Paper 0778-2017

Differentiate Effects from the Noise of Promotional Marketing Campaigns

Alex Glushkovsky and Matthew Fabian, BMO Financial Group

ABSTRACT

In highly competitive markets, the response rates to economically reasonable marketing campaigns are as low as a few percentage points or less. In that case, the direct measure of the delta between the average key performance indicators (KPIs) of the treated and control groups is heavily "contaminated" by non-responders. This paper focuses on measuring promotional marketing campaigns with two properties: (1) price discounts or other benefits, which are changing profitability of the targeted group for at least the promotion periods, and (2) impact of self-responders. The paper addresses the decomposition of the KPI measurement between responders and non-responders for both groups. Assuming that customers who rejected promotional offers will not change their behavior and that non-responders of both treated and control groups are not biased, the delta of the average KPIs for non-responders should be equal to zero. In practice, this component might significantly deviate from zero. It might be caused by an initial nonzero delta of KPI values despite a random split between groups or by existence of outliers, especially for non-balanced campaigns. In order to address the deviation of the delta from zero, it might require running additional statistical tests comparing not just the means but the distributions of KPIs as well. The decomposition of the measurement between responders and non-responders for both groups can then be used in differential modeling.

INTRODUCTION

Businesses run marketing campaigns in order to boost microeconomic results and logically face two questions on a backend:

- What are the results that have been achieved by the campaign?
- What should be done to run the campaign more effectively in the future?

Today, many publications can be found discussing marketing campaign measurement (Breur, 2007).

The topic of measuring campaign performance has been presented in previous SAS® Forums. Thus, practical methods of using gain and lift charts are presented in (Jaffery and Liu, 2009; Zhao, 2012; Dixon, 2015).

Running campaigns in the highly competitive market environment requires reservation of a control group in order to correctly estimate obtained effects. It is necessary for two reasons:

1. Oserved effects are usually not strong enough to be clearly attributed to the treatment of the campaign.
2. Macroeconomic and market environments are dynamic affecting the results of a campaign.

Fisher`s random split between treatment and control groups becomes a mandatory procedure if there are no operational or legal restrictions (Speed, 1992). However, even applying such methodology, and observing a small response rate, it may require some additional effort to estimate microeconomic effects given a vast majority of non-responders.

In highly competitive markets, the response rates to economically reasonable marketing campaigns are as low as a couple percent or even less. In that case, the direct measure of the effect as a delta between average KPIs of the treatment and control groups is heavily contaminated by non-responders.

This paper focuses specifically on promotional marketing campaigns with two properties:

- Treatment includes price discounts or other benefits, which are changing profitability of the targeted group for at least the promotion periods.
- There are self-responders in the control or treatment group before and after the promotional periods. This means that just focusing on response rate differentiation may not be used as an ultimate measure.

PROMOTIONAL CAMPAIGN RESPONSE FLOWS

Considering the second property of the promotional marketing campaigns, it means that there are five distinct response flows:

- Promotional responders in treatment group (received offer/treatment and responded as well as fulfill conditions of offer).
- Non-promotional responders in treatment group (received offer/treatment and responded but not per the offer).
- Non-responders in treatment group (received offer/treatment and did not respond at all).
- Non-promotional responders in control group (did not receive offer/treatment but acted similar to a response).
- Non-responders in control group (no offer/treatment and no response).

Illustration of promotional campaign response flows is presented in Figure 1.

