ECONOMETRICS FOR BUSINESS IN R AND PYTHON

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

How we are going to tackle each concept

What?	Why?	How?
Step 1	Step 2	Step 3
Use Cases	Intuition tutorial	Practice tutorial in R and Python

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Econometrics for Business in R and Python agenda

1. Difference-in-differences
2. Google‘s Causal Impact
3. Granger Causality
4. Propensity Score Matching
5. CHAID

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Difference-in-differences

0 Introduction

Use Cases

- Policy changes in countries or regions
- Impact of weather on sales
- Impact of M&A
- Geotests in marketing

Intuition tutorial

- Difference-in-differences framework
- Parallel trends assumptions and confounding policy change
- Linear Regression
- Logistic Regression
- Dummy variable trap
- Statistical Significance

Practice tutorial

- Take care of missing data
- Linear and logistic regression
- Present regression results (in R only)

Google's Causal Impact

0 Introduction

Use Cases	Intuition tutorial	Practice tutorial
• Policy changes in countries or regions • Impact of weather on sales • Impact of M&A • Geotests in marketing	• Causal Impact Framework • Value added of Causal Impact	• Load financial data • Basic Plotting • Correlations

Granger Causality

0 **Introduction**

Use Cases	Intuition tutorial	Practice tutorial
• Impact of economic drivers • Influencer marketing • Financial markets	• Granger Causality framework • Difference between correlation and causation • Stationarity	• Create Stationarity data • Plot time series • Apply Granger Causality • Create Loops (R only)

Propensity Score Matching

0 **Introduction**

Use Cases

- Referral Programs
- Mobile shopping
- New website languages
- People analytics

Intuition tutorial

- Propensity Score Matching framework
- Unconfoundness and Common Support Region
- T-tests

Practice tutorial

- Create segment summary statistics
- Apply t-tests to several variables at once
- Assess accuracy
- Plot Common Support Region
- Do Propensity Score Matching

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

CHAID

0 Introduction

Use Cases	Intuition tutorial	Practice tutorial

Use Cases

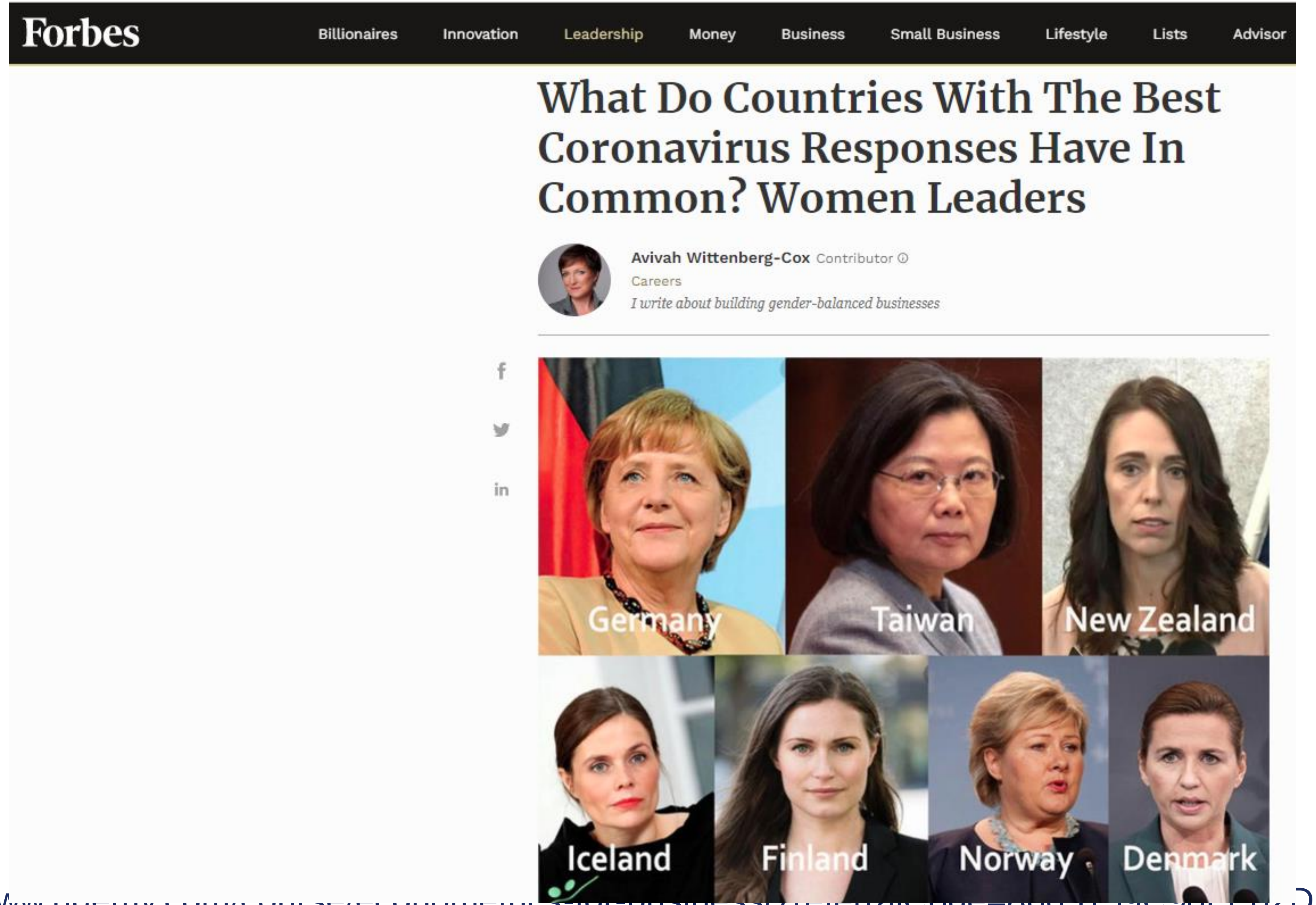
- Direct Marketing
- Customer Segmentation
- Customer satisfaction
- Employee Satisfaction

Intuition tutorial

- CHAID framework
- Confusion Matrix

Practice tutorial

- Create dataset based on data types
- Do and plot CHAID
- Create factors out of numerical variables
- Create density plots



Forbes

Billionaires Innovation Leadership Money Business Small Business Lifestyle Lists Advisor

What Do Countries With The Best Coronavirus Responses Have In Common? Women Leaders

Avivah Wittenberg-Cox Contributor

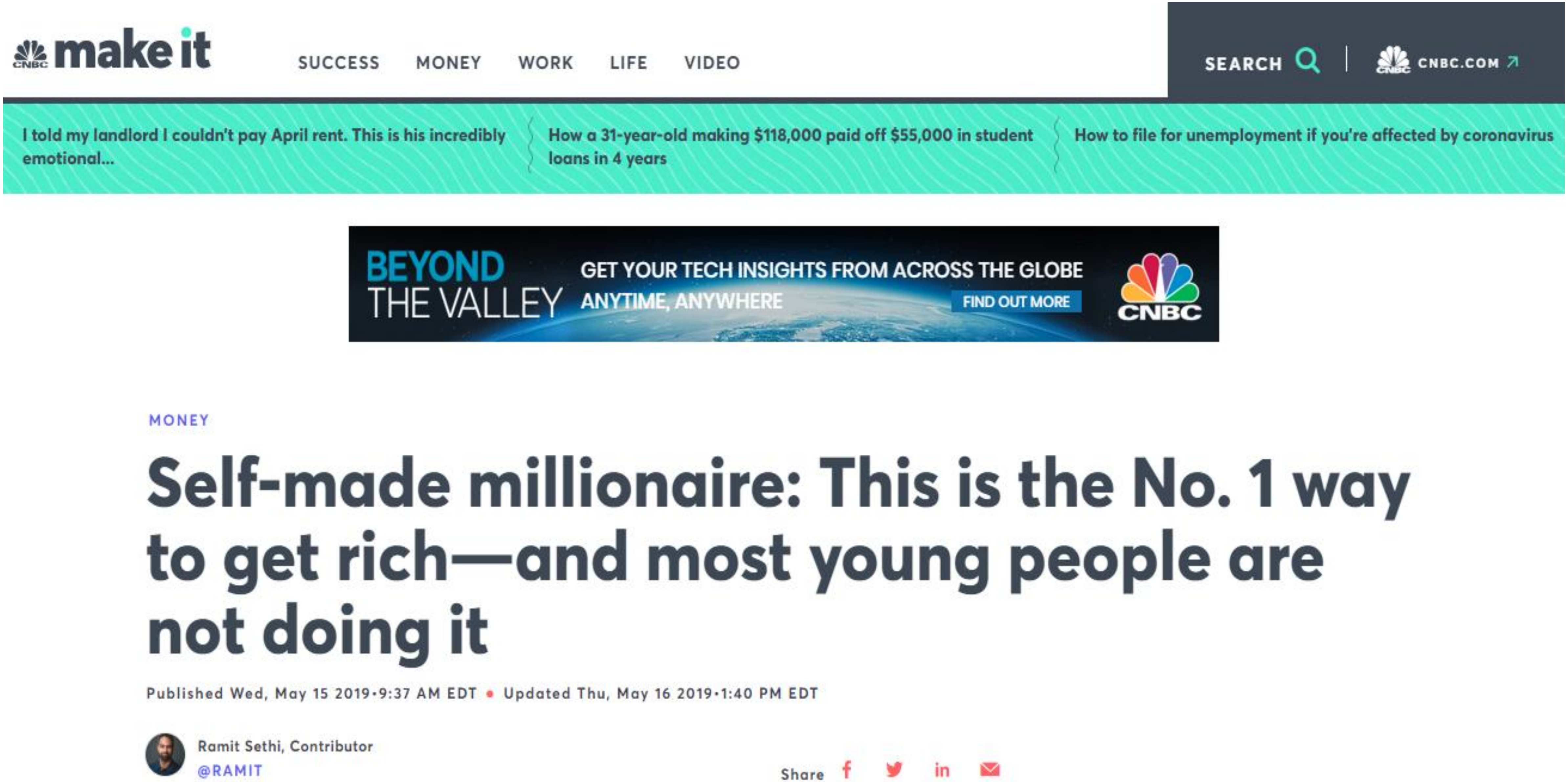
Careers

I write about building gender-balanced businesses

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A


SHARE
THE HISTORY OF
VACCINES
AN EDUCATIONAL RESOURCE BY THE COLLEGE OF PHYSICIANS OF PHILADELPHIA
Enter the terms you wish to search for.
TIMELINE
ACTIVITIES
ARTICLES
GALLERY
About
Blog
Educators
FAQ
Glossary
Media
Parents
Vaccine Science
History and Society
Vaccine Information
Do Vaccines Cause Autism?

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A



MONEY

Self-made millionaire: This is the No. 1 way to get rich—and most young people are not doing it

Published Wed, May 15 2019•9:37 AM EDT • Updated Thu, May 16 2019•1:40 PM EDT

Ramit Sethi, Contributor
@RAMIT

Share

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

December 10, 2015

The Troubling Link Between Parkinson's and Smoking: Can We Deny the Benefits?

Tori Rodriguez, MA, LPC

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

DIFFERENCE – IN – DIFFERENCES

https://www.udemy.com/course/econometrics-for-bu

Policy changes in a country / regions

Joshua D. Angrist, Alan B. Krueger, in Handbook of Labor Economics, 1999

Weather impact

Impact of M&A

Eero Lehto, Petri Böckerman,
Analysing the employment effects of mergers and acquisitions,
Journal of Economic Behavior & Organization,
Volume 68, Issue 1, 2008,
Pages 112-124,
ISSN 0167-2681

