AI for Economics AI60003

Module 2, Lecture 2

3rd October, 2023

Causality and Attribution

Adway Mitra

Causality between two variables

- Consider two variables X and Y (may be spatio-temporal)
- “X causes Y” = Value of X influences value of Y
- Eg. i) Smoking causes cancer

  ii) Clouds cause rainfall
- Spatial causality: X(s) causes Y(s’) where s, s’ may be same
- Temporal causality: X(t) causes Y(t’) where t’>=t
- Controlled process: if we can externally change the value of X, value of Y will change accordingly.

Bi-directional Causality

- Bi-directional causality: “X causes Y” and “Y causes X”!
- Self-replenishing or self-destructive
- i) High temperature (X) causes water evaporation

  ii) Water evaporation creates clouds

  iii) Clouds cause rainfall (Y)

  iv) Rainfall (Y) brings down temperature! (X)

- i) High temperature (X) -> people use air conditioners

  ii) Air conditioners release CO_2

  iii) CO_2 (Y) causes higher temperature (X)!!!!

Correlation and Causation

X	Y
12	105
25	176
13	109
19	140
23	168
37	225
16	115

Whenever X increases, Y increases too.
Whenever X decreases, Y decreases too.

X	Y
12	153
25	105
13	176
19	109
23	140
37	168
16	225

Whenever X increases, Y increases in next step.
Whenever X decreases, Y decreases in next step

X	Y
12	153
25	105
13	176
19	109
15	125
17	120
16	135

Whenever X in/decreases, Y de/increases.
Whenever Y increases, X increases in next step!

Correlation and Causation

X	Y
12	105
25	176
13	109
19	140
23	168
37	225
16	115

Whenever X increases, Y increases too.
Whenever X decreases, Y decreases too.

High Correlation

X	Y
12	153
25	105
13	176
19	109
23	140
37	168
16	225

Whenever X increases, Y increases in next step.
Whenever X decreases, Y decreases in next step

High lagged Correlation

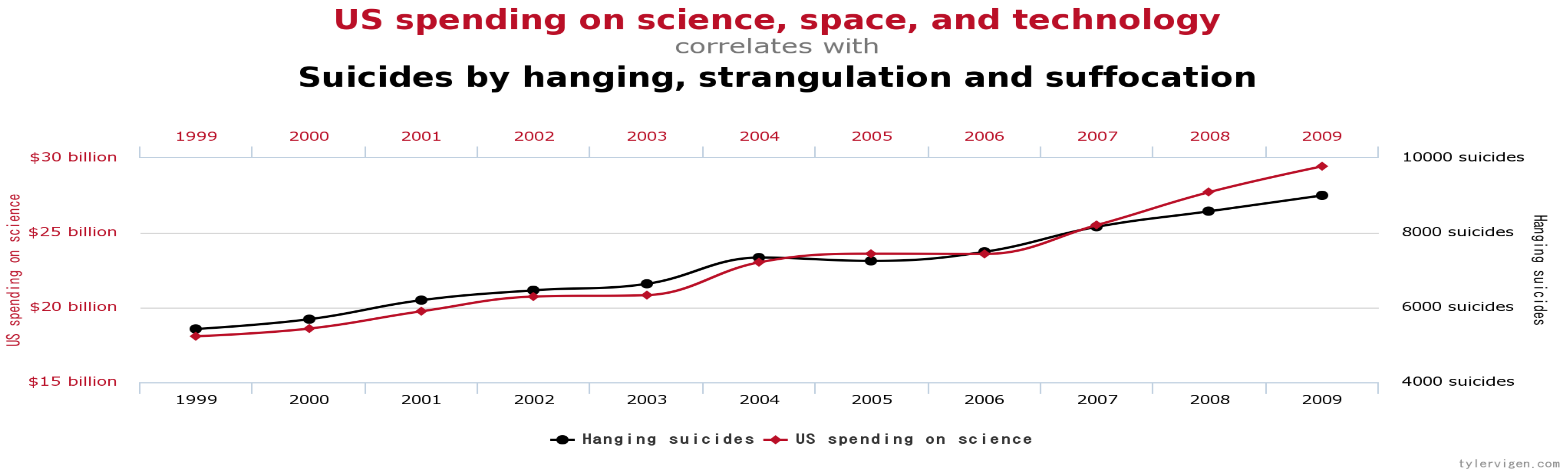
X	Y
12	153
25	105
13	176
19	109
15	125
17	120
16	135

Whenever X in/decreases, Y de/increases.
Whenever Y increases, X increases in next step!

High lagged Correlation, high anti-correlation!

