Causal AI in Business Practices

Victor S.Y. Lo

Presented at:
National Institute of Statistical Sciences (NISS)
Jan 2025

Content

- **Why we care about Causality?**
- **Three Major Schools of Thought**
- **Extension to Uplift Modeling**
- **Summary**

Disclaimer: The views expressed here are solely those of the speaker and do not in any way represent the views of Fidelity Investments

Turing Award and Nobel Econ Prizes on Causality

2010 Turing Award for contributions to AI through a calculus for probabilistic and causal reasoning
https://amturing.acm.org/award_winners/pearl_2658896.cfm

Judea Pearl, UCLA

2019 Nobel Econ for application of **Randomized Experiments** for poverty and health
https://www.nobelprize.org/prizes/economic-sciences/2019/popular-information/

Esther Duflo, MIT; Michael Kremer, Harvard; Abhijit Banerjee, MIT

2021 Nobel Econ for development and applications of **Causal Inference** techniques using **observational data**
https://www.nobelprize.org/uploads/2021/10/advanced-economicsciencesprize2021.pdf

David Card, UC Berkeley

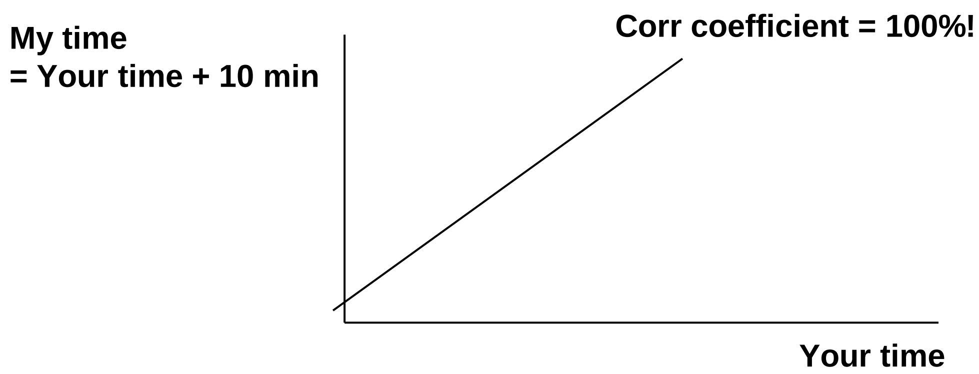
Joshua Angrist, MIT

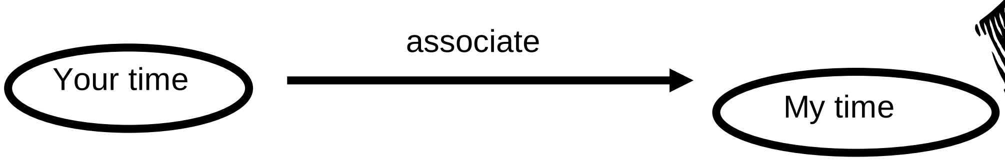
Guido Imbens, Stanford

Can anyone think of a causal question for any industry that can support decision making?

Correlation vs. Causality

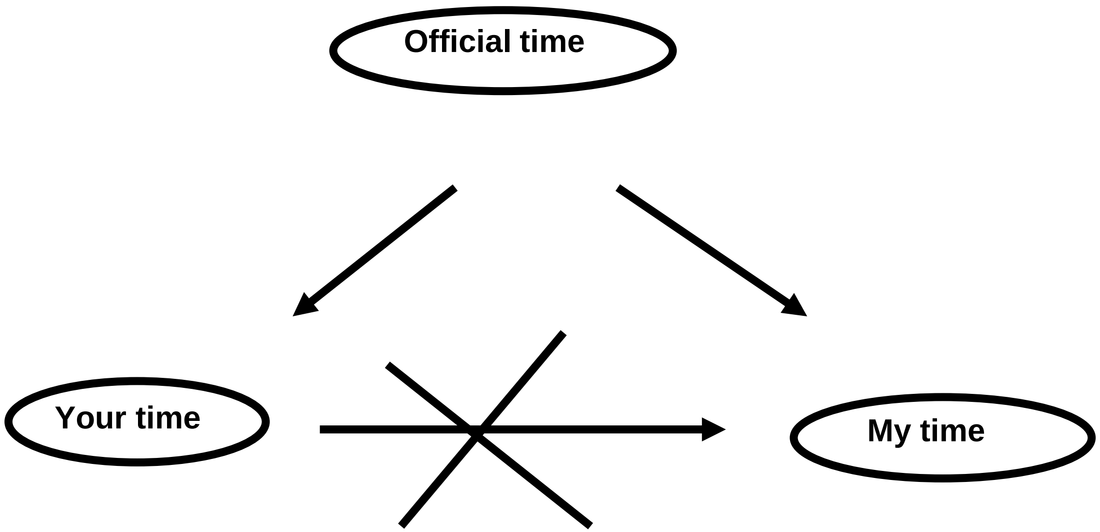
The time on your watch is highly correlated with the time on mine – so the movement on my watch is driven by yours…?





Correlation vs. Causality

Time and time (continued):



$$\text{Your time} \perp\!\!\!\perp \text{My time} \mid \text{Official time}$$

So What's the Point of Understanding Causality?

- **Attribution**
  - Impact of an intervention (e.g., marketing, medicine, product, policy)
  - Counterfactual questions: what-if (e.g., causal fairness)
  - Effects of Causes (EoC)

- **Explanation**
  - Why something happened? What are the possible causes?
  - What's the "causal mechanism"?
  - Causes of Effects (CoE) and Effects of Causes (EoC)

- **Optimization**
  - How can we do better in the future?