Examples

1 **Difference-in-Differences**

Geo tests in marketing

The New Jersey case is on of the most famous DiD studies

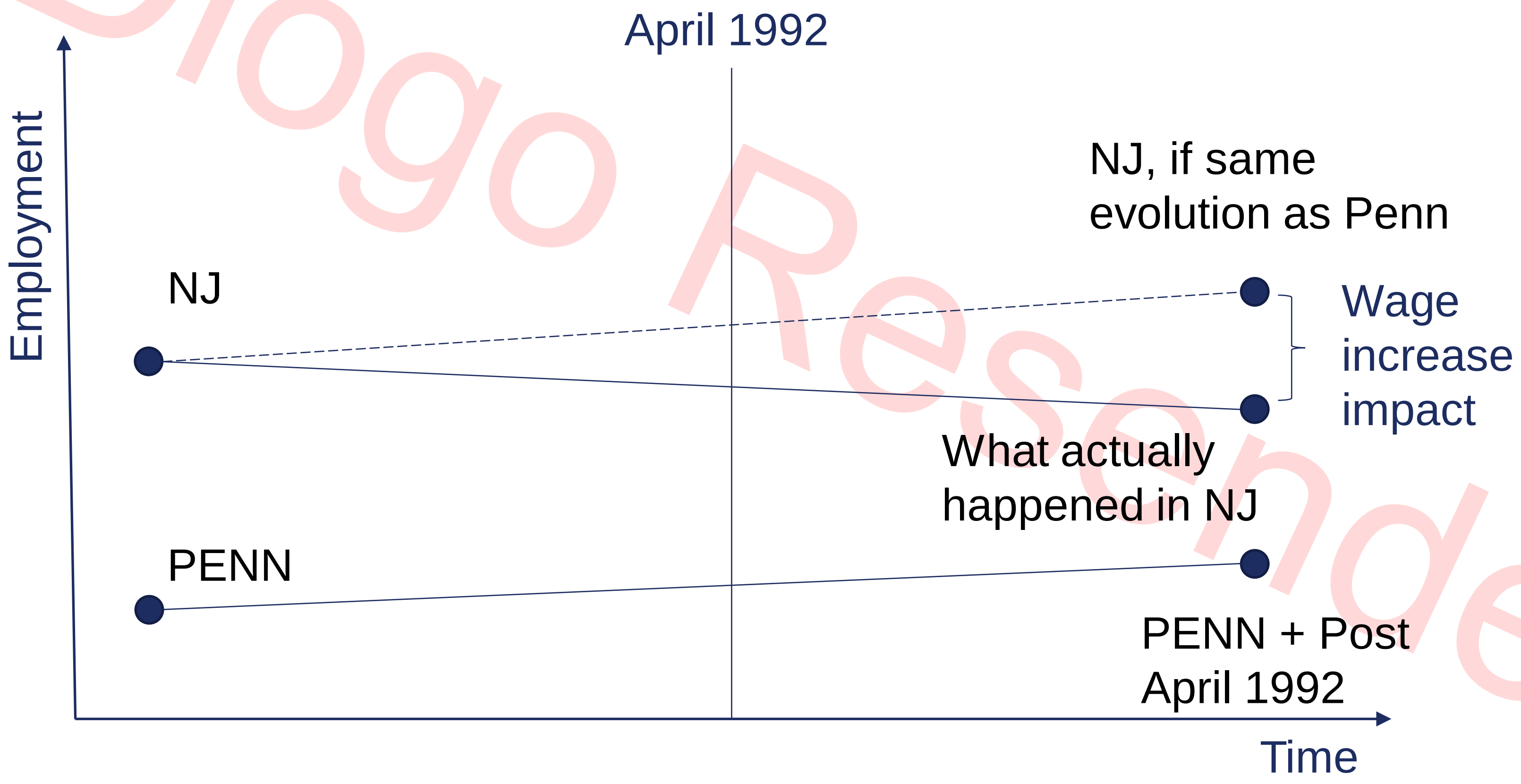
1 Difference-in-differences

- In April 1992, New Jersey rose the minimum wage from $4.25 to $5.05.
- Just comparing before and after would not be accurate, as it would fall into ommitted variable bias.
- In one of the most relevant Difference-in-Differences studies, Card and Krueger compared New Jersey to Pensylvania. This would resolve the ommitted variable bias mentioned above.
- Economy theory suggests that an increase in the minimum wage results in decreased unemployment.

Card D, Krueger AB. 1994. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. The American Economic Review 84: 772-793.

Concept explanation

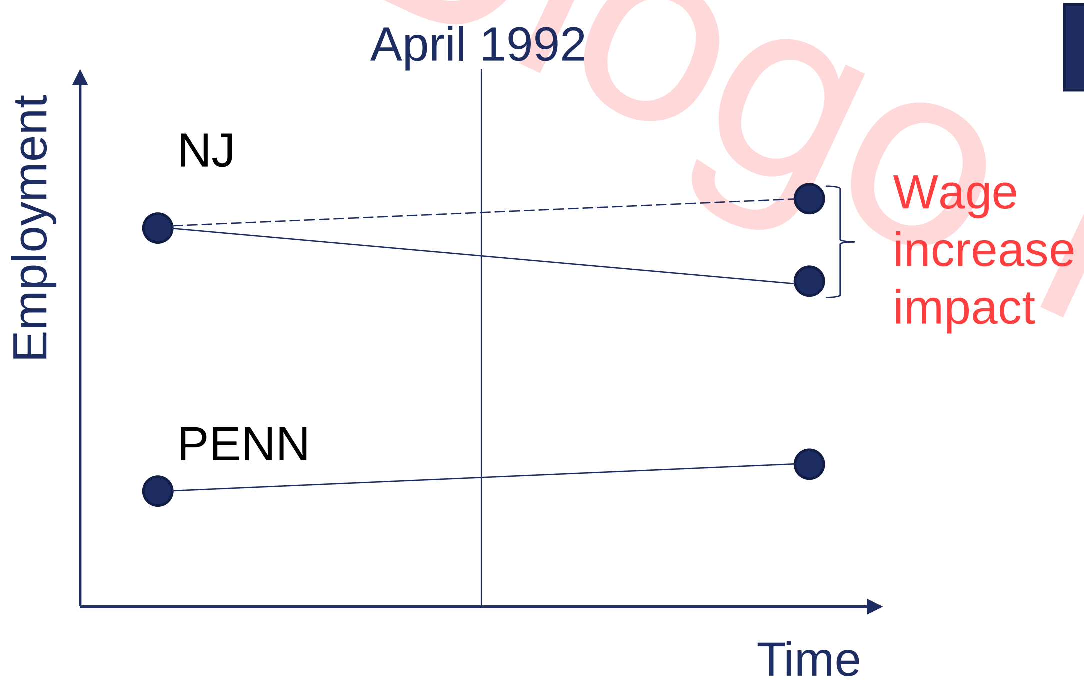
1 Difference-in-differences



https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Concept explanation

1 Difference-in-differences



How do we model? We need to define…

- Which fast food chains belong to New Jersey and which belong to Pennsylvania
  - We will use a dummy variable to flag whether a fast food chain belongs to NJ or PENN
- If the observation was recorded before or after April 1992
  - We will use a dummy variable to flag „after April 92“
- The wage impact on employment
  - We multiply the NJ variable by the „after April 92“

Which assumptions do we take and how to stengthen them

1 **Difference-in-differences**

Assumption	How to strengthen
• Parallel trends assumption • Confounding policy change	• Use more control groups • Use more time periods • Conduct a placebo test

Difference-in-differences Step by Step

1 Difference-in-differences

Define treatment, post period and treatment & post period variables

↓

Create a regression to calculate the impact

↓

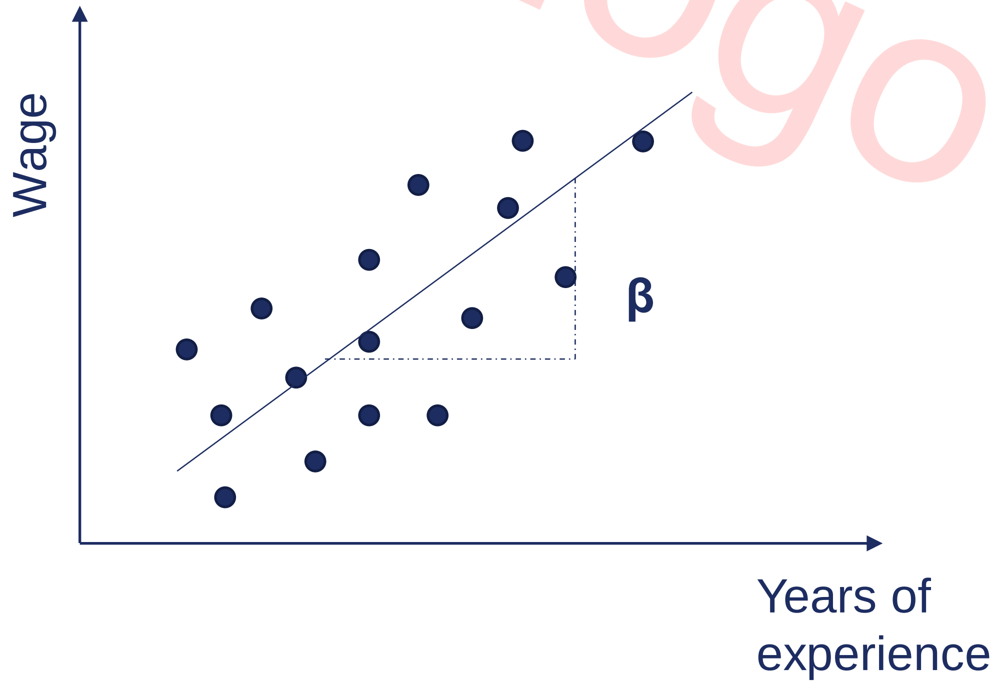
Add control variables to limit ommited variable bias

↓

Conduct Placebo test

(linear) Regression crash course

1 **Difference-in-differences**



What is it?

- It is the study of a relationship between an output or dependent variable and at least one independent variable or inputs

From an intuition perspective

- It is your method for “What is the impact of X on Y?

How is it different from a correlation?

- Correlation studies the direction
- Regression studies the impact

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Linear regression output

1 **Difference-in-differences**

```
Call:
lm(formula = fte ~ NJ + POST_APRIL92 + NJ_POST_APRIL92, data = da

Residuals:
    Min      1Q  Median      3Q     Max 
-21.162  -6.270  -0.773   4.338  64.543 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)       23.273      1.041  22.349   <2e-16 ***
NJ                -2.816      1.159  -2.430   0.0153 *  
POST_APRIL92      -2.111      1.473  -1.433   0.1522    
NJ_POST_APRIL92    2.681      1.639   1.636   0.1023    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.255 on 816 degrees of freedom
Multiple R-squared:  0.007206,  Adjusted R-squared:  0.003556 
F-statistic: 1.974 on 3 and 816 DF,  p-value: 0.1163
```

Estimates

- If continuous, it's the value Y increases per X unit
- If binary, it is the value when X = 1

Standard error

- The standard deviation of a sample

Confidence interval (95%)

- Estimate +- 2 time the standard error

Statistical Significance (5% level)

- When 0 is not part of the Confidence interval

P-value

- Statistical significance indicator. Probability of the coefficient being more than / less than 0

Dummy variable trap

1 **Difference-in-differences**

Observation	Coca cola	Pepsi
a	1	0
b	1	0
c	1	0
d	1	0
e	1	0
f	0	1
g	0	1
h	0	1
j	0	1

Multicollinearity

The Correlations between Coca cola and Pepsi is -1. Extremes are never good and regression models don't do well with multicollinearity. To avoid this, you should remove one dummy variable

Removing does not mean information is lost

When the algorithm goes row by row assessing the information, seeing only 0's is also information. In fact, the removed dummy variable becomes part of the intercept. You can see it as being your baseline.

DIFFERENCE – IN – DIFFERENCES

https://www.udemy.com/course/econometrics-for-bu

Background for second example

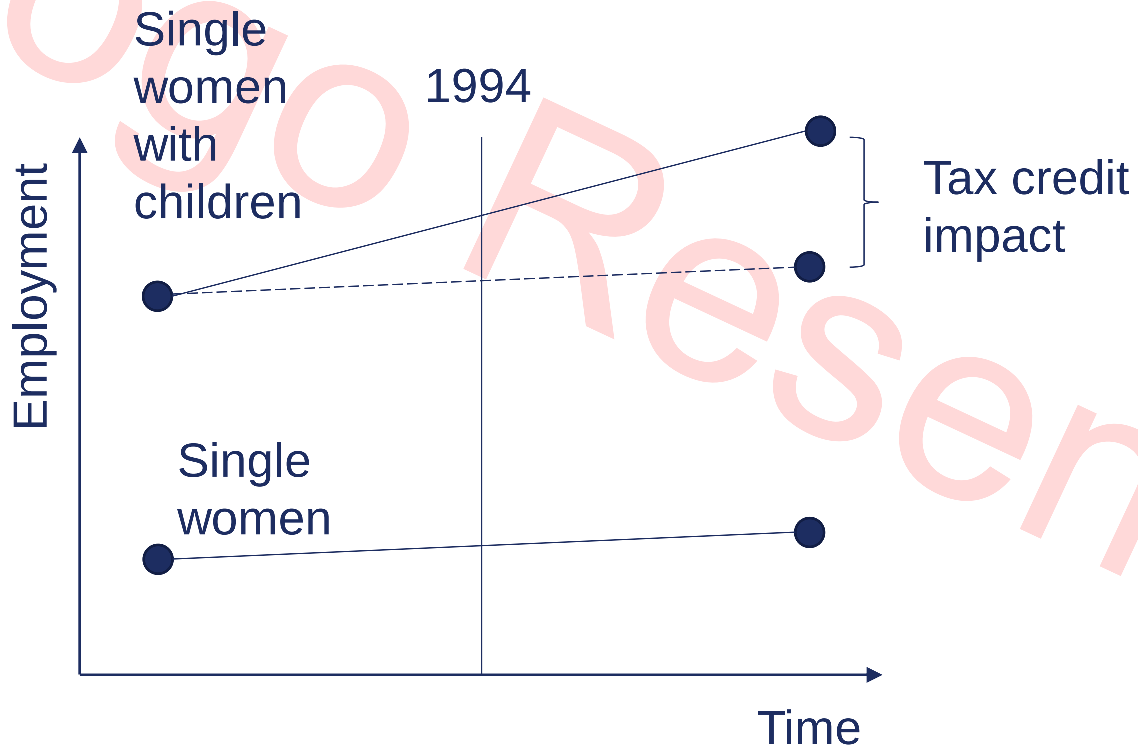
1 Difference-in-differences

- In 1994, the Earned Income Tax Credit was expanded to also include the employment of single women with children
- The United States federal earned income tax credit or earned income credit is a refundable tax credit for low to moderate-income working individuals and couples, particularly those with children.
- Standard labor supply theory does indeed predict that the EITC will encourage labor force participation. This occurs because the EITC is available only to taxpayers with earned income.
- Does this tax credit incentivize employment?

Eissa, Nada, and Jeffrey B. Liebman. 1996. Labor Supply Responses to the Earned Income Tax Credit. Quarterly Journal of Economics. **111**(2): 605-637.