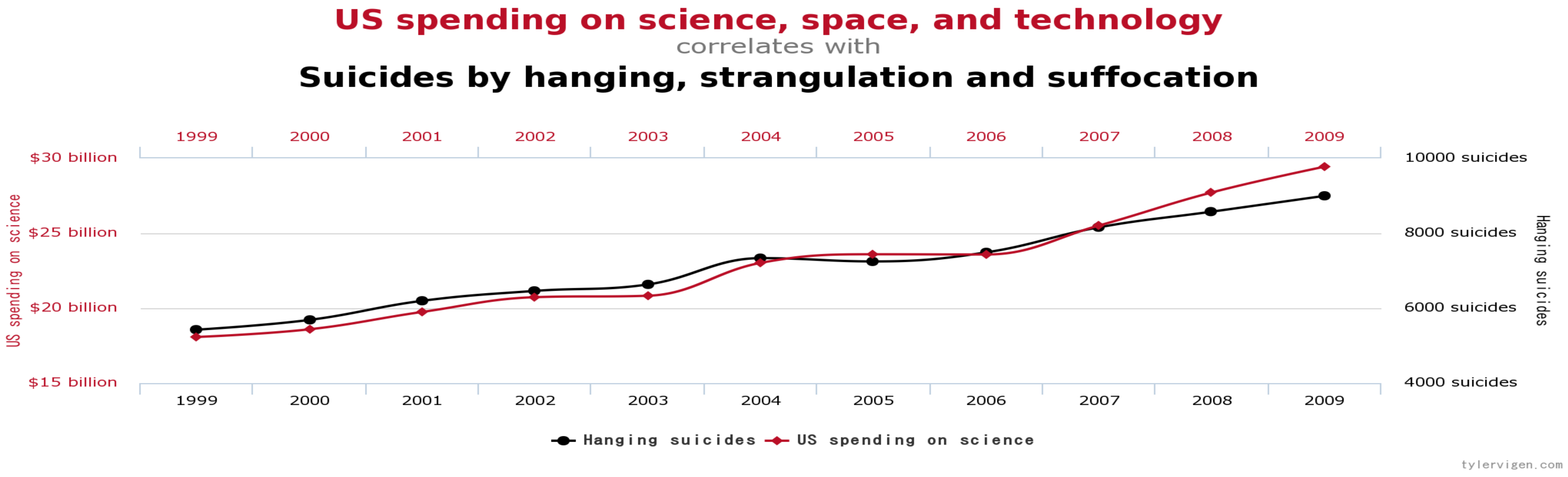
Correlation and Causation

- High correlation need not indicate causal relationship
- “Spurious correlation”: artificially enforced high correlation