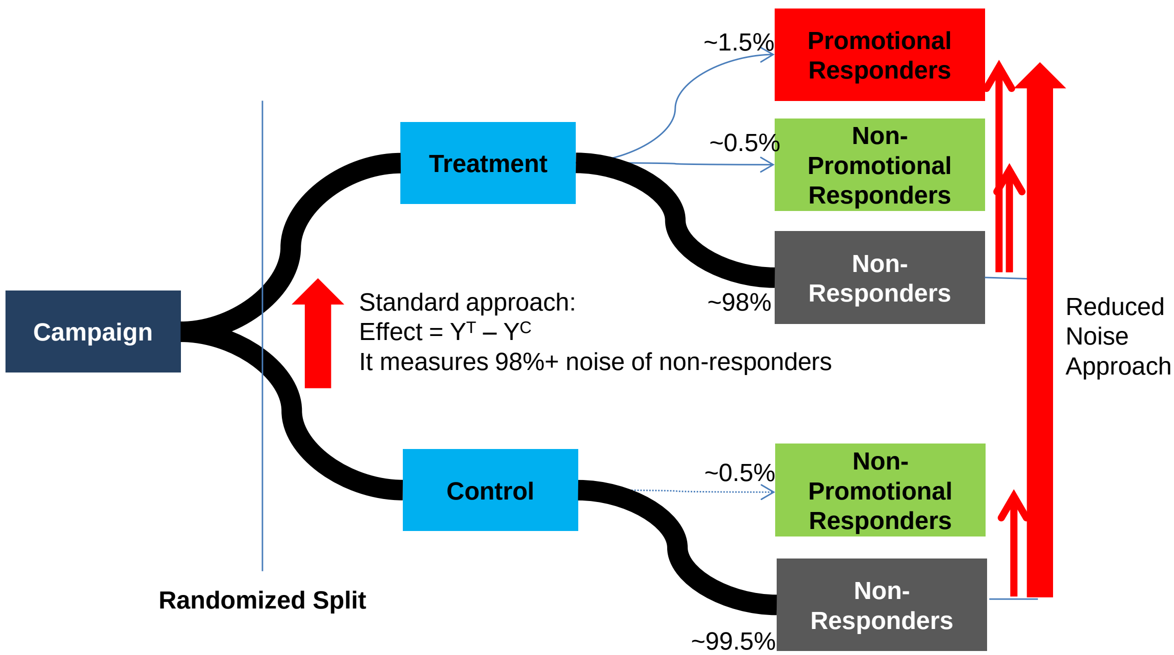

Figure 1. Illustrative Example of Promotional Campaign Response Flows

The thick flow lines in Figure 1 represent non-responders, which strongly dominate a standard measurement of a campaign effect comparing the results of two groups.

MEASUREMENT OF PROMOTIONAL CAMPAIGN EFFECTS

Having two groups of customers, the standard measurement approach, estimating an average microeconomic treatment effect, is a simple difference between average KPIs per customer of Treatment (T) and Control (C) groups:

$$\overline{TE}(t) = \overline{Y}^T(t) - \overline{Y}^C(t) \tag{1}$$

Where $\overline{Y}(t)$ is an average value of KPI of interest (such as, profit, revenue, balance, loss, etc.) per customer at time *t* after a promotional offer start date. Excepting measures that represent snapshot results, this value is usually a cumulative sum of monthly (weekly or daily) KPI values $y(\tau)$ covering monitoring period from the launch date (0) to the current one (t).

$$\overline{Y}(t) = \sum_{\tau=1}^{t} \overline{y}(\tau)$$

There are three issues with the Equation (1):

- The treatment effect (1) is a “gross” measure of the treatment effect and heavily contaminated by non-responders, which may be almost 100% of the targeted population in some cases. The stress testing of the formula reveals that even having no responders, it estimates some effects. These effects may be even statistically significant ones. In that case, the possible causes for some deviations from zero are imperfect randomization splitting between treatment and control groups, outliers, insufficient or unbalanced sample sizes of treatment and control groups, or results of quite exotic campaigns where the promotional offering itself has more impact than the offer acceptance.
- In practice, the treatment effect (1) usually does not equal zero just prior to a campaign starting point (i.e., at time prior to zero). The possible causes for some deviations from zero are the same as stated above except the latest one.
- The missing counterfactual means that $Y^T(t)$ and $Y^C(t)$ cannot be observed simultaneously on a customer level. Therefore, Equation (1) is applicable only on a segment level.

To mitigate a potential randomization bias, various approaches to randomization can be applied in addition to a simple one: blocked, stratified, or covariate adaptive randomization (Suresh, 2011).