Seeing the problem again through a graph

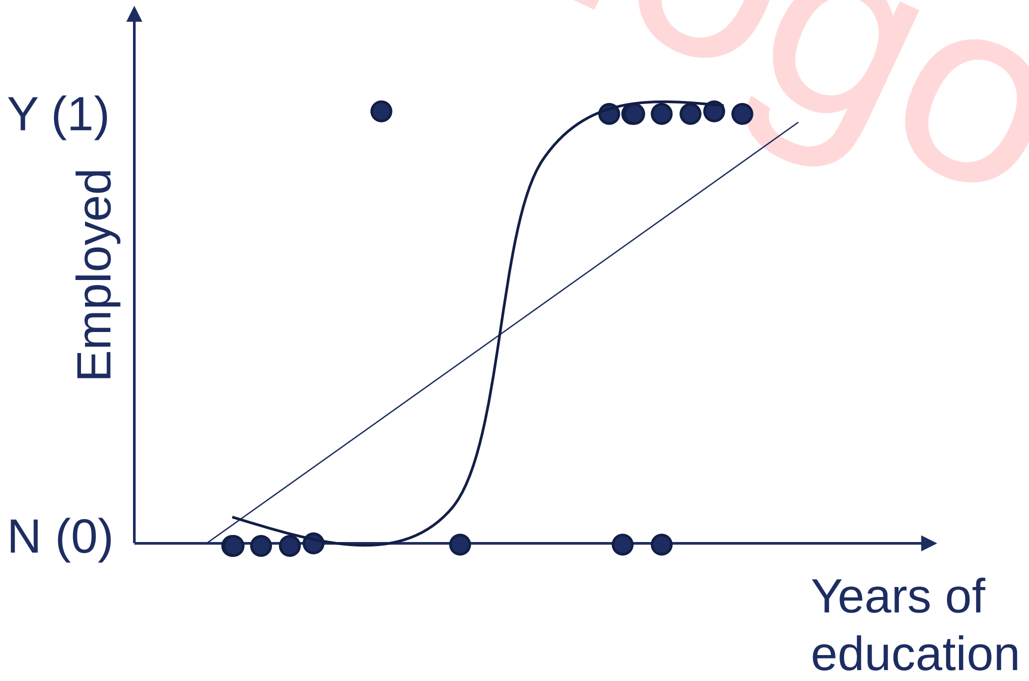
1 **Difference-in-differences**



https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

(Logistic) Regression crash course

1 Difference-in-differences



What is it?

- It is the study of a relationship between a discrete output or dependent variable and at least one independent variable or inputs

From an intuition perspective

- It is your method for "What is the impact of X on Y happening?

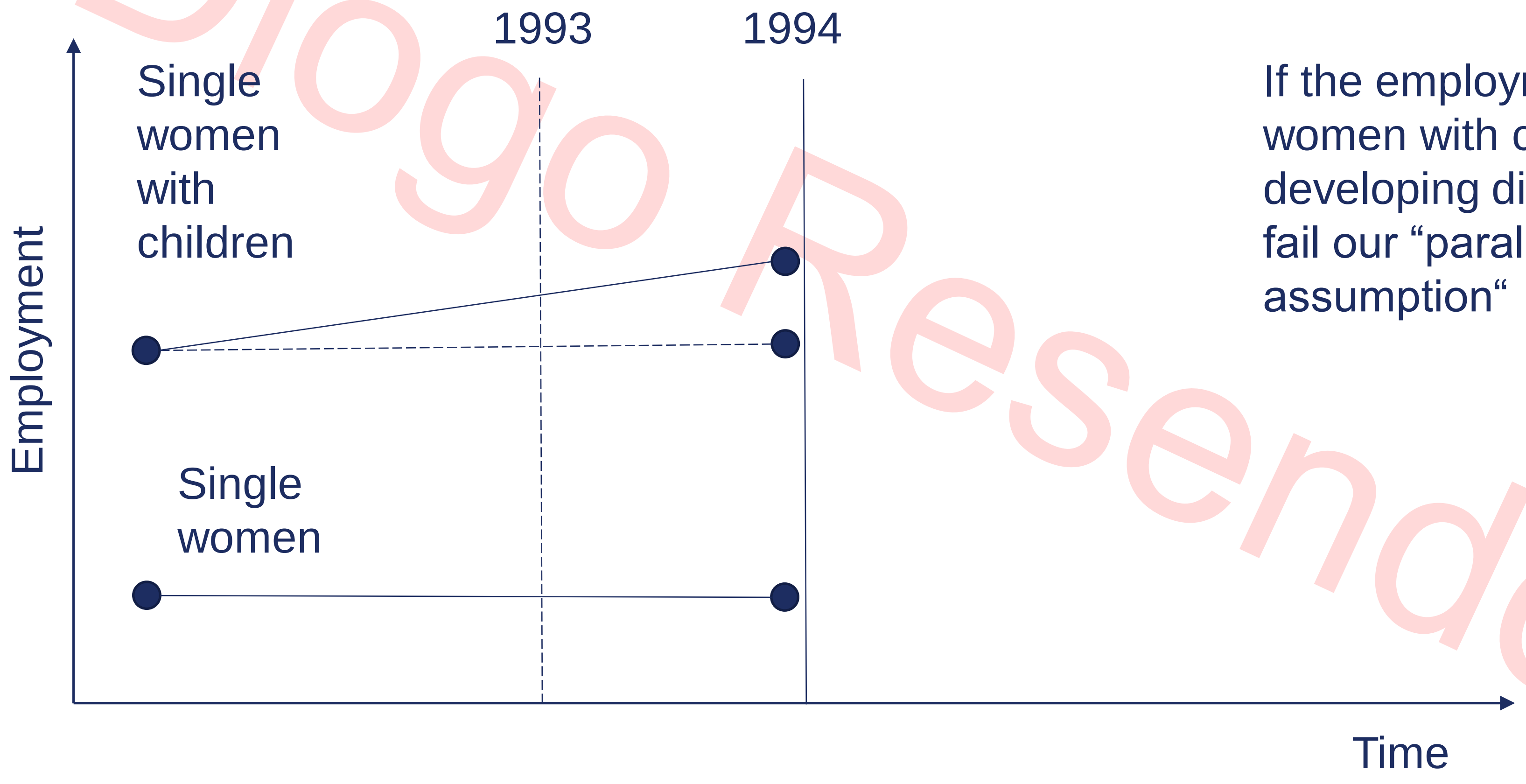
How is it different from a Linear Regression?

- Linear is for continuous, logistic is discrete
- Linear we fit a straight line, logistic a curve
- Linear assumes normal distribution, logistic a binomial distribution

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

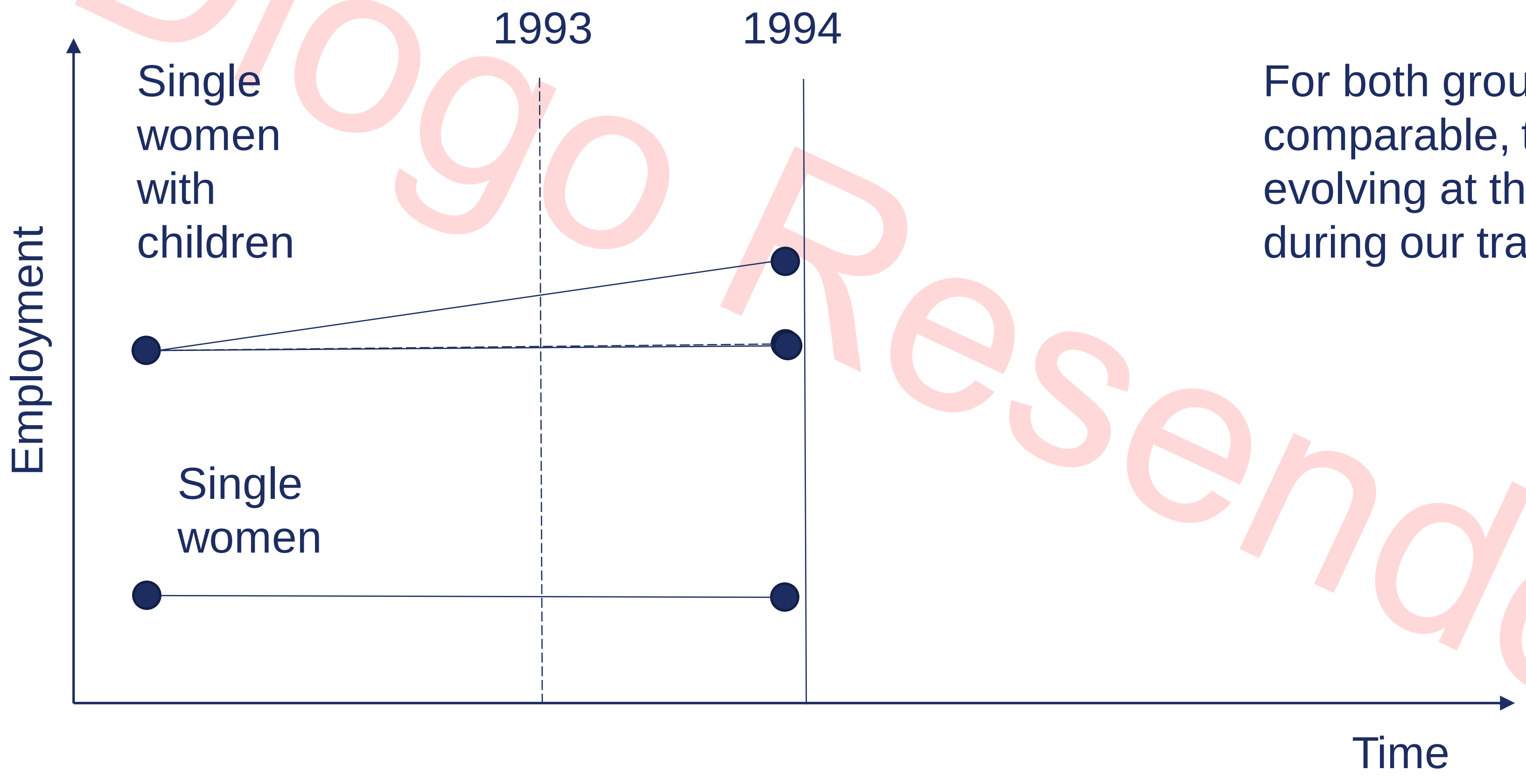
Let‘s look at the placebo test mechanics

1 Difference-in-differences



If the employment of single women with children is already developing differently, then we fail our “parallel trends assumption“

Let‘s look at the placebo test mechanics



For both groups to be comparable, they must be evolving at the same speed during our training period

Concept of difference-in-differences comes from c.a. 1850

1 **Difference-in-differences**

- Amid 19th century, London had an outbreak of cholera
- There were 2 popular theories at the time that were causing the cholera ruckus.
- Why is it relevant? In order to fight back against the outbreak would mean very different things pending on the cause.
- At the time, there were 2 major water suppliers – Southwark & Vauxhall Company and Lambeth Water Company. Both extracted water from the same part of the Thames in 1849. However, in 1852, Lambeth moved upstream.

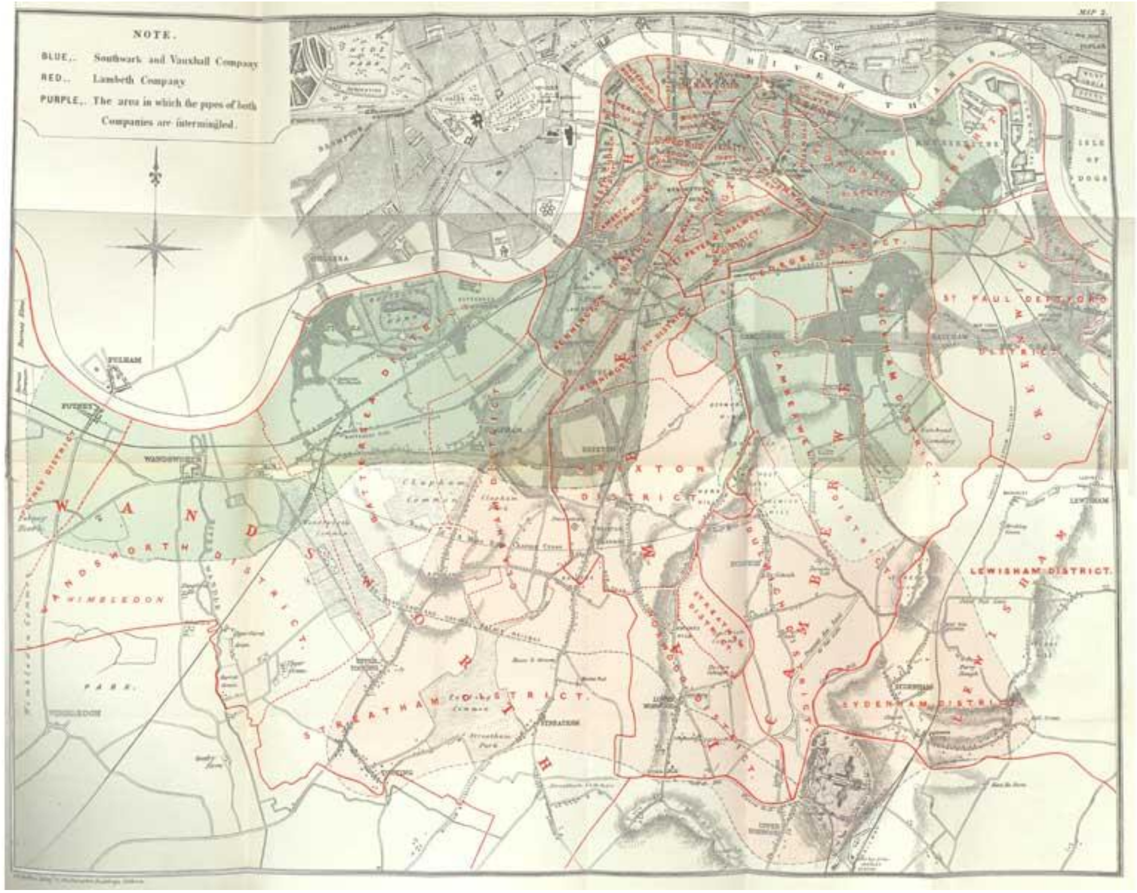
John Snow, 'Intimate Mixture of the Water Supply of the Lambeth with that of the Southwark and Vauxhall Company, 1854'

People with different house suppliers but living close enable a good counterfactual to one another.

