Causal Inference and Machine Learning for Outbound Strategy

Victor S. Y. Lo and Zhuang Li

Workplace Investing, Fidelity Investments

Abstract:

Telemarketing via outbound phone channel has been a popular channel for organizations to improve customer experience or increase revenue for decades. Randomized control trial (RCT) or A/B testing is often used to measure program effectiveness, and in this case, the impacts of phone conversations. However, because only a small portion of customers tend to pick up such calls, RCT can only adequately measure the average treatment effect (ATE) in an Intent-To-Treat (ITT) setting. In our talk, Instrumental Variable (IV) method is proposed to estimate the effect of phone interactions only for the reached audience which is formally referenced as the Local Average Treatment Effect (LATE). Armed with these additional insights (LATE), marketing managers can make more informed decisions on how to optimize future program effectiveness by either improving the phone scripts (if the treatment effects are small) or increasing the reach (if those effects are large but are diluted by a low reach rate).

In addition to measurement and attribution, organizations are naturally motivated to improve program effectiveness by targeting a subset of customers who would generate relatively high individual treatment effects (or lift values). To address this second objective, a new approach is proposed to optimize the lift values by combining a call pickup propensity model (to increase reach) and an uplift modeling approach (to identify heterogeneous treatment effect). This integrated analytics framework can be leveraged to optimize the overall effectiveness of any intervention in an environment where it has limited reach. The reached audience is often referred to as compliers in RCTs. Finally, a realistic example will be used to illustrate the proposed framework.

Key Words:

Telemarketing; Intent-To-Treat; Instrumental Variable (IV); Local Average Treatment Effect (LATE); Compliers; Uplift Modeling; Conditional Average Treatment Effect (CATE)

1. Introduction

Outbound telemarketing remains to be one of the common channels for marketing and sales even in this digital age. It has been reported that outbound telemarketing programs can generate higher response rates than email campaigns (Hadfield (2017)). Examples of a “response” include signing up as a new customer (acquisition), buying an additional product or service (cross-selling), purchasing a higher-end product or service (upselling), or continuing to be a customer (retention).

To justify the marketing spend, organizations are often interested in measuring the impact of an outbound telemarketing program, typically through A/B testing or randomized controlled trial (RCT) where the target group is randomly split into treatment and control. However, a phone conversion with a customer or potential customer, regardless of how effective it might be, can only happen if the customer actually picks up the phone (see Figure 1). Another usage of A/B testing is to use the data from the experiment to build statistical or machine learning models predicting which types of customers are more responsive so that we can improve future program outcome by targeting the right customers. This paper addresses both problems in a situation where an intermediate variable (call pickup in telemarketing) is inherently present as follows:

1) How to measure the effect of the actual treatment when an intermediate variable is present (the act of picking up the phone call or not is the intermediate variable in this case) .
2) How to improve future targeting when such an intermediate variable is available? Specifically, how additional information from the intermediate variable can be incorporated in standard uplift modeling method to improve targeting?