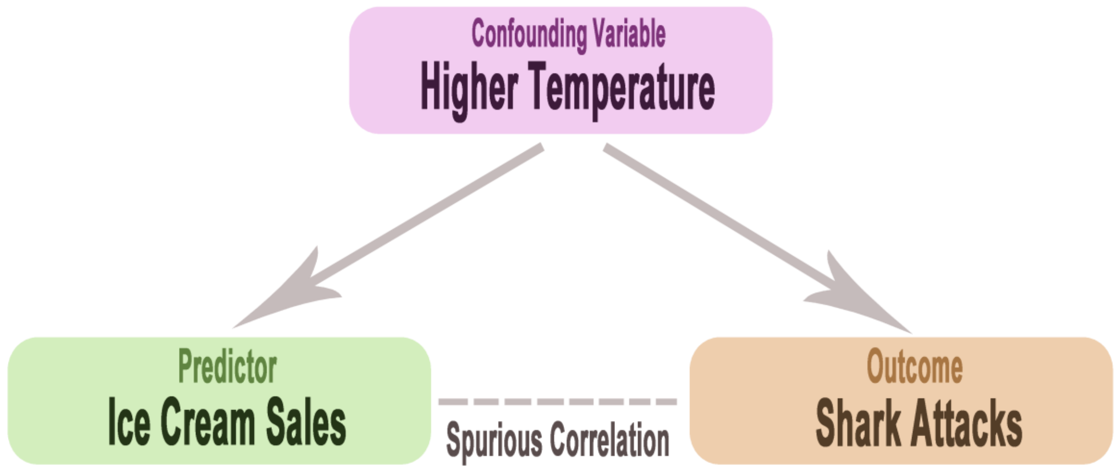
Correlation and Causation

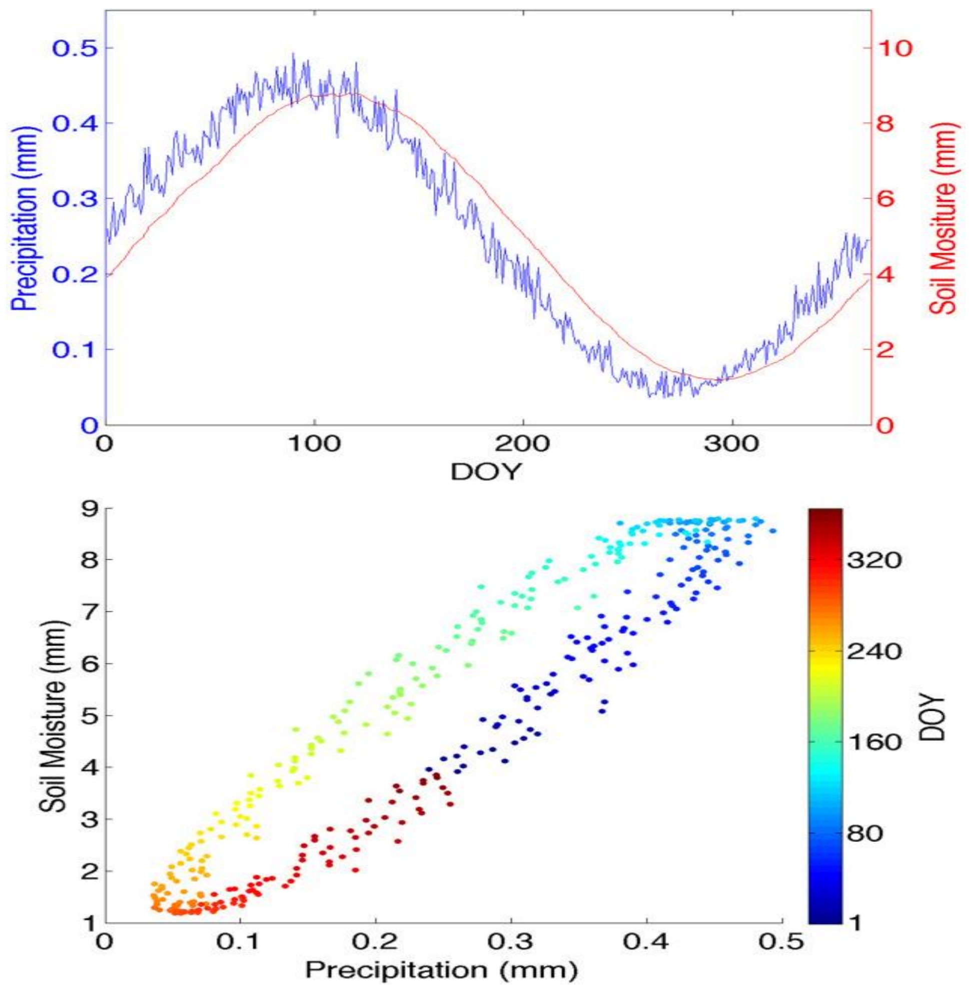
- High correlation need not indicate causal relationship
- “Spurious correlation”: artificially enforced high correlation
- System involves more “confounding” variables which are not observed



Correlation and Causation

- "Explaining away": identify new variable Z which has causal relationship with both X and Y!





Examples in Economics

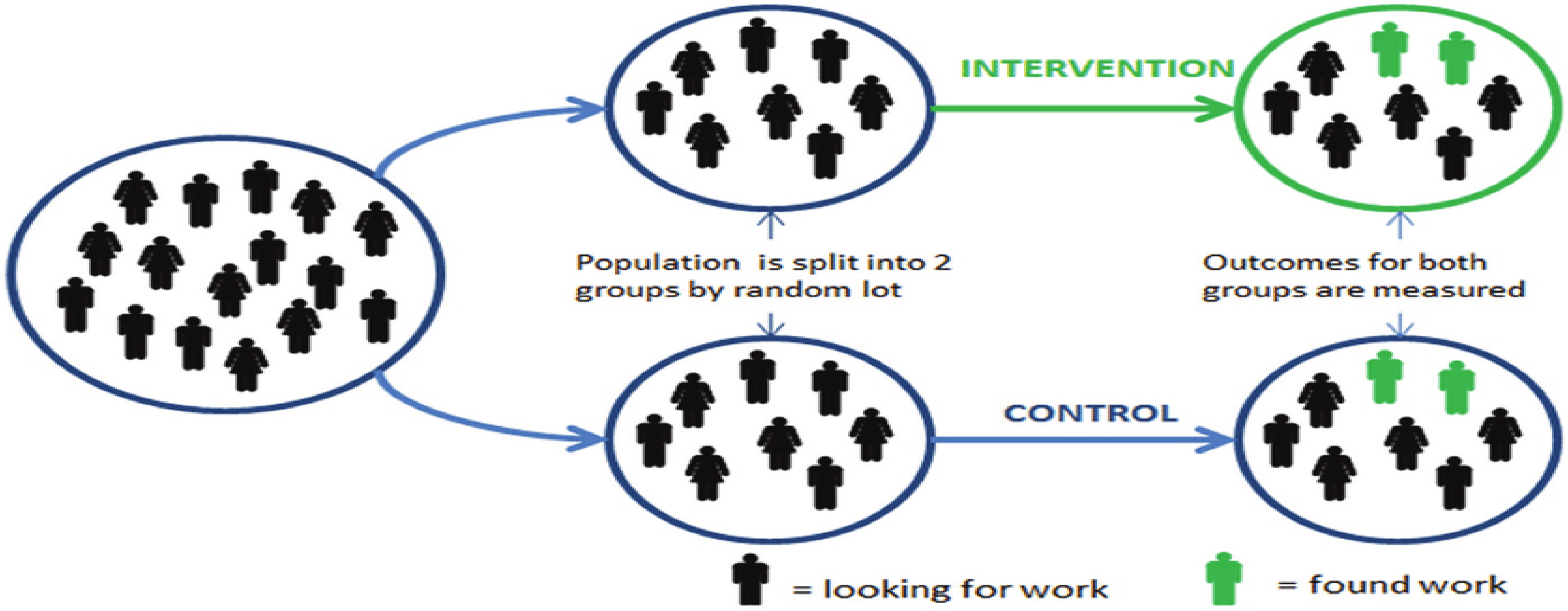
- **How Does Fertilizer Affect Crop Yields?**
- Data: (Fertilizer quantity X, Crop yield Y)
- We may be able to fit linear model: Y(t) = a1X(t-1) + a2X(t-2) + ...
- But what if there are confounders? Eg. soil fertility or weather?