1 Difference-in-differences

- Even though the water suppliers competed mostly alone, there were areas in which both competitors were present.

- John Snow had the intuition to go and look at the death from cholera per water company in the same areas of London



John Snow, 'Intimate Mixture of the Water Supply of the Lambeth with that of the Southwark and Vauxhall Company, 1854'

Wrapping up the John Snow example

1

	Number of houses	Deaths from Cholera
Southwark-Vauxhall Company	40,046	315
Lambeth Company	26,107	37
Rest of London	256,423	59

Snow, J. 1855. Table IX. *On the Mode of Communication of Cholera*, 86).

GOOGLE'S CAUSAL IMPACT

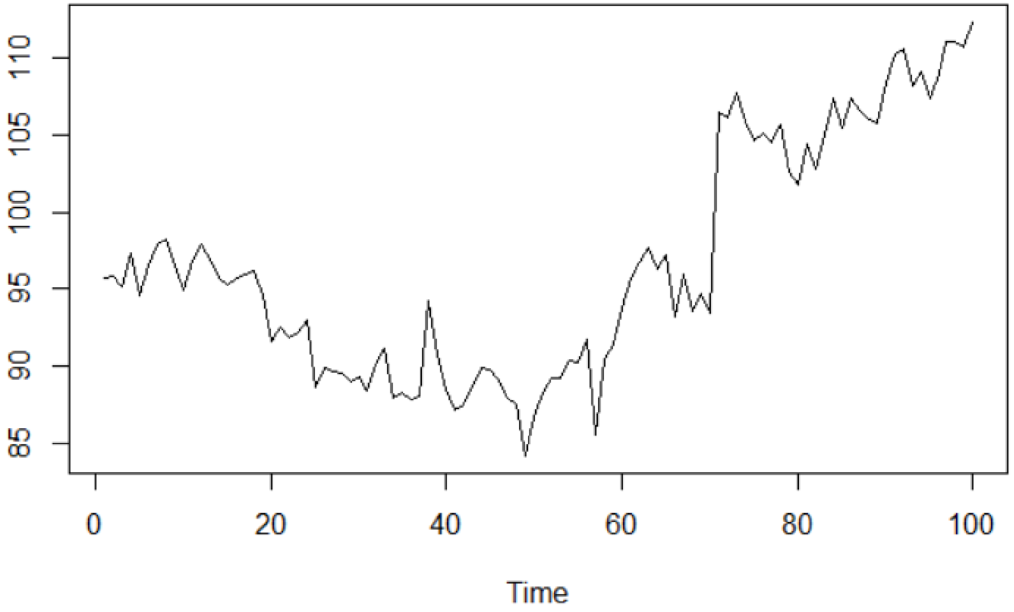
You were asked to assess the impact of your company’s latest brand campaign

2 **Google‘s Causal Impact**

- You do lots of TV, Social Media, Out of Home, Radio, etc… In the end, you want to understand whether it was worth it. Hence, how do you measure it?

- This graph shows the sales in the market. The campaign you launched started at the c.a. 70th day in the time axis.

- Comparing before and after would subject you to ommitted bias.



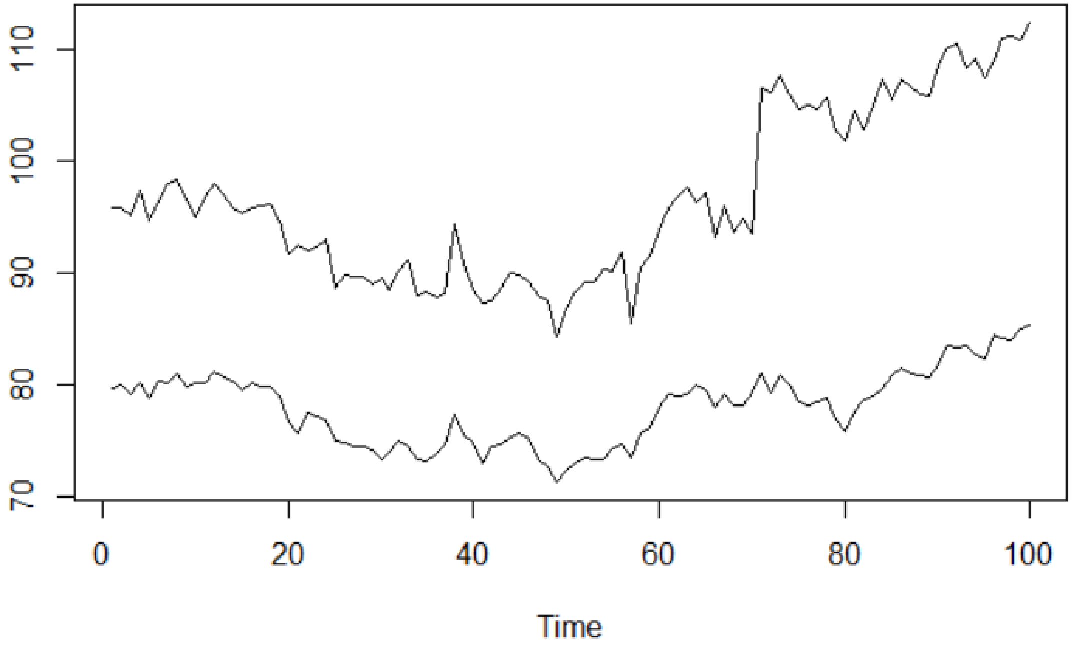
Brodersen, Kay H.; Gallusser, Fabian; Koehler, Jim; Remy, Nicolas; Scott, Steven L. Inferring causal impact using Bayesian structural time-series models. Ann. Appl. Stat. 9 (2015), no. 1, 247--274. doi:10.1214/14-AOAS788. https://projecteuclid.org/euclid.aoas/1430226092

Google’s Causal Impact intuition comes from Difference-in-Differences

2 Google‘s Causal Impact

- The idea is, similar to the last chapter is to compare with other markets, similar to DiD. Let‘s add a second market.

- Like in DiD, we should add more control groups to strengthen our results.



Brodersen, Kay H.; Gallusser, Fabian; Koehler, Jim; Remy, Nicolas; Scott, Steven L. Inferring causal impact using Bayesian structural time-series models. Ann. Appl. Stat. 9 (2015), no. 1, 247--274. doi:10.1214/14-AOAS788. https://projecteuclid.org/euclid.aoas/1430226092

Google's Causal Impact intuition comes from Difference-in-Differences

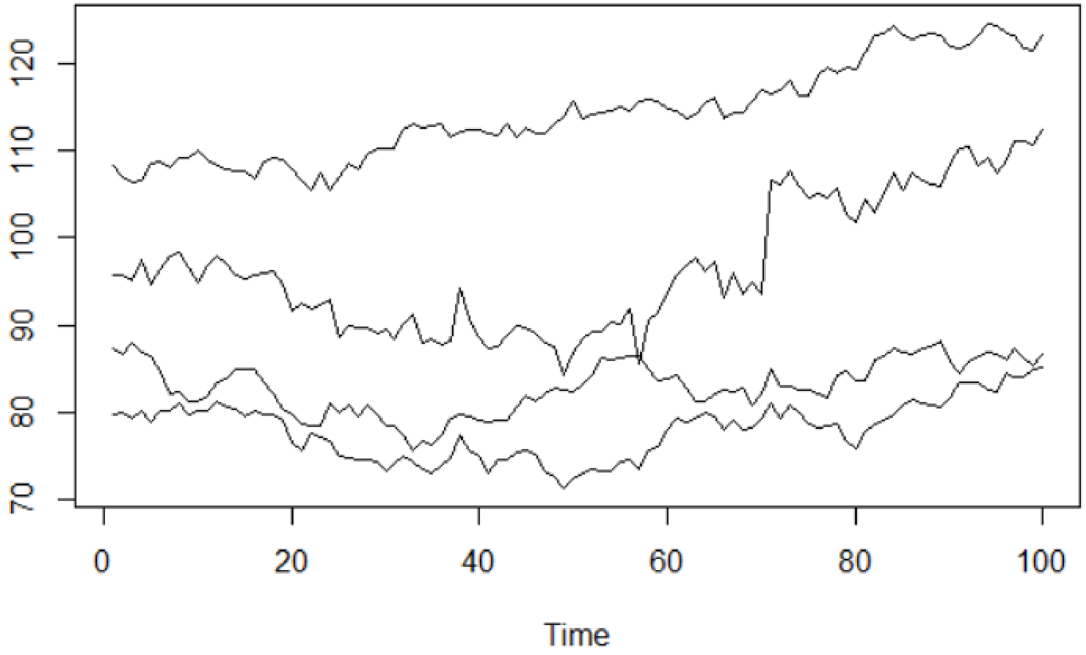
2 Google's Causal Impact

- The idea is, similar to the last chapter is to compare with other markets, similar to DiD. Let's add a second market.

- Like in DiD, we should add more control groups to strengthen our results.

- We can still very easily visualize that our market improved vs the other 3.

- What is the point then? Why should we have a more fancy solution if DiD worked well enough and apparently did the same thing

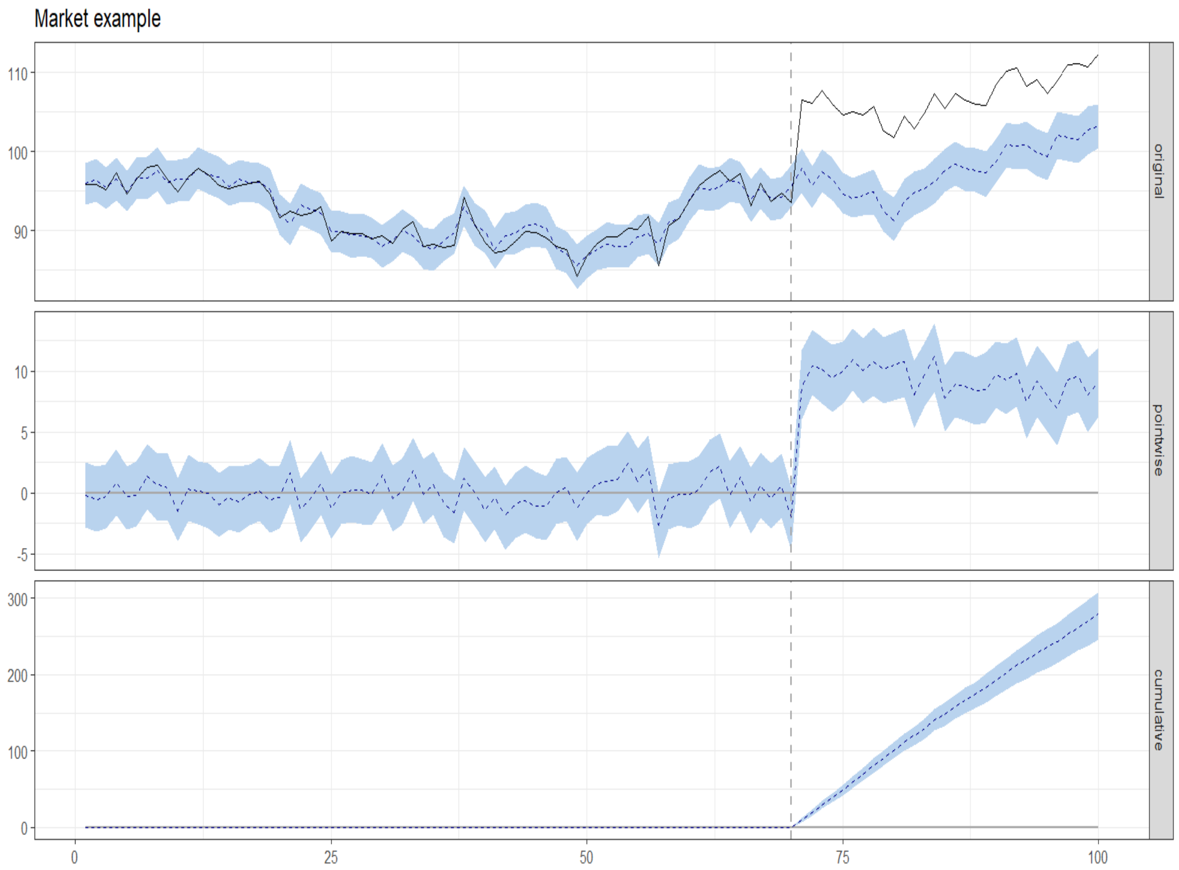


Brodersen, Kay H.; Gallusser, Fabian; Koehler, Jim; Remy, Nicolas; Scott, Steven L. Inferring causal impact using Bayesian structural time-series models. Ann. Appl. Stat. 9 (2015), no. 1, 247--274. doi:10.1214/14-AOAS788. https://projecteuclid.org/euclid.aoas/1430226092

Brand campaigns impact are still a bit of a mistery box

2 Google's Causal Impact

- Let's discuss what should be the impact of a major brand campaign:
  - Greater in the beggining
  - Impact gradually increases
  - You can also point out that the impact should continue after the campaign
- That is, in a nutshell, the value of Causal Implact. Whereas DiD would give you an average impact, CI allows the impact variations over time



Brodersen, Kay H.; Gallusser, Fabian; Koehler, Jim; Remy, Nicolas; Scott, Steven L. Inferring causal impact using Bayesian structural time-series models. Ann. Appl. Stat. 9 (2015), no. 1, 247--274. doi:10.1214/14-AOAS788. https://projecteuclid.org/euclid.aoas/1430226092

Causal Impact Step by Step

2 Google‘s Causal Impact

Define pre and post period

Retrieve the time series we need

Check whether the variables are correlated in the pre period

Use Causal Impact

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Before we go to the example, we need to wrap up

2 Google‘s Causal Impact

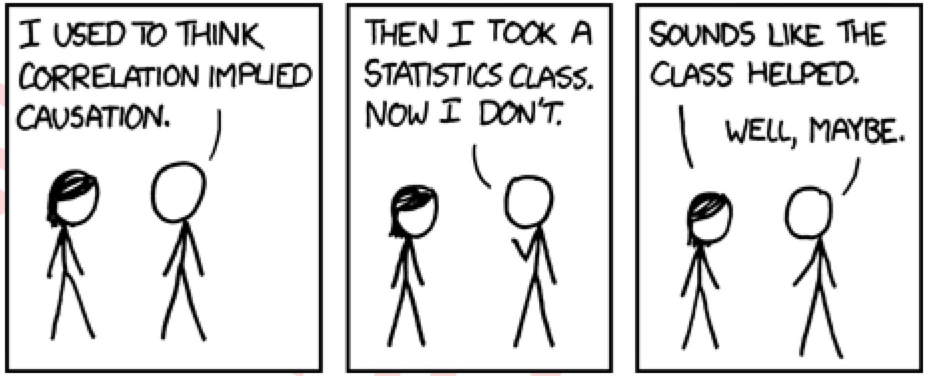
Assumption	How to strengthen	Why is it powerful	More info
• Parallel trends assumption • Confounding policy change	• Use more control groups • Use more time periods • Conduct a placebo test	• Allows for a powerful estimate even though A/B test is not feasible • Provides estimate of impact over time	• Presentation at Big Data Spain • https://google.github.io/CausalImpact/CausalImpact.html

What was the impact of the Cambridge Analytica scandal on Facebook stock price?

