chosen for the model and thus not zero (the product of the independent Bernoulli distributions per variable), and the slab for the prior distribution of the regression coefficient values (the wide-variance Gaussian prior).

This step of the Bayesian structural time series allows for the appropriate control markets to be selected while avoiding overfitting by promoting a sparseness during

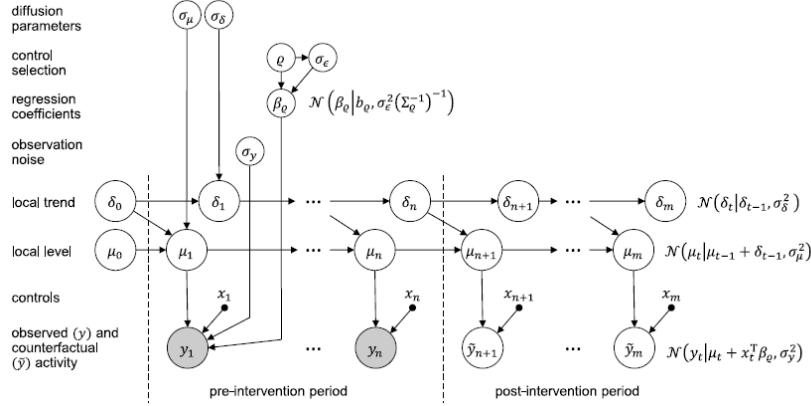


Figure 4: *Graphical representation of BSTS model*

selection [20]. Since the model is Bayesian, the coefficients follow a random distribution and allow for the incorporation of uncertain historical relationships between markets that often are correlated. This flexibility in forecasting is much more powerful than fixed coefficients as in the case of difference-in-differences. Figure 4 shows the graphical representation of the model [4]. Once the posterior distribution over the parameters and states ( $\alpha$ ) is calculated, it is possible to predict the counterfactual. By integrating out the parameter estimates and the state components (including the vectors determining the inclusion probabilities and the coefficients of the control markets), we are left with a posterior density for the counterfactual time series that is conditional only on the observed data (treated market before intervention and controls markets both before and after intervention):

$$p(\tilde{y}_{n+1:m} | y_{1:n}, x_{1:m}) \quad (7)$$

Due to the nature of the the Bayesian methodology, while the states and parameters are accounted for, the final model depends only on the observations.

### **3 Application of BSTS to Super Bowl ad campaigns**

#### **3.1 Statement of Problem**

The goal of this thesis is dual in nature. The task at hand is to infer the causal impact on digital buzz generated for various products and brands by Super Bowl related ad campaigns. Note the ad campaigns themselves may be digital, traditional (TV), or a mix of both. The two objectives are to test the technical applicability of the BSTS model while simultaneously drawing forward a number of business insights regarding one of the largest marketing campaigns of the year. For the purposes of this thesis, digital buzz has been measured in two ways: the amount of Google searches performed for the term or brand in question and the Twitter activity surrounding the brand.

These two facets of digital buzz, Twitter mentions and Google searches, were selected for a number of reasons. First, they will provide an interesting contrast as some brands focus heavily on their Twitter marketing in accompaniment of a TV commercial, while other brands do not. Twitter also can be viewed as a more public forum of Internet buzz where users need to consciously make the decision to post their opinion on a brand, knowing it will be viewed by many, as opposed to a Google search which can be done in private to satisfy a user's curiosity. This ties back to the element of brand lift previously discussed: favorability or the likelihood of recommending the brand or product. As was discussed in the background section, in the modern digital world the average customer journey - from learning about a product, through purchasing it, to potentially leaving a review or recommendation for peers - is increasingly grounded in technology and social media. Thus, measuring the digital buzz produced by marketing is critically important. Google and Twitter have been selected as proxies for digital buzz due to their being two of the most prevalent websites in use. According to Statista, the share of search traffic in the

U.S. that originates from Google is 81.4% [21] and as of the end of 2015, Twitter had roughly 305 million monthly active users [22]. These two websites, along with a few other major players, dominate the digital market.

Additionally, the existence of a positive relationship between Twitter mentions and revenue for certain types of products and markets, such as box office revenue, further motivates brands to understand the causal impact of marketing on this metric [23]. The final reason these two sources were selected was due to the availability of data. Both Google search and Twitter mentions data is publicly available, albeit in a somewhat limited manner (to be discussed in Section 3.2).

The two main criteria used to define level of buzz are the amount and the length of time the buzz remains. Unlike in marketing campaigns such as price discounts or a new slogan that can last for a much longer time frame, Super Bowl advertising is short. Millions of dollars are poured into marketing for an event lasting roughly 4 hours. There is some time before and after the actual game to push marketing agendas, but overall the window is quite short. Therefore, the buzz created must be massive in quantity to outweigh the brevity in length in order for the dollars spent to be justified. This hypothesis is one of many tested by this thesis.

In order to infer causality of Super Bowl marketing on digital buzz, a Bayesian structural time series model will be employed on relevant brands with control markets being consumer products with absolutely no association to the event. The BSTS model will be fit using data prior to the event, along with any given knowledge of trends and seasonality. The spike-and-slab method will aid in determining which control markets are best suited for inclusion. Based on the model, a counterfactual prediction of the given brand's digital buzz will be generated for the post-event period assuming no event took place. Finally, in a Bayesian fashion, the difference between the counterfactual and the observed data will be used to quantify the causal impact.

As stated above, the goal of the thesis is two fold: the first is technical and the second is business driven. First, we will test the use and applicability of the BSTS model, test its various model parameters, and compare it to a standard difference-in-differences approach to see how the results compare. Second, we will use the application of this machine learning technique and approach on Super Bowl marketing campaigns to analyze the findings with a business lens in order to hypothesize and answer a number of relevant questions such as:

- How much digital buzz does a Super Bowl ad typically generate?
- How long does the buzz last?
- Does buzz last longer on Twitter or Google?
- Do certain types of products (e.g., sports drinks, beer, candy, cars) get more of a push than others?
- How do different types of marketing campaigns (e.g., game day commercial, hashtag campaign, celebrity endorsement, Super Bowl product label, etc...) compare?

Data availability aside, the choice to infer the causal impact of Super Bowl marketing on digital buzz rather than on sales data directly was also influenced by the "Super Bowl Impossibility Theorem" coined by Lewis and Rao in 2012 [24]. Their hypothesis essentially states that since Super Bowl commercials cost roughly 3 cents per viewer (see section 2.1), if the ad were to have an impact on profits of 4 cents per viewer, the company would profit \$1.7M. But if the ad had an impact of 2 cents per viewer, the company would incur that very amount in losses. The line between profit and loss is so fine that the sales data noise makes it nearly impossible to infer whether a campaign was profitable or not. In this thesis, we instead focus on

the value of driving digital buzz as a means of increasing brand awareness, gaining new customers, staying ahead of competition, etc...

### 3.2 Data sources

There are three sources of data needed for constructing a control and a synthetic counterfactual:

1. Response variable time series data
2. Contemporaneous time series data for a predictive control
3. Available prior knowledge about the model parameters

The third item on the above list comes from previous research and general understanding of seasonality or trends. The first two items are the main types of data needed to run the Bayesian structural time series model. The first, the time series behavior data of the selected response variable, is essentially the Google search and Twitter mentions data for an effected product or brand, like Bud Light. The second, the time series behavior data for predictive control markets, put in simpler terms is the same Google and Twitter data for non-effected control products, such as Blue Moon. The data collected (both response variable and controls) should start some time before the Super Bowl (in order to create an accurate synthetic control) and end some time after (in order to obtain a counterfactual and measure its difference from the observed response variable).

