Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 162 (2019) 275–280

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Information Technology and Quantitative Management (ITQM 2019)

Difference in Differences in Marketing Performance Measurement

Wei Deng[a]*, Nan Xu*, Haotian Li[a], Fangyao Liu[a], Yong Shi[a,c]

[a]*University of Nebraska at Omaha, Omaha, NE, 68182*

[c]*Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China*

Abstract

Marketing performance measurement is important in retail companies, it is critical for the budget planning and adjustment of marketing strategies and tactics over time. Compared to digital online marketing, such as social media advertisers where tagging techniques [1] can be applied to track the lifecycle of the consumption and the customer's behaviours, traditional TV and radio advertising will be harder to evaluate at the individual level. Instead, common approach in this scenario is to study the relationship between ads spend and sales. This paper presents a casual inference methodology – "Difference in Differences"(DID) on the application of TV marketing measurement. Instead of building direct relationship between ads and sales, we attempt to create an experimental research design and study the differential effect on a 'treatment group' versus a 'control group'. In order to study the performance of advertising in a city, the treatment group is sales performance during and after TV advertisings and the control group is the statistical model that mimics the sales of the same city without ads influence. It is proved that the model can significantly predict the sales of city in nature by considering its past sales, the neighbourhood districts and weather conditions. And a case study in industrial data shows that DID methodology can effectively evaluate the TV advertising results on regions.

Keywords: *Advertising Measurement, Difference in Differences, Random Forest, Lasso Regression*

* Corresponding author. Tel.: +0-402-618-0863
E-mail address: wdeng@unomaha.edu, nanxu@umn.edu.

1877-0509
10.1016/j.procs.2019.11.285

1. Introduction

Marketing performance measurement is important in the retail industry, it is critical for the budget planning and adjustment of marketing strategies and tactics over time. Businesses have increasingly moved towards data driven decision making and started using incrementality measurement to help the business to make smart business decisions.

Major advertising companies such as Google, Facebook, and Adobe provide tools and services to measure incremental effect on sales and conversions of ads. Facebook provides a service to measure conversion lift to help advertisers to drive incremental sales and conversions. The measure of incremental effect of ads allows advertisers to gain insight into the causal impact of ads on true business value. The methodology typically used is the randomized experiment such as an A/B test. Users to visit the website are randomly split into a treatment group and a control group. Ads are delivered to the treatment group only. Then we can compare the conversions between treatment group and control group to calculate lift. A completely randomized experiment is the ideal way to measure incrementality because randomization makes the assignment of treatment be independent from the outcome that we try to measure. However, randomized experiment could be costly or not impossible to implement. Then we need to rely on observational data as in our case.

In this study, we evaluate the incremental impact of a TV marketing campaign for a retail company in the Midwest of the U.S. The company launched a TV advertising campaign in a city for a period of time and we want to know how many extra sales have been generated due to the expenditure on advertising. Traditionally, people use sales of the same week from previous year to measure the impact. However, this impact is not true incremental impact. We combine machine learning and difference-in-differences methodology in this study to identify the impact of ads.

Specifically, we use difference in differences (DID) as described in [2] to estimate the ATE and identify the causal impact of TV advertising on sales. we compare the change of sales before and after the advertising campaign between treatment group and control group. To construct a control group, we estimate the counterfactual sales for the treatment group. We leverage machine learning methods to predict the pre-treatment sales of control group. Machine learning provides a data driven approach to find the best control group as our baseline to compare with treatment group. Then we apply DID to identify the causal impact of TV ads on sales. We find that the incremental impact of the TV campaign to drive about 21k sales after four weeks of the end of the TV campaign.

The rest of the paper is organized as follows. Section 2 introduces the general methodology of difference in differences and machine learning modeling. Section 3 shows the results from our experiments. Section 4 concludes the paper.

2. Methods

2.1. Difference in Differences (DID)

The simplest approach to study the relationship between sales and ads is a linear regression, $y = bx + e$ where y denotes per capita sales and x denotes per capita ads. However, this formula ignores confounding variables that are important variables missed in the modeling [2]. The examples of confounding variables in this case include the variances of weather conditions, customers' preferences, seasonal trends etc. Instead, experiment is the best way to answer questions of causal inference [3]. In order to conduct an experiment, a treatment should be applied to selected subjects and the outcomes from the treatment will be compared with outcomes of untreated other subjects. The effect of casual inference can be derived by the formula:

$$\textit{Outcome from treated subjects} - \textit{Outcome from untreated subjects} \quad (1)$$