2 Google‘s Causal Impact

- For years, Cambridge Analytica harnessed Facebook users‘ data
- The data was used, most proeminently, in the 2016 United States Election
- In March 17th 2018, the New York Times and The Guardian, as well as the The Observer, which was working with a former Employee from Cambridge Analytica, broke the story.
- On April 10th, Mark Zuckerberg talks before Congress on the topic.
- In July 2018, Facebook is fined by both the UK and US government in over 5 billion euros

GRANGER CAUSALITY



Impact of economic drivers

Yi, Wen.
Granger Causality and Equilibrium Business Cycle Theory.
Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor],
2007-05-16

Examples

1 Difference-in-Differences

Influencer Marketing

Financial markets impact

de Oliveira, Erick Meira; Cyrino Oliveira, Fernando Luiz; Klötzle, Marcelo Cabus; Figueiredo Pinto, Antonio Carlos (2018),
“Data from: Dynamic Associations Between GDP and Crude Oil Prices in Brazil: Structural Shifts and Nonlinear Causality”,
http://dx.doi.org/10.17632/rxrsx28v9v.1

Do you agree with the following reasoning?

3 **Granger Causality**

- You are a Social Media Manager, responsible for the influencer program of your company.
- Feeling that it has a lot of potencial, you want to bring it to the next level. Hence, you ask for budget to diversify influencer activities.
- You go to the Director you report to and you present the following 3 premises:
  - Social Media is widely used by the customers
  - We see that impressions of our influencer campaigns is increasing
  - At the same time, we also see that sales are increasing
- Hence, because there is untapped potential influencer marketing contributes positively to sales, the company should further invest in influencer marketing.

If you were the Director, what would you reply?

3 **Granger Causality**

- Let's look at the reasoning in a diagram

Customers are more and more in Social Media

+

We have recently seen more impressions from the influencer program

+

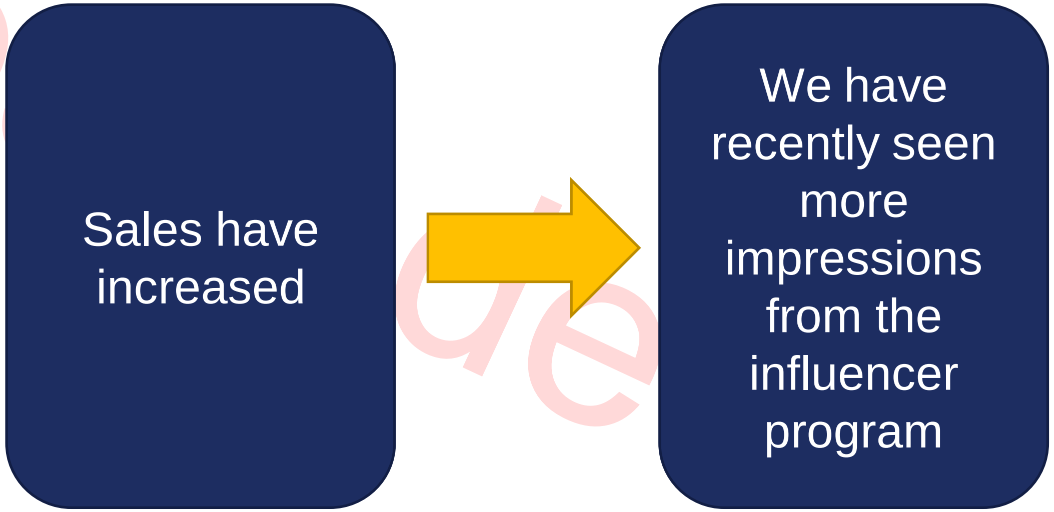
Sales have increased

=

Sales are coming from the extra impressions from the influencer program

The director could easily turn the reasoning the other way around…

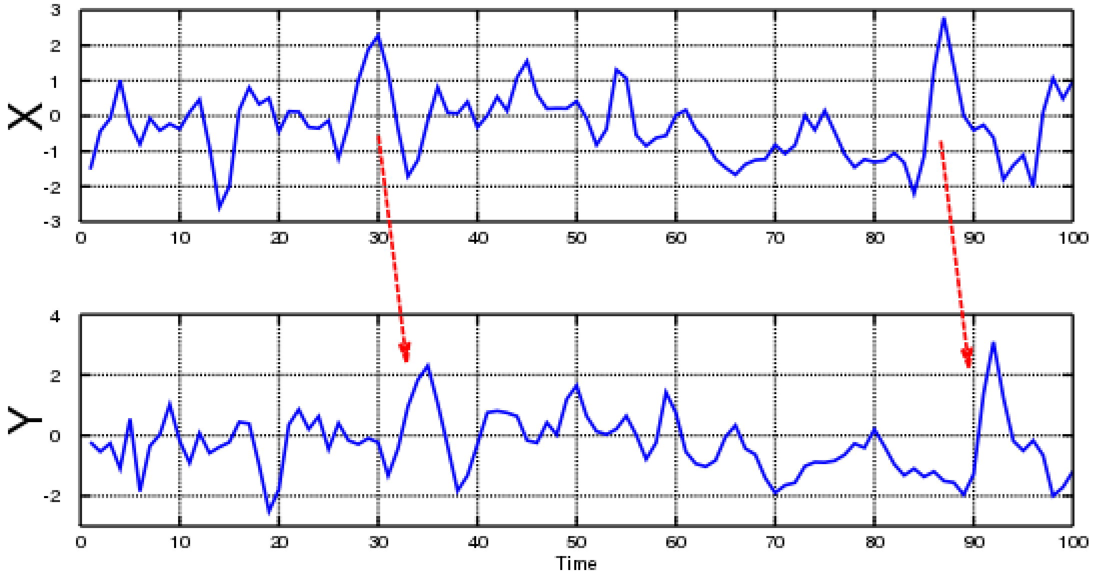
- Correlation is not causality!
- The same way you can argue that more impressions lead to more sales, the argument can be turned around and you can argue that the increase in sales can lead to an increase in impressions
- This now becomes a classic chicken and egg problem.
- Both factors are somewhat interconnected. The question is: which one started first?



Granger Causality gives insights into classic chicken and egg problems

- In order to convince the Director, we have to show that it is the impressions that come first and not other way around.

- This is where we apply our new technique. To have granger causality, we have to show:
  - That a certain lag of impressions is a statistically significant predictor of sales and…
  - Sales **is not** a statistically significant predictor of the lagged impressions

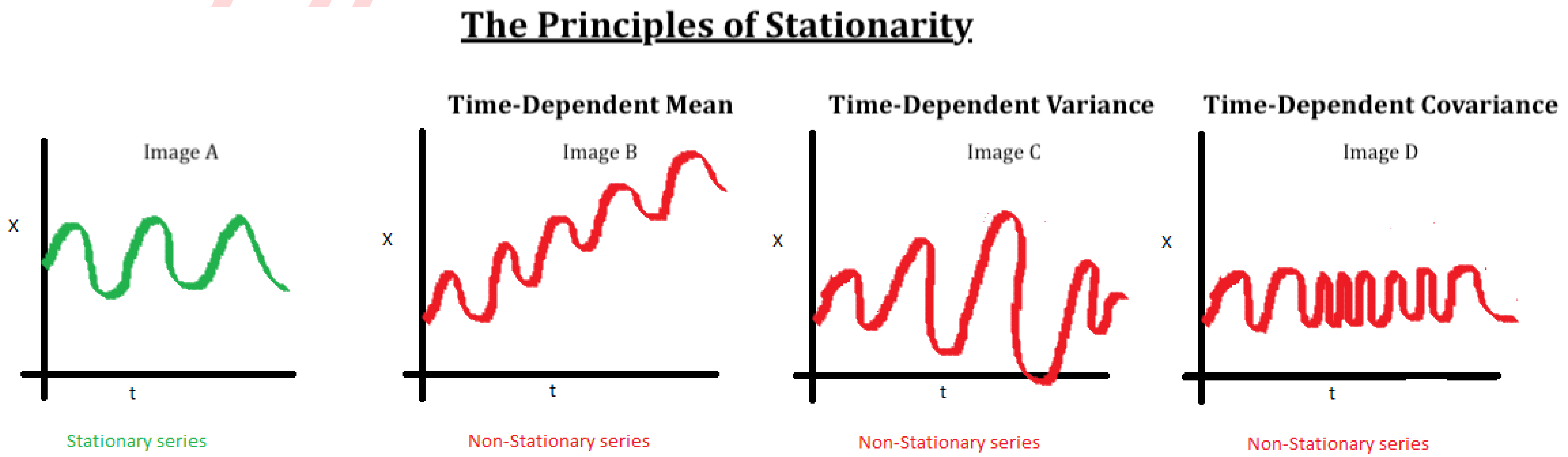


Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". Econometrica. 37 (3): 424–438. doi:10.2307/1912791. JSTOR 1912791.

https://www.wikiwand.com/en/Granger_causality

Stationarity

3 Granger Causality



Key idea: mean, variance and covariance are not time dependent

https://beingdatum.com/time-series-forecasting/

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Granger Causality Step by Step

3 **Granger Causality**

Have the 2 time series we want

↓

Check for stationarity

↓

Apply Granger Causality

↓

Find optimal number of lags

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Can you guess which will be our in practise example?

3 Granger Causality

Chicken

Or

Egg?

Thurman W.N. & Fisher M.E. (1988) Chickens, Eggs, and Causality, or Which Came First?, *American Journal of Agricultural Economics*, 237-238.

PROPENSITY SCORE MATCHING

Referral programs

Ina Garnefeld, Andreas Eggert, Sabrina V. Helm, Stephen S. Tax (2013),
“Growing Existing Customers’ Revenue Streams through Customer Referral Programs”.
Jounal of Marketing

Mobile shopping

Rebecca J. Wang, Edward C.Malthouse, Lakshman Krishnamurthi,
“On the Go: How Mobile Shopping Affects Customer Purchase Behavior”,
Journal of Retailing,
Volume 91, Issue 2,
2015, Pages 217-234.

New website languages

People analytics

As a Strategic HR/People manager, you propose a training program

4 **Propensity Score Matching**

- The focus of the training is on analytics, namely like the ones in this course.

- The goal of program is fourfold:
  - Give tools for better decision making in the company
  - Increase Employee Satisfaction
  - Decrease Employee Turnover
  - Internal success of the candidates

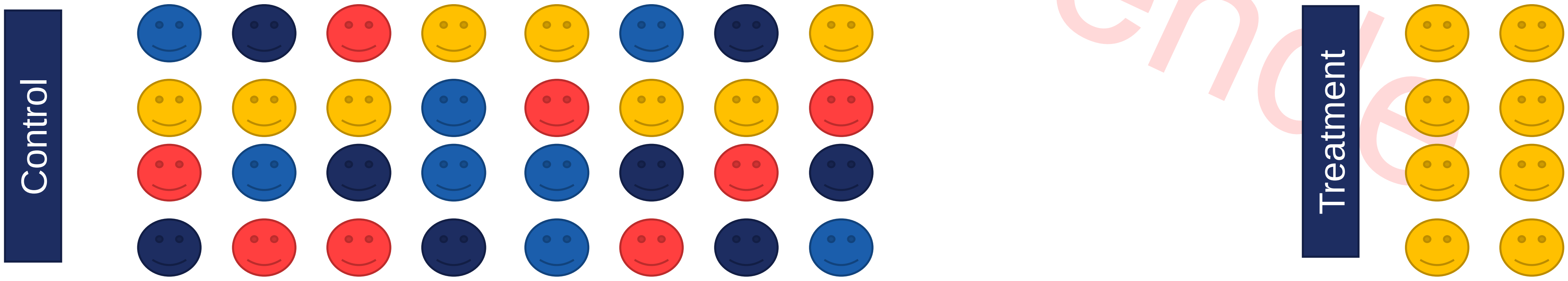
- The program is completely voluntary

- Now, 3 months after, you are asked to provide an overview of the program results. How do you do it?