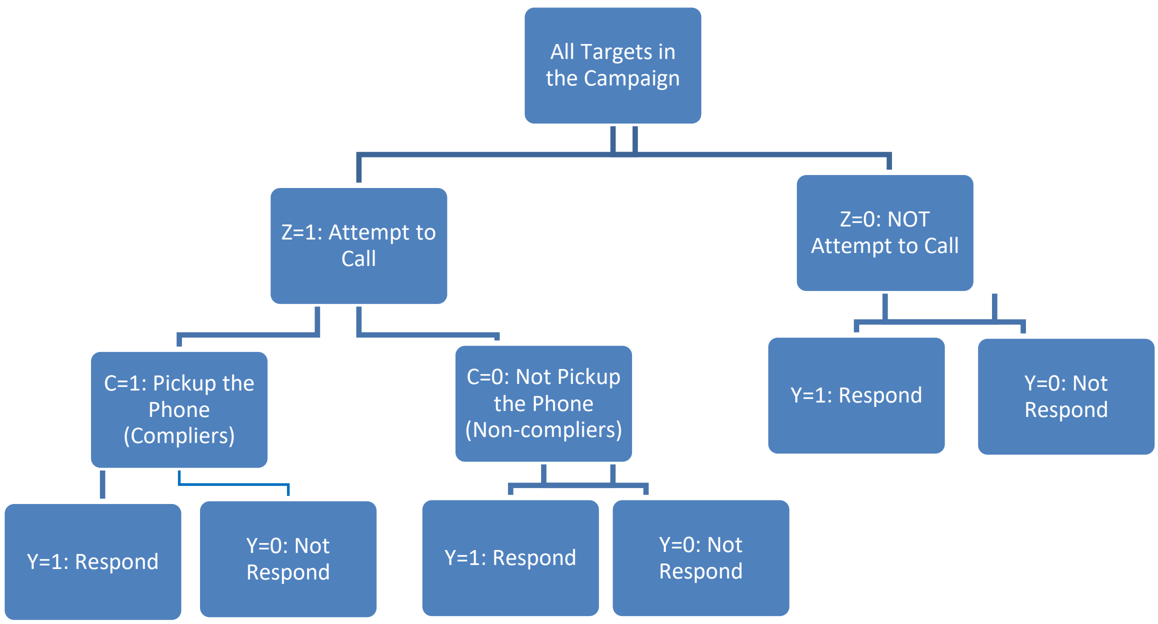

Figure 1. A Typical Telemarketing Program through A/B Testing.

This paper is organized as follows. Section 2 will discuss the campaign measurement problem and introduce an Instrumental Variable (IV)-based technique to measure the effect of the actual treatment (after having phone conversations). We will illustrate the technique using simulated data on non-profit charity donation. Section 3 will discuss the problem of improving future targeting through the usage of an intermediate variable (call pickup) and compare the proposed methodology with the standard uplift modeling method using the simulated example. Section 4 will introduce other applications where our proposed techniques can also be applied, followed by concluding remarks in Section 5.

2. Measuring the Effect of Actual Treatment

To measure the effect of a marketing campaign, a common method is through A/B testing or randomized controlled trial (RCT). As shown in Figure 1, campaign targets are randomly split into a treatment group (attempting to call them, Z=1) and a control group (not attempting to call them, Z=0). Since the treatment and control groups by design are similar in probability distributions of observed and unobserved variables, any difference in the outcome (Y) is attributable to the treatment (Z=1 vs 0) itself, resulting in measuring *causal* effects as opposed to associations. While this is the *gold standard* way of measurement (e.g., Hariton and Locascio (2018)), it does not provide a deeper insight organizations are often interested in: what is the effect of the "actual" treatment or the phone conversation itself? The phone conversation can only happen when customers pick up the call (C=1). Comparing the response rates in Z=1 (treatment group) vs Z=0 (control group) without conditioning on C will only measure the effectiveness of the program at the *Intent-To-Treat* (ITT) level, i.e., measuring the effectiveness based on the population we *intend* to treat rather than a portion of the population who end up *receiving* the treatment. The response outcome (Y=1 or 0) can happen whether they pick up the call (comply, C=1) or not (C=0) in the treatment group (Z=1), or they never receive the call in the control group (Z=0). This is analogous to the scenario in randomized clinical trials where patients are prescribed a medicine in the treatment group (Z=1), but they may or may not choose to take the medicine (i.e., comply, C=1, or not, C=0).

Measuring the effectiveness of the actual treatment (phone conversation) is equivalent to assessing the treatment effect among the compliers (i.e., those who pick up the call). A naïve approach is to compare the response rate among compliers (C=1) in the treatment group (Z=1) against the entire control group (Z=0). However, the complier population due to self-selection bias (e.g., the most loyal customers with high balance) is no longer a random sample of the treatment group, and this has become a causal inference problem with observational data rather than direct measurement with clean data from an RCT experiment. Although methodologies from statistics and social sciences exist to solve such causal inference problems, including the commonly used propensity score matching or weighting (e.g., Rubin (2006), Rubin & Waterman (2006), Rosenbaum (2002, 2010), and Imbens and Rubin (2015)), these methodologies would require us to identify an appropriate set of confounders which are not always fully observable. Given the business problem and the limitations of standard causal inference methodologies, we would introduce an Instrumental Variable (IV)-based methodology, inspired by a similar application in the biomedical settings, e.g., Carmody et al (2018) and Imbens and Rubin (2015, ch.23), as an alternative solution.

Instrumental Variable (IV) is defined as an exogeneous variable that only affects the outcome (Y) through an intermediate variable (Imbens and Rubin 2015, ch.23, Morgan & Winship 2015, ch.9 and Angrist and Pischke 2009, ch.4). This requirement is also known as "exclusion restriction." In our earlier example, the intermediate variable is having phone conversations (i.e., the actual treatment or complier, C), while being selected to the treatment group or not (i.e., attempt to call or not, Z) is the IV. It is also a perfect IV because it should not impact any outcome unless a phone conversation happens as seen in this causal path: $Z \rightarrow C \rightarrow Y$.

