Identifying Individuals Who Are Truly Impacted by Treatment: Introduction to + Recent Advances in Uplift Modeling

Conference Paper · April 2014

1 author:

Victor S. Y. Lo
Fidelity Investments / Bentley University

Identifying Individuals Who Are Truly Impacted by Treatment:

Introduction to + Recent Advances in Uplift Modeling

Victor S.Y. Lo, Fidelity Investments and Bentley University

Joint work with:
Kathleen Kane, Fidelity Investments, and
Jane Zheng, Focus Optimal

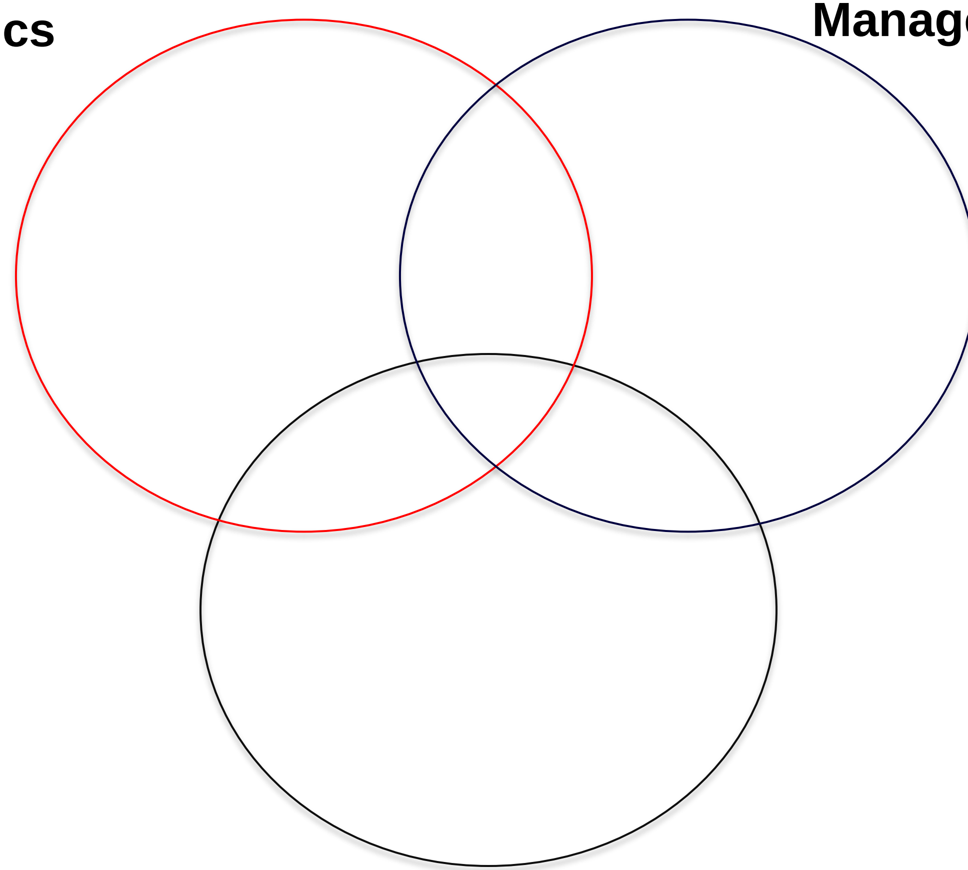
April, 2014

Presented to:

- Bentley Analytics Virtual Symposium,
- Boston Chapter INFORMS, and
- Boston Chapter of American Statistical Association

Statistics

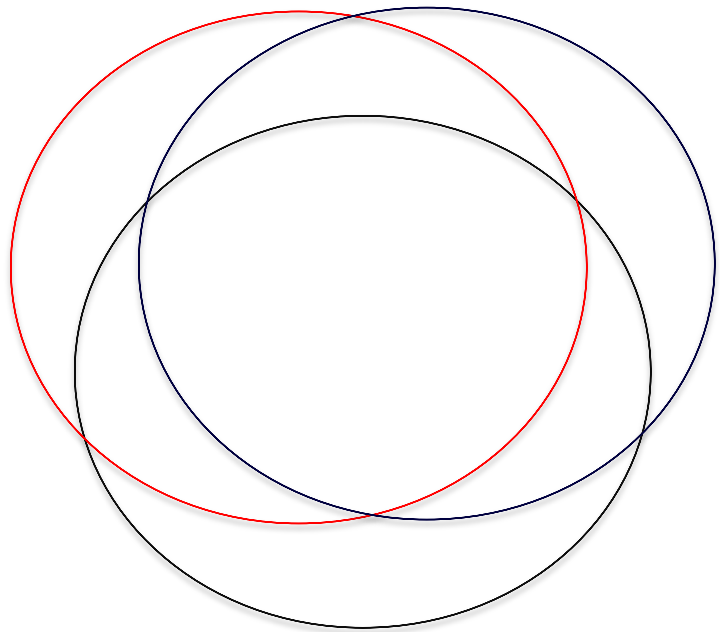
Operations Research /
Management Science



Business Analytics / Data Analytics
Data mining / Data Science

Operations Research /
Management Science

Statistics

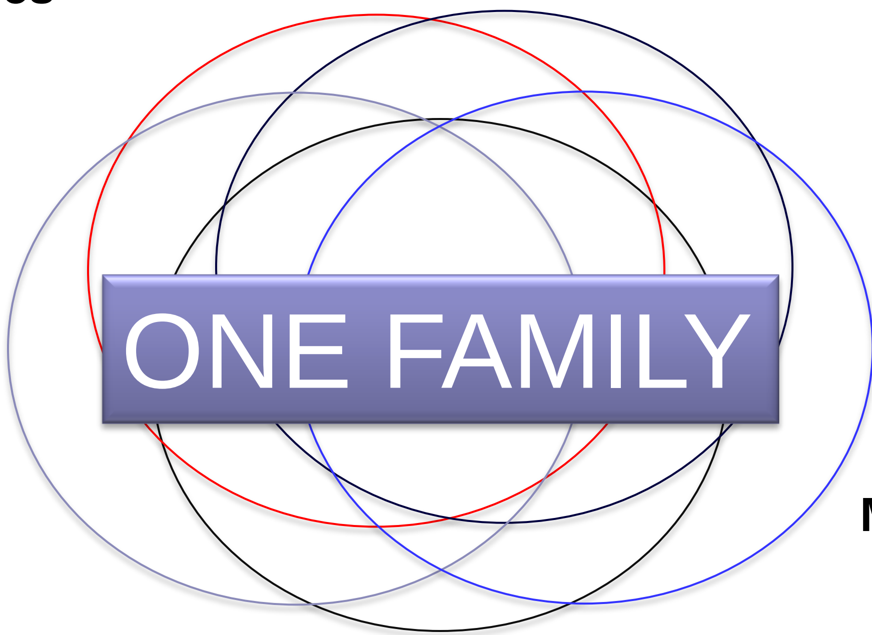


Business Analytics /
Data mining / Data Science

Statistics

Operations Research /
Management Science

Econometrics



Machine Learning / AI

And more…

Business Analytics /
Data mining / data science

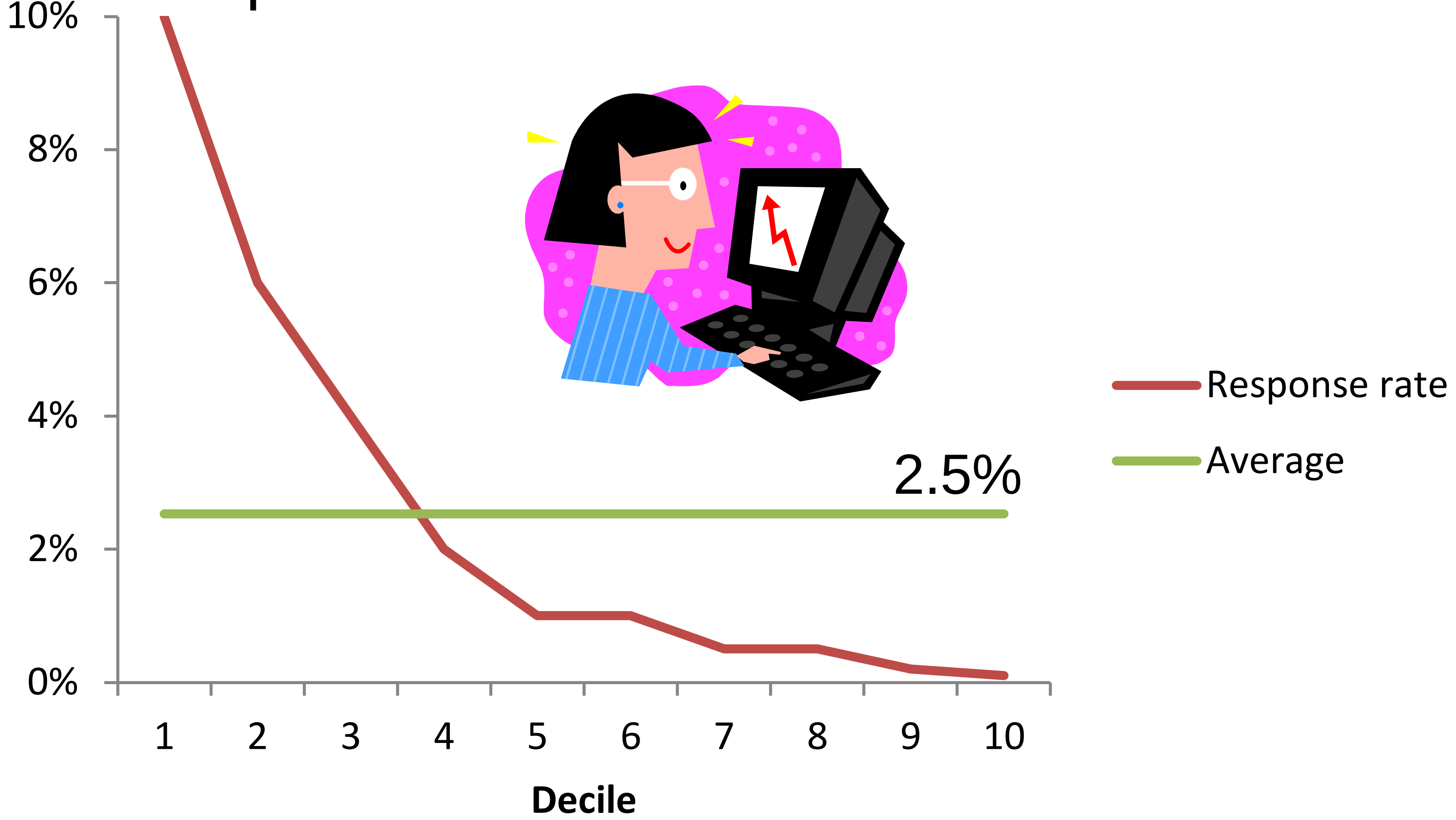
Outline

■ Why do we need Uplift modeling?	10 min
■ Various methods for Uplift modeling	25 min
■ Uplift with non-experimental data	10 min
■ Direct response vs. Uplift modeling	10 min
■ Questions	≥ 5 min

Disclaimer:

This presentation does not represent the views or opinions of Fidelity or Bentley

Top decile lift (over random) = 4 times
Top 3 deciles lift = 2.6 times



Big Lift
Modelers: VERY SUCCESSFUL MODEL!