You cannot just simply compare the average between trained and not

4 Propensity Score Matching

- Both groups may be inerently different from the start. Hence, they are not comparable.
- Beware of self-selection bias
- A possible solution is Propensity Score Matching.
- In a nutshell, you create a counterfactual group with similar characteristics to your treatment group

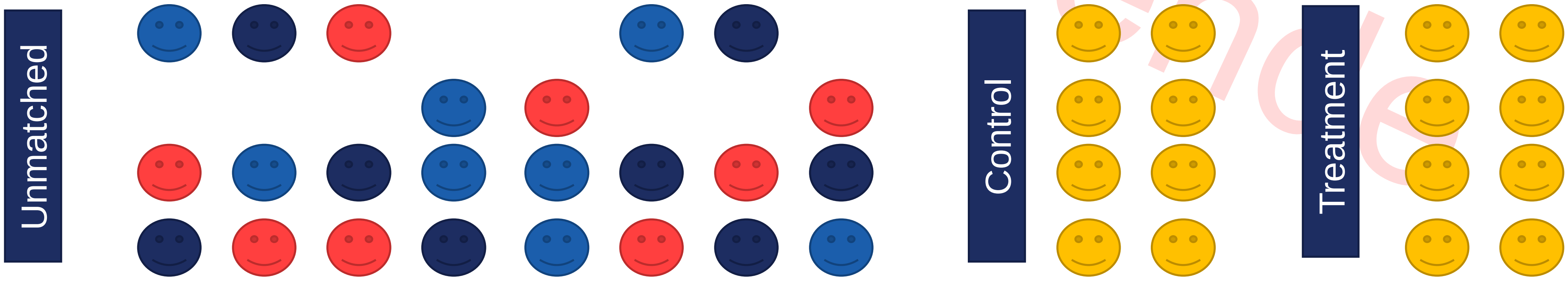


https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

You cannot just simply compare the average between trained and not

4 **Propensity Score Matching**

- Both groups may be inerently different from the start. Hence, they are not comparable.
- Beware of self-selection bias
- A possible solution is Propensity Score Matching.
- In a nutshell, you create a counterfactual group with similar characteristics to your treatment group



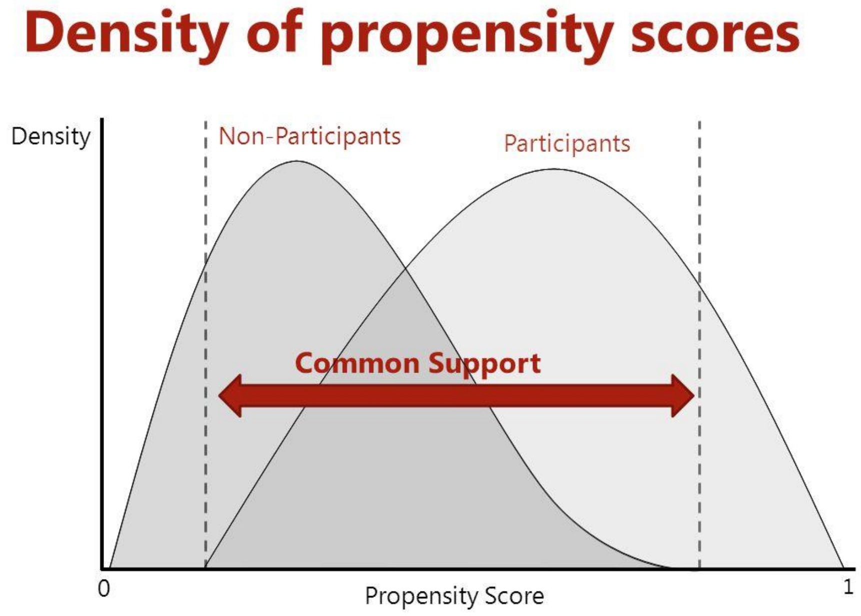
https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

You need to check two boxes to have a good PSM in place

4 **Propensity Score Matching**

- **Unconfoundness**
  - The control variables chosen and identified are enough to (almost) eliminate self selection-bias
  - Basically we aim at there being no difference between control and treatment group
  - In other words, it is like the control group is as good as if it was randomized

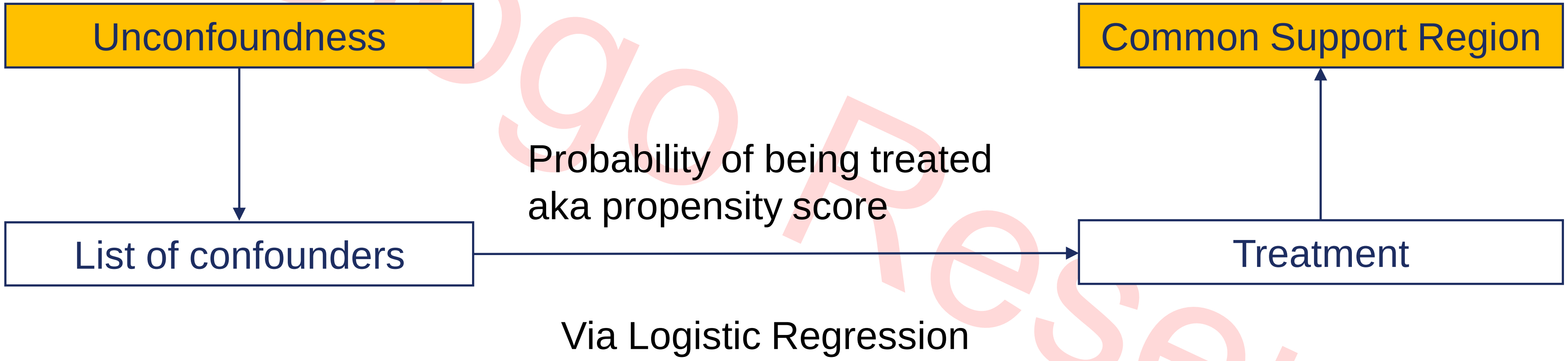
- **Common support region**
  - You can only match comparable individuals
  - To maximize this overlap, we should have a big enough control group



https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

How to determine the Common Support Region

4 **Propensity Score Matching**



Key ideas:

1. Finding how good you are at predicting whether someone is part of the treatment group
2. There will be people with super high likelihood of participating. You are not likely to find a control group for them.

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Propensity Score Matching Step by Step

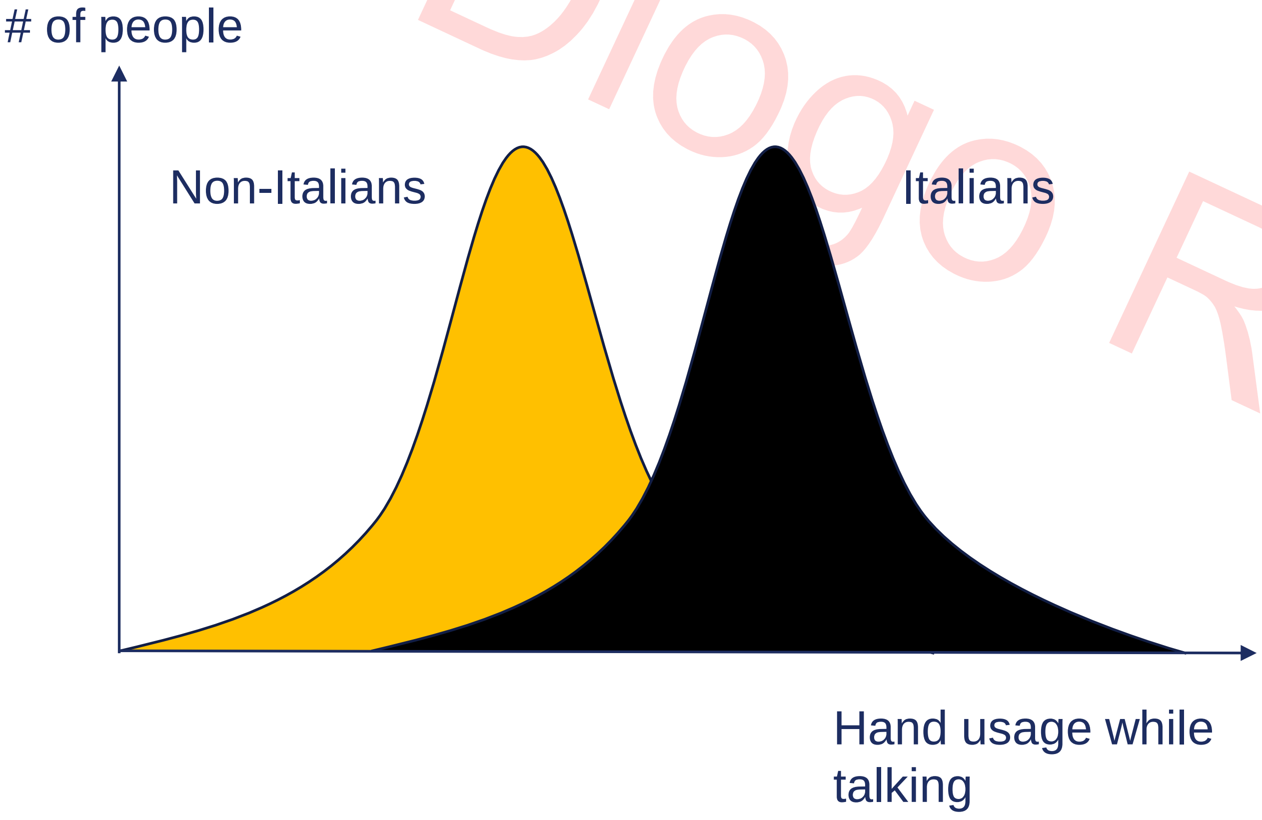
4 Propensity Score Matching

1. Variable selection
2. Summary statistics of the covariates
3. Logistic regression to assess Common Support Region
4. Matching
5. T tests to assess groups‘ comparability
6. Impact assessment

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

T-Tests

4 **Propensity Score Matching**



T Test formally

Test any statistical hypothesis in which the test statistic follows a Student's t-distribution under the null hypothesis.

In pratical terms

Helps us understand whether one group is different than the other

How do we know?

By looking at the p value of the test results

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Case Study Briefing

4 Propensity Score Matching

- In the 1970s, the National Support Work Demonstration held training programs for disadvanged workers
- Highly competent individuals were selected for the training (treatment)
- How do we measure the impact? We are in front of a selection bias problem

LaLonde, Robert J. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review*, vol. 76, no. 4, 1986, pp. 604–620. JSTOR, www.jstor.org/stable/1806062.

Background for second example

4 **Propensity Score Matching**

- Do students from catholic schools have better grades than the ones from public schools?

LaLonde, Robert J. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review*, vol. 76, no. 4, 1986, pp. 604–620. JSTOR, www.jstor.org/stable/1806062.

CHAID

Direct Marketing

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Customer Segmentation

Hsu, C. H. C., & Kang, S. K. (2007).
“CHAID-based Segmentation: International Visitors’ Trip Characteristics and Perceptions”.
Journal of Travel Research,
46(2), 207–216.
https://doi.org/10.1177/0047287507299571

Customer Satisfaction

Jinsoo Hwang & Jinlin Zhao (2010)
“Factors Influencing Customer Satisfaction or Dissatisfaction in the Restaurant Business”
Using Answer Tree Methodology,
Journal of Quality Assurance in Hospitality & Tourism,
11:2, 93-110

Employee satisfaction

Engin Üngüren, Rüya Ehtiyar (2016)
“Determination of the Demographic Variables Predicting Accommodation Business Employees’ Organizational Commitment and Job Satisfaction through CHAID Analysis”,
İşletme Araştırmaları Dergisi,
8/2016, 331-358,

You were hired to figure out which customers would be willing to sign up to a savings account in a bank via newsletter

5 CHAID

Why

- **Customer churn**: Sending a newsletter customer who cannot sign up can lead for he/she to unsubscribe
- **Opportunity cost**: sending to wrong product for the customer to sign up can create a loss in the case the customer would be interesting to sign up for another
- **Relevance**: Sending constinuously information that the customer is not interested can potentially lead for lower open rate willingness in the future

You go and look at previous savings newsletters to see in which customers performed better

5 **CHAID**

Age

		Young	Old
Gender	Male	Few	Medium
	Female	Medium	High

Has savings account

		Yes	No
Balance	High	Medium	High
	Low	Low	Medium

Other products

		Yes	No
Tenure	High	High	Medium
	Low	Medium	Low

Education

		High	Low
Job	Employed	High	Medium
	Freelancer	Medium	Low

Complexity increases as you deep dive in your problem

5 CHAID

What?

- **Problem depth**: Having more than 20, 50 or 100 drivers increases the complexity
- **Importance**: how do you know which driver actually matters most?
- **Analysis**: Diving deep will eventually result in having several buckets with few customers. How would you interpret them?