The Google search data was sourced from Google Trends which allows users to download normalized time series data for any search term. The data is unbiased and excludes repeated searches from the same person over a short period of time. The limitations are as follows:

- Minute by minute data is only available for the last 4 hours

- Eight minute data is available for the last day
- Hourly data is available for the last 7 days
- Anything older than 7 days can only be accessed as daily data

These limitations on how data can be downloaded created some hurdles. We discovered very early on that daily data is not particularly useful when looking at short term marketing campaigns with a duration of 24 hours. The search term "Super Bowl" itself only receives heightened Google search or Twitter attention for roughly 5-8 days after the game, making it rare that any advertising campaigns continue to rely on it for much longer than that. Daily time series data sets spanning from before to after the Super Bowl with roughly 20 points in total hardly serve for a machine learning model. Therefore the data used for this thesis is at least hourly if not more granular. Unfortunately, this limitation ruled out the possibility of pulling data from past Super Bowls. All data had to be collected live to ensure the proper granularity in the time series could be captured.

The Twitter mentions data was scraped from a social media analytics website called Talkwalker. Without purchasing a full license, the data is again somewhat limiting. This website only provides Twitter mentions data for any term for the last 7 days, in the form of hourly data. Given the restrictions, all analyses performed in this paper are based on Google search and Twitter mentions data for Super Bowl LI played on Sunday, February 5th, 2017 at 6:30pm EST.

The table below shows the response variables and control variables selected. It is important to note that while similar types of products were selected as both response and control, products that seem entirely unrelated can still provide useful control markets especially when looking at trends such as weekend peaks and troughs, nighttime decreases, and more. Since the spike-and-slab technique safely weights the various markets and decides which to include in the final synthetic con-

trol, we need not limit ourselves to only highly similar products. For example, Twix could serve as a perfectly viable control market for Buick. The key is only to make sure none of the control markets were impacted by Super Bowl advertising. This was done through internet research as well as close monitoring of Google, Twitter, and the control brand's website to confirm no relevant Super Bowl marketing campaigns were run.

Response markets (treated brands)	Control markets (untreated brands)
84 Lumber	Almond Joy
Audi	Blue Moon
Avocados from Mexico	Burger King
Bai	Corona
Bud Light	Dasani
Buick	Home Depot
Coca-Cola	Lowe's
Doritos	M&Ms
Fiji Water	Mountain Dew
Intel	Pringles
Kia	Shock Top
LIFEWTR	Smartwater
Microsoft	Twix
Skittles	
Snickers	
Sprite	
Tostitos	
Toyota	

### 3.2.1 Data treatment and normalization

The main treatment the data required was concatenating multiple pulls of data together due to the limited time frame that could be downloaded at any given time. As previously mentioned, the Google data was available in normalized form only, meaning that for any given pull of data, the highest number of searches in that time period was normalized to 100, which each data point scaling down from there. This required rescaling and adjusting when multiple separate data pulls were combined into one data set. The Twitter data was available as absolute number of mentions.

## 3.3 BSTS analysis

A Bayesian structural time series model was utilized to infer causal impact of the various types of Super Bowl marketing campaigns for the aforementioned response markets selected on their digital buzz. In order to properly develop and fit the model, a number of variables had to be tweaked and carefully adjusted. The first steps involved structuring the data in the appropriate way, i.e., properly formatting the date and time columns, as well as columns for each control market time series data and the affected brand in question. For the purposes of this thesis, the  $\alpha_t$  state vector, where  $t$  is from the start of the data to up until the treatment, was formed by concatenating three types of state components: the control markets (scaled by static linear regression coefficients), a local linear trend, and a "seasonality" model. Next the model had to be tuned on the proper pre and post-event windows. The following sections will describe in more detail this implementation. As previously mentioned, this portion was in fulfillment of the technical aspect of the thesis. After the model was fully developed and satisfactory, it was applied in a number of different ways to draw conclusions regarding the marketing and business benefits of Super Bowl advertising.

### **3.3.1 Implementation in R**

The Causal Impact R package developed by Brodersen et al. in 2015 served as the framework and method of implementation for this thesis [25]. The R package implements the approach discussed in section 2.3.3.1, beginning with the spike-and-slab model utilized to create a control combining multiple unaffected markets and ending with inferring causality as the difference between the generated counterfactual and the actual observed values for the response variable.

#### **3.3.1.1 Model assumptions**

Causal inference requires a number of strong assumptions to draw valid conclusions. For starters, arguably the most important assumption is that the markets chosen to be part of the control were not affected by the treatment themselves. If they were, we may underestimate the effect on the response variable or potentially even overestimate it if the treatment had a negative effect on what was thought to be a control market. We also assume that the relationship between the treated and untreated time series remain fairly stable as established during the pre-treatment period. Unlike a traditional difference in differences analysis, we do not need to hold to the assumption that the observations be i.i.d..

#### **3.3.1.2 Utilizing and adjusting the model**

The model requires two inputs, the response time series and at least one control time series. The R package then constructs a time-series model, performs posterior inference on the counterfactual, and returns a number of results summarizing the treatment's effect on the response variable. To estimate this effect, we first specify which portion of the data is the training, or pre-intervention, period and which period is for computing the post-intervention counterfactual.

The first decision that needed to be made was just how long the post-intervention period should be. As with any marketing campaign the impact has a specific win-

dow after which the response variable will either return to normal levels or become affected by other market conditions (e.g., a new campaign). It is important to identify a proper window of time during which the impact is expected as opposed to selecting an extended intervention period during which the effect has already worn off. Due to the nature of the Super Bowl and accompanying advertising, this window is fairly short yet the impact high. The first parameter tuning conducted in the process of creating an accurate model was to test various post-intervention lengths to determine what the ideal window should be. The various windows tested were from 9am EST on Super Bowl Sunday to:

1. The end of the game ( 11pm EST on Sunday)
2. 1 day after (Midnight PST Monday night)
3. 2 days after (Midnight PST Tuesday night)
4. 3 days after (Midnight PST Wednesday night)
5. 4 days after (Midnight PST Thursday night)
6. 5 days after (Midnight PST Friday night)
7. 6 days after (Midnight PST Saturday night)
8. 1 week after (5 pm EST Monday, February 13th)

The reason we stop there is due to the potential interference from Valentine's day marketing, especially for the candy or chocolate related products. The detailed results can be found in section 4.

Another adjustment made to the model was the number of control markets utilized. It is assumed that the more control markets are submitted as candidates to the spike-and-slab algorithm, the better the model will be. A brief test was done to see how the results varied by number of control markets. The model was tested with 1,