To obtain the Average Treatment Effect (ATE) of the compliers, let's start with the ATE of the entire program at the ITT level which can be expressed as follows using the Law of Iterated Expectation:

$$ATE\ (\text{ITT}) = ATE\ (\text{Complier}) * P(\text{Complier}) + ATE\ (\text{Non\_Complier}) * P(\text{Non\_Complier}) \qquad (1)$$

If we are willing to assume $ATE\ (\text{Non\_Complier}) = 0$, i.e., the response outcome of non-compliers is NOT impacted by "failed to treat" (failed to have a phone conversation), then from (1), ATE of the compliers can be derived by the *Wald Estimator*:

$$\textbf{ATE (Complier)} = \frac{\textbf{ATE (ITT)}}{\textbf{P(Complier)}} \qquad (2)$$

Equation (2), originated by Abraham Wald in 1940, is known as the Local Average Treatment Effect (LATE) or more intuitively, Complier Average Causal Effect (CACE). Note that the denominator of Equation (2), $\mathbf{P(Complier)}$ is simply the proportion of compliers in the "intent to treat" group. See Gordon et al (2019), DiazOrdaz et al (2018), Little et al (2009), and Schochet and Chiang (2009) for discussions of CACE.

Below are some common reasons why it is useful to estimate the average treatment effect among the compliers (LATE):

1) The marketers want to know if the intervention itself (call script) is effective in driving take-actions.
2) If the overall ATE (ITT) effect is low, data scientists and marketers would like to understand if it is caused by a) low complier rate or b) low impact even among compliers? These additional insights enable us to focus on the right component for future improvement.

Example: Charity Donation

We will illustrate the above method with a simulated example inspired by real use cases for a non-profit agency. Assuming the agency runs an annual fundraising campaign by outbound phone calls to past donors to encourage them to contribute again. Given their limited staff, the agency must select the most responsive donors to contact.

We assume a data set is available for a historical campaign based on an RCT design including the following information (details of the simulation assumptions are in the appendix):

- 80-20% split between treatment (400K) and control (100K)
- Randomly split into training (300K) and holdout (200K)
- Response variable (Y): Donate or not
- Predictors available:
    - Age of donor
    - Frequency – # times a donation was made in the past 3 years
    - Spent – average $ donation in the past
    - Recency – year of the last donation
    - Income
    - Wealth

- Call pickup probability (for C = 1) is a function of Recency, Age, and
- Response probability is a function of phone conversation (C), Age, Spent, and Frequency

As seen in Figure 2, the response rates in the treatment arm (Z=1) and control arm (Z=0) are 5.8% and 1.8%, respectively. As a result, the ATE (ITT) for the entire campaign is the difference between the two, which is 3.9%. Since the call pickup rate is 18.6%, the estimated LATE using Equation (2) is simply 3.9% / 18.6% = 21.0%. The actual LATE from our simulation is 20.9% so the estimate is very close to the actual result in this case. Note that if we simply compare the response rates in C=1 to Z=0 (a *naïve* estimate that ignores the potential selection bias of the compliers), it would be 21.7% - 1.8% = 19.9%, or if we compare the response rates in C=1 to C=0 (a *naïve* estimate that does not control for the differences between the two groups), it would be 21.7% - 2.1% = 19.6%, which are both slightly off from the true theoretical value.