- **How Does Education Affect Income?**
- Data: (Educational degrees and Marks obtained X, Income Y)
- Possible confounders: Parents income! (Rich parents-> better education-> better income)!

- **How does Advertising impact the sales of an item?**
- Data: places/times where/when advertising was higher, saw higher sales of the item
- Possible confounder: interest/relevance of item (eg. no one advertises winterwear in summer!)

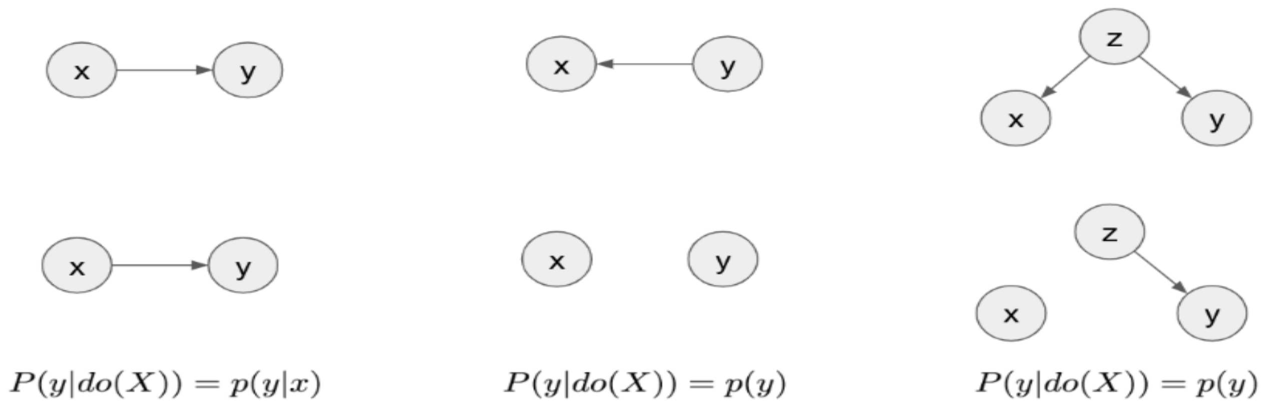
Randomized Control Trials

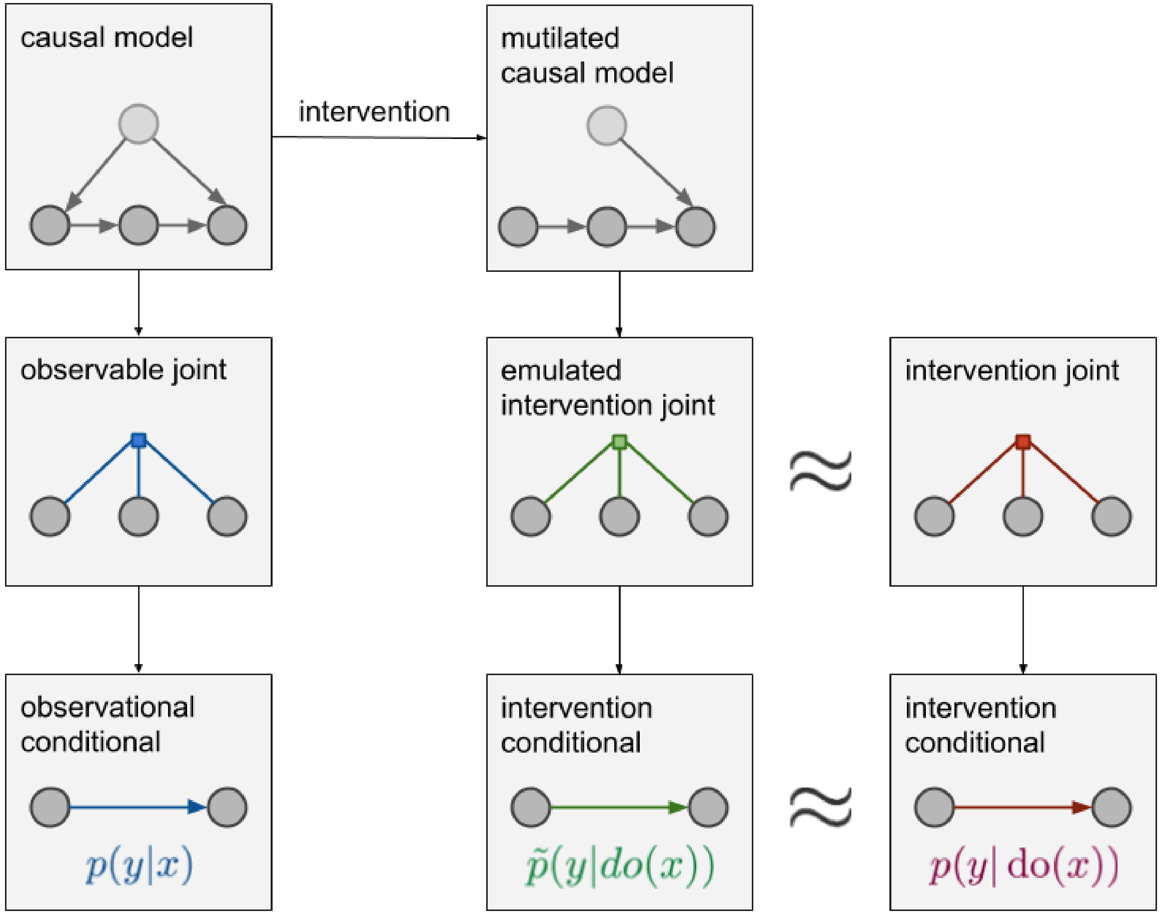
- A principled way to pull out causal effects of a “treatment” that takes “K” values
- Divide the population homogeneously into K parts
- Apply different values of the treatment variable to each part
- Observe the outcome of each part separately