Common Causality Related Questions in Business

- **Price: Would a price discount generate high demand?**

- **Promotion: What are the Impact of direct marketing and advertising?**

- **Place: What are the effects of new store location and appearance on business outcomes?**

- **Product: Would an improvement in product feature be valuable to customers?**

“Gold Standard”: Randomized Experiment, A/B Test, or Randomized Controlled Trial (RCT)

“A/B Testing is a **Randomized Experiment** with two variants, A and B. It includes application of statistical hypothesis testing.” Wikipedia

<u>Test two promotional messages</u>

Charles Sanders Peirce

Target Group

Randomization

<u>Treatment</u>
Focus on Durability

<u>Control</u>
Emphasize on multi-functional feature for business & leisure

Ronald A. Fisher

- **Ensures that the characteristics of the treatment and control groups are the same prior to the design**
  - **Balanced** in both ***observable*** and ***unobservable*** characteristics

What if Randomized Experiment is not possible?

►Three Major Schools of Thought:

I. Statistics

II. AI / CS

III. Economics / Econometrics

Rather than directly competing, they share some *similarities* and can mostly *complement* each other…, despite some “philosophical” differences

Donald Rubin, Harvard, Tsinghua, and Temple

I. Statistical School of Thought: Rubin Causal Model (RCM)

- Followed by statisticians, social scientists (e.g., economists, political scientists, sociologists), and medical scientists (e.g., epidemiologists, biostatisticians)

What are Potential Outcomes

Customer Retention Program
Y(C) and Y(T) are potential outcomes for each customer under the scenarios of no call and call, resp.

Customer	Treatment Assignment	Y(C) = Response if not called	Y(T) = Response if called
A	C	0	1
B	C	0	0
C	T	1	1
D	T	0	1

In reality, only one outcome is observable (***Fundamental Problem of Causal Inference***)

Customer	Treatment Assignment	Y(C)	Y(T)
A	C	0	?
B	C	0	?
C	T	?	1
D	T	?	1

Statistical adjustment can be applied through matching or weighting

Blocking the "Back-Door" Path

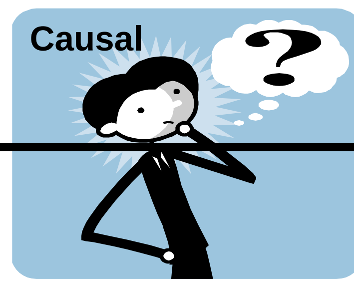
Historical Sales Data

Confounder (if known)

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

Sales reps cherry-picked by demographics

Treatment (or Control) Assignment



Outcome

E.g. Sales Effort

E.g. Buy or not

By <u>breaking the Confounder-Treatment link</u>, we can have a cleaner estimate of **Treatment Effect**, similar situation applies to measuring product value

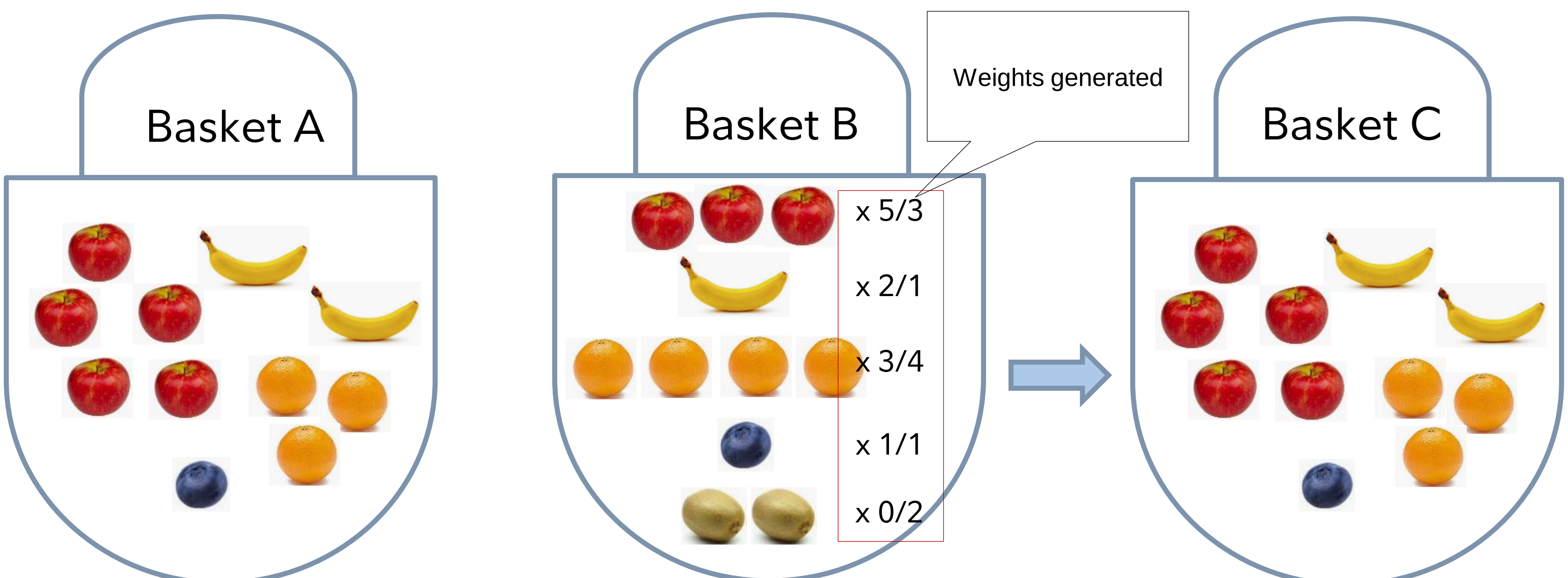
Propensity Score Matching (PSM)