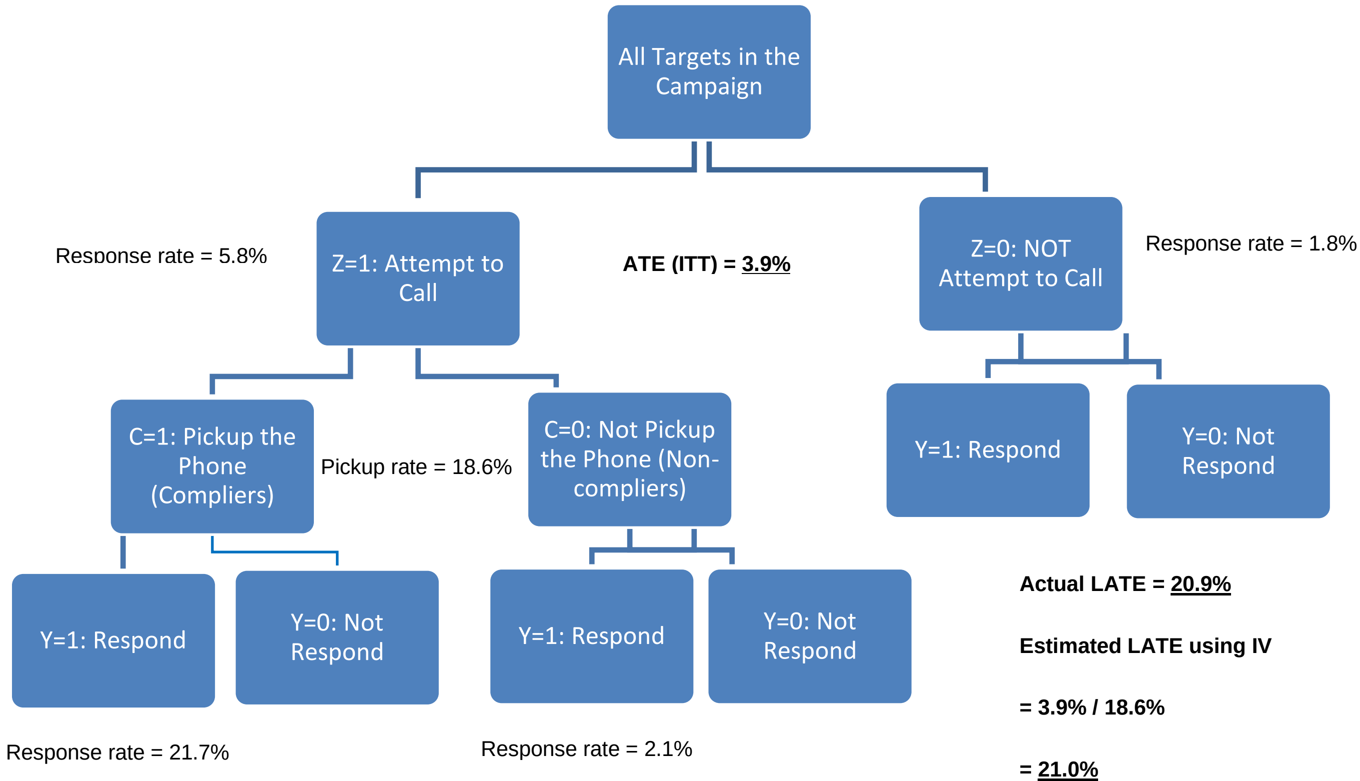

Figure 2. Historical Outbound Campaign with Simulated Results

3. Targeting Improvement using Model Decomposition and Integration for Uplift Modeling

As one of the key applications of predictive analytics and machine learning, models are often developed at the individual customer level to predict which customers are likely to respond to a marketing campaign based on the characteristics of the customers in the modelling population. Traditional response models are designed to identify *likely responders* regardless of whether these customers are contacted. A paradigm shift happened when Radcliffe & Surry (1999) and Lo (2002) proposed the Uplift or True lift modeling method to discover customers whose decisions would be ***positively influenced*** by marketing interaction instead of finding those who would respond *naturally*. Subsequent methodological developments and case studies include Radcliffe (2007a,b), Radcliffe & Surry (2011), Kane et al (2014), Athey & Imbens (2015), Lo & Pachamanova (2015), Pachamanova et al (2021). Gutierrez & Gerardy (2016) and Devriendt et al (2018) provide surveys of uplift modeling methodologies. An end-to-end coverage of uplift modeling, from sampling and design of experiment to model development and treatment optimization, is included in Haughton et al (2021, expected).

Uplift modeling relies on data from RCT or A/B testing in which a randomly-assigned control group is available. Similar research and applications in biomedicine for personalized or precision medicine include Cai et al. (2011) and Yong (2015, ch.3), and applications in political election are documented in Porter (2013), Scherer (2012), Siegel (2013a), and Samuelson (2013) which began with the 2012 presidential campaign as a key case study.

The main objective of uplift modeling is to predict individual customers lift values (or individual treatment effects), defined as:

$$Lift\,(x) \coloneqq P(Y = 1 \mid Z = 1,\, X = x) - P(Y = 1 \mid Z = 0,\, X = x) \qquad (3)$$

where Y is the response outcome, Z is the intent-to-treat variable, and X is a set of available covariates or features describing individual characteristics. This modeling technique is enabled by the differentiating ability of the available covariates or predictors, X, which typically include individual demographics and past behaviors. Similar to the previous section, this estimate is at the ITT (intent-to-treat) level but is individual specific rather than a group average. Since this is conditional on covariates, it is also known as Conditional Average Treatment Effect (CATE), as opposed to the overall group-level treatment effect ATE discussed in the previous section.

We now address the scenario where an intermediate variable, C (pick up a phone call or not in our case), is available. Referring to Figure 1 or 2, for those in the treatment group (Z = 1), the campaign process happens in this sequence: intent-to-treat (attempt to call customers), customers pick up the calls, and customers take actions, i.e., $Z \rightarrow C \rightarrow Y$. We can *decompose* this process into two steps:

1) $Z \rightarrow C$ describes the call pickup (or complier), and
2) $C \rightarrow Y$ describes the response among those picked up the call, i.e., among the compliers.