One of the CHAID‘s benefits is that figures out which drivers are more important

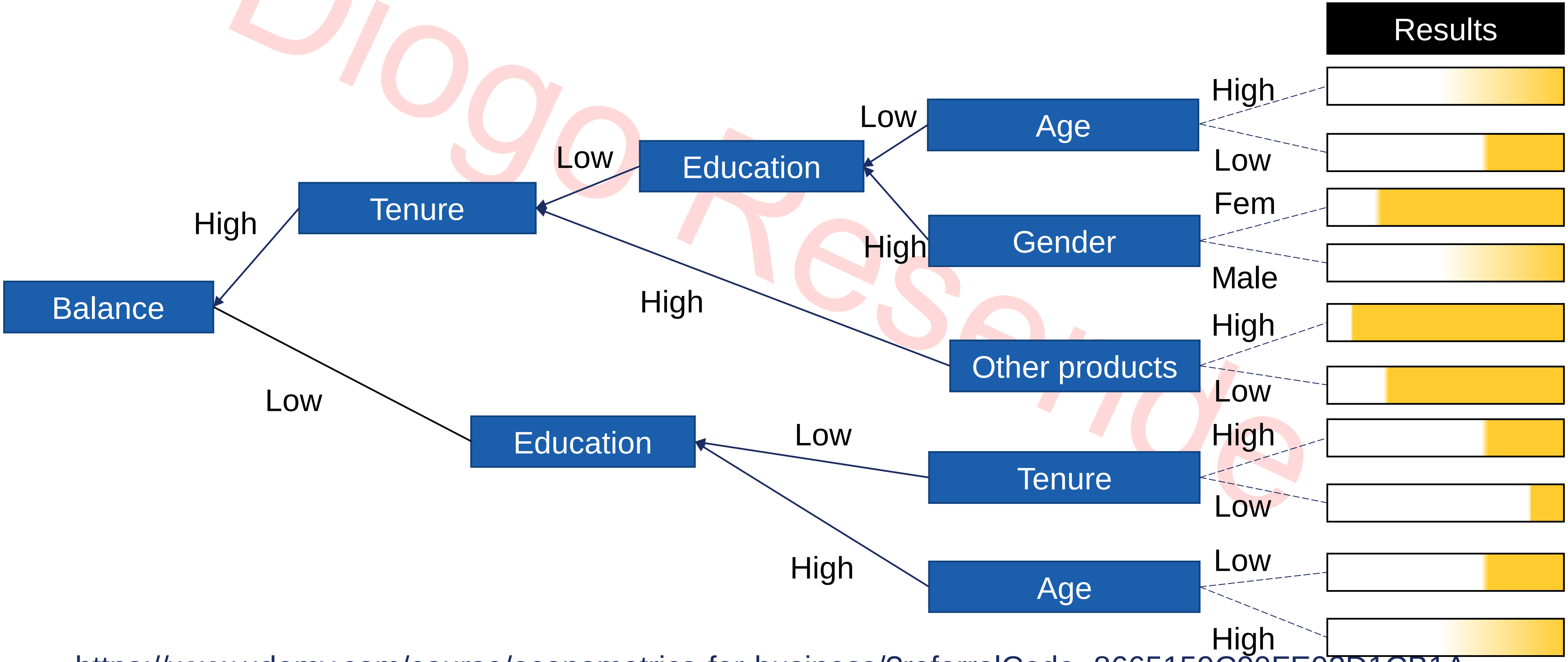
5 CHAID

Which?

- **Importance ranking**: CHAID figures out which drivers matter more, by doing significance tests
- **Aggregation**: If a certain bucket has few elements, CHAID aggregates it with another creating less noise
- **Interpretability**: CHAID provides easy to read graphs with customer segments

Let's see how it works visually

5 CHAID



How CHAID processes

5 CHAID

Balance (Signs up)

Balance	Yes	No
High	A lot	Few
Low	Few	A lot

Of the people who have high balance:

Tenure (Signs up)

Tenure	Yes	No
High	A lot	Few
Low	Few	A lot

How Does it start?

- CHAID looks at all predictors and tries to find the one where the “yes” is most different from the “no”

How does it work

- CHAID performs a Chi-square test. It shows whether the frequencies of the categorical variables are different or not. Very similar to t-test, but focus on caregorical variables

And then?

- After it finds the first segment split, tries to find another for each branch

The Confusion Matrix allows to access the results of a classifier

		Truth	
		False	True
Predicted	False	True negative	False Negative
	True	False Positive	True positive

Accuracy

- Accuracy = (True positive + True negative) / All
- Used when we have balanced dataset

Sensitivity or Recall or True Positive Rate

- True positive / (true positive + false negative)
- Used when we are skewed towards False values

Specifiticy or False Positive Rate

- True negative / (true negative + false positive)
- Used when we are skewed towards True values

Last few things consider

5 CHAID

Which?

- **Tree size**: You can choose how many levels the tree will have
- **Bucket size**: You can choose a minimum threshold that you want your buckets to have
- **Continuous variables**: CHAID accepts only categorical variables

CHAID Step by Step

5 CHAID

Variable selection

↓

Transforming continuous variables into categorical

↓

Do your first tree

↓

Prune it for better interpretability

Practical example

5 CHAID

- You have been hired to understand why employees quit
- You are given a dataset by IBM with more than 30 drivers
- Let‘s apply CHAID ☺

STRUCTURAL EQUATION MODELLING

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

Customer Satisfaction

Understanding Customer Behavior

Dakduk, S., González, &., & Portalanza, A. (2019).
Learn about structural equation modeling in smart PLS with data from the customer behavior in electronic commerce study in Ecuador (2017).
London, United Kingdom: SAGE Publications, Ltd.

Impact of leadership

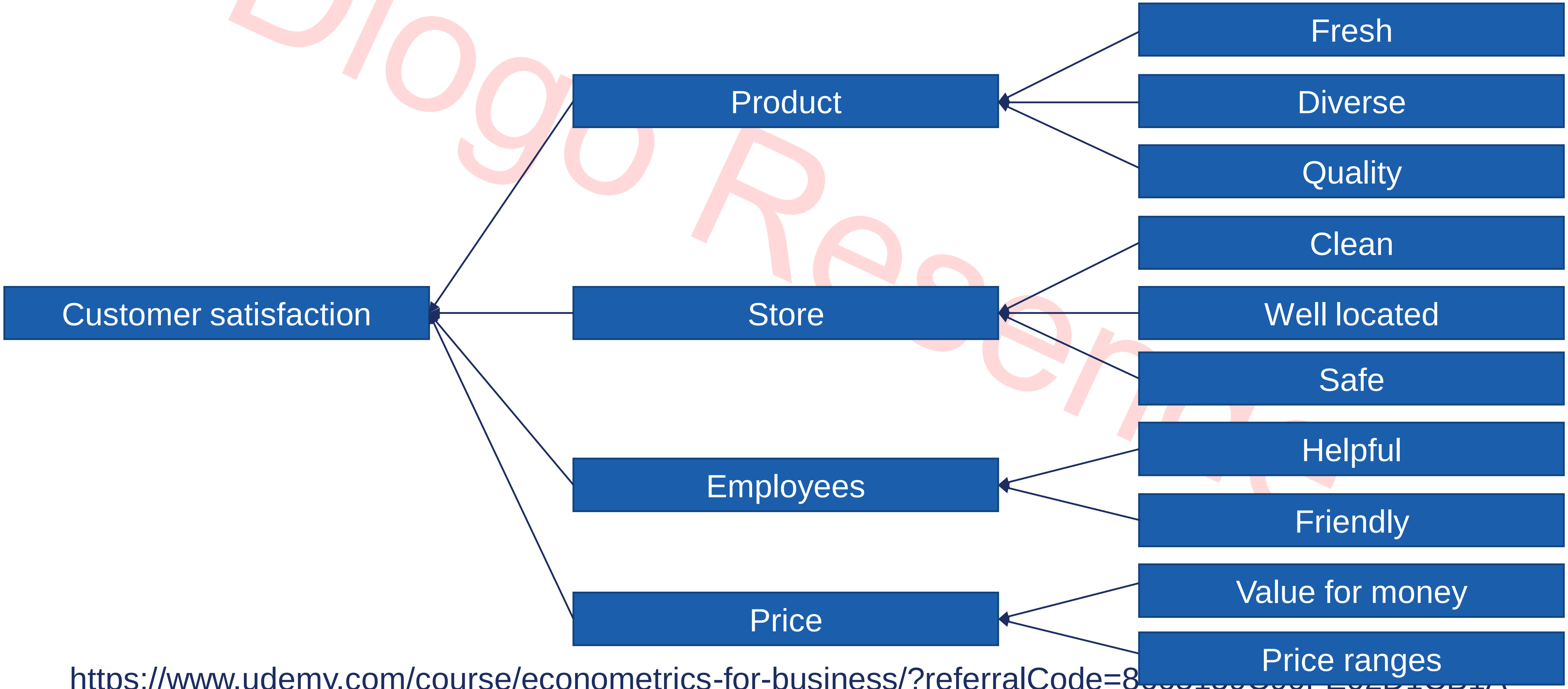
Tojari, Farshad & Sheikhalizadeh, Mahboub & Zarei, Ali. (2011).
Structural equation modeling analysis of effects of leadership styles and organizational culture on effectiveness in sport organizations

People analytics

Triguero-Sánchez, Rafael; Peña-Vinces, Jesús; Guillen, Jorge (2018):
How to improve firm performance through employee diversity and organisational culture.
SciELO journals.

You have just been hired to understand what drives customer satisfaction in a supermarket chain