Illustration of Weighting Method

Measures treatment effect by neutralizing audience biases (confounders)

- Baskets A and B are not comparable due to different composition of fruit types
- Construct weights ($\#A_i/\#B_i$ for all i where i represents a fruit type)
- Apply weights to Basket B to form Basket C – now compatible with Basket A
- PSM can handle biases due to *multi-dimensional* variables

Single Variable Illustration



Very similar:
Horvitz-Thompson estimator from survey sampling, Covariate Shift in machine learning, Domain Adaption from transfer learning

Judea Pearl, UCLA
Turing 2011

II. AI/CS School of Thought:
Pearl Causal Model (PCM)

- Followed by a *subset* of computer scientists, data scientists, philosophers, and epidemiologists

Directed Acyclic Graphs (DAG)

- **Path: variables ('vertices') and links ('edges')**
- **No cyclic paths**
  - Given its parents, each variable is independent of its non-parents and non-descendants (Markov condition, Spirtes et al 2000)
  - **Joint distribution of variables is encoded in its DAG**

$$P\left(X_1, \dots, X_p | \Theta\right) = \prod_{i=1}^{p} P(X_i | pa(X_i); \Theta)$$

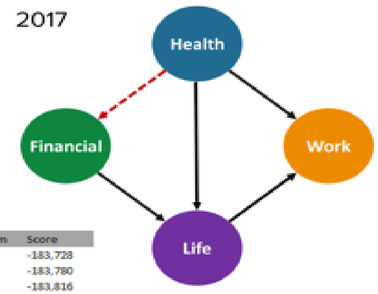
How are we able to come up with a DAG for the FULL dependence structure?

If not, can we "estimate" a DAG based on data?

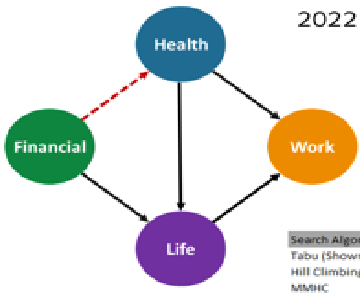
Total Well-Being Analysis

Results

An employer making investments in financial or health benefits could potentially expect increased well-being in Life and Work. The Health and Financial domains influence Life and Work. Work is influenced by all three other domains, but Financial acts on Work through Health and Life in 2022, but just Life in 2017. In both sets of survey data we found the Tabu algorithm produced the best scoring DAG. The arc between Financial and Health reversed in 2022, thebootstrapping results in the bottom figure offer some insight.

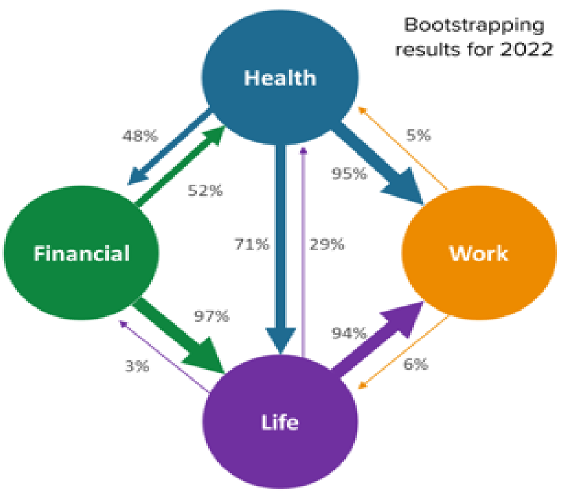


Search Algorithm	Score
Tabu (Shown)	-183,728
Hill Climbing	-183,780
MMHC	-183,816



Search Algorithm	Score
Tabu (Shown)	-144,968
Hill Climbing	-144,970
MMHC	-144,989

The DAG to the right shows how often each arc appeared in each direction in the 500 simulations. We can see that the arc between **Financial and Health nodes is almost evenly split.** Because BNs are acyclic, we can only represent one arc in the final graph but **in the real world these domains may be inter-related.** There is no arc between Financial and Work, suggesting they are conditionally independent given Health and Life.



Source:
Arnost, William and Victor S.Y. Lo (2023), “Prioritizing Workplace Benefits Using Bayesian Networks,” Joint Statistical Meetings, Toronto.