Campaign Results

	Top 3 Deciles	Random
Treatment	6.7%	2.5%
Control	6.7%	2.5%
Lift	0.0%	0.0%

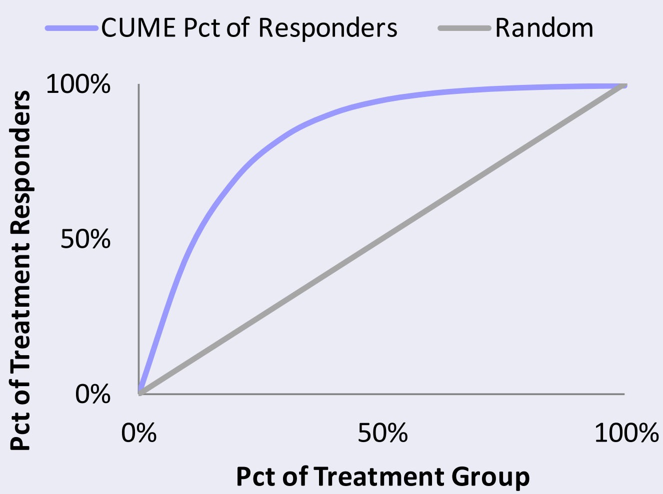
No Lift
Marketers: VERY DISAPPOINTING!

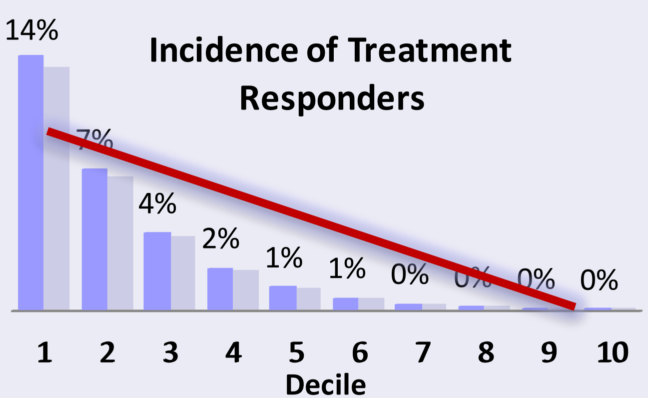
Modelers:
Not my problem, it is the mail design!

So, Who is Right?

What's wrong with this picture?

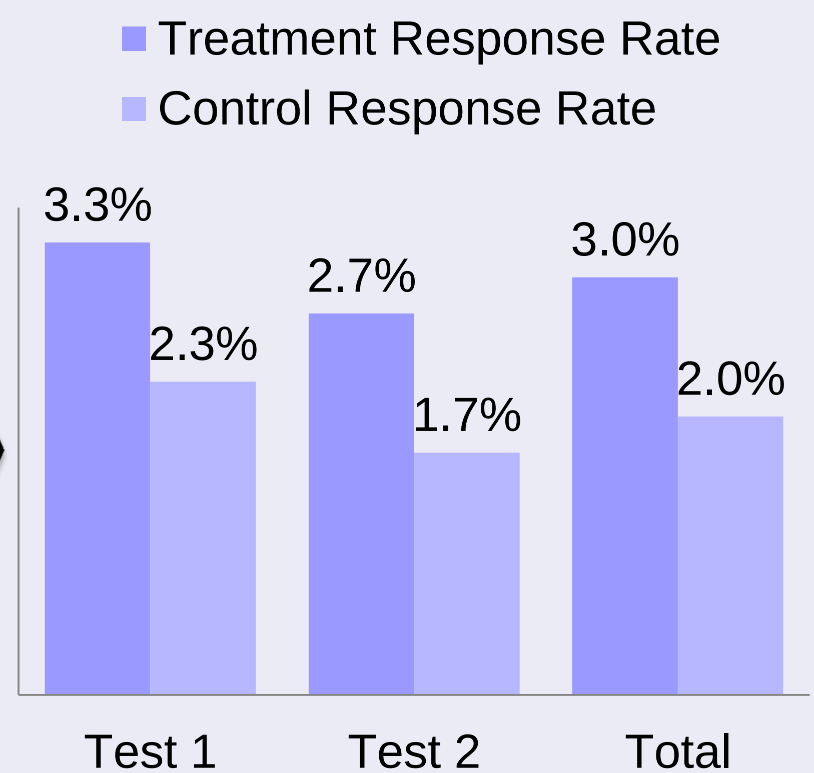
DM LIFT?





A successful response model

DM LIFT?



A successful marketing campaign

Motivation

- Based on the following campaign result, which of the customer groups is the best for future targeting ?

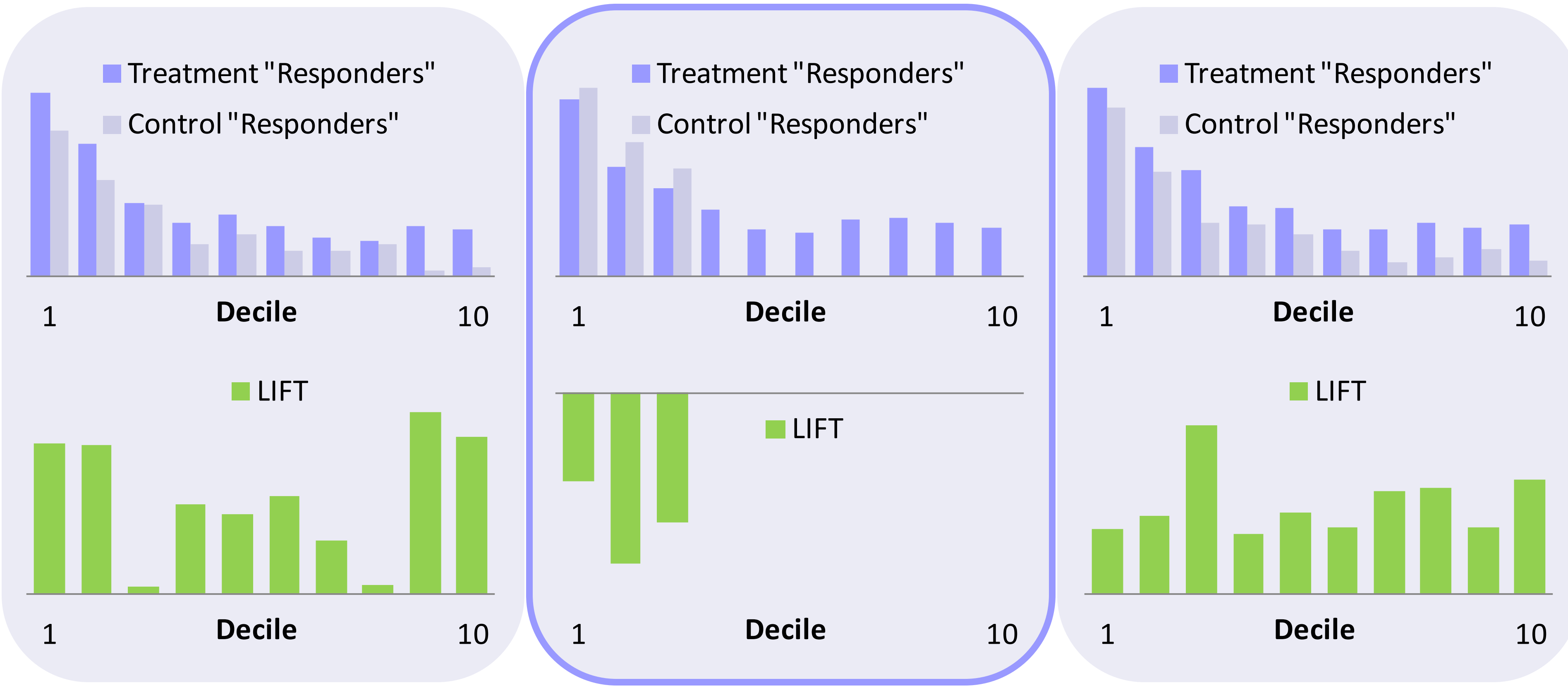
Response Rate By Age and Treatment/Control

Age		Treatment	Control	Difference
	<35	0.5%	0.2%	0.3%
	35-60	2.5%	0.5%	2.0%
	>60	3.5%	2.5%	1.0%

• **>60 has the highest response rate** – treatment-only focus (common practice)

• **35-60 has the highest Lift** (highest likelihood to be *positively influenced* by the treatment)

Similarly, Measure Response Models by Lift over Control



Simulated data

Why do we need Uplift Modeling?

Look-alike model: find people who will respond

- P(R)

Standard response model: find people who will respond after receiving a treatment

- P(R | Treatment)

Uplift model: find people who will respond because of the treatment

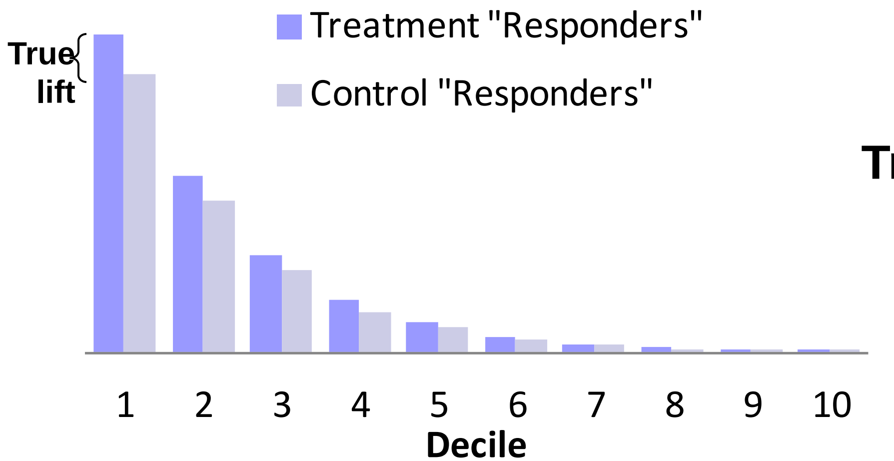
- P(R | Treatment) – P(R | no Treatment)

- Standard response models often behave more like Look-alike models

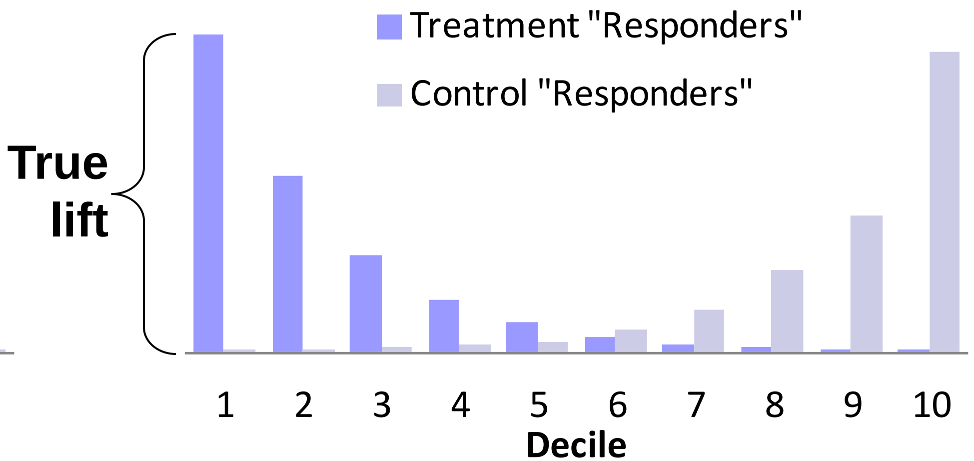
- Why spend marketing $ and resources on people who would “respond” anyway?