In the case of sales and ads, based on the formula 1, we can derive the marginal gain of ads spend of a city by comparing the sales outcome after advertising and the sales without advertising. However, since it is impossible to track into individual level for broadcasting medias, the challenge here is how to construct a control group if treatment applied to tested city. “Difference in Differences” provides feasible solution. Difference in differences [4, 5] is a classic statistical method that addresses the question of casual inference, and to study differential effect of a 'treatment group' versus a 'control group' by an experimental research design. DID applies treatment on the treatment group, and leave control group grows in nature, then two average changes of both groups are compared over time. The following image shows the relationship between treatment and control group:

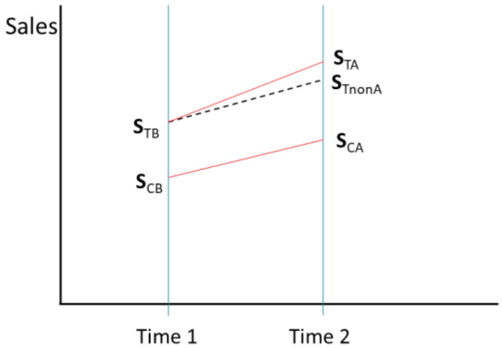

Fig. 1. the Difference in Differences on the sales

$\mathbf{S}_{TB}$ = sales before ads campaign for treated groups
$\mathbf{S}_{TA}$ = sales after ads campaign for treated groups
$\mathbf{S}_{TnonA}$ = sales without ads campaign for treated groups
$\mathbf{S}_{CB}$ = sales before ads campaign for control groups
$\mathbf{S}_{CA}$ = sales after ads campaign for control groups
effect of treatment on treated= (STA − STB) – (SCA − SCB)

The requirement of control group is randomly selected from population as the same as treatment group. However, the limit of broadcasting advertising does not allow to select part of customers and block the other part precisely. Instead, we use the sales in neighboring cities around advertising target to mimic as control group through statistical modeling.

2.2. Statistical modeling for control group

The modeling of this study is to create a control group which mirrors the sales of treatment group before the treatment is applied. Therefore, the model can be used to mimic the average growth of the period during the treatment. The better the model matches with the previous pattern, the more confident we can use it as control group to compare the effect of treatment. There is no consensus on which statistical modeling approaches are the best in a given context. In this study, the authors decided to use two methods: random forest and stepwise lasso regression. Random forest is an ensemble bagging algorithm [6] where Bootstrap sampling applies to select part of attributes and rows of record, and then the traditional decision tree method [7] will be used to build weak model. An ensemble decision is made by the results from a series of subtrees, Collective decision-making guarantees not only the accuracy but also the stability of prediction. The model is created by using the library

RandomForestClassifier in sciki-learn [8], 100 subtrees were built for the purpose of ensemble. The other algorithm we choose is lasso regression. Lasso regression [9] is a linear regression with L1 regulation that is absolute value of magnitude of coefficient as a penalty item to the loss function. The key difference to ordinary linear regression is that lasso shrinks the weights of less important features by decrease the coefficient values of these attributes. This works well for feature selection and fits to this study case since the variances of the sales of some cities in control group are severe and better minimize the impact of these instable factors in prediction. The Lasso in sciki-learn[8] is the library for lasso regression in this study.

3. Experiment results

The data of the experiment is from a retail company, it is a study for the proof of concept and experiment tested on the customers from a small city. The traditional way to study the performance of the advertising is to compare the same week of last year or pervious weeks. However, two methods have problems.