Do-Calculus

- p(Y|X) vs p(Y|do(X))! [observational vs interventional]
- In second case, you FORCE treatment variable to particular value, leaving the rest unchanged!
- Estimating these distributions helps us quantify the impacts of interventions!
- Do-calculus: a set of rules and theorems for such calculations!





Counterfactuals and Predictive Models

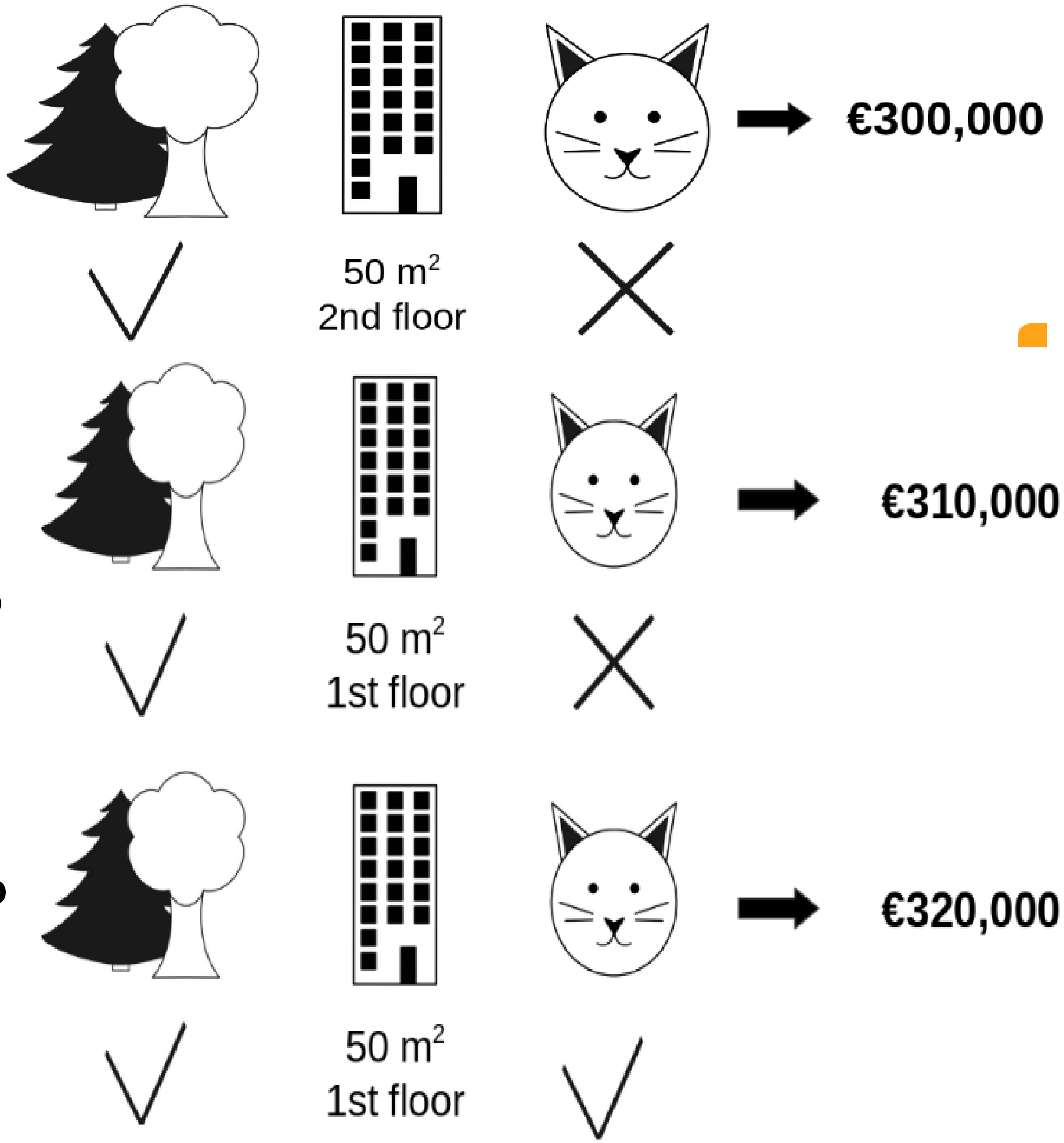
- Not all causality questions can be answered by interventions!
- Observation: rising temperatures all over the world
- Scientist’s Claim: human-induced climate change is the cause of it!
- CC Denier’s Claim: it is part of natural process and fluctuations
- Scientist’s Challenge : if industrialization HAD NOT HAPPENNED, then we would not see such high temperatures today! [COUNTERFACTUAL]
- We have neither any observations, nor any interventions to know counterfactual scenarios
- Possible Solution: Structural Causal Models!
- We use ML-based predictive models as approximations of SCM