The uplift model objective

- Maximize the Treatment responders while minimizing the control "responders"



A standard response model



A uplift response model (Ideal)

Hypothetical data

Uplift model solutions

0. Baseline results: Standard response model – treatment-only (as a benchmark)

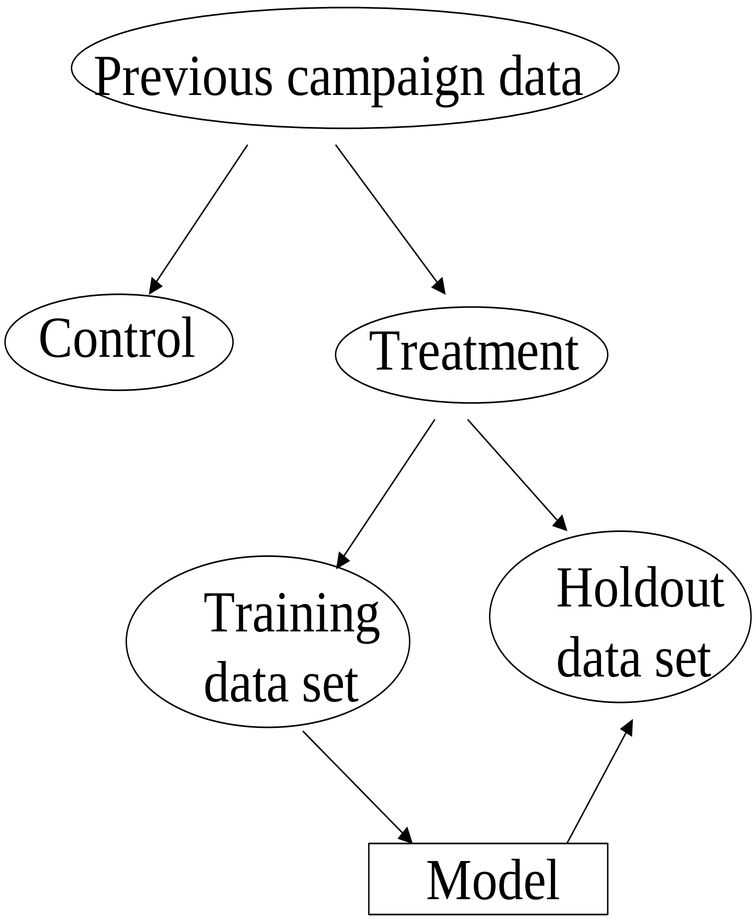
1. Two Model Approach: Take difference of two models, Treatment – Control

2. Treatment Dummy Approach: Single combined model using treatment interactions

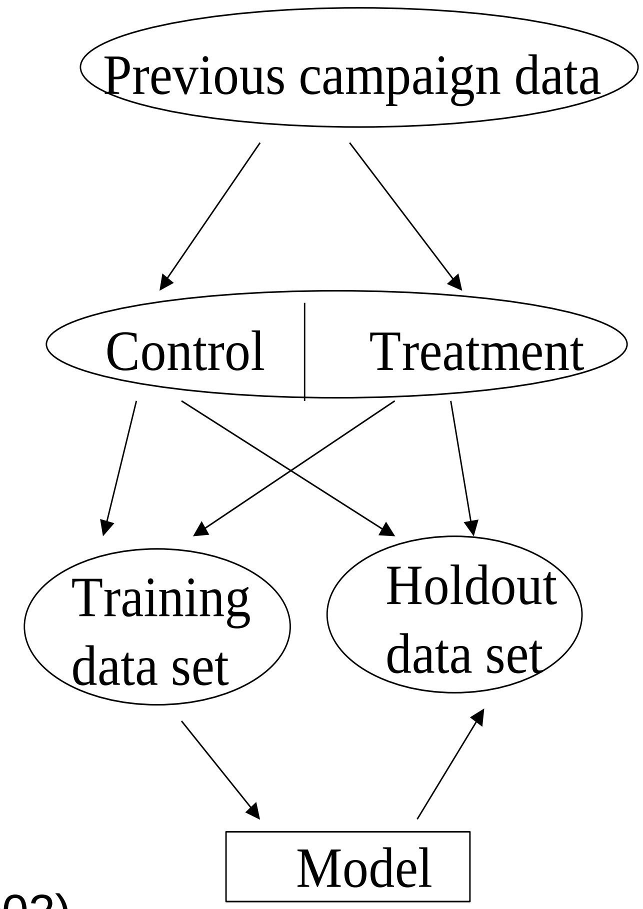
3. Four Quadrant Method

Uplift Approaches

- Traditional Approach

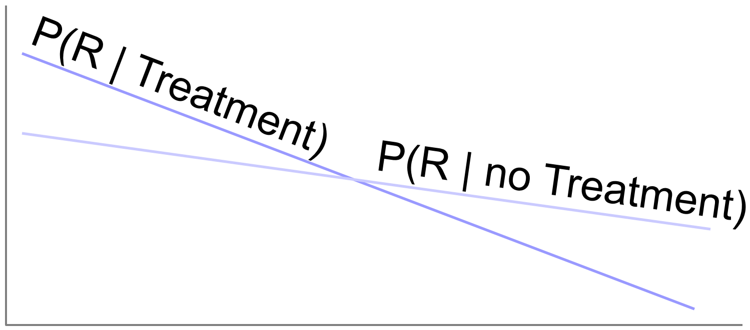


- Uplift Modeling



Source: Lo (2002)

Method 1: Two Model Approach: Treatment - Control



- Model 1 predicts P(R | Treatment)
  - Model Sample = Treatment Group
- Model 2 predicts P(R | no Treatment)
  - Model Sample = Control Group

- Final prediction of lift =

Treatment Response Score – Control Response Score

- Pros: simple concept, familiar execution (x2)
- Cons: indirectly models uplift, the difference may be only noise, 2x the work, scales may not be comparable, 2x the error, variable reduction done on indirect dependent vars

Method 2: Treatment Dummy Approach, Lo (2002)

- 1. Estimate both *E(Y_i|X_i;treatment)* and *E(Y_i|X_i;control)* and use a dummy *T* to differentiate between treatment and control:
  - Linear logistic regression:

$$P_i = E(Y_i \mid X_i) = \frac{exp(\alpha + \beta' X_i + \gamma T_i + \delta' X_i T_i)}{1 + exp(\alpha + \beta' X_i + \gamma T_i + \delta' X_i T_i)}$$

- 2. Predict the lift value (treatment minus control) for each individual:

$$Lift_i = P_i \mid treatment - P_i \mid control$$

$$= \frac{exp(\alpha + \gamma + \beta' X_i + \delta' X_i)}{1 + exp(\alpha + \gamma + \beta' X_i + \delta' X_i)} - \frac{exp(\alpha + \beta' X_i)}{1 + exp(\alpha + \beta' X_i)}$$

- Pros: simple concept, tests for presence of interaction effects
- Cons: multicollinearity issues

Method 3: Four Quadrant Method

- Model predicts probability of being in one of four categories
  - Dependent variable outcome (nominal) = TR, CR, TN, or CN
  - Model Population = Treatment & Control groups together

		Response	
		Yes	No
Treatment	Yes	TR	TN
	No	CR	CN

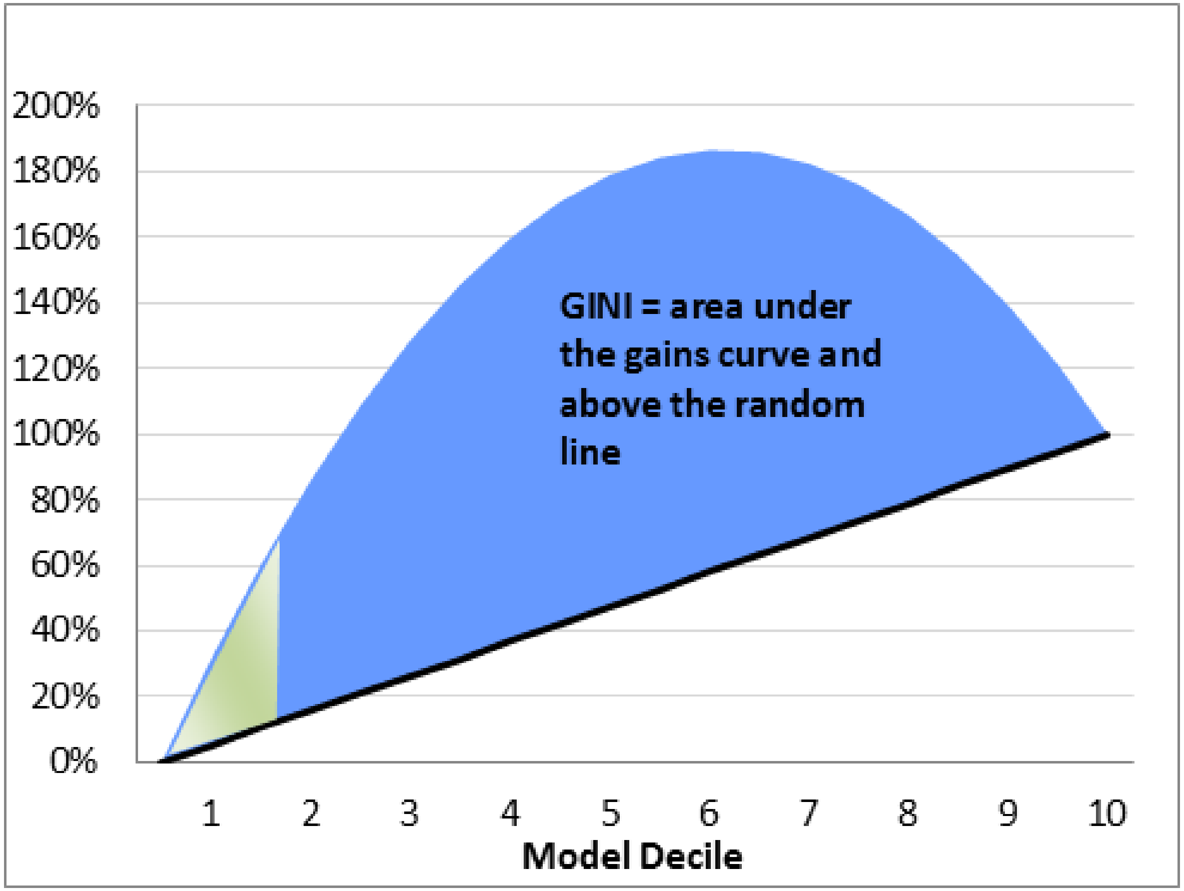
- Prediction of lift:

$$Z(x) = \frac{1}{2}\left[\frac{P(TR|x)}{P(T)} + \frac{P(CN|x)}{P(C)} - \frac{P(TN|x)}{P(T)} - \frac{P(CR|x)}{P(C)}\right]$$

Generalized Lai (2006)

- Pros: only one model required; more “success cases” to model after
- Cons: not that intuitive…