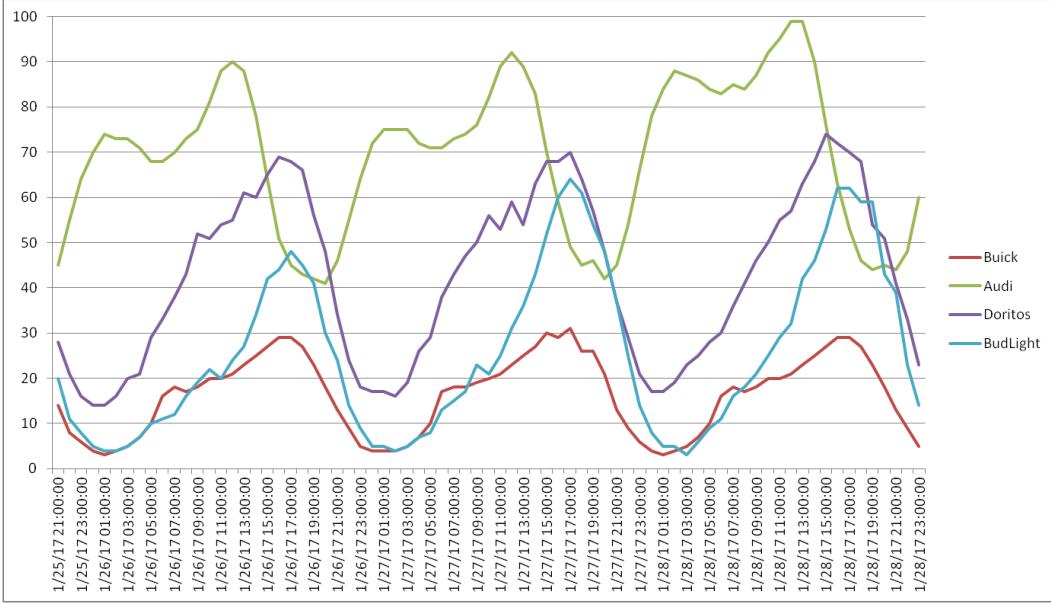


Figure 5: *Pre-Super Bowl Google search trends*

5, and 11 control markets. The results can be found in section 4. These learnings can be informative for future users of the model, in whatever application of BSTS.

Aside from these adjustments, the model was also fit specifically to our Super Bowl campaign application. There are a number of tunable parameters built into the R Causal Impact package. The first change made was to add a "seasonality" trend for weekends vs. weekdays. As can be seen in Figure 5, much of the Google search data has variability between weekdays and weekends as well as peaks for time zones and night vs. day. By informing the model that the data is hourly with 7 day cycles we can further improve the accuracy of the model.

Another parameter chosen was whether to use static or dynamic coefficients. Since the overall time frame of the Super Bowl is quite short, the relationship between the control markets and the treated market is fairly stable leading up to the event. This makes static coefficients a good option. Had we been examining a longer period of time leading up to the event, and a longer event overall, it may have been useful to incorporate dynamic coefficients.

By performing these parameter adjustments and tests, we can get a sense of

how the model can be properly fitted. From there, each product requires its own customized model because parameters such as the post-intervention period need to be tweaked slightly differently depending on the type of marketing campaign run.

### **3.3.2 Comparison to difference-in-differences method**

As discussed in section 2.3.2.1, the traditional method for performing causal inference analysis used to be a difference in differences model. To see just how the results varied for our particular use case, a BSTS model and a DD model were run on the same data sets. Given the nature of the analysis being performed, it is difficult to judge whether one method is more accurate than another, however we can at least see if the results are directionally consistent (i.e., is the estimated causal impact similar, always off in the same direction, always off by a consistent amount, etc...) and how the confidence intervals compare.

## 4 Results

### 4.1 BSTS model testing

The output of the Causal Impact R package contains the following information regarding the response variable time series during the post intervention period:

	Average	Cumulative
Actual	#	#
Prediction (s.d.)	# (#)	# (#)
95% Confidence Interval	[#, #]	[#, #]
Absolute effect (s.d.)	# (#)	# (#)
95% Confidence Interval	[#, #]	[#, #]
Relative effect (s.d.)	% (%)	
95% Confidence Interval	[% , %]	
Posterior probability of effect	%	

When measuring a response variable like sales, the "cumulative" column is most meaningful, while a variable like stock price would require analysis based on the "average" column. For the purposes of this paper, since the Google search data is only available as normalized values (i.e., the highest number of searches in a given time period is assigned the value of 100, and the rest is scaled down from there), neither the average or cumulative values are of particular interest. Instead, we focus on the % relative effect and the accompanying standard deviation. This tells us that an ad campaign resulted in an x% increase in searches. The Twitter mentions data on the other hand is in absolute number of mentions per hour. In this case we can consider the average number of mentions as well as cumulative mentions over the given post intervention period, in addition to looking at the % relative effect.

Figure 6 shows the aforementioned output data in visual form, using Avocados from Mexico as the selected product. The first panel of the three panel graph

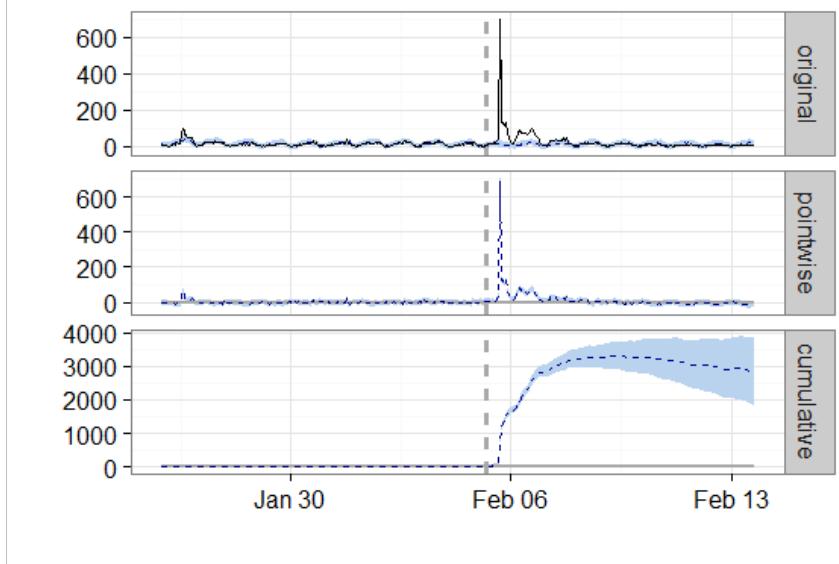


Figure 6: *Avocados from Mexico causal impact output*

shows the actual time series Google search data as a solid line and the counterfactual prediction as a dotted line. The second panel shows the pointwise difference between the actual data and the prediction, i.e., the causal impact over time of the Super Bowl marketing ad. The third panel shows the cumulative effect of the campaign over time. For all three panels the blue shaded area represents the confidence interval.

#### 4.1.1 Varying the post period window

It was found that the length of the post period window had a significant impact on the calculated causal impact of the marketing event. The Causal Impact R package and the accompanying paper describing the algorithm and its use did not relate details regarding how best to choose the length of the post period window [4], therefore this variable was varied and tested to better understand how to make this decision in using the model.

It is not always clear how long the impact of a marketing campaign will last before fading. By varying the post period time until the 95% confidence interval included zero, the length of time could be determined for each product. From this

information, two statements can be made: at what period of the time post event does the response variable reach it's maximum impact, and how long does the impact last before settling back to normal pre-event levels. Graphically this can be seen in Figure 6 as the time at which the second panel reverts to a consistent zero and the slope of the third panel flattens.

#### 4.1.2 Controls

A powerful aspect of the BSTS model is the ability to combine information from multiple control markets to calculate a more accurate counterfactual prediction of the response variable. Testing the model with 1, 5, and then 11 control markets not only changed the counterfactual prediction (either making it lower or higher depending on which controls markets were selected) but always resulted in a narrower confidence interval. However the reason for this increase in accuracy is not due to simply adding *more* control markets to the final model, but rather adding the *right* control markets.

Contrary to common intuition, the best control markets for a certain response variable are not always similar products. The spike-and-slab algorithm is crucial for identifying the underlying trends that correlate markets, and weighting the inclusion probability of the controls appropriately. For example, it was found that Corona was actually a poor control for Bud Light, but Mountain Dew was quite strong (see Figure 7). The value of providing a high number of control markets comes from the ability of the spike-and-slab method to identify the best control markets to pull out and use. Often it may turn out that only 3 of the provided controls are selected, but by allowing the model to identify which 3, we improve the accuracy of our results.