Inferring Causality by Predictive Models

- Many possible inputs/predictors -> One outcome
- How to choose “useful” predictors to predict the outcome?
- Useful predictors: cause, outcome: effect!
- Essentially a “feature selection problem” of Machine Learning!
- Machine Learning algorithms:
  - LASSO Linear Regression
  - Decision Trees/Random Forests!
- Struggle in presence of Confounders!
- Solution 1: Feature Importance quantification
- Solution 2: Granger Causality (for time-series)
- Solution 3: Double Machine Learning (DML)

Feature Contribution

- For different combinations of feature values, we get different outputs
- Which features cause deviation from "mean" outcome for a specific datapoint?
- Can we "quantify" the contribution of different features in biasing the outcome?
- Contributions: positive or negative!



Christopher Molnar, Interpretable Machine Learning

Shapley Value Analysis

- For linear regression, $y_i = w_0 + w_1X_{i1} + \ldots\ldots + w_DX_{iD}$
- Contribution of feature ‘j’ on the output’s deviation: $w_jX_{ij} – w_jE(X_j)$
- Not all relations are linear!
- Alternative models: nonlinear y = f(w.x)
- f: can be a neural network (any function can be represented by NN)
- Other models: Decision Tree/Random Forest/SVM
- How to quantify feature quantifications in case of other predictive models?

Shapley Value Analysis

- How to quantify in case of other predictive models?
- Shapley Value: Concept from Game Theory!
- Aims to quantify the “contribution” of each member in a “team”

Bowler	Average	Economy	Strike Rate
Shami	25.5	5.57	27.4
Bumrah	24.31	4.67	31.2
Siraj	20	4.87	24.6
Shardul	30.34	6.24	29.1

Career Record (Expected case)

Bowler	Overs	Runs	Wickets
Shami	10	48 (-)	2 (.)
Bumrah	8	51 (+)	3 (+)
Siraj	7	23 (-)	0 (-)
Shardul	6	44 (+)	1 (.)

Match Performance

Shapley Value Analysis

- Shapley Value: Concept from Game Theory
- Aims to quantify the "contribution" of each member in a "team"
- Each member can push the total team output
- How would the team have done if that member didn't join?