Key Steps for Causal Inference under the Pearl Causal Model

1. **Construct a DAG** as much as possible based on domain expertise
2. Apply **Structure Learning** on a data set to complete the DAG
3. **Identify confounders** for the treatment to outcome relationship of interest
4. **Estimate the treatment effect** by controlling for the identified confounders, using any or multiple techniques

Structure Learning (Causal Discovery)

- **Constraint-based algorithms**
  - Use conditional independence tests to check whether two variables are independent conditional on a third variable
  - Link variables (nodes) that are not found to be independent
  - Classic: the PC algorithm

- **Score-based algorithms**
  - Optimizing a score function, typically BIC:

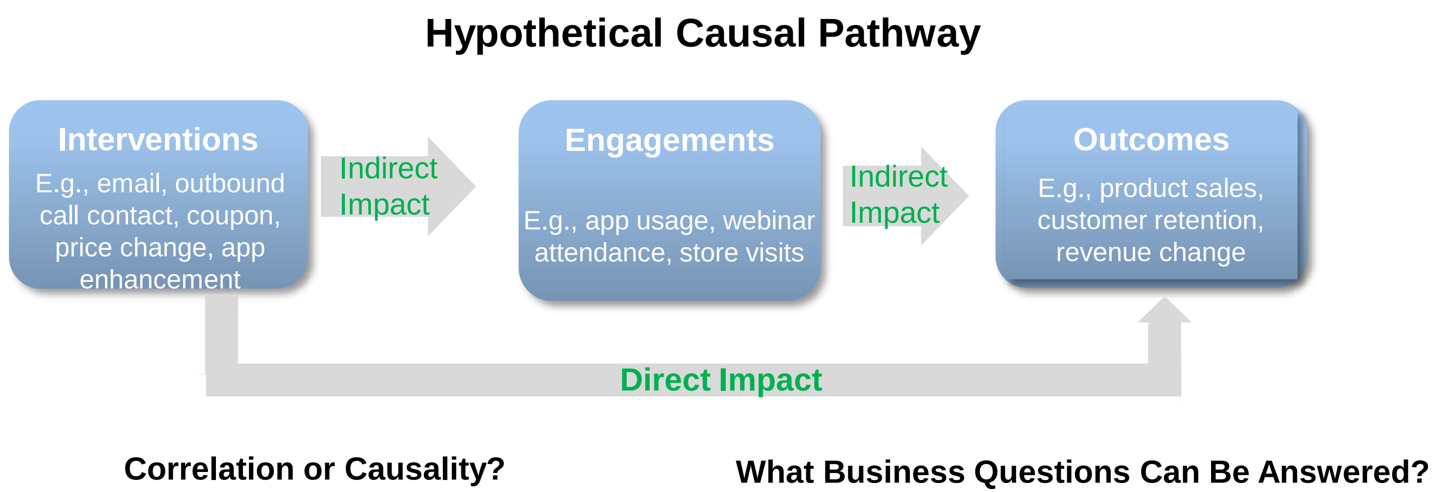
$$BIC = \log \text{likelihood} + \text{penalty term} = \sum_{i=1}^{p} \log P(X_i|pa(X_i);\Theta) - \frac{\#\, parameters}{2} \log n$$

Source: Scutari et al (2018) "Who Learns Better Bayesian Network Structures"

- **Hybrid algorithms**
  - Combine the above

Tools: Tetrad, bnlearn, DoWhy, and more

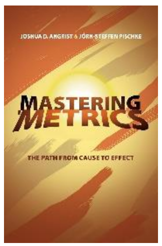
Marketing Funnel Analysis

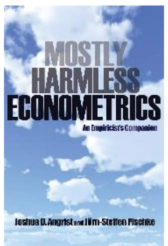


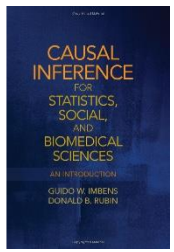
Separating direct from indirect effects enables better optimization

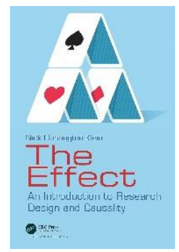
III. Economics School of Thought:

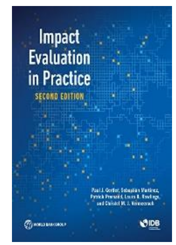
- Difference-In-Difference (DID)
- Synthetic Control Method (SCM)
- Regression Discontinuity Design (RDD)
- Instrumental Variable (IV)