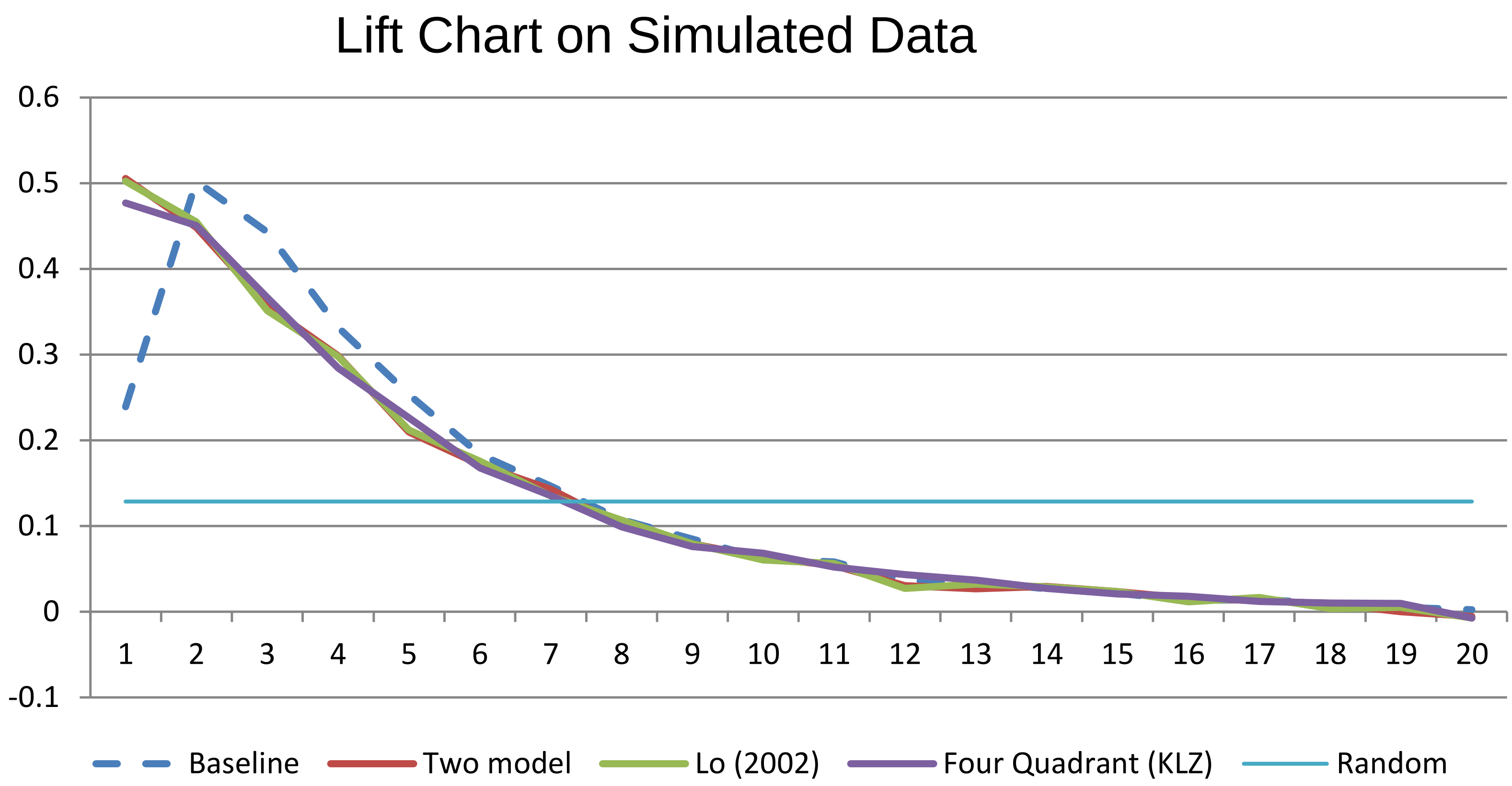
Gini and Top 15% Gini in Holdout Sample



Simulated Example: Charity Donation

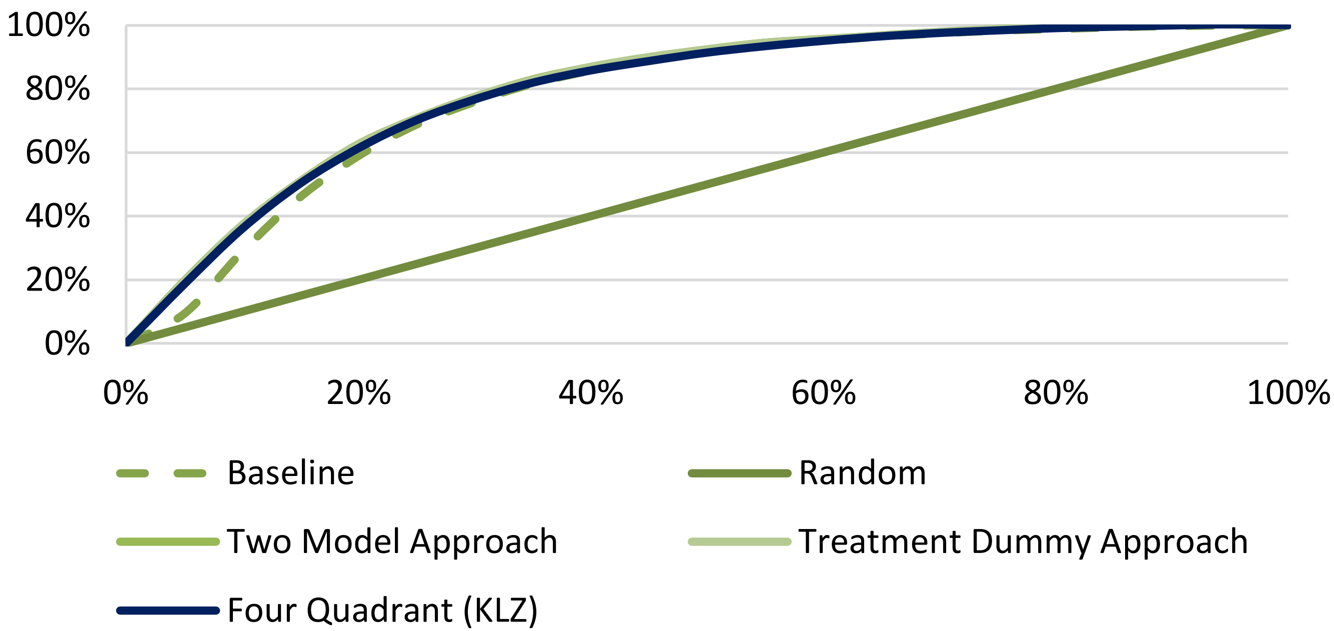
- 80-20% split between treatment and control
- Randomly split into training (300K) and holdout (200K)
- Predictors available:
  - Age of donor
  - Frequency – number of times a donation was made in the past
  - Spent – average amount of donation in the past
  - Recency – year of the last donation
  - Income – estimated annual income
  - Wealth – estimated wealth

Holdout Sample Performance



Theoretical model: Two logistics for treatment and control

Gains Chart on Simulated Data



	Gini	Gini 15%	Gini repeatability (R^2)
Baseline	5.6420	0.5412	0.7311
Method 1: Two Model approach	6.0384	0.7779	0.7830
Method 2: Lo(2002), Treatment Dummy	6.0353	0.7766	0.7836
Method 3: Four Quadrant Method (or KLZ)	5.9063	0.7484	0.7884

Online Merchandise Data

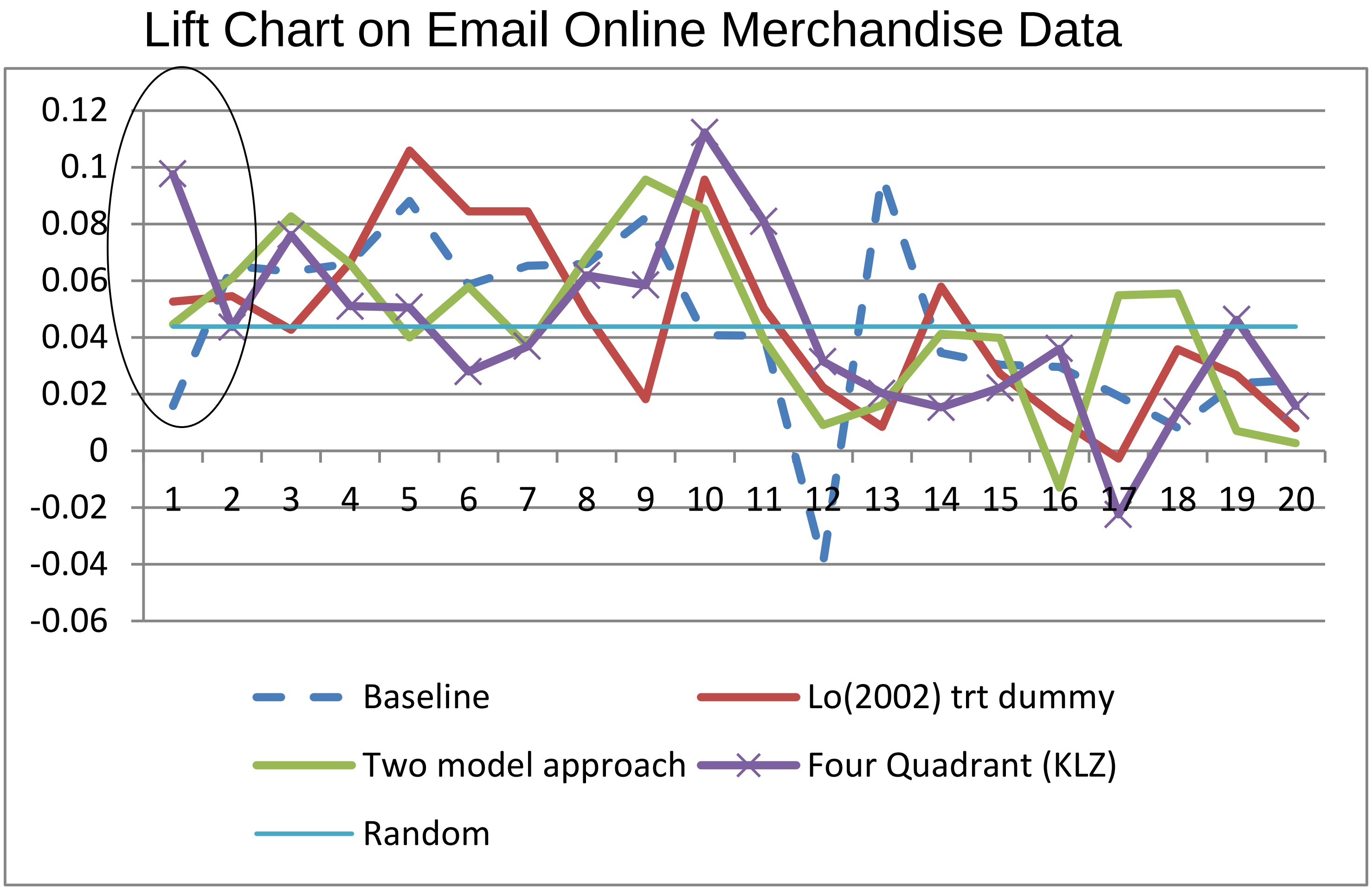
- From blog.minethatdata.com, with women's merchandise online visit as response

- 50-50% split between treatment and control (43K in total)