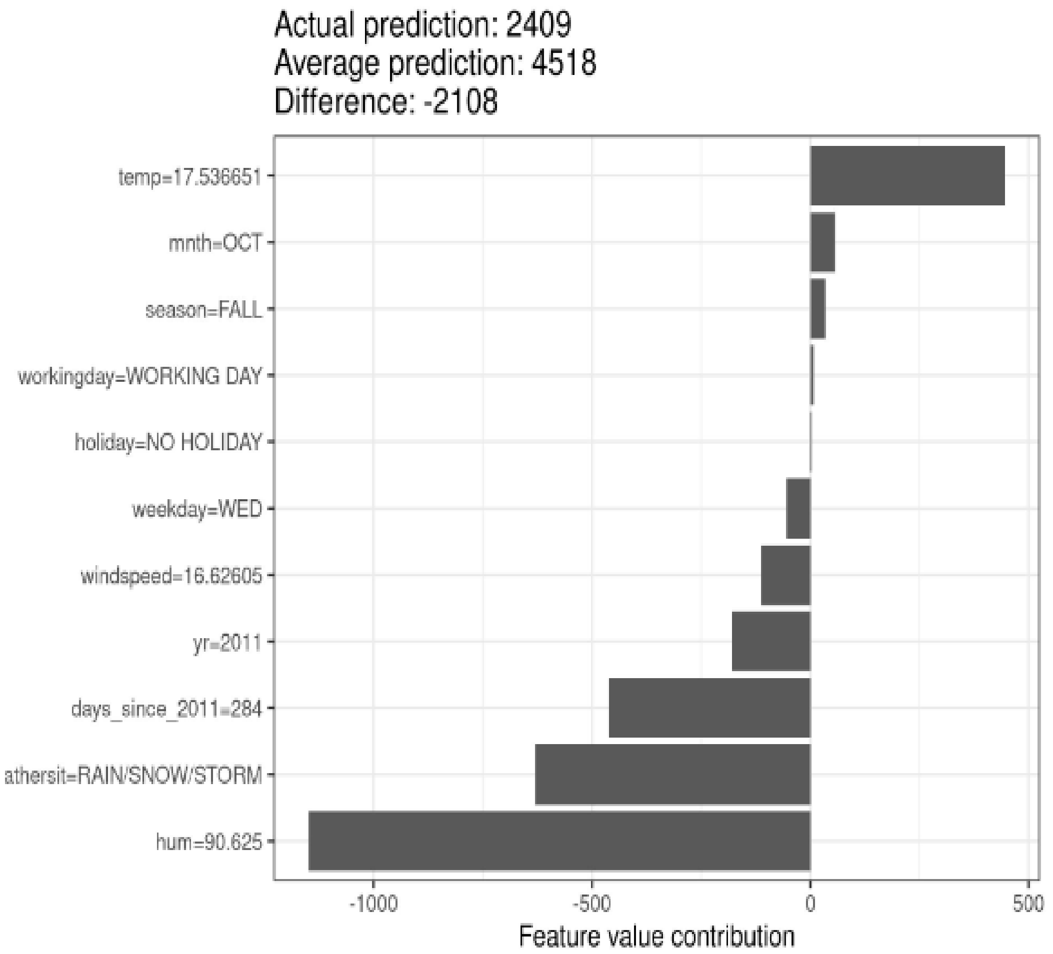
$$\phi_j(val) = \sum_{S \subseteq \{1,\ldots,p\} \setminus \{j\}} \frac{|S|!\,(p-|S|-1)!}{p!}\left(val\left(S \cup \{j\}\right) - val(S)\right) \qquad val_x(S) = \int \hat{f}(x_1,\ldots,x_p)\,d\mathbb{P}_{x \notin S} - E_X(\hat{f}(X))$$

- Consider each feature as a member of a team!
- Team output: $y_i - E(y_i)$ where $y_i = f(x_{i1}, \ldots, x_{iD})$
- Problem: too many feature combinations, integration intractable!

Shapley Value Computation

Approximate Shapley estimation for single feature value:

- Output: Shapley value for the value of the j-th feature
- Required: Number of iterations M, instance of interest x, feature index j, data matrix X, and machine learning model f
  - For all m = 1,...,M:
    - Draw random instance z from the data matrix X
    - Choose a random permutation o of the feature values
    - Order instance x: $x_o = (x_{(1)}, \ldots, x_{(j)}, \ldots, x_{(p)})$
    - Order instance z: $z_o = (z_{(1)}, \ldots, z_{(j)}, \ldots, z_{(p)})$
    - Construct two new instances
      - With j: $x_{+j} = (x_{(1)}, \ldots, x_{(j-1)}, x_{(j)}, z_{(j+1)}, \ldots, z_{(p)})$
      - Without j: $x_{-j} = (x_{(1)}, \ldots, x_{(j-1)}, z_{(j)}, z_{(j+1)}, \ldots, z_{(p)})$
    - Compute marginal contribution: $\phi_j^m = \hat{f}(x_{+j}) - \hat{f}(x_{-j})$
- Compute Shapley value as the average: $\phi_j(x) = \frac{1}{M}\sum_{m=1}^{M} \phi_j^m$



ɌE 9.21: Shapley values for day 285. With a predicted 2409 rental bikes, this day is -2108 below the age prediction of 4518. The weather situation and humidity had the largest negative contributions. emperature on this day had a positive contribution. The sum of Shapley values yields the difference of actual and average prediction (-2108).

Christopher Molnar, Interpretable Machine Learning

Granger Causality in Time-series

- Express X as a linear function of past values of itself
- $X(t) \sim a_1X(t-1) + a_2X(t-2) + a_3X(t-3) + \ldots\ldots$
- Does the estimate improve if we include past values of Y?
- $X(t) \sim a_1X(t-1) + a_2X(t-2) + a_3X(t-3) + \ldots\ldots + b_1Y(t-1) + b_2Y(t-2) + b_3Y(t-3) + \ldots\ldots$
- If yes, then Y Granger-causes X.