As a result, the overall probability of response for those who received a call (i.e., intended to be called) can be decomposed using the Law of Total Probability as follows:

P(response | received a call)

= *P*(pick up the call | received a call) *P*(response | pick up the call)

+ *P*(do not pick up the call | received a call) *P*(response | do not pick up the call).

Mathematically, the problem is expressed as:

$$
\begin{aligned}
&P(Y = 1 \mid Z = 1,\ X = x) \\
&= P(C = 1 \mid Z = 1,\ X = x)\, P(Y = 1 \mid C = 1,\ Z = 1,\ X = x) \\
&\quad + (1 - P(C = 1 \mid Z = 1,\ X = x))\, P(Y = 1 \mid C = 0,\ Z = 1,\ X = x) \qquad (4)
\end{aligned}
$$

This decomposition serves an important purpose in that it enables us to model the two steps in equation (4) separately. A major advantage of this approach is that it allows for a different set of features to enter each component model, further increasing the predictive accuracy of the combined outcome.

To arrive at the lift value (Equation (3)) which is the key decision metric for targeting, we will need to subtract the model score for the control response probability, $P(Y = 1 \mid Z = 0,\ X = x),$ from equation (4). Note that this probability will need to be estimated through a separate model.

Finally, we assume that the response rate is the same for those in the control ($Z = 0$) and those who did not pick up the call in the treatment group ($C = 0$ and $Z = 1$), i.e.,

$$P(Y = 1 \mid C = 0,\ Z = 1,\ X = x) = P(Y = 1 \mid Z = 0,\ X = x).$$

This assumption is intuitive and reasonable because both groups never had an opportunity for a phone conversation[1]. This enables us to combine the two groups, i.e., $\{Z = 0\} \cup \{C = 0,\ Z = 1\}$, in a single model to predict $Y = 1$.

In summary, the approach we propose includes estimating the following three models separately:

1) Call pickup probability in the treatment group: $P(C = 1 \mid Z = 1,\ X = x)$
2) Response probability among the compliers in the treatment group: $P(Y = 1 \mid C = 1,\ Z = 1,\ X = x)$
3) Response probability among the non-compliers in the treatment group which is assumed to be the same as the response probability in the control group: $P(Y = 1 \mid C = 0,\ Z = 1,\ X = x) = P(Y = 1 \mid Z = 0,\ X = x)$

Once the above model estimates are available, they will be *assembled* using Equations (4) and then (3) to compute the lift estimate at the individual level for campaign prioritization. In summary, this new *Integrated Model* not only uses data required by standard uplift models, but also takes advantage of the additional information contained in the intermediate variable through probability decomposition.

<u>Example: Charity Donation (continued)</u>

To continue with the example in the previous section, we apply the above *Integrated Model* to the simulated data in the charity donation case and compare this new method with the following more standard methodologies:

1) *Baseline*: Treatment only model, which is a predictive model built based on the treatment group only, i.e., directly modeling $P(Y = 1 \mid Z = 1,\ X = x)$ without using any data from the control group. This

[1] This is similar to the assumption of assigning ATE (Non_Complier) $= 0$ in Equation (1) in the previous section with the LATE.

approach used to be the standard technique prior to the invention of uplift modeling and is still routinely used today for campaign targeting; and

2) *Standard Uplift Model*: This set of methodologies is to model the individual lift in Equation (3) (without any decomposition proposed in this section). While there are alternative techniques available, for illustration purposes, we will use the two-model approach which models the treatment and control response probabilities separately: $P(Y = 1 \mid Z = 1, X = x)$ and $P(Y = 1 \mid Z = 0, X = x)$, before taking the difference as in Equation (3), as covered in Kane et al (2014), Lo & Pachamanova (2015), and Pachamanova et al (2020).

The results over the test (holdout) sample (Figures 3-4) indicate that that the Integrated Model posts the strongest performance, followed by the uplift model, with the baseline conventional model finishing in last place. Summary model performance measures in terms of Gini confirm this conclusion. See Kane et al (2014) for a discussion of Gini and Gini 15% (which focuses on the top 15% or 1.5 decile as target audience in campaigns).