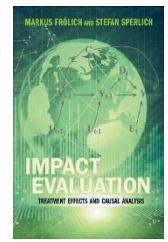


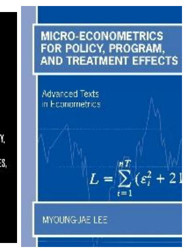


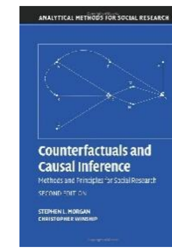












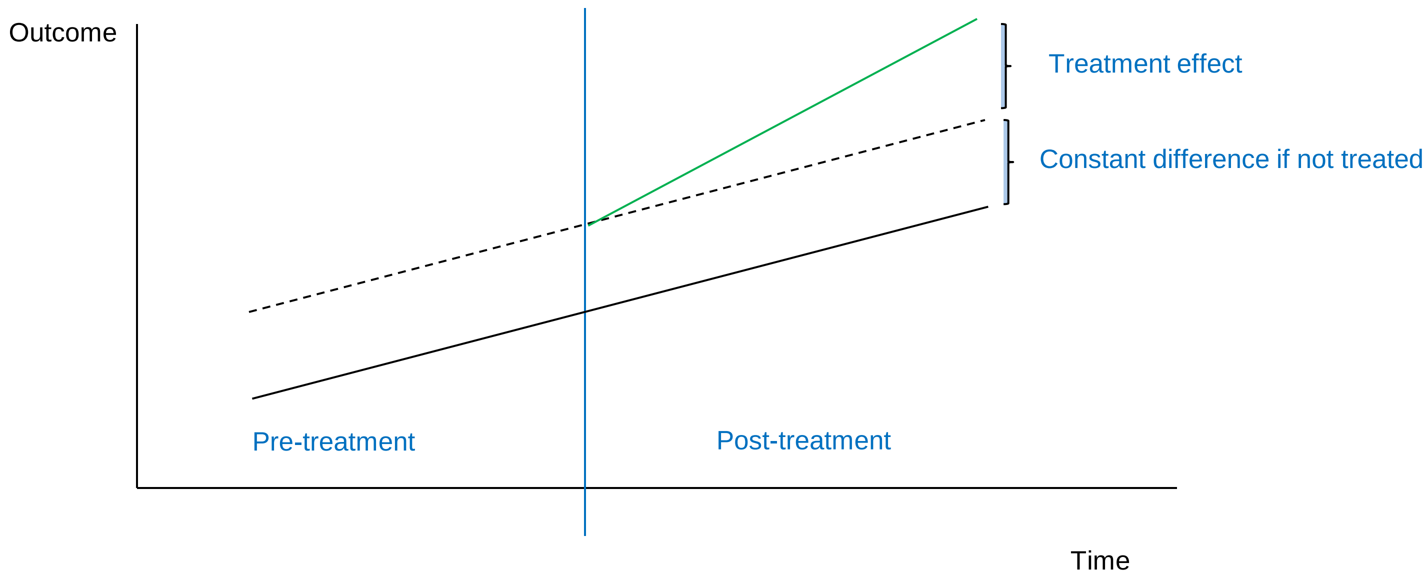
- Embraced by economists and other social scientists

[Story of Imbens, Angrist, and Rubin]

Difference In Difference (DID):

Simplest Causal Inference Method?

Assumption: Parallel Trend



(Treatment_post - Treatment_pre) - (Control_post - Control_pre)

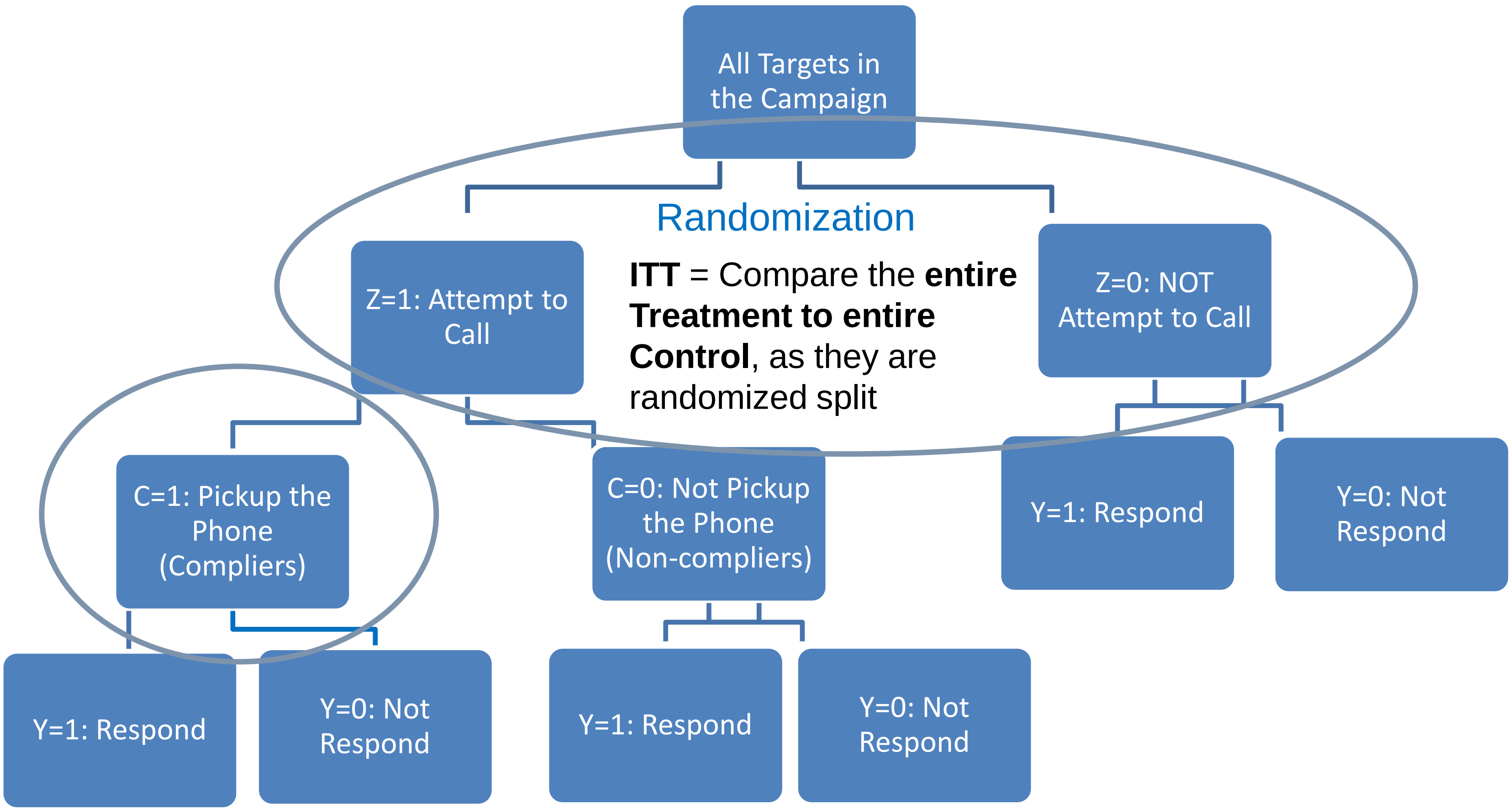
Perhaps the simplest causal technique for model monitoring (without A/B testing)