Figure 8 shows how different the control markets selected for Audi versus Buick were. Likewise, Figure 9 shows Skittles versus Snickers. The controls selected in all these cases were not intuitive and could not have been chosen without the spike-and-slab.

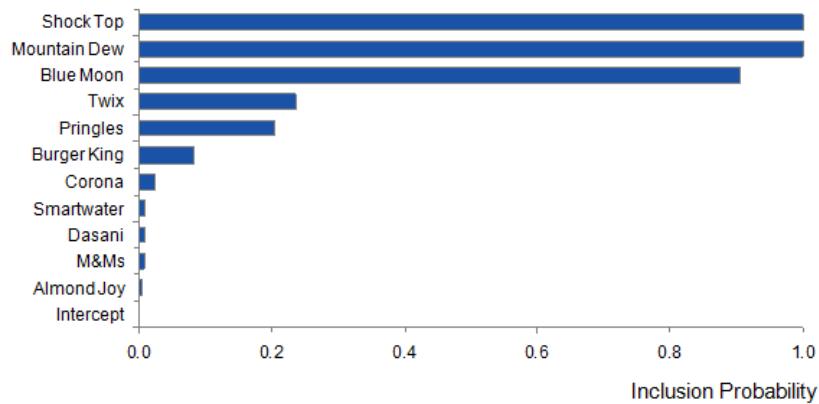


Figure 7: *Bud Light controls as weighted by the spike-and-slab algorithm*

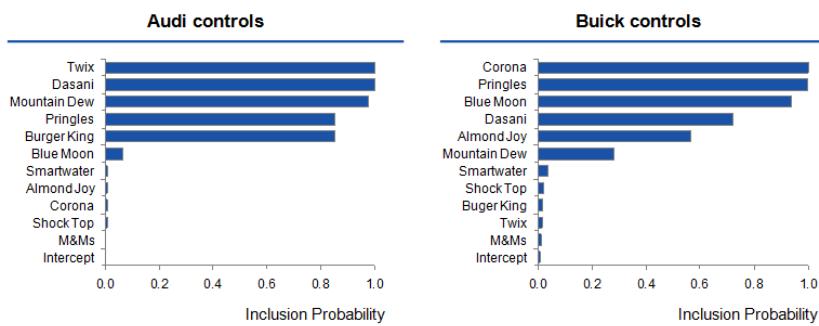


Figure 8: *Audi control weighting versus Buick*

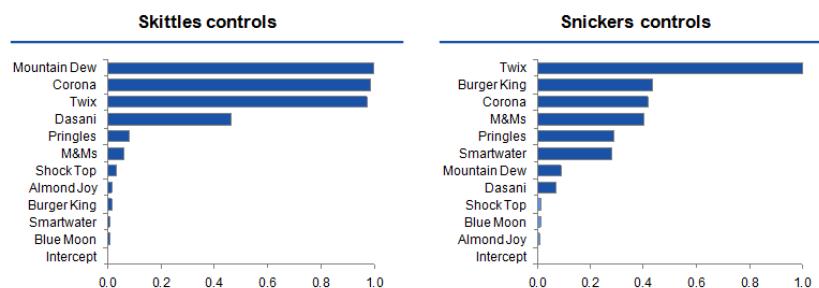


Figure 9: *Skittles control weighting versus Snickers*

		All controls		Controls with prob. > 0.4	
<b>Snickers</b>	Actual	216	3243	216	3243
	Prediction (s.d.)	71 (2.5)	1068 (36.8)	71 (2.2)	1066 (33)
	95% CI	[67, 76]	[999, 1137]	[67, 76]	[1003, 1134]
	Absolute effect (s.d.)	145 (2.5)	2175 (36.8)	145 (2.2)	2177 (33)
	95% CI	[140, 150]	[2106, 2244]	[141, 149]	[2109, 2240]
	Relative effect (s.d.)	204% (3.4%)		204% (3.1%)	
	95% CI	[197%, 210%]		[198%, 210%]	
	Posterior prob. of effect	99.89%		99.89%	
<b>Skittles</b>	Actual	230	3452	230	3452
	Prediction (s.d.)	72 (3.1)	1075 (46.5)	71 (2.9)	1070 (43.7)
	95% CI	[66, 78]	[985, 1167]	[66, 77]	[985, 1153]
	Absolute effect (s.d.)	158 (3.1)	2377 (46.5)	159 (2.9)	2382 (43.7)
	95% CI	[152, 164]	[2285, 2467]	[153, 164]	[2299, 2467]
	Relative effect (s.d.)	221% (4.3%)		223% (4.1%)	
	95% CI	[213%, 229%]		[215%, 231%]	
	Posterior prob. of effect	99.89%		99.89%	

Figure 10: *Causal impact results with threshold on control market inclusion probability*

A threshold was then applied to the inclusion probability of the control markets. Controls with a probability of less than .4 were excluded in the BSTS model in an attempt to reduce overfitting. There was a slight improvement in the confidence interval but overall the impact was small. Figure 10 shows sample results for Snickers and Skittles. As can be seen, the standard deviations decreased slightly with the removal of the additional control markets.

## 4.2 Super Bowl marketing impact

Figure 11 shows the results of BSTS model in determining the causal impact of Super Bowl marketing campaigns on Google searches for each of the analyzed products. From this data, it is possible to hypothesize regarding the questions previously mentioned in section 3.1.

The amount of buzz generated by a Super Bowl ad campaign varies greatly, as do the types of campaigns, the quality of the marketing, and the products being marketed. The data appears to show a 25-40% increase in Google searches for car brands, with the exception being Toyota, who may have just have had a bad year in terms of marketing quality in 2017. Candy brands such as Skittles and Snickers received a 200% increase in searches, significantly higher than cars. Software

Product	% increase in Google searches at peak	Peak buzz	Length of buzz
Bud Light	129%	After game ends	4 days
Bai	0%	None	None
LIFEWTR	413%	After game ends	8 days
Fiji Water	325%	After game ends	8 days
Coca-Cola	74%	After game ends	8 days
Sprite	8%	After game ends	1 day
Avocados from Mexico	1009%	After game ends	8 days
Doritos	221%	After game ends	8 days
Tostitos	95%	After game ends	1 day
Snickers	204%	After game ends	8 days
Skittles	221%	After game ends	8 days
Buick	41%	After game ends	5 days
Audi	35%	After game ends	3 days
Kia	25%	After game ends	7 days
Toyota	0%	None	None
Intel	17%	Consistent for 4 days	8 days
Microsoft	30%	2 days after	5 days
84 Lumber	59007%	After game ends	8 days

Figure 11: Results of BSTS model in determining the causal impact of Super Bowl marketing on Google searches

brands such as Intel and Microsoft again are on the lower side of the scale at 15-30%.

Avocados from Mexico, a brand who returned to Super Bowl marketing for the third consecutive year in 2017, had a significant boost in Google searches with their 30 second spot which aired during the first commercial break of the game. According to Kevin Hamilton, director of brand marketing at Avocados from Mexico, this early spot was chosen because "everybody is paying attention early on, while audience attention may lag later in the evening if the game play is not captivating, the spot has the benefit of being independent of the action of the game" [26]. Additionally, the brand remained socially active throughout the game with continuous Twitter postings of real time game commentary with avocado puns and a hashtag to accompany their TV commercial, further boosting their digital buzz on both Google and Twitter.