Granger Causality in Time-series

- Express X as a linear function of past values of itself
- $X(t) \sim a_1X(t-1) + a_2X(t-2) + a_3X(t-3) + \ldots\ldots$
- Does the estimate improve if we include past values of Y?
- $X(t) \sim a_1X(t-1) + a_2X(t-2) + a_3X(t-3) + \ldots\ldots + b_1Y(t-1) + b_2Y(t-2) + b_3Y(t-3) + \ldots\ldots$
- If yes, then Y Granger-causes X.
- Does X Granger-cause Y?
- $Y(t) \sim c_1Y(t-1) + c_2Y(t-2) + c_3Y(t-3) + \ldots\ldots$
- $Y(t) \sim c_1Y(t-1) + c_2Y(t-2) + c_3Y(t-3) + \ldots\ldots + b_1X(t-1) + b_2X(t-2) + b_3X(t-3) + \ldots\ldots$
- If both yes, then bidirectional causality!

Causality in Time-Series

A Granger Causality Analysis between the GDP and CO_2 Emissions of Major Emitters and Implications for International Climate Governance

This paper empirically examines the relationship between carbon emissions and economic growth by applying the co-integration analysis and Granger causality test to the time series data of carbon emissions and gross domestic product (GDP) of the world's top 20 emitters from 1990 to 2015. Co-integration analysis shows that there is a long-term equilibrium relationship between carbon emissions and economic growth in most countries; Granger causality test verifies a one-way causal link between carbon emissions and economic growth in most major emitters. In developed countries, economic growth is the Granger cause of carbon emissions, while the opposite is true in developing countries. The results reflect different characteristics regarding carbon emission reduction in developed and developing countries as they are at different developing stages. Carbon emission reduction exerts much greater adverse effects on the economic growth of developing countries than it does on that of developed countries. Based on the results of the Granger causal analysis, it is found that the requirements for developing countries to substantially reduce emissions are not in line with the characteristics in their current developing stage and therefore may pose

Impact of FDI on GDP per capita in India using Granger causality

Anand Nadar

Abstract

This research investigates the causality between FDI and GDP per capital in the context of India. Using WDI data from 1970-2019, We applied two types of Granger causality tests: long-run causality and short-run causality tests. For the long-run causality, we applied pairwise Granger causality test, and for short-run, we performed the Wald test approach under VECM (Vector Error Correction Model). The long-run causality test indicates that there is a unidirectional causality running from FDI to GDP per capita, implying that FDI causes the GDP per capita to change and not vice-versa. The short-run causality test indicates that there is no causality between FDI and GDP per capita, suggesting that, in the short-run, FDI and GDP per capita does not cause each other. The central policy conclusion from this study is that although FDI does not cause GDP per capita in the short-run, it causes in the long-run. Therefore, according to our study, India should attract FDI to sustain a long-run growth of GDP per capita.

Keywords: GDP per capita, Granger causality, FDI, India, VECM,

Sparse Regression for Feature Selection

- Case 1: we want "w" such that most of its elements are small
- Case 2: we want "w" such that most of its elements are 0
- Can we convert these demands into mathematical formulations?
- General recipe: find a regularization function f(w)
- f(w) should have low value for suitable "w", high value for unsuitable "w"

- Find (w,b) to minimize $L(w,b) + \lambda f(w)$
- First term to find w that fits data, second term to find "w" that is suitable, λ to balance them!

Double LASSO/Double ML

- Main problem with LASSO: confounders! (some of the predictors may influence other predictors!)
- Possible solution:

  1) Choose huge pool of predictors, fix a predictor as “treatment”

  2) Use LASSO to identify predictors including treatment

  3) Try to identify the predictors that influence the treatment with LASSO

$$
\begin{aligned}
y_i &= d_i\alpha_0 + \underbrace{x_i'\beta_{g0} + r_{gi}}_{g(z_i)} + \xi_i \\
d_i &= \underbrace{x_i'\beta_{m0} + r_{mi}}_{m(z_i)} + v_i
\end{aligned}
$$

- Identifiability issues arise in LASSO, unable to differentiate between intervention and confounder!!

The Frisch–Waugh–Lovell Theorem

- Let's say we want to estimate the following model using OLS:
  - $Y = \beta_0 + \beta_1 D + \beta_2 X + U$
- The Frisch–Waugh–Lovell Theorem shows us that we can recover the OLS estimate of β_1 using a residuals-on-residuals OLS regression:
  1. Regress D on X using OLS
     - Let $\widehat{D}$ be the predicted values of D and let the residuals $\widehat{V} = D - \widehat{D}$
  2. Regress Y on X using OLS
     - Let $\widehat{Y}$ be the predicted values of Y and let the residuals $\widehat{W} = Y - \widehat{Y}$
  3. Regress $\widehat{W}$ on $\widehat{V}$ using OLS
- The estimated coefficient on $\widehat{V}$ will be the same as the estimated coefficient $\hat{\beta}_1$ from regressing Y on D and X using OLS!