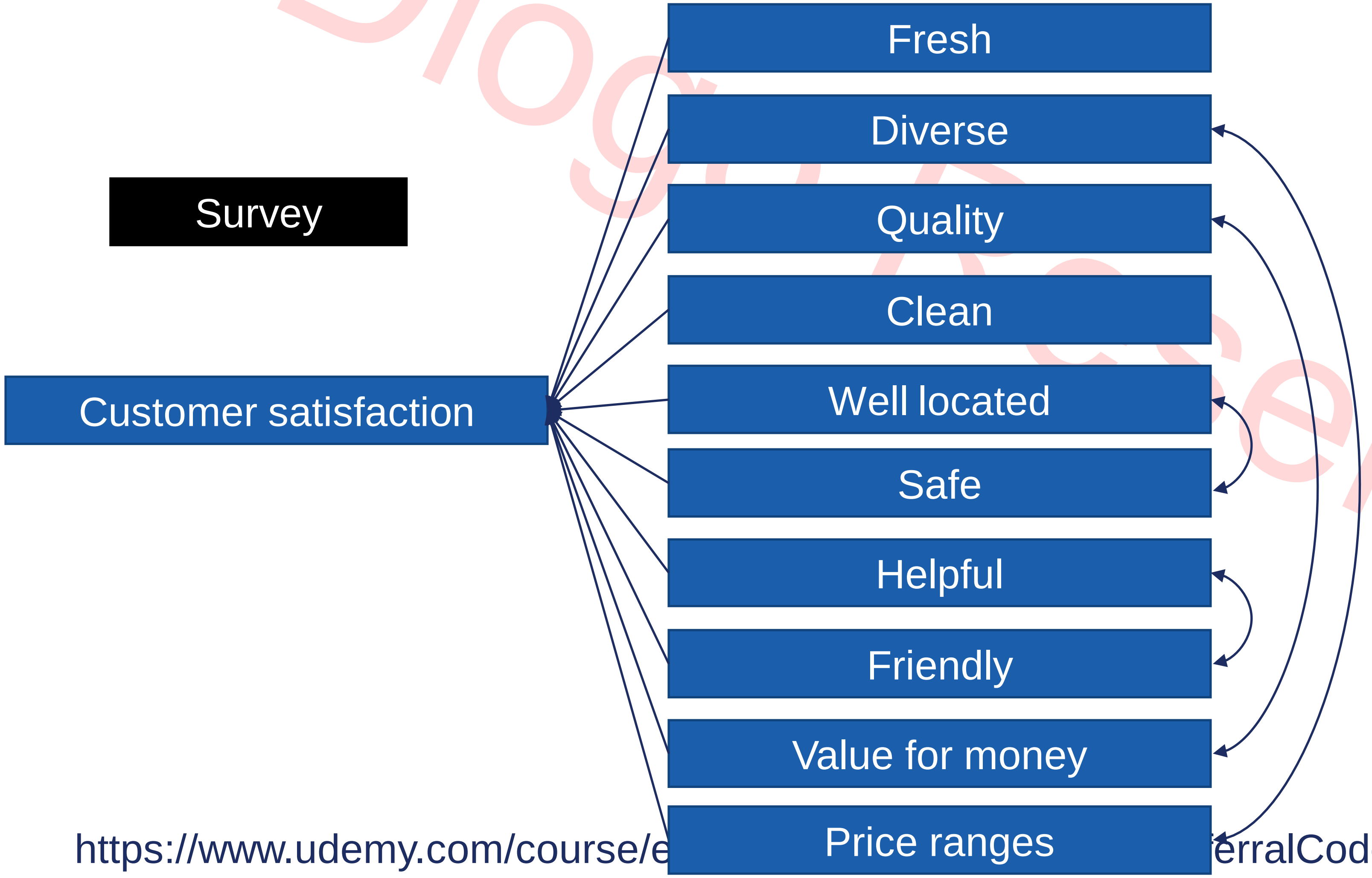
5 Structural Equation Modelling



The easy solution is always to run a regression

5 **Structural Equation Modelling**

Survey



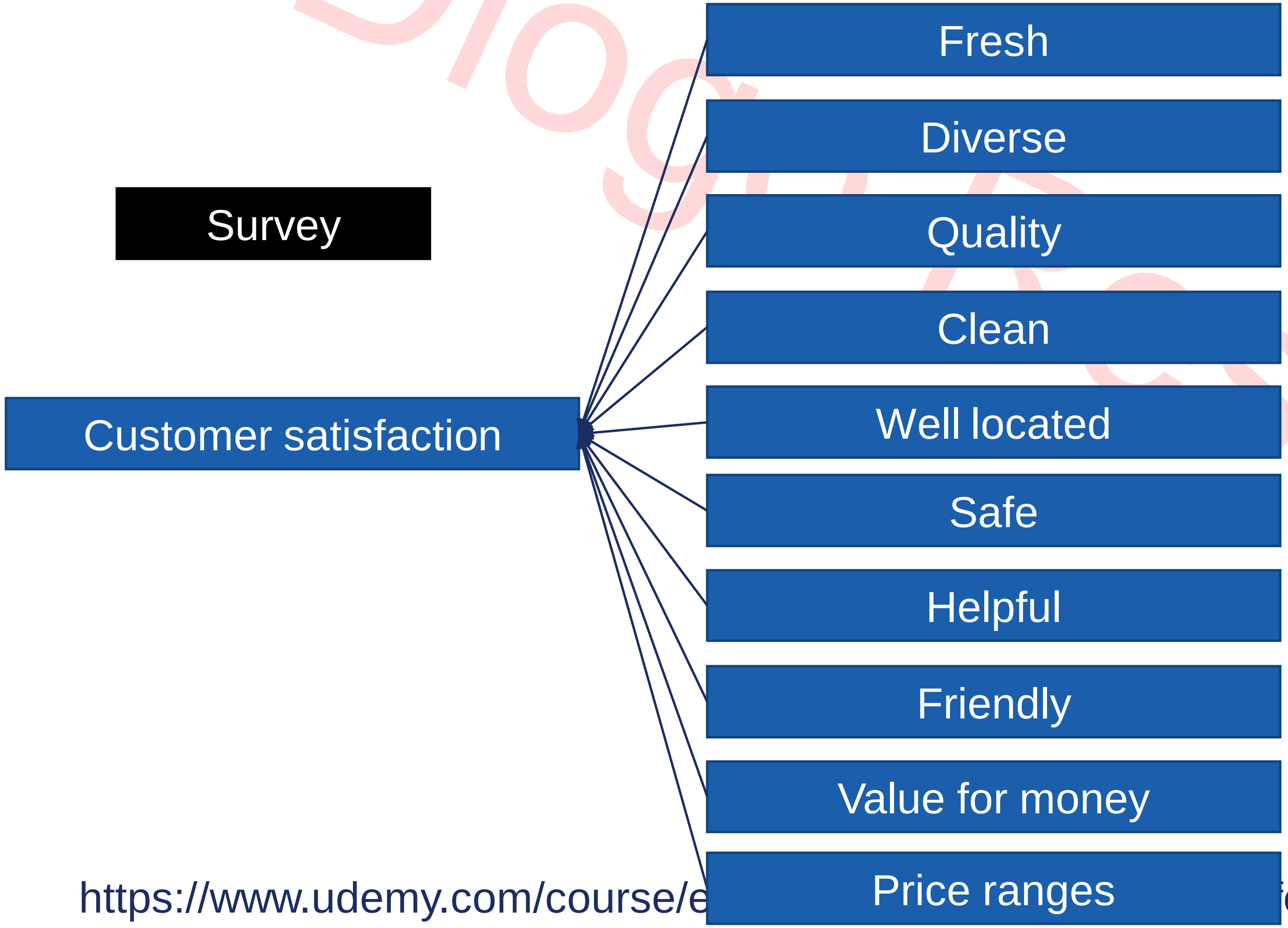
Regression

- Can measure impact
- Provides significance
- Multicollinearity
- No causal relationships

https://www.udemy.com/course/e ferralCode=8665159C90FE02D1CB1A

Random forest solves the multicollinearity issue but not the causal relationships

Survey



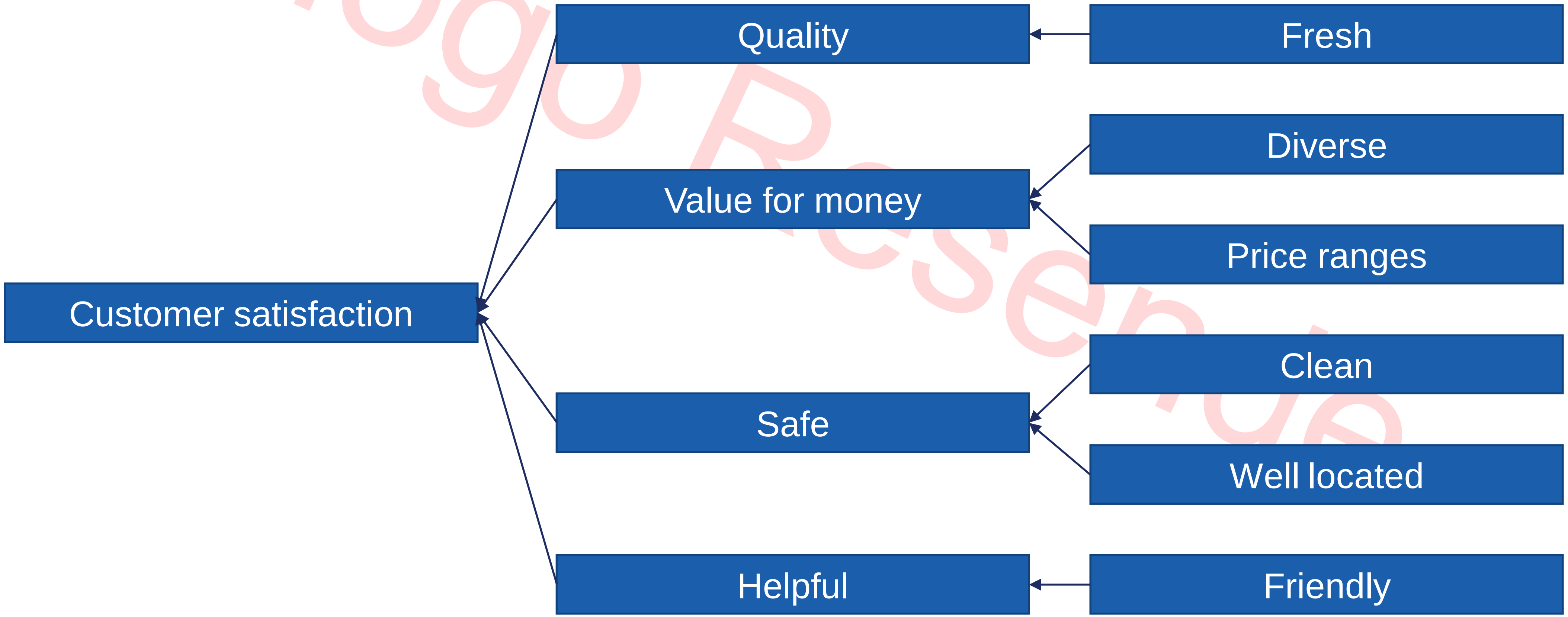
Random forest

- No Multicollinearity
- Provides significance
- Provides importance
- No causal relationships
- No impact estimate

https://www.udemy.com/course/e ferralCode=8665159C90FE02D1CB1A

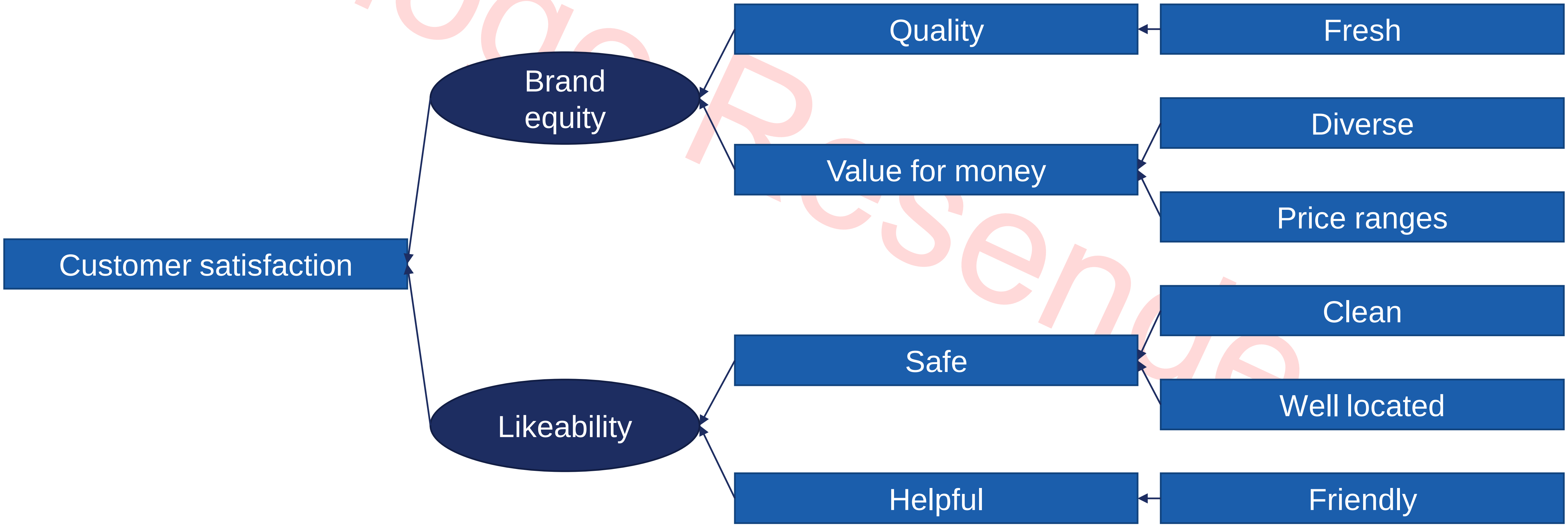
SEM helps understand causal relationships among drivers

5 **Structural Equation Modelling**



https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

SEM also helps with unmeasurable drivers



https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

What drives airlines' customer satisfaction

5 Structural Equation Modelling

Gender	Departure/arrival time convenient	On-board service
Customer Type	Ease of online booking	Leg room service
Age	Gate location	Baggage Handling
Type of travel	Food and drinks	Checkin service
Class	Online boarding	Inflight service
Flight distance	Seat comfort	Cleanliness
Inflight wifi service	Inflight entertainment	Departure delay
Arrival delay		

What drives airlines' customer satisfaction

5 **Structural Equation Modelling**

- Gender
- Customer Type
- Age
- Type of travel
- Class
- Flight distance
- Inflight wifi service
- Arrival delay

- Departure/arrival time convenient
- Ease of online booking
- Gate location
- Food and drinks
- Online boarding
- Seat comfort
- Inflight entertainment

- On-board service
- Leg room service
- Baggage Handling
- Checkin service
- Inflight service
- Cleanliness
- Departure delay

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

What drives airlines' customer satisfaction

5 Structural Equation Modelling

Gender	Departure/arrival time convenient	Gate location
Customer Type	Ease of online booking	Baggage Handling
Age	Online boarding	Checkin service
Type of travel	Food and drinks	Cleanliness
Class	Inflight entertainment	Seat comfort
Flight distance	Inflight service	Leg room service
Departure delay	Inflight wifi service	
Arrival delay	On-board service	

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

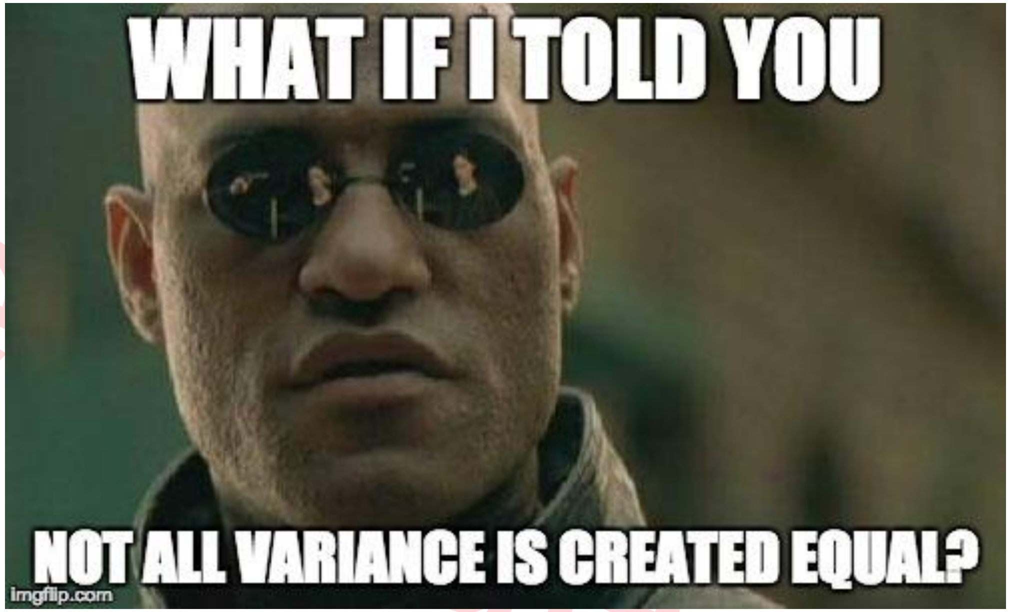
Structural Equation Modelling Step by Step

5 Structural Equation Modelling

Variable selection

https://www.udemy.com/course/econometrics-for-business/?referralCode=8665159C90FE02D1CB1A

PANEL DATA


WHAT IF I TOLD YOU
NOT ALL VARIANCE IS CREATED EQUAL?
imgflip.com

Pricing

Wolak, F. A. (2007).
Residential Customer Response to Real-time Pricing: The Anaheim Critical Peak Pricing Experiment.
UC Berkeley: Center for the Study of Energy Markets.
Retrieved from https://escholarship.org/uc/item/3td3n1x1

Stock Market

Anderson, E. W., & Mansi, S. A. (2009).
Does Customer Satisfaction Matter to Investors? Findings from the Bond Market.
Journal of Marketing Research, 46(5), 703–714.
https://doi.org/10.1509/jmkr.46.5.703

Willingness to pay

Anderson, E.W. Market Lett (1996)
Customer satisfaction and Price Tolerance 7: 265.
https://doi.org/10.1007/BF00435742

Customer Retention

van Triest, S., Bun, M.J.G., van Raaij, E.M. et al. Mark Lett (2009)
The impact of customer-specific marketing expenses on customer retention and customer profitability
20: 125.
https://doi.org/10.1007/s11002-008-9061-2

You have been asked to by an eletronic retailer to decide in which stores you should discount more

7 **Panel Data**

Why

- **Profitability**: your discounts need to generate enough volumes to compensate for the lower prices
- **Opportunity cost**: Your budget will most likely be limited so you need to optimize
- **Long term**: Discounting in the wrong areas can lead undesirable customer expectations

Factors that Panel Data helps control for

Which

- **Inter-entity variance**: There are factors that may vary accross entities, like cities, countries, population. However, these factors do not vary accross time
- **Unobserved/ unmeasured**: such factors would not be able to be measured through a regression and could lead fall into ommitted variable bias

Key idea

- **If a certain factor does not change over time, then any change in our Y variable, cannot be caused by ommitted variable bias**