The one brand that performed well above all others was 84 Lumber, an American building supplies company. The company's game day commercial featured a

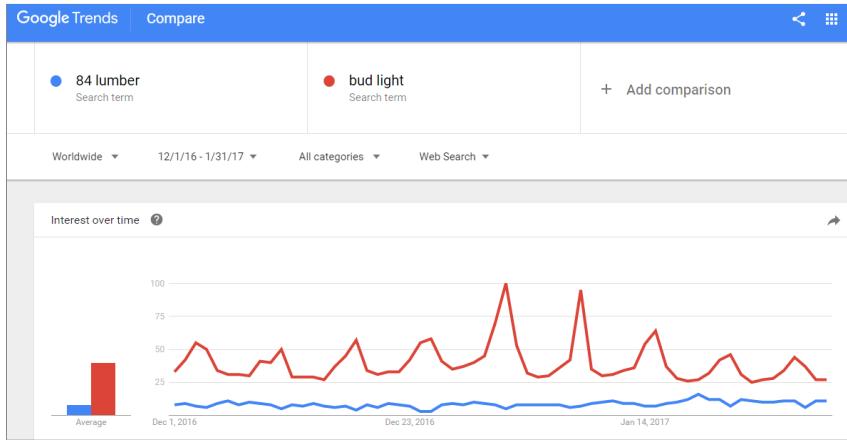


Figure 12: December through January Google search data for Bud Light and 84 Lumber (pre-Super Bowl marketing campaigns)

politically controversial story regarding illegal immigration from Mexico and President Trump's plans surrounding the wall to be built on the border. The first half of the story was aired during the game, ending with a cliffhanger and the message to visit the company's website to see the ending. The website crashed shortly after due to the influx of visitors trying to view the page. The spike in Google searches for this brand were the highest of any of the Super Bowl marketing done, both as a relative % increase for themselves as well as relative to other brands. Figure 12 below shows the December through January Google search data for Bud Light and 84 Lumber on the same scale to show the relationship between the two. Pre-Super Bowl marketing, Bud Light was receiving roughly 5 times as much Google buzz as 84 Lumber. Figure 13 shows the comparative view of Bud Light, which received a 129% push, and 84 Lumber during the Super Bowl marketing period. The spike in 84 Lumber searches is about 100 times higher than the spike Bud Light received, and lasts significantly longer. 84 Lumber's 2017 Super Bowl marketing strategy can serve as a case study for many companies in the future. Their commercial was not only engaging and relevant under today's political landscape, but also drew large audiences directly to their website.

Along with analyzing how buzz across types of products compares, it is also

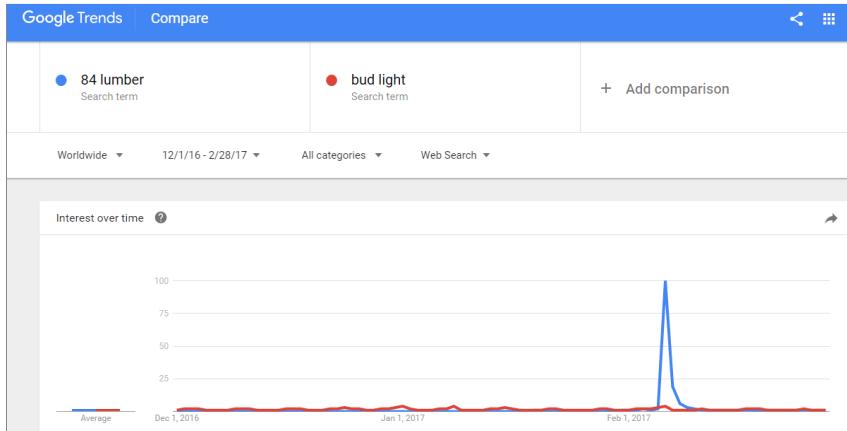


Figure 13: *December through February Google search data for Bud Light and 84 Lumber (post-Super Bowl marketing campaigns)*

possible to compare various types of Super Bowl marketing strategies. 84 Lumber presented a politically charged commercial which they tied directly to their website, bringing in the highest volume of Google searches of all the brands tested. Snickers' strategy involved filming a commercial in real time, using the most recent score in the commercial as proof that it had not been shot ahead of time. Tostitos focused on altering their product label itself for the Super Bowl. Partnering with Uber, Tostitos' marketing campaign involved "party safe bags" which allowed consumers to blow on the bag to see whether they had any alcohol on their breath and should call an Uber instead of driving. This strategy generated a 100% increase in Google searches, most accompanied with the phrases like "Super Bowl bag" and questions on where to buy one. However, this strategy also led to one of the shortest lasting impacts, ending roughly 1 day after the Super Bowl. Kia's Super Bowl commercial stuck out as having the celebrity Melissa McCarthy starring in the commercial. While celebrity endorsements can help boost buzz, Kia only received a 25% increase. Microsoft and Intel display a different marketing approach for the Super Bowl, focusing on sponsorship rather than TV commercials. Both brands' logos appeared throughout the game in the form of Microsoft Surface tablets used by the coaches or screens displaying "Sponsored by Intel" during the half time show.

Product	% increase in Google searches at peak	Peak buzz (Google)	Length of buzz (Google)	% increase in Twitter mentions at peak	Total cumulative causal Twitter mentions	Peak buzz (Twitter)	Length of buzz (Twitter)
Bud Light	129%	After game ends	4 days	0%	None	None	None
Doritos	221%	After game ends	8 days	65%	9,440	1 day after	8 days
Snickers	204%	After game ends	8 days	160%	4,370	After game ends	6 days
Kia	25%	After game ends	7 days	53%	5,569	After game ends	2 days
Toyota	0%	None	None	0%	None	None	None
84 Lumber	59007%	After game ends	8 days	3188%	34684	1 day after	8 days

Figure 14: *Causal impact of Super Bowl marketing on Google searches versus Twitter mentions for select brands*

This strategy did not result in a steep peak directly after or during the game as was common with other brands, but rather a delayed increase in Google searches. One hypothesis for this may be that rather than create immediate buzz, these tech companies plant the brand name in the back of viewer's minds so that when the time comes for a consumer to purchase a new device, they remember the tablets or other electronics from the Super Bowl.

LIFEWTR was another brand that received significant Google attention from its Super Bowl marketing. A purified bottled water launched by PepsiCo in February of 2017, LIFEWTR's commercial aired during the Super Bowl creating buzz that lasted for at least a week after the game. Timing the product launch with the Super Bowl proved to be a smart tactic given the 400% spike in Google searches.

Almost all the products reached a peak in digital buzz directly after the game ended, with the effects lasting various lengths of time. Figure 14 shows a comparison of the Google buzz and the Twitter buzz generated for a select number of products. Most of the brands tested received less of a causal impact on Twitter mentions than Google searches, with the exception being Kia. Typically buzz on Twitter lasted a shorter length of time than Google as social media moved on from the Super Bowl onto other topics of interest.