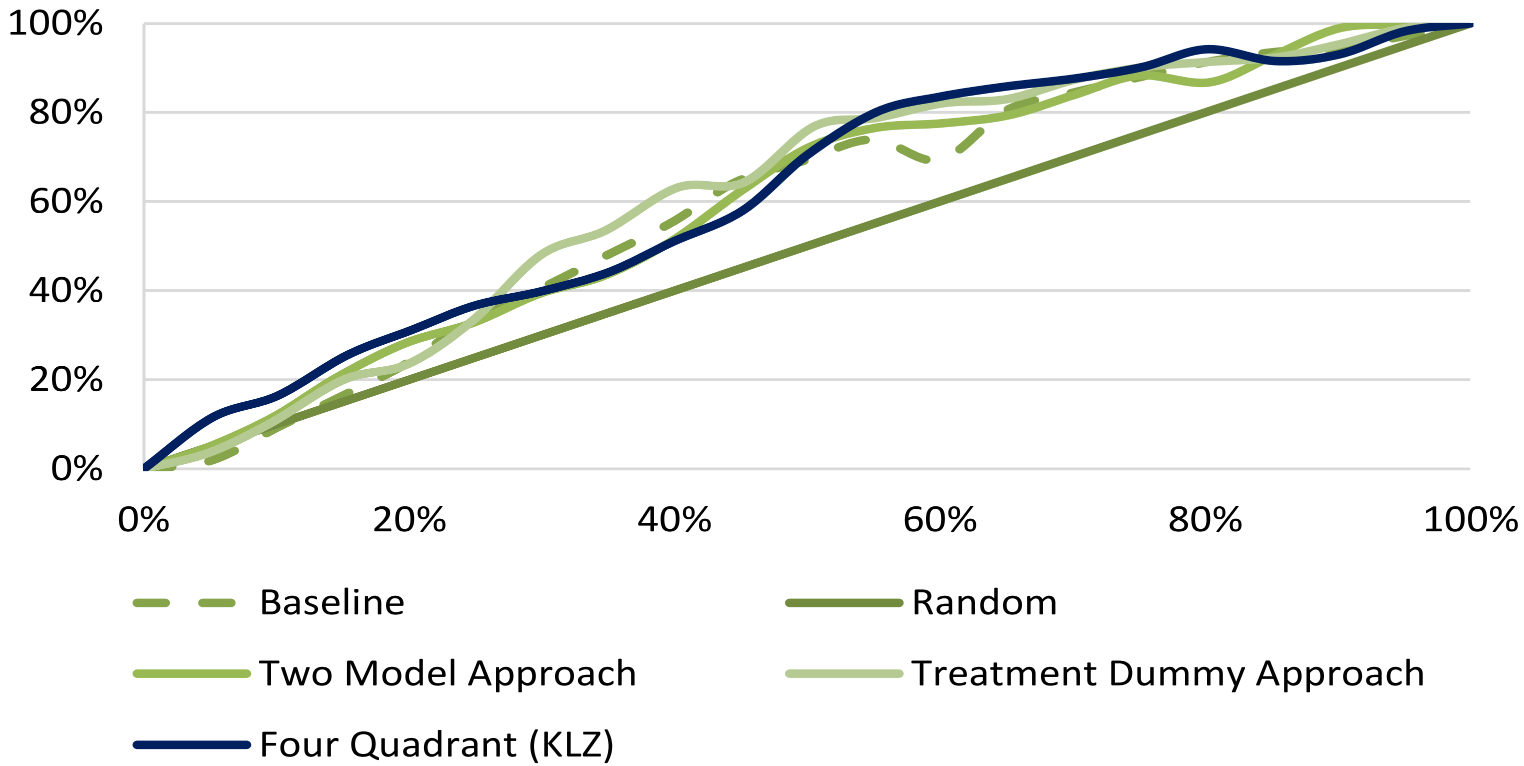
- Randomly split into training (70%) and holdout (30%)

- Predictors available:
  - Recency
  - Dollar spent last year
  - Merchandise purchased last year (men's, women's, both)
  - Urban, suburban, or rural
  - Channel – web, phone, or both for purchase last year

Holdout Sample Performance



Gains Chart on Email Online Merchandise Data

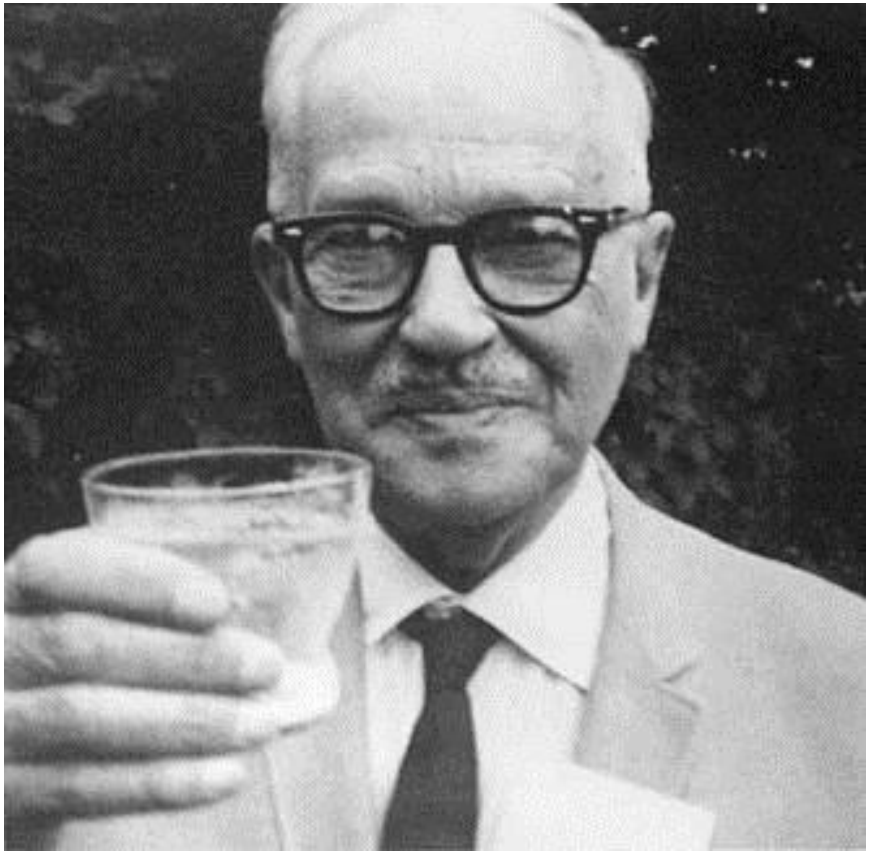


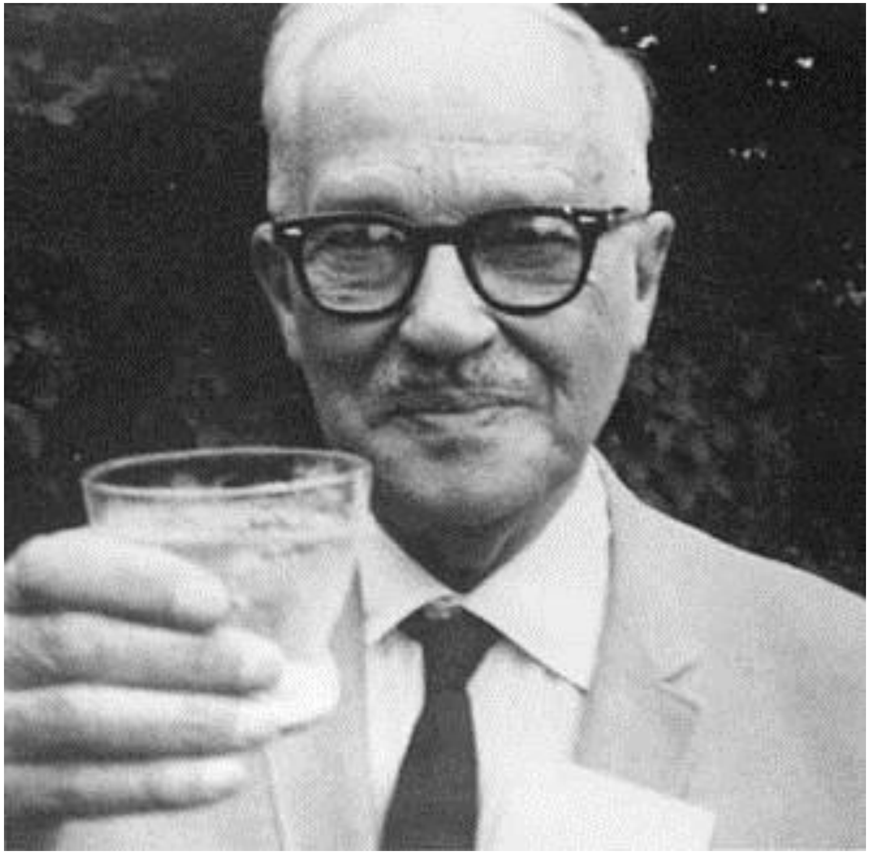
	Gini	Gini 15%	Gini repeatability (R^2)
Baseline	1.8556	-0.0240	0.2071
Method 1: Two Model approach	2.0074	0.0786	0.2941
Method 2: Lo(2002), Treatment Dummy	2.4392	0.0431	0.2945
Method 3: Four Quadrant Method (or KLZ)	2.3703	0.2288	0.3290

Ideal Conditions for Uplift Modeling

- A randomized control group is withheld!
- Treatment does not cause all “responses,” i.e. control response rate > 0
- Natural Response is not highly correlated to Lift
- Lift Signal-to-Noise ratio (Lift/control rate) is large enough

Recognize these faces?





A link between Statistics and Operations Research?

Case I:
Non-randomized Experiments

Uplift for Non-Randomized Experiments

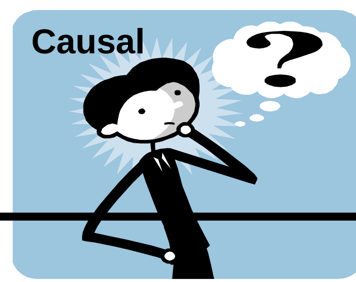
- Sales program
- Car safety program
- Talent development
- Pre-clinical / pre-experimental analysis
- College value

Blocking the “Back-Door” Path

E.g. age, income, geography, etc.