Extension: Synthetic Control Method (SCM)

- Create a control group that is a weighted average of multiple potential groups

Another: PSM + DID, see Li, Lo, Liu, and Smith (2022) for a health policy application

Decision Tree of Outbound Call Program



Question:

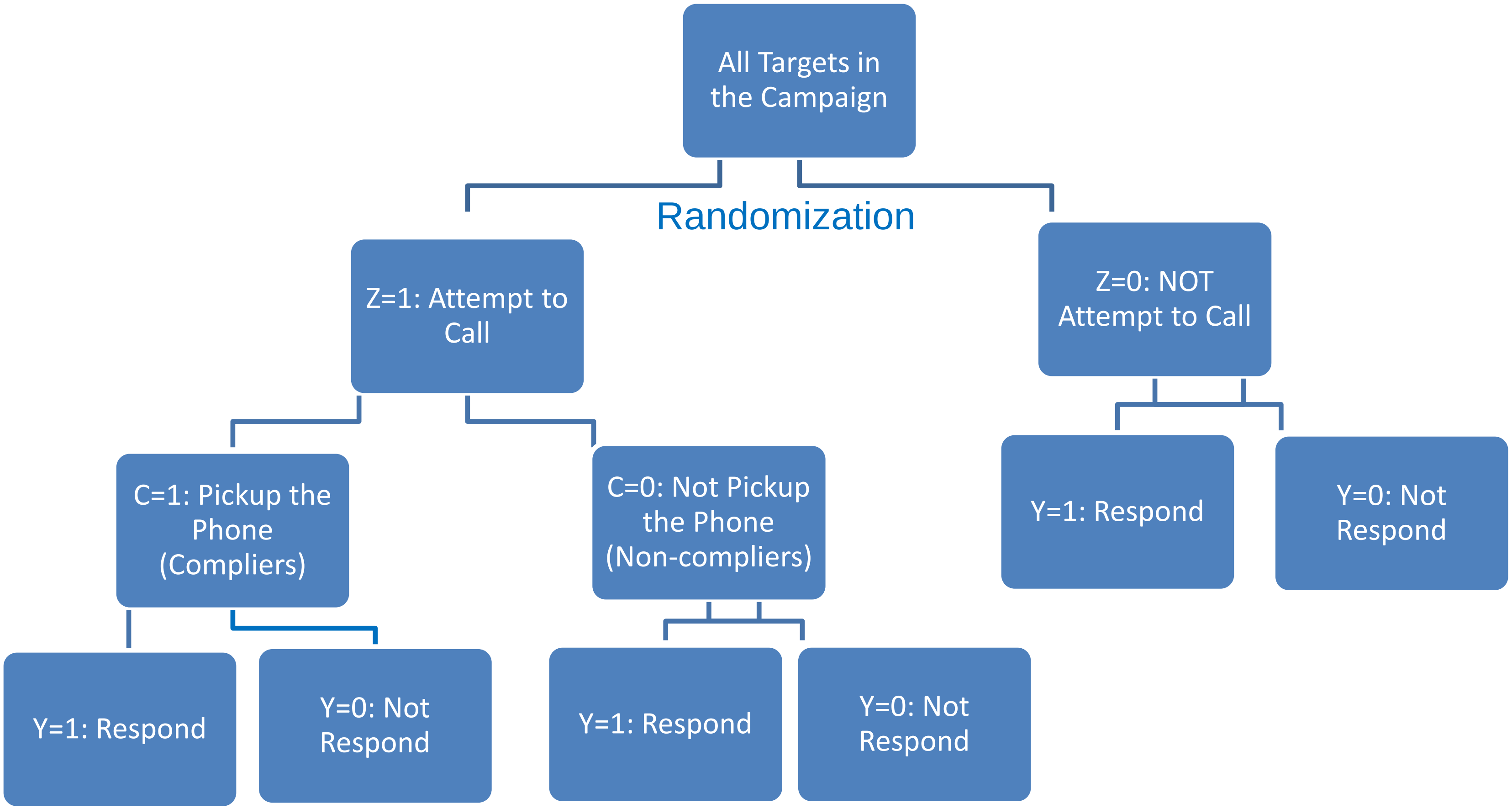
What is the Average Treatment Effect among the ***Compliers***?

I.e., Compare the response rate in C=1 with its *counterfactual* group if they did not pick up the phone

How about using Propensity Score Matching (PSM) to find a group from C=0 that looks like C=1?
What are the confounders?

Source: Lo & Li (2021)

Introducing Instrumental Variable (IV)



Definition of Instrumental Variable (IV):
- ❑ IV -> Treatment (T) -> Outcome (Y)
- ❑ IV does not have any other path to impact Y (Exclusion Restriction)
- ❑ Z is a perfect IV, and C is the actual treatment

See Imbens and Rubin (2015) and the Story of Imbens, Angrist, and Rubin

LATE (Local Average Treatment Effect): Wald Estimator

$$ATE\ (\text{ITT}) = ATE\ (\text{Complier})\ * P(\text{Complier}) + \overbrace{ATE\ (\text{Non}-\text{Complier})}^{0} * P(\text{Non}-\text{Complier})$$

If we assume $ATE\ (\text{Non}-\text{Complier})\ =\ 0$, i.e., non-compliers are NOT impacted by "failed to treat" though attempted, we have the *Wald Estimator*:

$$\boxed{\mathbf{ATE\ (Complier)} = \frac{\mathbf{ATE\ (ITT)}}{\mathbf{P(Complier)}}}$$, known as **LATE** or **CACE** (**Complier Average Causal Effect**).