		BSTS	DD
		After game ends	After game ends
Brand	Metric	Average	Average
Bud Light	Actual	122	
	Absolute effect (s.d.)	69 (2.8)	77 (6.3)
Doritos	Actual	210	
	Absolute effect (s.d.)	145 (2.4)	144 (7.0)
Kia	Actual	64	
	Absolute effect (s.d.)	13 (1.5)	4 (5.6)
Toyota	Actual	81	
	Absolute effect (s.d.)	-2.2 (1.4)	-9 (5.6)
Snickers	Actual	216	
	Absolute effect (s.d.)	145 (2.5)	127 (7.1)
84 Lumber	Actual	1215	
	Absolute effect (s.d.)	1213 (0.13)	1127 (74)

Figure 15: *Causal impact of Super Bowl marketing on Google searches for select brands as calculated by the BSTS model versus the Difference-in-Differences model*

### 4.3 Comparison to difference-in-differences method

The difference-in-differences model performed similarly to the BSTS model as can be seen in Figure 15. Overall the absolute effect calculated by the DD method was directionally consistent in all cases and almost always slightly lower than the BSTS method, with the exception being Bud Light. The major difference between the results of the two methods is in the standard deviation which in every case was much larger for the DD method, resulting in significantly wider confidence intervals. In the particular instance of Kia, the BSTS method predicted an absolute impact of 13 with a standard deviation of 1.5, meaning it is safe to assume the marketing had an impact. However, given the same data, the DD model predicted an impact of 4 with a standard deviation of 5.6, meaning 0 would be included in the confidence interval and therefore the impact is questionable. In marketing campaigns where the causal impact is low, close to 10-15% perhaps, the wide confidence intervals and slight underestimation of the DD model could lead to the misinterpretation that the advertising had no impact, while the BSTS model may be more accurate.

## 5 Conclusions and Future Work

### 5.1 Conclusions

While more analysis is needed, it appears the BSTS model is a useful tool for measuring causal impact of marketing events on digital buzz. Applying daily and weekly trends to hourly data further improves the model as does feeding in a large number of control markets for the spike-and-slab algorithm to select. The most accurate control markets for a given product are often counter-intuitive and not simply similar products, make the spike-and-slab method quite important. The confidence intervals can be narrowed by further applying a threshold to the control market weighting to prevent overfitting.

The parameter driving a significant amount of variance in the results is the length of the post-event window. Further analysis is needed here to determine the best way to select this window.

For the brands selected in this study, Super Bowl marketing campaigns almost always drove short term significant impact on digital buzz. Certain types of products tend to get more of a boost while other products may have longer lasting buzz. Buzz typically lasts longer on Google than on Twitter, although more data would be needed to confirm this.

The BSTS model as compared to the difference-in-differences model displays directionally consistent results in terms of inferring causal impact, however with much narrower confidence intervals, resulting in a more reliable result. BSTS models appear to be a better choice especially in the cases of lower impact marketing campaigns (e.g., 15% impact) where the DD wide confidence intervals often make it difficult to determine if an impact even exists.

## 5.2 Future Work

Improvements can be made to the model including but not limited to the addition of more potential control markets, further testing of the post-event period, or the application of more trends and prior known information.

The largest limiting factor for the analysis presented here was availability of data. If the data for past Super Bowl marketing campaigns could be made available, a number of trends and comparisons could be analyzed year to year, for a larger number of brands, to truly understand which types of marketing campaigns generate the most buzz. Additionally the Google data currently used was only available normalized, making it difficult to get a sense of the impact in absolute numbers as well as to compare results from brand to brand. By pulling digital buzz data for other social media platforms or other metrics besides digital buzz, the impact of Super Bowl marketing could continue to be validated.

The BSTS model unlocks a large variety of possible applications besides marketing. Future analysis could involve the application of the model on a number of current issues in politics or society such as the causal impact of gun shootings on gunmaker stock prices or the causal impact of President Trump's tweets, to name a few examples.

## References

- [1] V. Letang and L. Stillman, “Global advertising forecast,” *Magna Global*, 2016. [Online]. Available:  
<http://magnaglobal.com/wp-content/uploads/2016/12/MAGNA-December-Global-Forecast-Update-Press-Release.pdf>
- [2] G. Powell, *Return on Marketing Investment: Demand More From Your Marketing And Sales Investments*. RPI Press, 2003.
- [3] A. Abadie, “Semiparametric difference-in-differences estimators,” *Review of Economic Studies*, vol. 72, 2005. [Online]. Available:  
<https://economics.mit.edu/files/11869>
- [4] K. Brodersen, F. Gallusser, J. Koehler, N. Remy, and S. Scott, “Inferring causal impact using bayesian structural time-series models,” *The Annals of Applied Statistics*, vol. 9, 2015. [Online]. Available:  
<http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/41854.pdf>
- [5] B. Steinberg, “Tv ad prices: Football hikes, walking dead stumbles, chicago fire, goldbergs on the rise,” = <http://variety.com/2016/tv/news/tv-ad-prices-football-walking-dead-empire-advertising-1201890660/>, October 2016.
- [6] L. Watters, “The history of super bowl commercial costs,” American Marketing Association, January 2017. [Online]. Available:  
<https://www.ama.org/publications/marketingnews/pages/history-of-super-bowl-commercial-costs.aspx>
- [7] Nielsen, “Super bowl 50 draws 111.9 million tv viewers, 16.9 million tweets,” = <https://www.ama.org/publications/marketingnews/pages/history-of-super-bowl-commercial-costs.aspx>, February 2016.
- [8] M. Glassman, “Super bowl ad stats,” Bloomberg, February 2015. [Online]. Available: <https://www.bloomberg.com/news/articles/2015-02-02/super-bowl-ad-stats-115-ads-13-minutes-of-cars-two-screaming-goats>
- [9] “How to succeed at super bowl marketing,” February 2017. [Online]. Available: [http://www.huffingtonpost.com/entry/how-to-succeed-at-super-bowl-marketing\\_us\\_58937ebce4b0b4d6092105a4](http://www.huffingtonpost.com/entry/how-to-succeed-at-super-bowl-marketing_us_58937ebce4b0b4d6092105a4)
- [10] “Building for the next moment,” Google, May 2015. [Online]. Available: <https://adwords.googleblog.com/2015/05/building-for-next-moment.html>

- [11] “Super bowl ads 2017: Analyzing value and cost of top commercials,” Bleacher Report, February 2017. [Online]. Available: <http://bleacherreport.com/articles/2691090-super-bowl-ads-2017-analyzing-value-and-cost-of-top-commercials>
- [12] “The marketing land #hashtag bowl,” Marketing Land, February 2017. [Online]. Available: <http://marketingland.com/hashtag-bowl>
- [13] N. Borden, “The concept of the marketing mix,” *Journal of Advertising Research*, 1964. [Online]. Available: [http://www.guillaumenicaise.com/wp-content/uploads/2013/10/Borden-1984\\_The-concept-of-marketing-mix.pdf](http://www.guillaumenicaise.com/wp-content/uploads/2013/10/Borden-1984_The-concept-of-marketing-mix.pdf)
- [14] A. McKenzie, “Measuring brand lift,” = <https://www.viget.com/articles/measuring-brand-lift>, March 2014.
- [15] “Quarterly retail e-commerce sales,” U.S. Census Bureau News, February 2017. [Online]. Available: [https://www.census.gov/retail/mrts/www/data/pdf/ec\\_current.pdf](https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf)
- [16] G. M. T. Jr., “Building the buzz in the hive mind,” *Journal of Consumer Behavior*, 2004. [Online]. Available: <http://onlinelibrary.wiley.com/doi/10.1002/cb.158/abstract>
- [17] P. Danaher and R. Rust, “Determining the optimal return on investment for an advertising campaign,” *European Journal of Operational Research*, vol. 95, 1996.
- [18] D. Sontag and U. Shalit, “Icml 2016 tutorial: Causal inference for observational studies,” = <http://www.cs.nyu.edu/~shalit/tutorial.html>, 2016.
- [19] P. Holland, “Statistics and causal inference,” *Journal of the American Statistical Association*, December 1986.
- [20] K. Larsen, “Making causal impact analysis easy,” = <http://multithreaded.stitchfix.com/blog/2016/01/13/market-watch/>, January 2016.
- [21] Statista, “Statistics and facts about google,” = <https://www.statista.com/topics/1001/google/>, January 2016.
- [22] ——, “Statistics and facts about twitter,” = <https://www.statista.com/topics/737/twitter/>, January 2016.
- [23] S. Asur and B. Huberman, “Predicting the future with social media,” *IEEE*, August 2010.