Confounder (if known)

Treatment (or Control) Assignment



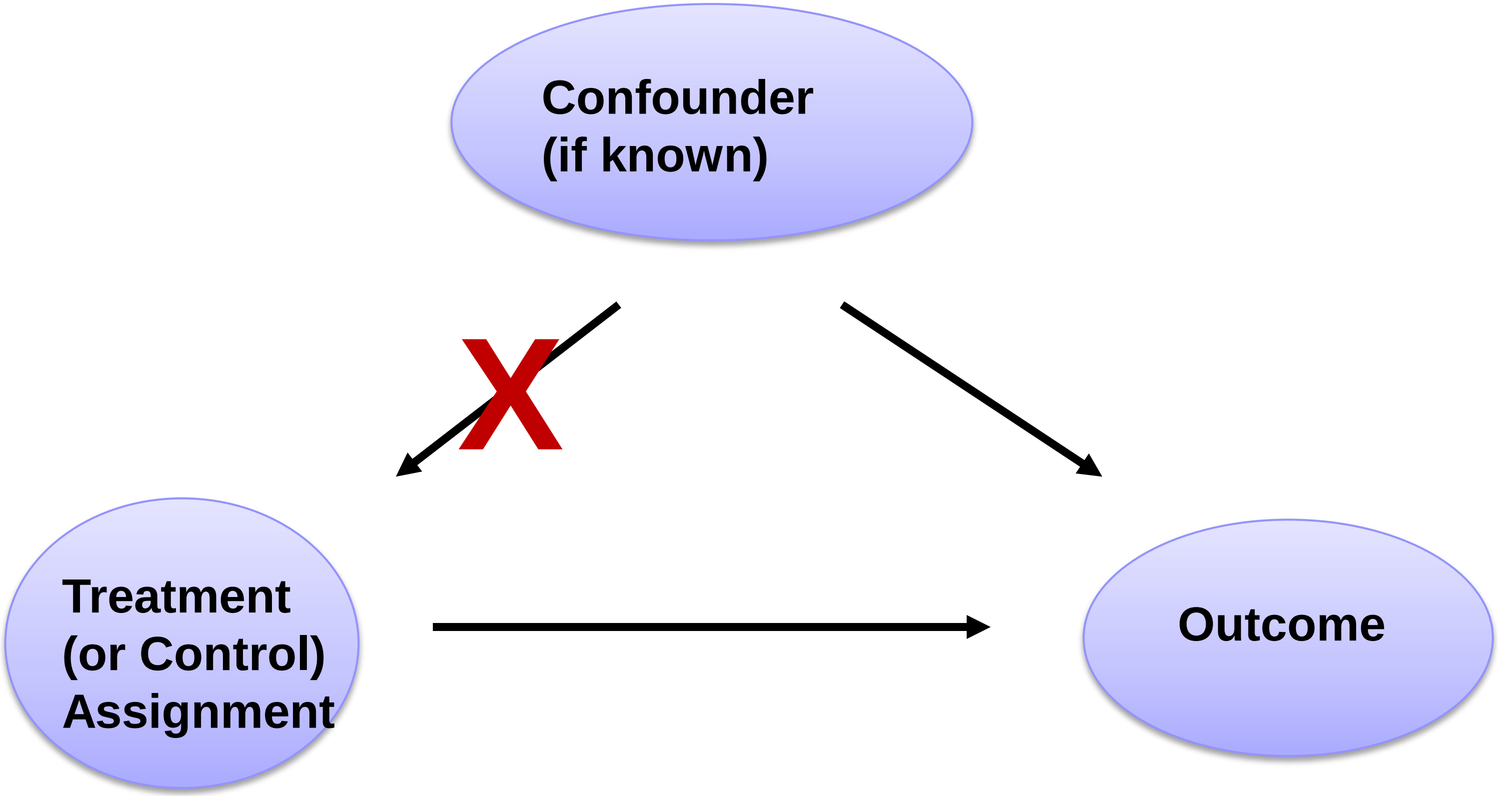
Outcome

E.g. Sales Effort

E.g. Buy or not

By blocking the Confounder, we can have a cleaner estimate of Treatment Effect

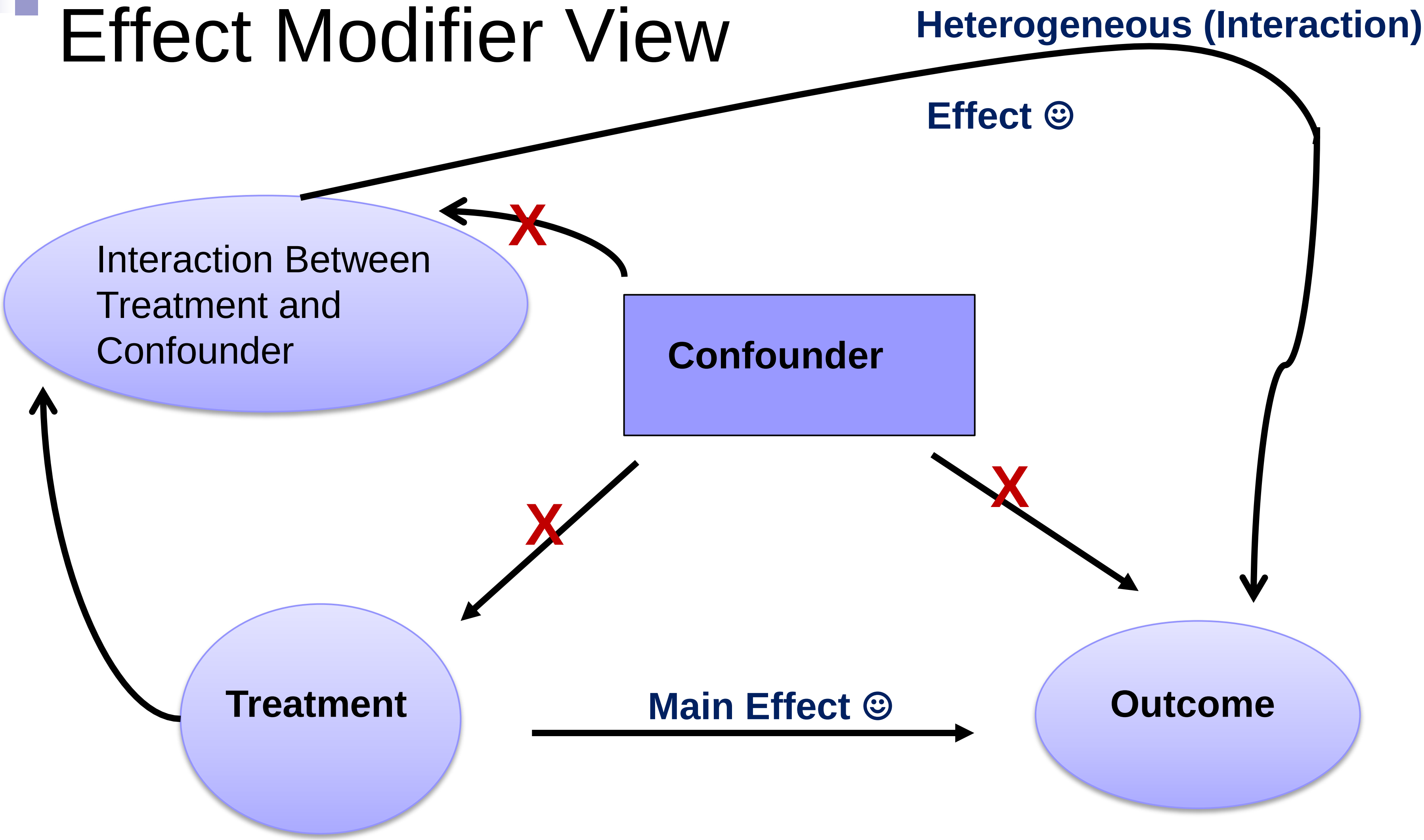
Blocking the “Back-Door” Path with Propensity Score



Break the confounder-treatment link

Usual assumptions: Conditional Ignorability, Positivity, etc.

Effect Modifier View



Interests: Main and Interaction Effects

~ Doubly Robust Estimation for Non-experimental Uplift

Propensity Score Matching

- Find a group of untreated individuals ("control") that are similar to the treatment group in all *appropriate* pre-treatment characteristics

- Matching with multiple covariates is difficult – propensity score (PS) is a good summary of multiple covariates:
  - PS: $P(T = 1 \mid x) = f(x)$
  - $X \coprod$ T | $_{\text{PS}}$

- Once PS is available for every individual, there are alternative procedures for adjustment

Propensity Score Matching (cont.)

Inverse Probability Weighting Method (IPW)

$$w_i = \begin{cases} \frac{p_t}{P(T_i=1|x_i)}, & \text{if } i \text{ is in treatment;} \\ \frac{p_c}{1-P(T_i=1|x_i)}, & \text{if } i \text{ is in control.} \end{cases}$$

Creates a pseudo-population with no confounding

Adjust the Two Model Approach, Treatment Interaction Approach, or Four Quadrant Approach

Four Quadrant Method with Adjustment

- Model predicts probability of being in one of the four categories
  - Dependent variable outcome = TR, CR, TN, or CN
  - **With IPW weights** in estimation

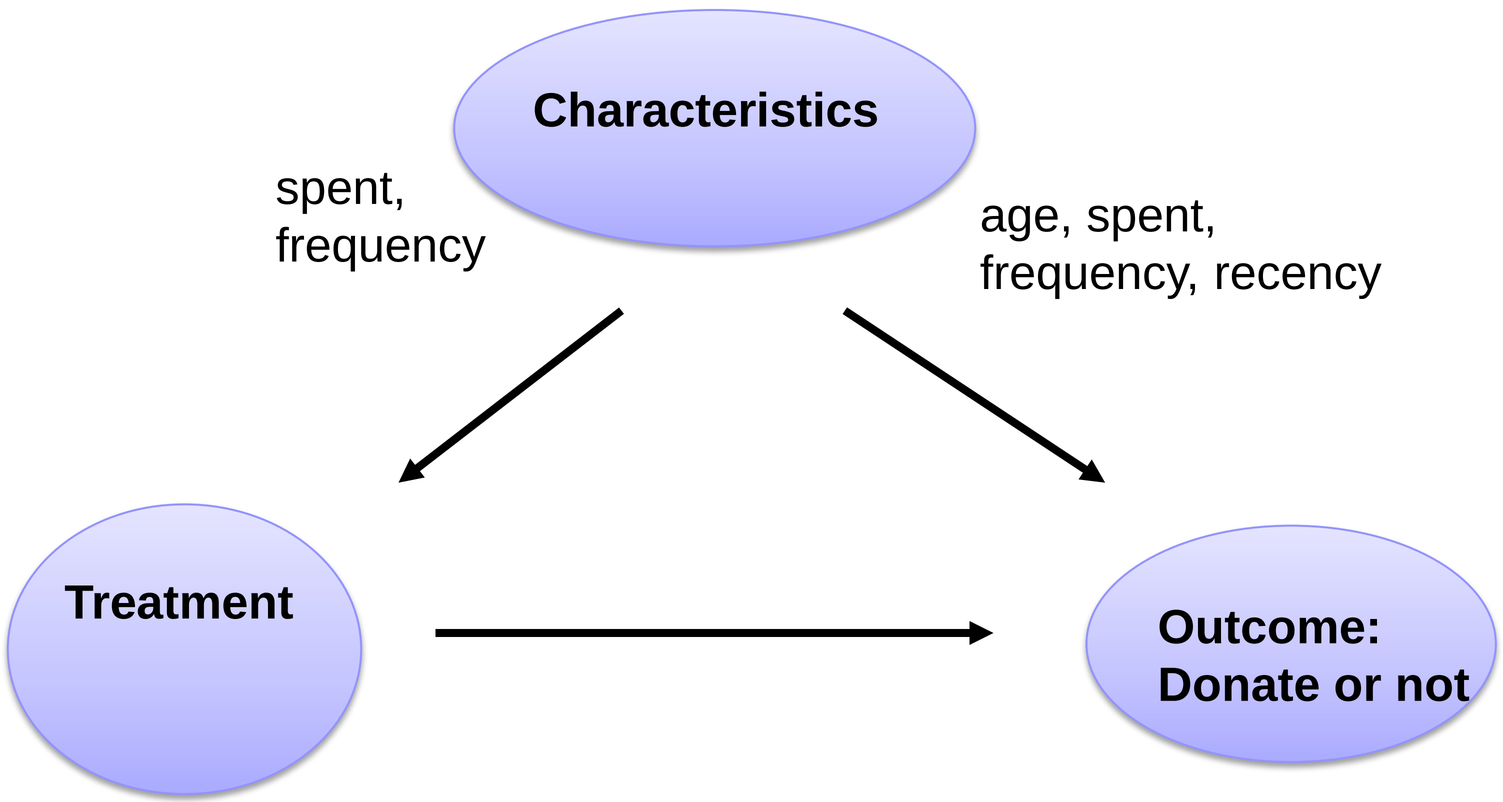
		Response	
		Yes	No
Treatment	Yes	TR	TN
	No	CR	CN