As a post randomization adjustment, it is sensible to apply a simple incremental measure:

$$\overline{ITE}(t) = \overline{IY}^T(t) - \overline{IY}^C(t) \tag{2}$$

Where $\overline{IY}(t)$ is an average value of incremental over time KPI per customer for both treatment and control groups:

$$\overline{IY}(t) = \sum_{\tau=1}^{t} (\overline{y}(\tau) - \overline{y}(0\_))$$

The formula (2) ensures that at time just prior to zero there is no effect:

$$\overline{ITE}(0_-) = 0$$

When applying incremental measurement, it is still important to focus on $\overline{TE}(0_-)$ as a measure of an initial error. It provides an additional possibility to judge significance of the obtained results.

Also, if there is a significant difference between $\overline{Y}(0_-)$ of responders and non-responders, then that variable may be an important factor of the response model and it should be included as a potentially predictive variable in a propensity model.

The total incremental effect of the campaign is a trivial calculation:

$$TotITE(t) = N^T \cdot \overline{ITE}(t) \tag{3}$$

It is an estimation based on average incremental effects per customer multiplied by the number of treated customers N^T.

Having small sample sizes for the control group, reduces the robustness of the effect estimations $\overline{ITE}(t)$ dramatically, especially when observing rare events. Furthermore, multiplication of effects by a very big number of treated customers amplifies the noise. As a result, we can observe very high volatility of the metric (3) even on a campaign level.

To address the above mentioned challenges, the decomposition of the treatment effects by response flows is applied (Figure 1).

The average value of incremental KPI per treated customer can be simply presented as:

$$\overline{IY}^{T}(t) = R_{P}^{T}(t)\cdot\overline{IY}^{T}_{P}(t) + R_{NP}^{T}(t)\cdot\overline{IY}^{T}_{NP}(t) + (1 - R_{P}^{T}(t) - R_{NP}^{T}(t))\cdot\overline{IY}^{T}_{NR}(t)$$

Where *R* is the cumulative response rate, and subscripts *P*, *NP* and *NR* identify values for promotional response, non-promotional response, and no response, respectively.

It should be noted that there are rare cases when the same customer responded first to the promotional offer and subsequently became a self-triggered non-promotional responder as well. For simplicity, these cases were combined with promotional ones. However, they can be isolated as a special segment considering their frequency and amplitude of the effects.

Having self-triggered responders in the control group means that the promotion should compete against this flow. It requires logically decomposing control performance by responders and non-responders as well – same as for the treatment group:

$$\overline{IY}^{C}(t) = R_{NP}^{C}(t)\cdot\overline{IY}^{C}_{NP}(t) + (1 - R_{NP}^{C}(t))\cdot\overline{IY}^{C}_{NR}(t)$$

The Equation (2) can be transformed now to:

$$\begin{aligned}\overline{ITE}(t) = {} & R_{P}^{T}(t)\cdot\left(\overline{IY}^{T}_{P}(t) - \overline{IY}^{T}_{NR}(t)\right) + R_{NP}^{T}(t)\cdot\left(\overline{IY}^{T}_{NP}(t) - \overline{IY}^{T}_{NR}(t)\right) - \\ & R_{NP}^{C}(t)\cdot\left(\overline{IY}^{C}_{NP}(t) - \overline{IY}^{C}_{NR}(t)\right) + \\ & \overline{IY}^{T}_{NR}(t) - \overline{IY}^{C}_{NR}(t)\end{aligned}$$

Or:

$$\overline{ITE}(t) = \overline{ITY}_{R}(t) + \overline{ITY}_{NR}(t) \tag{4}$$

Where, Incremental Treatment Effect of responders equals:

$$\begin{aligned}\overline{ITY}_{R}(t) = {} & R_{P}^{T}(t)\cdot\left(\overline{IY}^{T}_{P}(t) - \overline{IY}^{T}_{NR}(t)\right) + R_{NP}^{T}(t)\cdot\left(\overline{IY}^{T}_{NP}(t) - \overline{IY}^{T}_{NR}(t)\right) - \\ & R_{NP}^{C}(t)\cdot\left(\overline{IY}^{C}_{NP}(t) - \overline{IY}^{C}_{NR}(t)\right)\end{aligned} \tag{5}$$

And, Incremental Treatment Effect of non-responders equals:

$$\overline{ITY}_{NR}(t) = \overline{IY}^{T}_{NR}(t) - \overline{IY}^{C}_{NR}(t) \tag{6}$$

Therefore, the reduced noise approach decomposes the total treatment effect into two parts of responders and non-responders (Equation (4)).