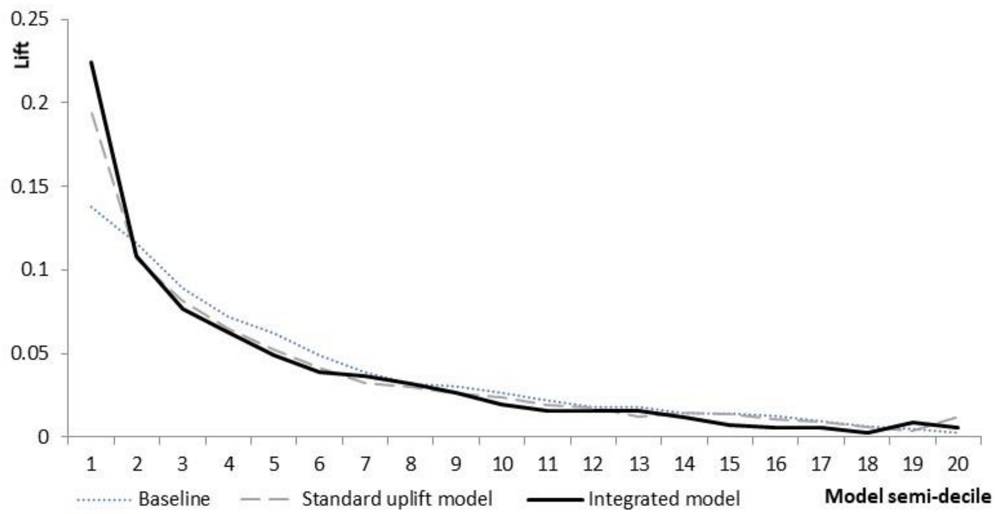

Figure 3. Lift Chart for the Charity Donation Simulated Example

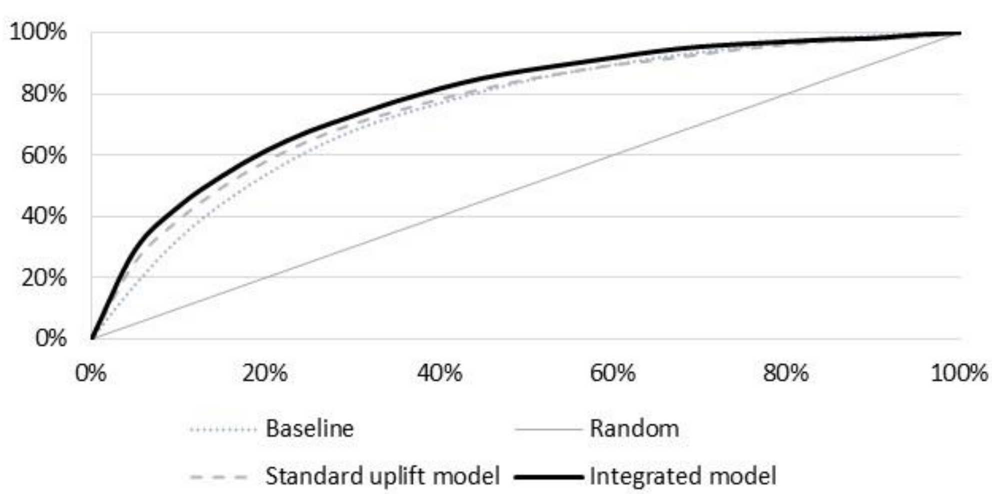

Figure 4. Gains Chart for the Charity Donation Simulated Example

Model	Gini	Gini 15%
Baseline: Treatment Only	4.93	0.65
Standard Uplift Model	5.20	0.83
Integrated Model	5.69	0.95

Table 1. Performance Metrics for the Charity Donation Simulated Example

4. Other Applications of Our Proposed Methodologies

While our paper focuses on outbound telemarketing program, the proposed methodologies based on instrumental variable and probability decomposition can be applied more broadly in other similar situations such as:

1) Webinar invitation: the goal in this case is to invite customers to a webinar through A/B testing so they may take a desirable action after attending the webinar but typically not everyone invited will show up (Z = invited, C = attend, Y = take action).
2) Email marketing: when emails are sent out to customers in an A/B test, they will have to open the email or click a link before being directed to the website to take an action (Z = email sent, C = open/click, Y = take action).
3) Digital display marketing: in a typical online digital display A/B test, the goal may be for customers to click on a banner ad but they will have to be able to see it first (Z = customer slated for an ad impression, C = ad impression is rendered, Y = click on the ad).
4) Pharmaceutical Clinical Trials: in a randomized clinical trial setting where some patients are assigned to receive a drug to determine its impact on their health outcome, they will have to take the medicine first (Z = drug prescription, C = take the medicine, Y = health outcome).
5) Digital health: when individuals are recommended to follow health-related advice (e.g., nutrition, exercise, or sleep) in an A/B test in order to measure the impact on health outcome, they will have to follow the advice first (Z = message recommended, C = follow advice, Y = health or cost outcome).