Technically, it is ATT, average treatment effect on the "treated"

Proportion of Compliers in the "attempt to treat" group

Practical usage of LATE:

1) Direct attribution of effect to the "real treatment"

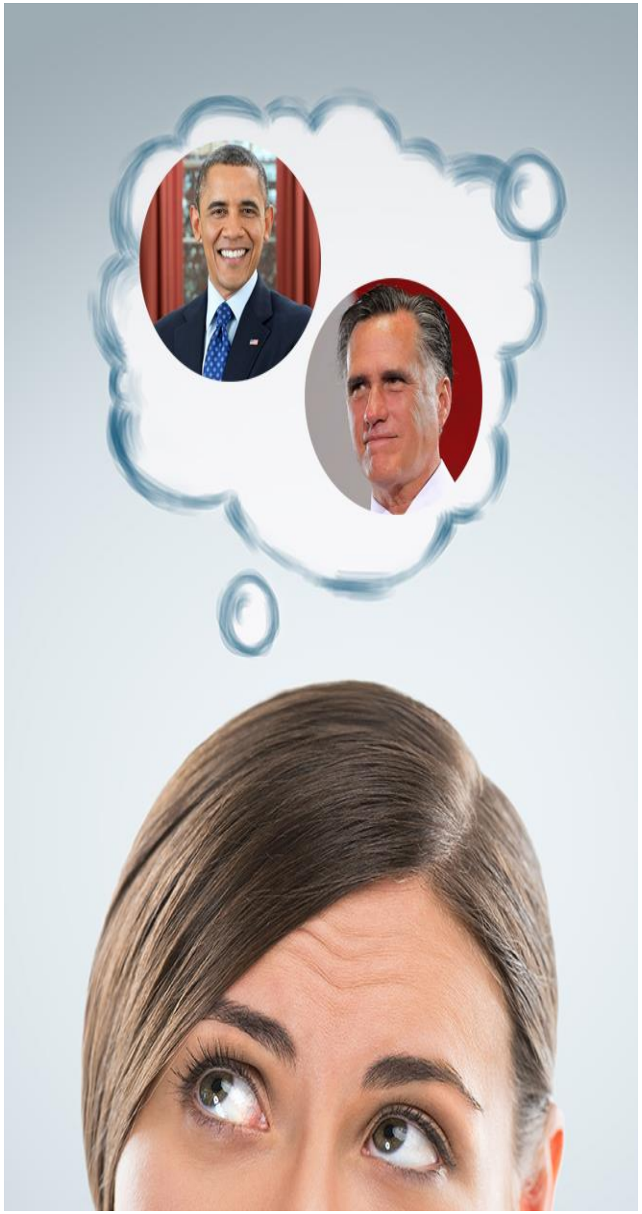
2) If the overall ATE (ITT) is low, is it driven by a) low complier rate or b) low LATE? This enables us to focus on the right component for improvement

See Lo & Li (2021) for further description

▶ Extension to Uplift Modeling or Conditional Average Treatment Effect (CATE)

Uncover the **most responsive** customers (or voters, patients) **to treatment** and **optimize** future targeting

Uplift Modeling: Find The Most Persuadable Voters



Personalized Medicine:

Uplift Modeling to find the right treatment for each patient

		Clinical Benefit achieved if Receiving Placebo or no treatment	
		YES	NO
Clinical Benefit achieved if Receiving Active Treatment	YES	Wasteful [Over-Treat]	Beneficial [Should-Treat]
	NO	Harmful [Do-Not-Treat]	Futile [Do-Not-Treat]

Source: Chapter 3 of [Yong (2015),](#) with permission

Uplift Modeling Software

Open Source or Commercial	Uplift on RCT Data	Uplift on RCT or Observational Data
Open Source: R	• Uplift (Gruelman 2015) – tree, RF • Tools4uplift (2019) – 2-model, interaction • quint (2020) – tree, based on effect size • mr_uplift (2020) – multiple treatments, neural net	• grf (Athey et al 2019 and Tibshirani et al 2023) – transformed outcome, tree, RF, PSM-IPW • rlearner (2020) – meta learner, PSM-IPW
Open Source: Python	• Pylift (Wayfair 2019) – transformed outcome • scikit uplift (2022) – single model, 2-model, transformed outcome	• CausalLift (2019) – 2-model, PSM-IPW • Causal ML (Uber 2023) – tree, RF, meta learners, PSM-IPW • Econ ML (Microsoft 2023) – RF, meta learners, DML
Commercial Software	• JMP Pro: Uplift Model – tree • SAS Enterprise Miner: Incremental Response Model – 2-model, interaction	

Prediction versus Causal AI

(Pure) Prediction Problem	Causal AI
Estimating $E(Y\|T=t)$ **or** $P(Y=1\|T=t)$	**Estimating** $\frac{\partial E(y)}{\partial x}$
Prediction	**Attribution**
Forecasting Y_t	**Explanatory Model**
Classification	**Causal Impact of *Intervention* or *Treatment***
Supervised Learning	$E(Y\|\mathrm{do}(T=1)) - E(Y\|\mathrm{do}(T=0))$
Pattern Recognition	**Direct and Indirect Effects**

"Predictive Modeling" can achieve either or both but the techniques are slightly different

Summary

- Causal AI enables **attribution, explanation,** and **optimization**
- Three key schools of thought: **Statistics, AI, and Economics** – not really competing
- From average treatment effect (**ATE**) for measuring overall impact to conditional average treatment effect (**CATE**) or **uplift modeling for prioritization / optimization**
- **Plenty of business applications**!