- [24] R. Lewis and J. Rao, “On the near impossibility of measuring the returns to advertising,” *Unpublished paper, Google Inc. and Microsoft Research*, 2013. [Online]. Available: [http://justinmrao.com/lewis\\_rao\\_nearimpossibility.pdf](http://justinmrao.com/lewis_rao_nearimpossibility.pdf)
- [25] “Causalimpact 1.1.0, brodersen et al., annals of applied statistics (2015).” [Online]. Available: <http://google.github.io/CausalImpact/>
- [26] “Avocados from mexico kicks off nutritional campaign with super bowl spot,” February 2017. [Online]. Available: <http://adage.com/article/super-bowl/avocados-mexico-s-super-bowl-spot/307791/>
- [27] J. Hoeting, “Bayesian model averaging: a tutorial,” *Statistical Science*, vol. 14, 1999. [Online]. Available: <http://projecteuclid.org/euclid.ss/1009212519>
- [28] S. Scott and H. Varian, “Predicting the present with bayesian structural time series,” *International Journal of Mathematical Modeling and Numerical Optimization*, June 2013. [Online]. Available: [http://people.ischool.berkeley.edu/\\_hal/Papers/2013/pred-present-with-bsts.pdf](http://people.ischool.berkeley.edu/_hal/Papers/2013/pred-present-with-bsts.pdf)
- [29] “Feedback control systems,” = [https://ocw.mit.edu/courses/aeronautics-and-astronautics/16-30-feedback-control-systems-fall-2010/lecture-notes/MIT16\\_30F10\\_lec05.pdf](https://ocw.mit.edu/courses/aeronautics-and-astronautics/16-30-feedback-control-systems-fall-2010/lecture-notes/MIT16_30F10_lec05.pdf), September 2010.
- [30] Imbens and Wooldridge, “Difference-in-differences estimation,” = [http://www.nber.org/WNE/lect\\_10\\_diffindiffs.pdf](http://www.nber.org/WNE/lect_10_diffindiffs.pdf), June 2007.
- [31] S. Stephens-Davidowitz, H. Varian, and M. Smith, “Super returns to super bowl ads?” *Unpublished paper*, December 2016. [Online]. Available: [http://people.ischool.berkeley.edu/\\_hal/Papers/2015/super.pdf](http://people.ischool.berkeley.edu/_hal/Papers/2015/super.pdf)
- [32] J. Mooij, J. Peters, D. Janzing, J. Zscheischler, and B. Scholkopf, “Distinguishing cause from effect using observational data: Methods and benchmarks,” *Journal of Machine Learning Research*, vol. 17, 2016.
- [33] L. Bottou and J. Peters, “Counterfactual reasoning and learning systems: The example of computational advertising,” *Journal of Machine Learning Research*, vol. 14, 2013.

## A Code Sample

```
## DIFFERENCE IN DIFFERENCES ANALYSIS ON GOOGLE DATA (EXAMPLE: 84 LUMBER)
# Create dataset
setwd("C:/Users/Aggarwal Neema/Documents/Thesis/Data")
AllData = read.csv("2017DDdataLumber.csv", header = TRUE)
AllData$Brand = as.character(AllData$Brand)
AllData$value = as.numeric(as.character(AllData$value))

# Create Date-Time column
AllData$date = as.POSIXct(paste(AllData$date, AllData$time), format = "%m/%d/%Y %H:%M:%S", tz = "America/Los_Angeles")

# Set controls to 0, and pre-event window to zero, then run analysis
AllData$timenew = ifelse(AllData$date >= "2017-02-05 05:00:00 PST", 1, 0)
AllData$treated = ifelse(AllData$Brand == "Lumber", 1, 0)

did = lm(value ~ treated + timenew + timenew*treated, data = AllData)
summary(did)

## BSTS ANALYSIS ON TWITTER DATA (EXAMPLE: 84 LUMBER)
library(CausalImpact)
rm(list=ls())

# Create dataset for Twitter 7 DAY DATA
setwd("C:/Users/Aggarwal Neema/Documents/Thesis/Data/7 day")
AllDataT = read.csv("Combined_twitter.csv", header = TRUE)

datetimeT = as.POSIXct(paste(AllDataT$date, AllDataT$time), format = "%m/%d/%Y %H:%M:%S",
tz = "America/Los_Angeles")

# Uneffected Control markets
AlmondJoy = as.numeric(as.character(AllDataT$AlmondJoy))
Corona = as.numeric(as.character(AllDataT$Corona))
Pringles = as.numeric(as.character(AllDataT$Pringles))
Twix = as.numeric(as.character(AllDataT$Twix))
HomeDepot = as.numeric(as.character(AllDataT$HomeDepot))

# Effected market
Lumber = as.numeric(as.character(AllDataT$Lumber))

# Combine into set
Lumberset = zoo(cbind(Lumber, AlmondJoy, Corona, Pringles, Twix, HomeDepot), datetimeT)

# Determine pre and post event time frames varying post period time
pre.period <- as.POSIXct(c("2017-02-01 00:00:00 PST", "2017-02-05 05:00:00 PST"))

# Immediately after game ends
post.period <- as.POSIXct(c("2017-02-05 06:00:00 PST", "2017-02-05 20:00:00 PST"))
# ~ 1 day after
post.period2 <- as.POSIXct(c("2017-02-05 06:00:00 PST", "2017-02-07 00:00:00 PST"))
```

```

# ~ 2 days after
post.periodTWO <- as.POSIXct(c("2017-02-05 06:00:00 PST", "2017-02-08 00:00:00 PST"))
# ~3 days after (end of "Super Bowl" buzz)
post.period3 <- as.POSIXct(c("2017-02-05 06:00:00 PST", "2017-02-09 00:00:00 PST"))
# ~4 days after (end of "Super Bowl" buzz)
post.period4 <- as.POSIXct(c("2017-02-05 06:00:00 PST", "2017-02-10 00:00:00 PST"))
# ~5 days after (end of "Super Bowl" buzz)
post.period5 <- as.POSIXct(c("2017-02-05 06:00:00 PST", "2017-02-11 00:00:00 PST"))
# ~6 days after (end of "Super Bowl" buzz)
post.period6 <- as.POSIXct(c("2017-02-05 06:00:00 PST", "2017-02-12 00:00:00 PST"))
# All data
post.period7 <- as.POSIXct(c("2017-02-05 06:00:00 PST", "2017-02-13 21:00:00 PST"))

# Run CASAUL IMPACT analysis and see results
impactGame = CausallImpact(Lumberset, pre.period, post.period, model.args =
list(nseasons=7,season.duration=24))
impactGameWeekend = CausallImpact(Lumberset, pre.period, post.period2, model.args =
list(nseasons=7,season.duration=24))
impactGame2 = CausallImpact(Lumberset, pre.period, post.periodTWO, model.args =
list(nseasons=7,season.duration=24))
impactGame3 = CausallImpact(Lumberset, pre.period, post.period3, model.args =
list(nseasons=7,season.duration=24))
impactGame4 = CausallImpact(Lumberset, pre.period, post.period4, model.args =
list(nseasons=7,season.duration=24))
impactGame5 = CausallImpact(Lumberset, pre.period, post.period5, model.args =
list(nseasons=7,season.duration=24))
impactGame6 = CausallImpact(Lumberset, pre.period, post.period6, model.args =
list(nseasons=7,season.duration=24))
impactGame7 = CausallImpact(Lumberset, pre.period, post.period7, model.args =
list(nseasons=7,season.duration=24))

summary(impactGame)
summary(impactGameWeekend)
summary(impactGame2)
summary(impactGame3)
summary(impactGame4)
summary(impactGame5)
summary(impactGame6)
summary(impactGame7)

# Plot outputs
matplot(AllControls, type = "l")
plot(impactGame7)
plot(impactGame7$model$bsts.model, "coefficients")

## VARY NUMBER OF CONTROLS WITH GOOGLE DATA (EXAMPLE: SKITTLES)
# Google 7 Day Data - test 1 control vs. 5 vs. all 11
library(CausallImpact)
rm(list=ls())

# Create dataset for GOOGLE 7 DAY DATA
setwd("C:/Users/Aggarwal Neema/Documents/Thesis/Data/7 day")