5. Concluding Remarks

We have proposed two methodologies for outbound telemarketing where customers will have to pick up the phone to receive an actual treatment (phone conversation). First, an instrumental variable (IV) based methodology is proposed to *causally* estimate the effect of the phone conversation on business outcome among those who picked up the call (i.e., compliers). These insights enable marketing managers to make more informed decisions on how to optimize future program effectiveness by either improving the phone scripts (if the treatment effects are small) or increasing the reach (if those effects are large but are diluted by a low reach rate). Second, we have proposed a new methodology for improving future campaign targeting by integrating standard uplift modeling technique with the intermediate variable (call pickup). Using simulated data, we have demonstrated that the new approach generates a stronger performance than conventional approaches. Finally, we have provided a list of potential application areas where our proposed methodologies can be utilized to measure program effectiveness and improve ability of targeting.

Appendix: Details of the Charity Donation Simulation Example

The following assumptions are used in the simulation example in this paper for the response variable, call pickup, and predictors:

$Age \sim N(55,\ 10^2)$, $\ln(income) \sim N(4 + 0.007\ Age,\ 0.5^2)$, $frequecy \sim Binomial(10{,}0.25)$,

$$Spent' \sim 5\ * N(8\ icome,\ 150^2), Spent = \max(10,\ Spent')\ ,$$

$$Wealth' \sim N(4\ income + 0.04\ income^2 + 200,\ 100^2),\ Wealth = \max(0.001, Wealth'),$$

and $Recency \sim Poisson(1)$;

$Pickup\ probability = \frac{1}{1+\exp\left(-(-6.5\ -0.3\ Recency\ -0.15\ Age)\right)}$,

$Pickup \sim Binomial(1, Pickup\ probability)$;

IF $Pickup = 1,\ then\ Response\ probability = \frac{1}{1+\exp\left(-\left(-16+3.3+\frac{150\ Age+Spent+100\ frequency}{1000}\right)\right)}$,

Else $Response\ probability = \frac{1}{1+\exp\left(-\left(-16+\frac{90\ Age+Spent+100\ frequency}{1000}\right)\right)}$.

References

Angrist, J.D. and J.-S. Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*," Princeton University Press.

Athey, S. and G.W. Imbens (2015), "Machine Learning Methods for Estimating Heterogeneous Causal Effects," *Working Paper*, Graduate School of Business, Stanford University.

Cai, T., Tian, L., Wong, P., and Wei, L.J. (2011), "Analysis of Randomized Comparative Clinical Trial Data for Personalized Treatment," *Biostatistics*, 12:2, p.270-282.

Carmody, T., T.L. Greer, R. Walker, C.D. Rethorst, and M.H. Trivedi (2018), "A Complier Average Causal Effect Analysis of the Stimulant Reduction Intervention using Dosed Exercise Study," *Contemporary Clinical Trials Communications*, 10, p.1-8.

Devriendt, F., D. Moldovan, and W. Verbeke (2018), "A Literature Survey and Experimental Evaluation of the State-of-the-Art in Uplift Modeling: A Stepping Stone Toward the Development of Prescriptive Analytics," *Big Data*, v.6, no.1, p.13-40.

DiazOrdaz, K., A. J. Franchini, and R. Grieve (2018), "Methods for Estimating Complier Average Causal Effects for Cost-Effectiveness Analysis," *J. R. Statist. Soc. A*, 181, Part 1, p.277-297.

Gordon, B.R., F. Zettelmeyer, N. Bhargava, and D. Chapsky (2019), "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook," *Marketing Science*, v.38, No.2.

Gutierrez, P. and J.-Y. Gerardy (2016), "Causal Inference and Uplift Modeling: A Review of the Literature," *Journal of Machine Learning Research: Workshop and Conference Proceedings*, 67:1-13.

Hadfield, R. (2017), "53 Sales Follow Up Statistics," zoominfo, https://blog.zoominfo.com/sales-follow-up-statistics/?utm_medium=is&utm_source=link&utm_campaign=cold-calling/email-prospecting-vs-cold-calling/

Hariton, Ed. And J.J. Locascio (2018), "Randomised Controlled Trials – The Gold Standard for Effectiveness Research," *BJOG*, 125 (13), https://obgyn.onlinelibrary.wiley.com/doi/10.1111/1471-0528.15199

Haughton, D., Haughton, J., and Lo, V.S.Y. (2021, expected) *Cause-and-Effect Business Analytics*, CRC/Chapman & Hall.

Imbens, G. W. and D. B. Rubin (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press.