Assuming that customers who rejected promotional offers will not change their behaviour and that non-responders of both treatment and control groups are not biased, the latest should be equal to zero.

$$\overline{ITY}_{NR}(t) \to 0 \tag{7}$$

In this case, the measure of the Incremental Treatment Effect is equal to $\overline{ITY}_{R}(t)$ (Equation 5).

In practice, the component (6) may be significantly deviated from zero. Even small deviations from zero can dramatically change estimations of the total campaign effects. It may be caused by an initial bias despite a random split between groups, imperfect randomization, highly volatile customer performances, existence of outliers, insufficient sample sizes, or unbalanced splits between treatment and control groups. Thus, the treatment group usually has more outliers due to a larger sample size than the control group and it introduces significant noise to the standard effect measurements. In some campaigns, the number of outliers is comparable to the number of responders!

One of the arguments to apply the standard measure is that non-responders potentially represent biased populations comparing treatment and control groups and, therefore, the $\overline{ITY}_{NR}(t)$ should be part of the estimation. The concern about biased populations has some justification based on simple logic that if responders and non-responders have different performances prior to the treatment, then removing responders may potentially bias the population of non-responders even though they represent very small number of customers.

To clarify this concern, it is rational to perform a statistical test on hypothesis that $\overline{ITY}_{NR}(t) = 0$ or to run additional statistical tests, for example, such as Kolmogorov-Smirnov comparing not just the means but entire cumulative distributions of KPIs of both treatment and control groups (https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test).

This test should take into consideration the potential existence of outliers and unbalanced sample sizes of treatment and control groups (Cousineau and Chartier, 2010). Also, when performing a statistical test and considering the p-value, the focus should be on its context and purpose (Wasserstein, 2016).

Consider components of the Equation (5). There are two groups of variables: response rates *R(t)* and KPIs *IY(t)*. Illustration of dynamic curves of cumulative response rates after the campaign launch time are presented in Figure 2.