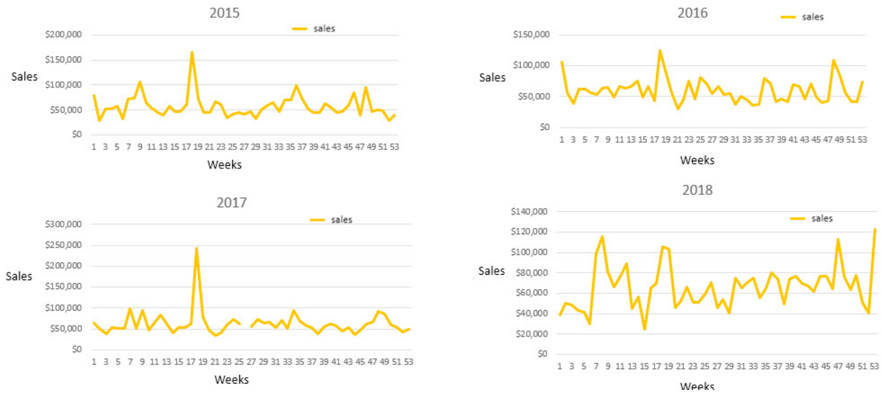

Fig. 2. the weekly sales of a city

The figure 2 shows the sales of a city in the last four years. Even though no strong seasonal trend, the variance of weekly sales is obvious, partly is caused by the conditions of weather and promotions. It indicates the infeasibility to take performance of previous weeks as reference. The other problem is that a confound conclusion likely happens due to the carryover effect. Carryover effect [10] is the portion of marketing eff ect that occurs in time periods following the pulse of advertising. In the case study, we choose 2 and 4 weeks as the length of carryover effect, the conclusion is different by comparing with the sales of last year in the same week.

Ads Carryover Effect		2 weeks
Year	After Ads	comparison after Ads
2015	$90,033.00	
2016	$85,988.00	-4,045.00
2017	$81,631.00	-4,357.00
2018	$98,679.00	17,048.00
2019	$82,458.00	-16,221.00

Ads Carryover Effect		4 weeks
Year	After Ads	comparison after Ads
2015	$217,610.00	
2016	$206,124.00	-$11,486.00
2017	$182,555.00	-$23,569.00
2018	$215,450.00	$32,895.00
2019	$234,033.00	$18,583.00

Table 1. The comparison of sales from last year

The above problems can be solved by DID method through the control group. The sales of close cities nearby target and the sales of previous year at target city are used as independent variables and the sales of target city as dependent variable. The sales from 2015 to 2018 is used to train the model. The test has been applied to the sales of 2019 before ads, which was the sales history of 8 weeks. The result is shown as following:

	Coefficient rate	R square	MAE
Lasso Regression on training	0.76	0.68	14,521
Random Forest on training	0.94	0.88	4,570
Lasso Regression on testing	0.86	0.7	15,210
Random Forest on testing	0.65	0.69	16,120

Table 2. Control group modelling performance

Ads Carryover Effect			4 weeks
Year	**Mapping**	**Target City**	**Difference**
2018	$237,384	$215,450	
2019	$234,212	$234,033	
DID	-$3,172	$18,583	$21,755

Table 3. DID value of an ads on 2019

Table 2 shows that both lasso regression and random forest have positive effects on sales prediction. Even though the random forest uses more complicated structure to mimic sales, and much better performance if the model itself is applied to training dataset, but the performance on test dataset will be a little bit worse than lasso regression. The simpler model has consistent result and better solution to the problem of overfit. The table 3 represents the application model to the target city with 4 weeks as carryover effect. The result gives more credit to ads effect, which meets our expectation by considering the relative decreasing sales of nearby cities.

4. Conclusion

In this paper, we present DID method on the application of TV marketing measurement and propose a novel approach to create control group by using statistical modeling. The result shows that combination of both the sales of neighbor cities and that of same time last year will provide strong signal on the sales prediction of target city. The feasibility of sales prediction shed light on DID method where control group benefits the evaluation of extract gain from ads. In order to continue future research, the other factors may be included such as weather condition, and how to improve the accuracy of modeling as well as overfit prevention are our interests.

References

[1] Blumenau, Trevor. "Tracking of internet advertisements using banner tags." U.S. Patent No. 6,434,614. 13 Aug. 2002.
[2] Varian, Hal R. "Causal inference in economics and marketing." Proceedings of the National Academy of Sciences 113.27 (2016): 7310-7315.
[3] BoxGEP, HunterJS,Hunter WG(2005) Statistics for Experimenters(Wiley-Interscience, New York).
[4] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How much should we trust differences-in-differences estimates?." The Quarterly journal of economics 119.1 (2004): 249-275.
[5] Abadie, Alberto. "Semiparametric difference-in-differences estimators." The Review of Economic Studies 72.1 (2005): 1-19.
[6] Liaw, Andy, and Matthew Wiener. "Classification and regression by randomForest." R news 2.3 (2002): 18-22.
[7] Breiman, Leo. Classification and regression trees. Routledge, 2017.
[8] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of machine learning research 12.Oct (2011): 2825-2830.
[9] Tibshirani, Robert. "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society: Series B (Methodological) 58.1 (1996): 267-288.
[10] Stone, Walter J. "The carryover effect in presidential elections." American Political Science Review 80.1 (1986): 271-279.