References

Ahmed, Arif (2016), *Evidence, Decision and Causality*, Cambridge University Press: Cambridge, UK.

Arnost, William and Victor S.Y. Lo (2023), “Prioritizing Workplace Benefits Using Bayesian Networks,” Joint Statistical Meetings, Toronto.

Bertsimas, Dimitris and Nathan Kallus (2020), “From Predictive to Prescriptive Analytics,” *Management Science*, v.66, 3, p.1025-1044. http://www.p2-analytics.com/papers/PredToPresc.pdf

Dekimpe, M.G. and D.M. Hanssens (2002), “Time Series Models in Marketing: Past, Present, and Future,” *International Journal of Research in Marketing*, p.183-193, at: http://www.anderson.ucla.edu/faculty/dominique.hanssens/content/timeseries.pdf

Efron, Bradley (2020), “Prediction, Estimation, and Attribution,” *Journal of the American Statistical Association*, v.115, no.530, p.646-655.

Elwert, Flix (2013), “Graphical Causal Models ,” In Morgan, S.L. ed. (2013), *Handbook of Causal Analysis for Social Research*, Springer.

Fabozzi, Frank J., Petter N. Kolm, Dessislava A. Pachamanova, and Sergio M. Focardi (2007), *Robust Portfolio Optimization and Management*, Wiley.

Hagmayer, York and Philip M. Fernback (2017), “Causality in Decision Making.” In Waldmann, Michael R. ed. (2017) *The Oxford Handbook of Causal Reasoning*.

Hagmayer, York and Steven A. Sloman (2005), “Causal Models of Decision Making: Choice as Intervention.” In *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society,* Stresa, Italy.

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

Hanssens, D.M., D. Thopre, and C. Finkbeiner (2008), “Marketing when Customer Equity matters,” *Harvard Business Review*, May 2008.

Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer.

Haughton, D., Haughton, J., and Lo, Victor S.Y. (2025, expected) *Cause-and-Effect Business Analytics and Data Science*, CRC/Chapman & Hall.

Joyce, James M. (2008) *The Foundations of Causal Decision Theory*, Cambridge University Press: New York, NY.

Kane, K., Lo, Victor 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.

Li, Mingfei, Victor S.Y. Lo, Piaomu Liu, and Eric Smith (2022), “The Impact of Timothy’s Law on Hospitalization among Mental Health Conditions in New York State,” *International Journal of Mental Health Systems*, v.16, 25.

Lo, Victor 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.

Lo, Victor 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, Victor S.Y., D. Haughton, and J. Haughton (2020), “Essential Data Science for Business: Causal Inference and Uplift Modeling,” *National Institute of Statistical Sciences (NISS) Tutorial*.

Lo, Victor 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.

Lo, Victor S.Y. and D. Pachamanova (2023), “From Meaningful Data Science to Impactful Decisions: The Importance of Being Causally Prescriptive, “ *Data Science Journal*, v.22.

References (continued)

Lo, Victor S.Y. and Zhuang Li (2021) "Causal Inference and Machine Learning Techniques for Outbound Calls Strategy," *Joint Statistical Meetings (JSM) Proceedings*, American Statistical Association.

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

Pachamanova, D., Victor 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.

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

Pearl, Judea (2012), "The Do-Calculus Revisited," In Nando De Freitas and Kevin Murphy (eds.), *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, OR, AUAI Press. https://ftp.cs.ucla.edu/pub/stat_ser/r402.pdf

Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell (2016), *Causal Inference in Statistics: A Primer*, Wiley.

Pearl, Judea and Dana MacKenzie (2018), *The Book of Why: The New Science of Cause and Effect.* Basic Books.

Poornima, S. and M. Pushpalatha (2020), "A Survey on Various Applications of Prescriptive Analytics," *International Journal of Intelligent Networks.*

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.

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

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

Sloman, Steven A. (2005) *Causal Models: How People Think About the World and its Alternatives*. Oxford University Press.

Sloman, Steven A. and York Hagmayer (2006), "The Causal Psycho-Logic of Choice," *TRENDS in Cognitive Sciences*, v.10, no.9, p.407-412.

Spirtes, P., C. Glymour, and R. Scheines (2001), *Causation, Prediction, and Search*. 2nd ed., MIT Press.

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

APPENDIX

Other Variations, Extensions, or Related Techniques of the Statistical School of Thought

- Doubly Robust (DR) Estimation
- Generalized Propensity Score (GPS)
- Target Maximum Likelihood Estimation (TMLE)
- Double Machine Learning (DML)
- Exact Matching and Almost Exact Matching

Other statistical methods for causal inference:

- Regression adjustment
- Time series regression
- Mixed effects models
- Structured equation models

Future Applications?

- More Personalization and Optimization
- More AI-generated contents ⇒ more demand for measurement & optimization
- Impact of AI models (when A/B testing is difficult)
- Causal Fairness in AI ethics: causal effects of sensitive variables on model score
- How about LLM-based causal analysis or causal based LLM?