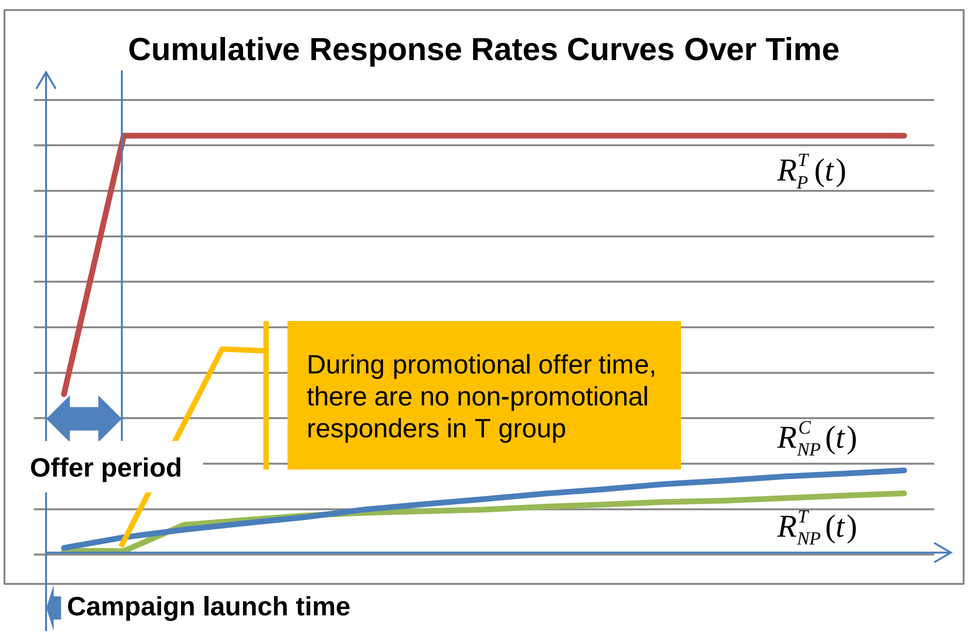

Figure 2. Illustration of Cumulative Response Rates Curves over Time

The cumulative curve of promotional responders of the treatment group has significantly elevated as a result of some discounts or other incentives. However, the curve has a plateau after expiration of the offer period. In contrast, curves of non-promotional responders are trending up even after that promotional period.

Figure 3 illustrates dynamic curves of cumulative KPIs of the Equation (5).

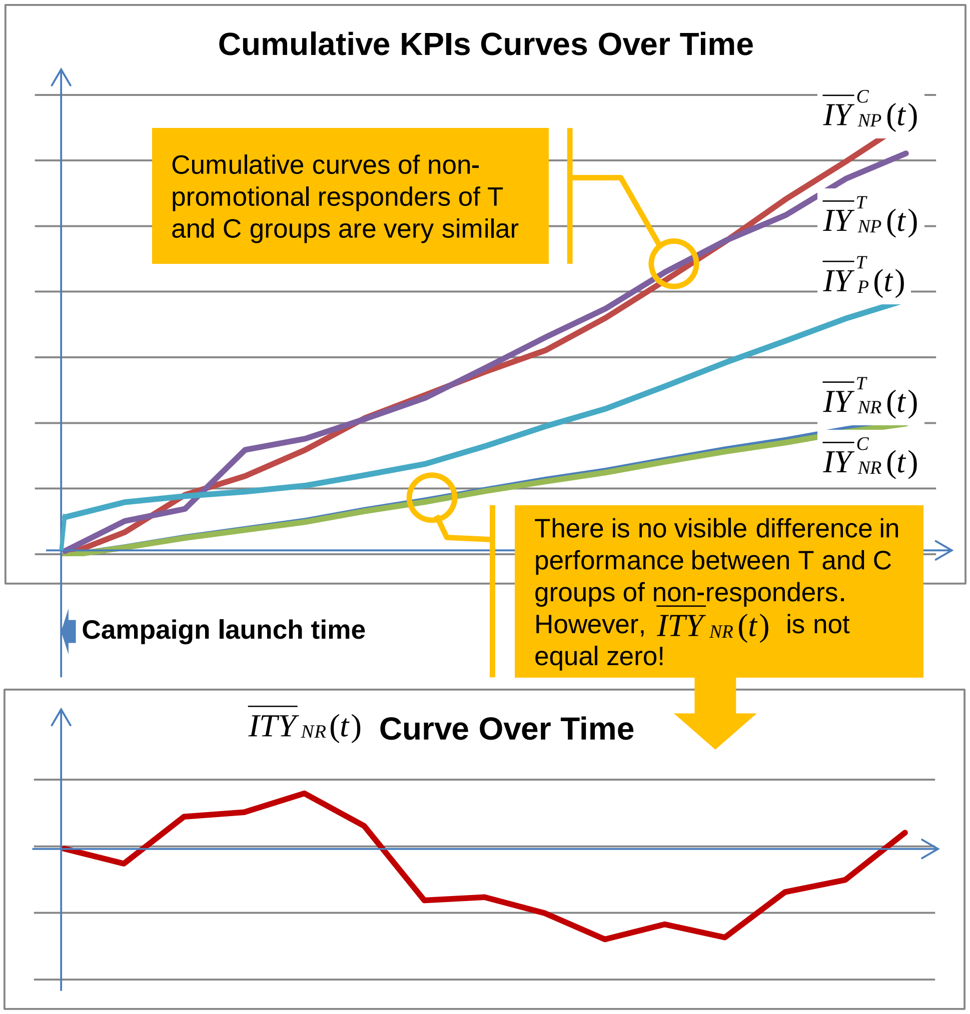

Figure 3. Illustration of Cumulative KPIs Curves over Time

Even though, there is an invisible difference between KPIs of non-responders of treatment and control groups, the $\overline{ITY}_{NR}(t)$ does not equal zero and it has a leveraged effect considering low response rates. The latest works as an amplifier. This effect is illustrated in Figure 4.