Kane, K., Lo, V.S.Y., and Zheng, J. (2014) "Mining for the Truly Responsive Customers and Prospects Using True-Lift Modeling: Comparison of New and Existing Methods." *Journal of Marketing Analytics*, v.2, Issue 4, p.218-238.

Little, R.J., Q. Long, and X. Lin (2009), "A Comparison of Methods for Estimating the Causal Effect of a Treatment in Randomized Clinical Trials Subject to Noncompliance," *Biometrics*, 65, p.640-649.

Lo, V.S.Y. (2002) "The True Lift Model – A Novel Data Mining Approach to Response Modeling in Database Marketing," *SIGKDD Explorations* 4, Issue 2, p.78-86, at: http://www.acm.org/sigs/sigkdd/explorations/issues/4-2-2002-12/lo.pdf

Lo, V.S.Y. (2008), "New Opportunities in Marketing Data Mining," in *Encyclopedia of Data Warehousing and Mining*, Wang (2008) ed., 2nd edition, Idea Group Publishing.

Lo, V.S.Y. (2020), "Top 10 Essential Data Science Topics to Real-World Application from the Industry Perspectives," *Harvard Data Science Review*, Issue 2.3, at: https://hdsr.mitpress.mit.edu/pub/diub13so/release/2

Lo, V.S.Y. and D. Pachamanova (2015), "A Practical Approach to Treatment Optimization While Accounting for Estimation Risk," *Journal of Marketing Analytics*, v.3, Issue 2, p.79-95.

Morgan, S.L. and Winship C. (2015). *Counterfactuals and Causal Inference.* 2nd edition, Cambridge University Press.

Pachamanova, D., V.S.Y. Lo, and N. Gulpinar (2020) "Uncertainty Representation and Risk Management in Direct Segmented Market," *Journal of Marketing Management,* v.36, Issue 1-2.

Porter, Daniel (2013) Pinpointing the Persuadables: Convincing the Right Voters to Support Barack Obama. Presented at Predictive Analytics World; Oct, Boston, MA; http://www.predictiveanalyticsworld.com/patimes/pinpointing-the-persuadables-convincing-the-right-voters-to-support-barack-obama/ (available with free subscription).

Radcliffe, Nicholas J. (2007a) Using Control Groups to Target on Predicted Lift. *DMA Analytic Annual Journal*, Spring, 14-21.

________ (2007b) Generating Incremental Sales: Maximizing the incremental impact of cross-selling, up-selling and deep-selling through uplift modelling. Stochastic Solutions Limited.

________ and Patrick D. Surry (1999) Differential response analysis: modeling true response by isolating the effect of a single action. In: *Proceedings of Credit Scoring and Credit Control VI,* Credit Research Centre, U. of Edinburgh Management School.

________ and ________ (2011), Real-World Uplift Modelling with Significance-Based Uplift Trees. Portrait Technical Report TR-2011-1 and Stochastic Solutions White Paper 2011, http://stochasticsolutions.com/pdf/sig-based-up-trees.pdf.

Rosenbaum, P.R. (2002), *Observational Studies*. Springer.

Rosenbaum, P.R. (2010), *Design of Observational Studies*. Springer.

Rubin, D.B. (2006), *Matched Sampling for Causal Effects*. Cambridge University Press.

Rubin, D.B. and Waterman, R.P. (2006), "Estimating the Causal Effects of Marketing Interventions Using Propensity Score Methodology," *Statistical Science*, p.206-222.

Samuelson, Douglas A. (2013) Analytics: Key to Obama's Victory. *OR/MS Today*, Feb, 20 - 24.

Scherer Michael (2012) How Obama's Data Crunchers Helped Him Win. CNN News at http://www.cnn.com/2012/11/07/tech/web/obama-campaign-tech-team/index.html?hpt=hp_bn5, accessed 7 November 2012.

Schochet, P. Z. and H. Chiang (2009), "Estimation and Identification of the Complier Average Causal Effect Parameter in Education RCTs," *Technical Methods Report* NCEE 2009-4040, U.S. Dept. of Education.

Siegel, Eric (2011) Upilft Modeling: Predictive Analytics Can't Optimize Marketing Decisions Without It. Prediction Impact white paper sponsored by Pitney Bowes Business Insight.

________ (2013a) The Real Story Behind Obama's Election Victory. *The Fiscal Times* 01/21/2013, at http://www.thefiscaltimes.com/Articles/2013/01/21/The-Real-Story-Behind-Obamas-Election-Victory.aspx#page1, accessed 31 January 2013.

________ (2013b) *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. New Jersey: Wiley.

Yong, Florence H. (2015), "Quantitative Methods for Stratified Medicine," *PhD Dissertation*, Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University.