```

```

AllDataG = read.csv("Combined 7 day v2.csv", header = TRUE)

datetimeG = as.POSIXct(paste(AllDataG$Date,AllDataG$Time), format = "%m/%d/%Y
%H:%M:%S", tz = "America/Los_Angeles")

# Uneffected Control markets
AlmondJoyG7day = as.numeric(as.character(AllDataG$AlmondJoy))
BlueMoonG7day = as.numeric(as.character(AllDataG$BlueMoon))
BurgerKingG7day = as.numeric(as.character(AllDataG$BurgerKing))
CoronaG7day = as.numeric(as.character(AllDataG$Corona))
DasaniG7day = as.numeric(as.character(AllDataG$Dasani))
MnMsG7day = as.numeric(as.character(AllDataG$MnMs))
MountainDewG7day = as.numeric(as.character(AllDataG$MountainDew))
PringlesG7day = as.numeric(as.character(AllDataG$Pringles))
ShockTopG7day = as.numeric(as.character(AllDataG$ShockTop))
SmartwaterG7day = as.numeric(as.character(AllDataG$Smartwater))
TwixG7day = as.numeric(as.character(AllDataG$Twix))

# Effected market
SkittlesG7day = as.numeric(as.character(AllDataG$Skittles))

# Combine into control set of various sizes
OneControl = zoo(cbind(SkittlesG7day,CoronaG7day), datetimeG)
FiveControls =
zoo(cbind(SkittlesG7day,CoronaG7day,BlueMoonG7day,PringlesG7day,TwixG7day,MountainDew
G7day),datetimeG)
AllControls =
zoo(cbind(SkittlesG7day,AlmondJoyG7day,BlueMoonG7day,BurgerKingG7day,CoronaG7day,Dasa
niG7day,MnMsG7day,MountainDewG7day,PringlesG7day,ShockTopG7day,SmartwaterG7day,Tw
ixG7day), datetimeG)

# Determine pre and post event time frames
pre.period <- as.POSIXct(c("2017-01-25 21:00:00 PST", "2017-02-05 05:00:00 PST"))
post.period <- as.POSIXct(c("2017-02-05 06:00:00 PST", "2017-02-07 00:00:00 PST"))

# Run CASAUL IMPACT analysis and see results
impactOneControl = CausalImpact(OneControl[1:292], pre.period, post.period, model.args =
list(nseasons=7,season.duration=24))
summary(impactOneControl)

impactFiveControls = CausalImpact(FiveControls[1:292], pre.period, post.period, model.args =
list(nseasons=7,season.duration=24))
summary(impactFiveControls)

impactAllControls = CausalImpact(AllControls[1:292], pre.period, post.period, model.args =
list(nseasons=7,season.duration=24))
summary(impactAllControls)

# Plot outputs
matplotlib(OneControl[1:292], type = "l")
plot(impactOneControl)
plot(impactOneControl$model$bsts.model, "coefficients")
plot(impactFiveControls$model$bsts.model, "coefficients")

```

```

plot(impactAllControls$model$bsts.model, "coefficients")

## SAMPLE GOOGLE DATA SCRAPE CODE
# Download 7 day data - HOURLY (168 values)
rm(list=ls())
downloadDir="C:/Users/Aggarwal Neema/Documents/Thesis/2017_data_scrape"
setwd(downloadDir)

URL_GT=function(keyword=""){

  start="http://www.google.com/trends/trendsReport?hl=en-US&q="
  end="&date=now%207-d&cmpt=q&content=1&export=1"

  queries=keyword[1]
  if(length(keyword)>1) {
    for(i in 2:length(keyword)){
      queries=paste(queries, "%2C ", keyword[i], sep="")
    }
  }

  URL=paste(start, queries, end, sep="")
  URL<- gsub(" ", "%20", URL)
  return(URL)
}

downloadGT=function(URL, downloadDir){

  #Determine if download has been completed by comparing the number of files in the download directory to the starting number
  startingFiles=list.files(downloadDir)
  browseURL(URL)
  endingFiles=list.files(downloadDir)

  while(length(setdiff(endingFiles,startingFiles))==0) {
    Sys.sleep(3)
    endingFiles=list.files(downloadDir)
  }
  filePath=setdiff(endingFiles,startingFiles)
  return(filePath)
}

keywords=c("superbowl")
url = URL_GT(keywords)
filePath=downloadGT(url,downloadDir)

```

## B Data Sample

Date	Time	AlmondJoy	AvocadosFM	Bai	BlueMoon	BudLight
1/25/2017	21:00:00	28	10	61	27	20
1/25/2017	22:00:00	24	6	62	21	11
1/25/2017	23:00:00	9	5	57	19	8
1/26/2017	0:00:00	7	3	52	17	5
1/26/2017	1:00:00	3	2	47	16	4
1/26/2017	2:00:00	1	1	41	17	4
1/26/2017	3:00:00	9	1	41	16	5
1/26/2017	4:00:00	15	4	46	19	7
1/26/2017	5:00:00	19	5	49	21	10
1/26/2017	6:00:00	22	12	45	24	11
1/26/2017	7:00:00	28	8	34	26	12
1/26/2017	8:00:00	38	9	23	28	16
1/26/2017	9:00:00	27	19	15	33	19
1/26/2017	10:00:00	38	13	11	32	22
1/26/2017	11:00:00	43	15	9	32	20
1/26/2017	12:00:00	63	64	8	35	24
1/26/2017	13:00:00	52	100	9	36	27
1/26/2017	14:00:00	67	86	14	43	34
1/26/2017	15:00:00	67	65	26	47	42
1/26/2017	16:00:00	73	64	40	49	44
1/26/2017	17:00:00	71	45	52	50	48
1/26/2017	18:00:00	86	46	65	52	45
1/26/2017	19:00:00	53	46	73	46	41
1/26/2017	20:00:00	38	46	75	37	30
1/26/2017	21:00:00	25	28	69	32	24
1/26/2017	22:00:00	16	13	65	26	14
1/26/2017	23:00:00	11	13	63	17	9
1/27/2017	0:00:00	9	1	59	17	5
1/27/2017	1:00:00	5	6	55	17	5
1/27/2017	2:00:00	14	2	51	16	4
1/27/2017	3:00:00	13	8	51	18	5
1/27/2017	4:00:00	16	9	53	18	7
1/27/2017	5:00:00	24	19	52	21	8
1/27/2017	6:00:00	36	19	55	25	13
1/27/2017	7:00:00	36	23	65	29	15
1/27/2017	8:00:00	27	25	100	28	17
1/27/2017	9:00:00	29	26	47	33	23
1/27/2017	10:00:00	35	22	20	33	21