The example in Figure 4 illustrates, the cumulative incremental treatment effect of responders $\overline{ITY}_{R}(t)$ is less noisy and has a clearer pattern, while the total effect $\overline{ITY}(t)$, as expected, is highly correlated and significantly impacted by the effect of non-responders $\overline{ITY}_{NR}(t)$.

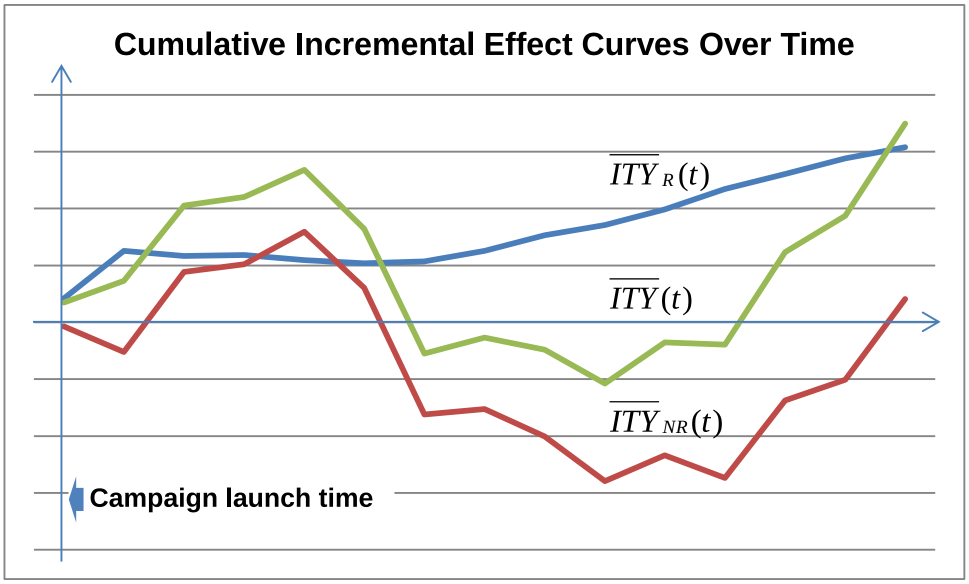

Figure 4. Illustration of Cumulative Incremental Effect Curves over Time

PROMOTIONAL CAMPAIGN UPLIFT MODELING

To answer the question: "What should be done to run the campaign more effectively in the future?", it is necessary to develop an uplift model (also known as a differential or net lift).

Uplift modeling is a powerful methodology identifying segments of customers with positive, neutral, and negative treatment effect results. It is done by comparing treatment and control results given missing counterfactual on a segment level (Radcliffe and Surry, 1999; Radcliffe, 2007; Rzepakowski and Jaroszewicz, 2011). Results of such models provide support of future campaign optimization.

The decomposition of the measurement between responders and non-responders for both groups can be used in such modeling focusing on effects of responders of Equation (5).

Considering KPIs curves of responders in Figure 3, we can reasonably assume that:

$$\overline{IY}^{T}_{NP}(t) \approx \overline{IY}^{C}_{NP}(t) = \overline{\overline{IY}}_{NP}(t) \tag{8}$$

Where the double bar means that estimation is based on an entire population of non-promotional responders regardless of their assigned groups.

If assumption (8) is correct, then we are getting a trivial case for $\overline{ITY}_R(t)$ that has only two components:

- lift of cumulative response rates of the treatment group against the control one, and
- incremental effect of a KPI comparing promotional responders versus non-responders.