- Prediction of lift:

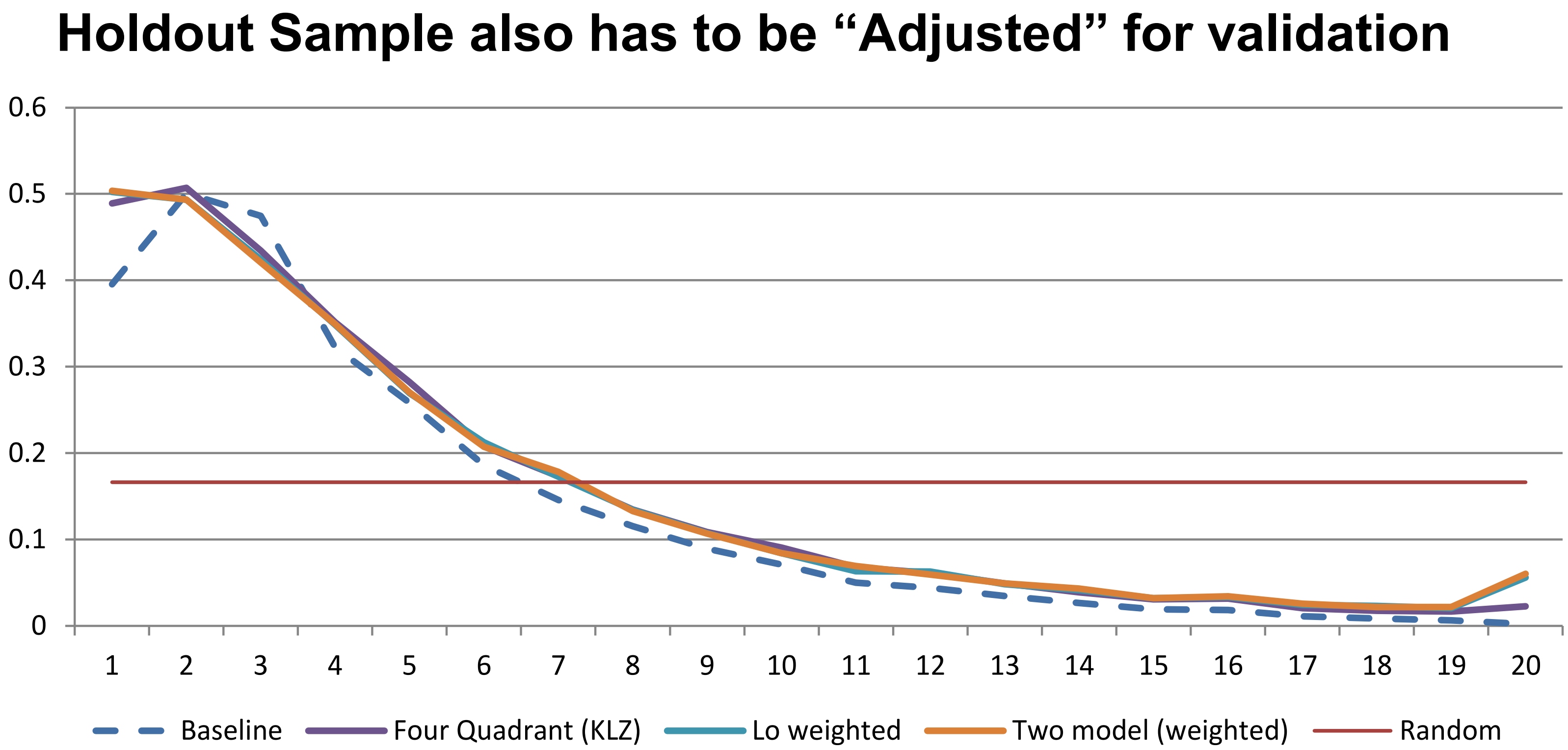
$$Z(x) = \frac{1}{2}\left[\frac{P(TR|x)}{P(T|x)} + \frac{P(CN|x)}{P(C|x)} - \frac{P(TN|x)}{P(T|x)} - \frac{P(CR|x)}{P(C|x)}\right]$$

$$\text{where } P(T|x) = P(TR|x) + P(TN|x), etc.$$

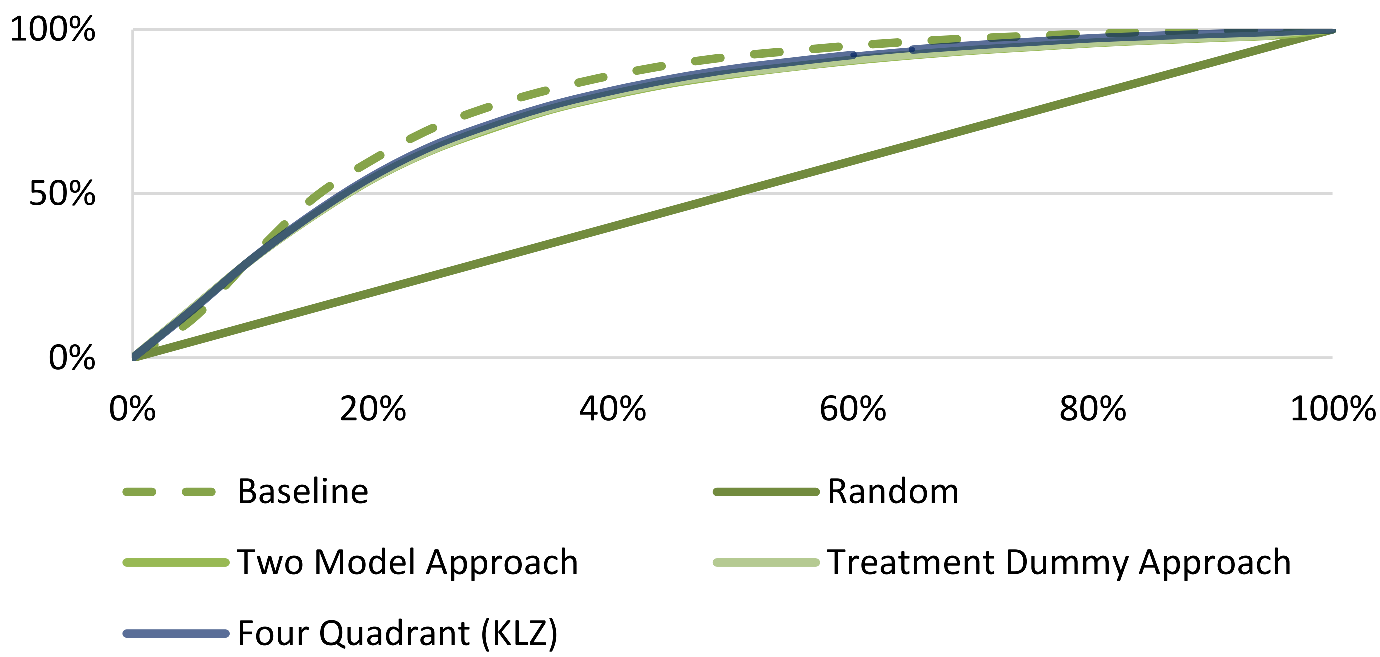
Simulated Charity Donation Data with a *Twist*



Holdout Sample Performance



Gains Chart on Twisted Simulated Charity Donation Data



	Gini	Gini 15%	Gini 5%	Gini repeatability (R^2)
Baseline	5.7456	0.6125	0.0690	78.22%
Two Model Approach	4.9928	0.5849	0.1014	78.65%
Treatment Dummy, Lo (2002)	5.0271	0.5851	0.1009	78.84%
Four Quadrant Method (KLZ)	5.2185	0.5872	0.0969	81.25%