I.e., Equation (5) can be approximated by dropping second-order small quantities as:

$$\overline{ITY}_R(t) \approx \left(R^T_P(t) + R^T_{NP}(t) - R^C_{NP}(t)\right) \cdot \left(\overline{IY}^{T}_{P}(t) - \overline{\overline{IY}}_{NR}(t)\right) = \Delta R(t) \cdot \overline{\Delta ITE}(t) \tag{9}$$

Where $\Delta R(t) = R^T_P(t) + R^T_{NP}(t) - R^C_{NP}(t)$

$$\overline{\Delta ITE}(t) \approx \overline{IY}^{T}_{P}(t) - \overline{\overline{IY}}_{NR}(t)$$

This approximation means that uplift modeling, such as decision tree, can be applied just against the cumulative incremental effects of the response rates $\Delta R(t)$. It dramatically simplifies uplift modeling considering all eight components of the Equation (5).

Applying uplift modeling, the expected result is an identification of segments with discriminatory power against $\Delta R(t)$. It is reasonable to assume that some final nodes of the trained decision tree will have positive effects, while some negative, or neutral. After training the uplift model, the estimation of $\overline{ITY}_R(t)$ is required for each final node, and then two lift charts arranging segments by descending order of predicted $\Delta R(t)$ can be produced:

(1) for $\overline{ITY}_R(t)$ that calculated based on Equation (5), and

(2) for approximation $\Delta R(t) \cdot \overline{\Delta ITE}(t)$

In case, that the first lift chart produces reasonable lift and it is not significantly underperforming the second one, the results can be recognized as adequate, and the approximation (9) is acceptable. Otherwise, more sophisticated uplift modeling approaches should be applied considering Equation (5).

CONCLUSION

When measuring treatment effect of marketing promotion campaigns and providing that this effect may be strongly affected by noise of non-responders, it is important to focus on decomposed elements of the effect estimations and consider underlying assumptions. The decomposition of the measurement between responders and non-responders for both treatment and control groups can then be used in uplift modeling to find profitable segments for the future promotional campaigns.

DISCLAIMER

The paper represents the views of the authors and do not necessarily reflect the views of the BMO Financial Group. All charts and data are simulated for illustrative purposes only and do not reflect the actual business state.

REFERENCES

Breur, T., 2007, “How to evaluate campaign response - The relative contribution of data mining models and marketing execution”, *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 15, Num. 2

Cousineau, D. and Chartier, S. 2010. “Outliers detection and treatment: a review“, *International Journal of Psychological Research*, Vol. 3. No. 1

Dixon, P., 2015. “The Business of Campaign Response Tracking”, *Proceedings SAS Institute Inc*, SAS Paper 1825

Jaffery, T. and Liu, S., 2009, “Measuring Campaign Performance by Using Cumulative Gain and Lift Chart”, *Proceedings SAS Institute Inc*, SAS Paper 196

Kolmogorov-Smirnov Test, 2017, https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test

Radcliffe, N. and Surry, P. 1999, *Differential response analysis: Modelling true response by isolating the effect of a single action*, in *Proceedings of Credit Scoring and Credit Control VI*, Credit Research Centre, University of Edinburgh Management School

Radcliffe, N., 2007. “Using Control Groups to Target on Predicted Lift: Building and Assessing Uplift Models”, *Direct Marketing Analytics Journal, Direct Marketing Association*

Rzepakowski, P. and Jaroszewicz, S., 2011. "Decision trees for uplift modeling with single and multiple treatments", *Knowledge and Information Systems*. 32

Speed, T., 1992. *Introduction to Fisher (1926) The Arrangement of Field Experiments. Breakthroughs in Statistics*, Springer Series in Statistics

Suresh, KP., 2011 "An overview of randomization techniques: An unbiased assessment of outcome in clinical research", *Journal of Human Reproductive Science*. Jan-Apr; 4(1): 8–11.

Wasserstein, R., 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose", *The American Statistician*, Vol. 70, Issue 2

Zhao, R., 2012. "Improve Marketing Campaign ROI using Uplift Modeling", *SAS Toronto Data Mining Forum*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Alex Glushkovsky
Alex.Glushkovsky@bmo.com

Matthew Fabian
Matthew.Fabian@bmo.com