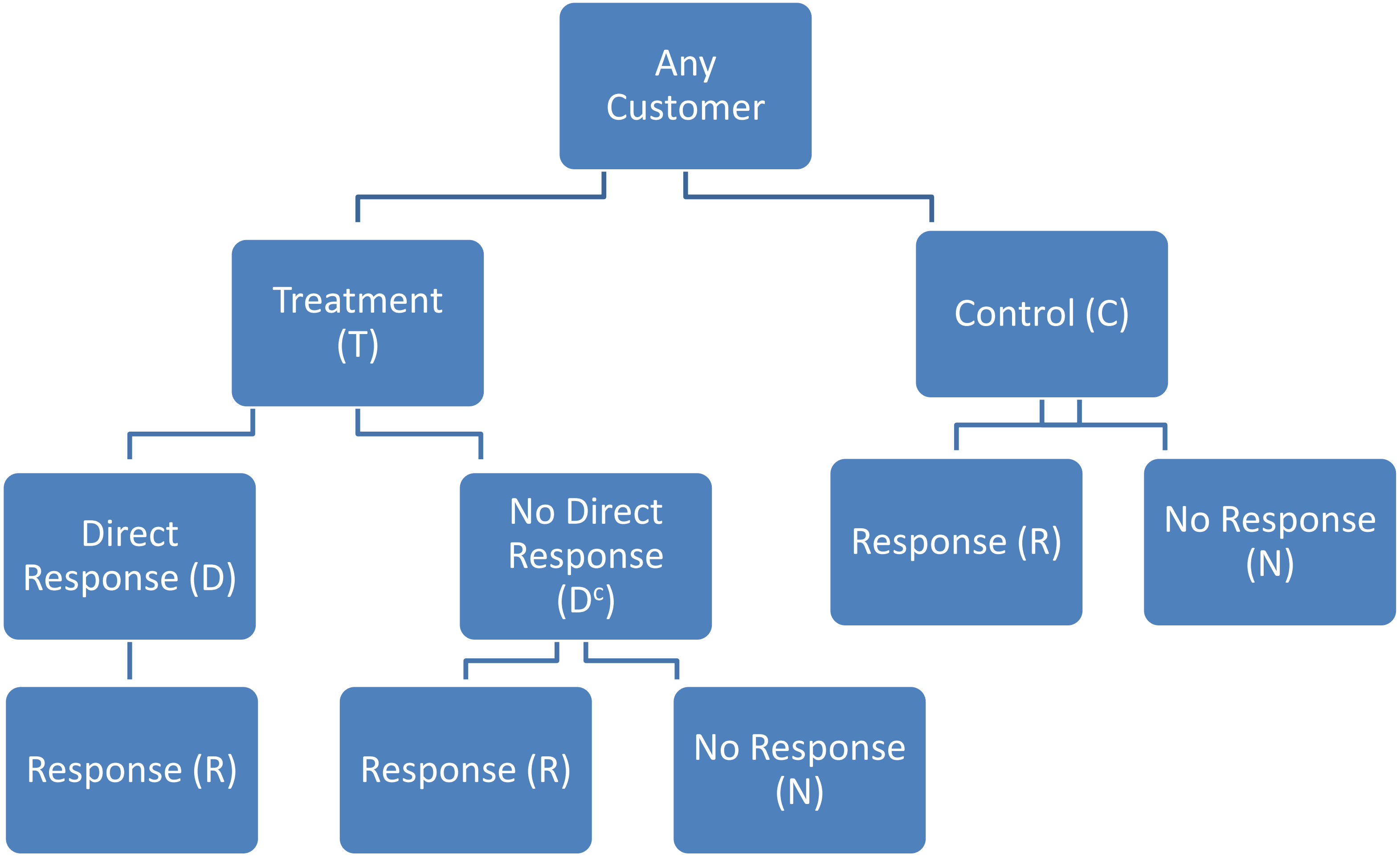
Case II:
Direct Response versus Uplift

Direct Response vs. Uplift Modeling

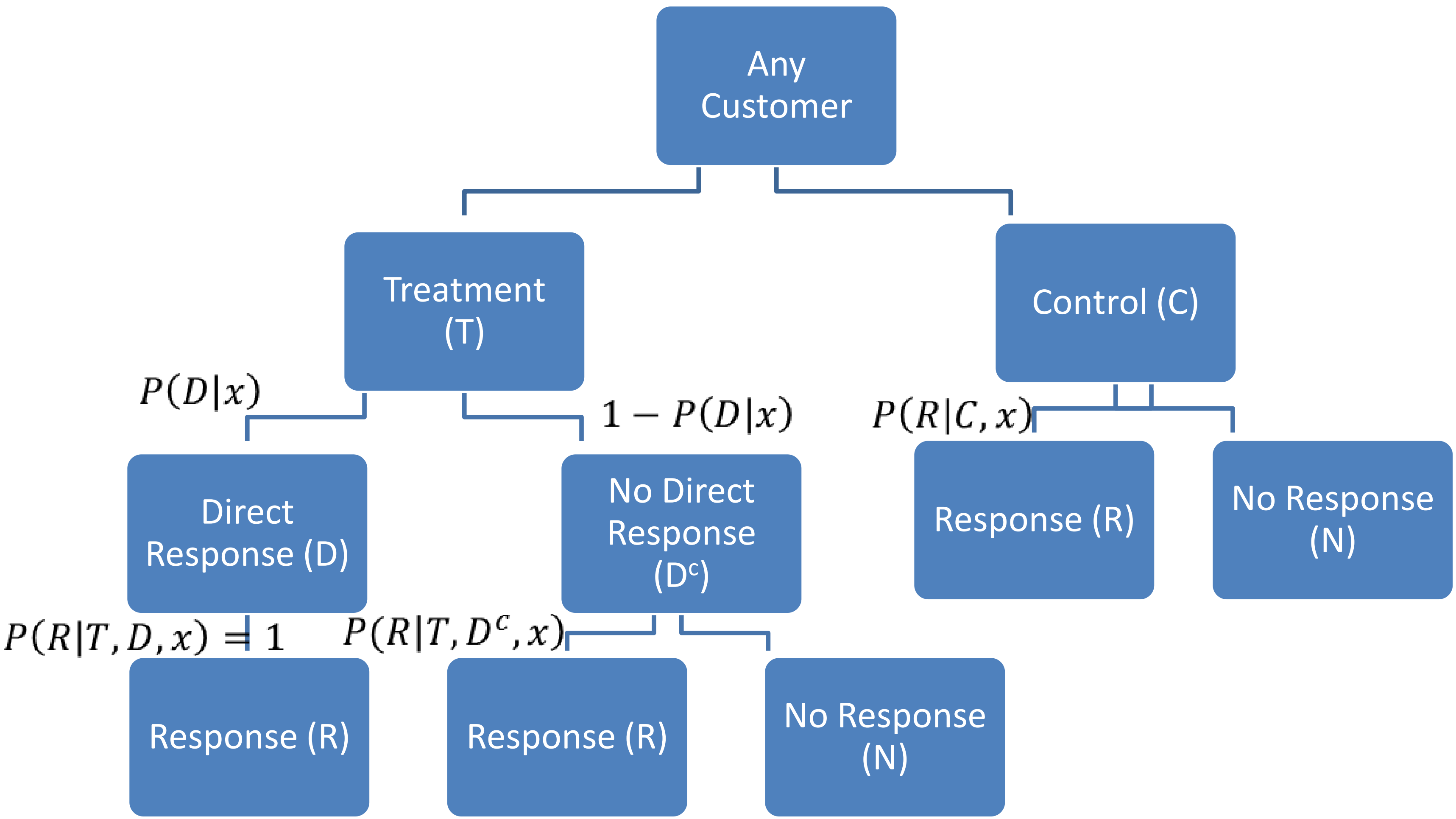
- Retailer couponing

- E-mail click-through

Decision Tree of Campaign and Customers

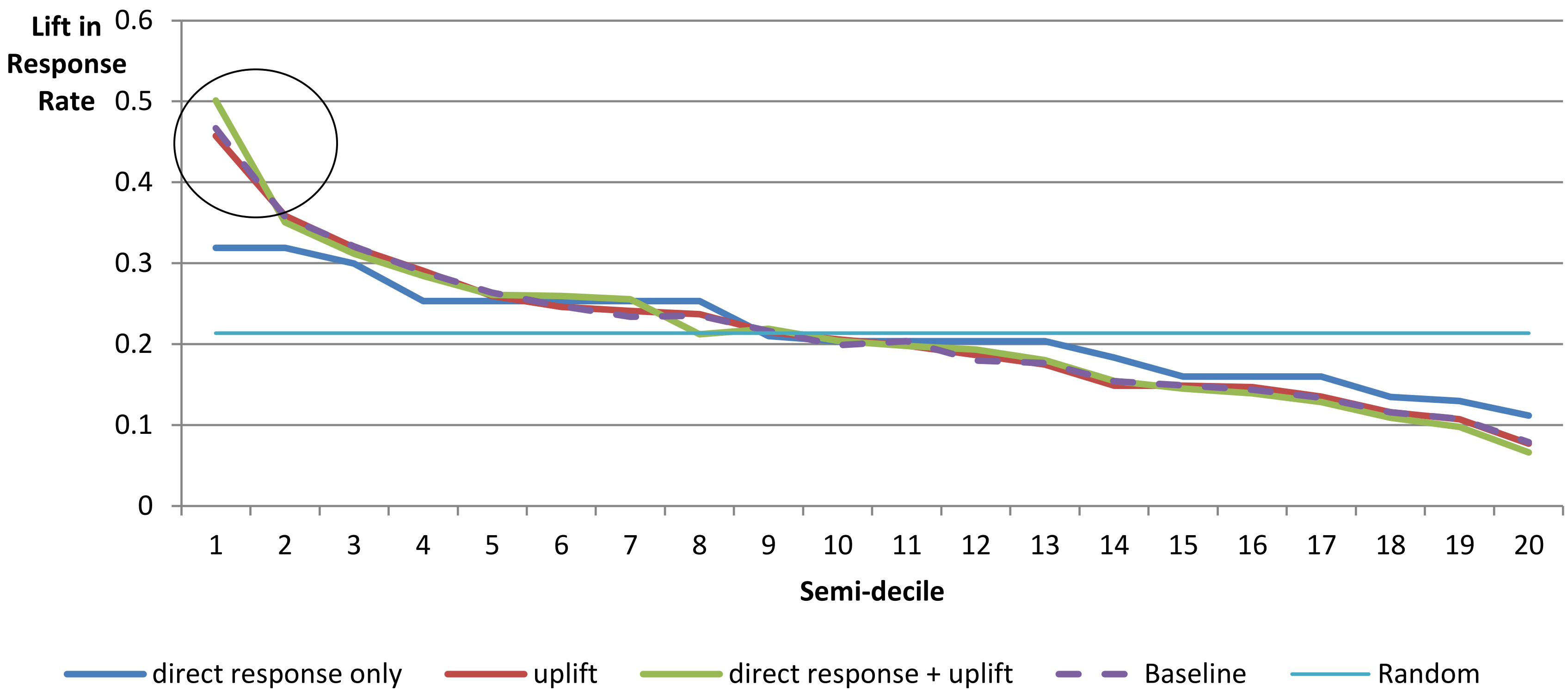


Decision Tree of Campaign and Customers



$$Lift(x) = P(D|x) + P(R|T,D^c,x)\big(1 - P(D|x)\big) - P(R|C,x)$$

Holdout Sample Validation of Simulated Data



	Gini	Gini 15%	Gini repeatability (R^2)
Baseline	2.3282	0.2698	88.8%
Direct response	1.5771	0.1439	95.2%
Uplift	2.3166	0.2644	89.8%
Integrated (Uplift+direct response)	2.4557	0.2896	86.3%

Conclusion

- Uplift is a very impactful emerging subfield
  - Deserves more R&D

- Potential extensions are tremendous:
  - Multiple treatments
  - Optimization
  - Non-randomized experiments
  - Direct tracking

- Applications in other fields

References

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,

Collins, F.S. (2010), *The Language of Life: DNA and the Revolution in Personalized Medicine*, HarperCollins.

Conrady, S. and Jouffe, L. (2011). "Causal Inference and Direct Effects," Bayesia and Conrady Applied Science, at http://www.conradyscience.com/index.php/causality

Freedman, D. (2010). *Statistical Methods and Causal Inference*. Cambridge.

Hamburg, M.A. and Collins, F.S. (2010). "The path to personalized medicine." *The New England Journal of Medicine*, 363;4, p.301-304.

Haughton, D. and Haughton, J. (2011). *Living Standards Analytics*, Springer.

Holland, C. (2005). *Breakthrough Business Results with MVT*, Wiley.

Lai, Lilly Y.-T. (2006) Influential Marketing: A New Direct Marketing Strategy Addressing the Existence of Voluntary Buyers. Master of Science thesis, Simon Fraser University School of Computing Science, Burnaby, BC, Canada.

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.

McKinney, R.E. et al. (1998),"A randomized study of combined zidovudine-lamivudine versus didanosine monotherapy in children with sympotomatic therapy-naïve HIV-1 infection," *J. of Pediatrics,*133, no.4, p.500-508.

Mehr, I.J. (2000), "Pharmacogenomics and Industry Change," *Applied Clinical Trials*, 9, no.5, p.34,36.

Morgan, S.L. and Winship C. (2007). *Counterfactuals and Causal Inference.* Cambridge University Press.

Pearl, J. (2000), *Causality*. Cambridge University Press.

Potter, 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, N.J. and Surry, P. (1999). "Differential response analysis: modeling true response by isolating the effect of a single action," *Proceedings of Credit Scoring and Credit Control VI*, Credit Research Centre, U. of Edinburgh Management School.

Radcliffe, N.J. (2007). "Using Control Groups to Target on Predicted Lift," *DMA Analytics Annual Journal*, Spring, p.14-21.

Robins, J.M. and Hernan, M.A. (2009), "Estimation of the Causal Effects of Time-Varying Exposures," In Fitzmaurice G., Davidian, M,, Verbeke, G., and Molenberghs, G. eds. (2009) *Longitudinal Data Analysis*, Chapman & Hall/CRC, p.553 – 399.

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. (2008), "For Objective Causal Inference, Design Trumps Analysis," *The Annals of Applied Statistics*, p.808-840.

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

Russek-Cohen, E. and Simon, R.M. (1997), "Evaluating treatments when a gender by treatment interaction may exist," *Statistics in Medicine, 16, issue 4, p.455-464.*

Signorovitch, J. (2007), "Estimation and Evaluation of Regression for Patient-Specific Efficacy," Harvard School of Public Health working paper.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd edition, MIT Press.

Wikipedia (2010), "Uplift Modeling," at http://en.wikipedia.org/wiki/Uplift_modelling

Zhao,L., Cai, T., Tian, L., Uno, H., Solomon, S.D., and Wei, L.J. (2010), "Stratifying Subjects for Treatment Selection with Censored Event Time Data from a Comparative Study," Harvard University Biostatistics Working Paper 122.