What is Causal Inference

and why should Data Scientists know?

Ludvig Hult / 2019-10-31 / Stockholm

Reichenbach's Common Cause Principle

If two things – A and B - are dependent, then either: A causes B Or B causes A Or something else causes both A and B

↳ If A and B are correlated, they will not be spontanously correlated, there must be a reason why they are correlated.

The three tasks of Data Science

1. Description
What is there?

Where do we sell the most lager beers?

The answer is in the data

2. Prediction
What will happen?

How much beer will we sell in Germany in April?

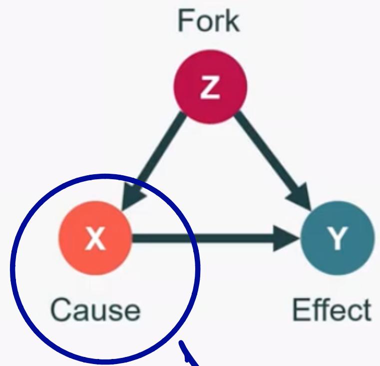
The answer is not directly in the data

3. Causal Inference
What would happen?

How much more beer will we sell if we buy more google ads?

Structural Causal Models: Example

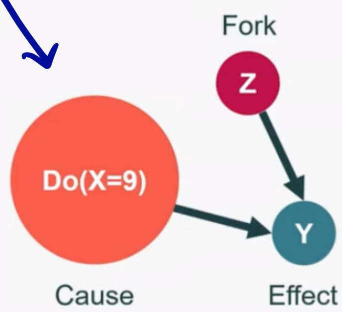
$Z := f_Z(U_Z)$
$X := f_X(Z, U_X)$
$Y := f_Y(X, Z, U_Y)$



we can interven the system

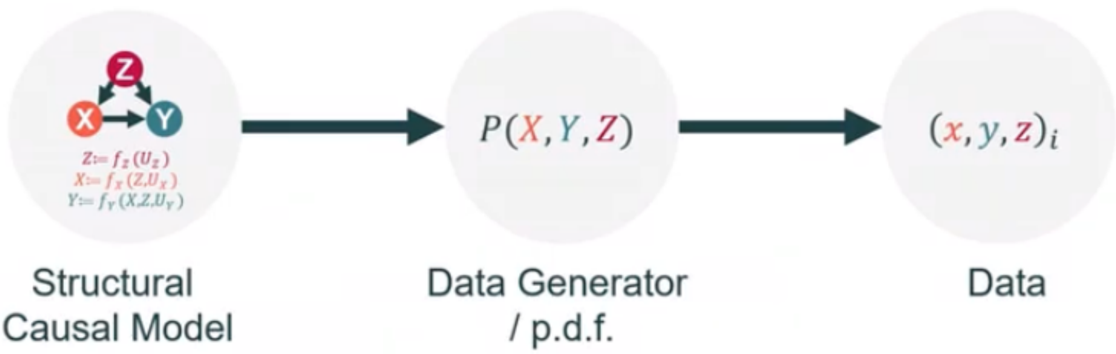
Structural Causal Models: Example

$Z := f_Z(U_Z)$
$X := 9$
$Y := f_Y(X, Z, U_Y)$



which in turn generates a probabilistic distribution

Structural Causal Models vs probabilistic model



The three tasks of Data Science

1

Description
What is there?

Where do we sell the most lager beers?

2

Prediction
What will happen?

How much beer will we sell in Germany in April?

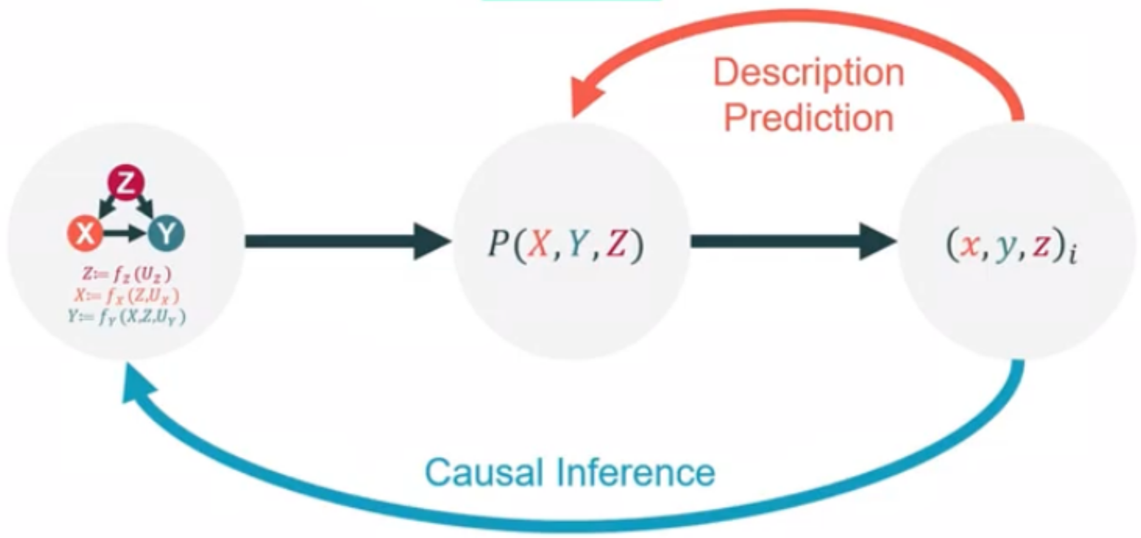
3

Causal Inference
What would happen?

How much more beer will we sell if we buy more google ads?

$\Downarrow$

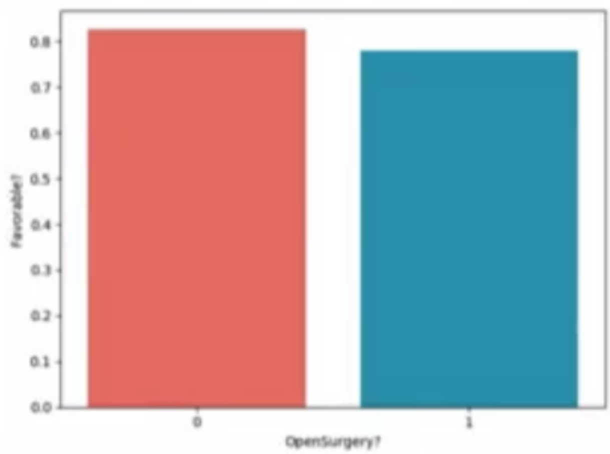
Structural Causal Models vs probabilistic model



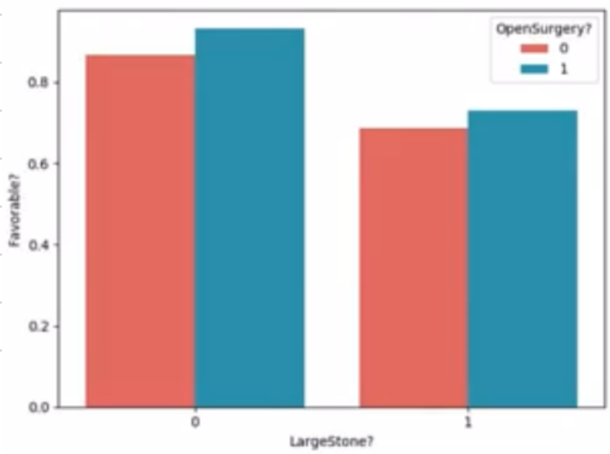
What is the Average Causal Effect?


```
In[2]: df.head()
Out[2]:
   OpenSurgery?  LargeStone?  Favorable?
0             0            0           0
1             0            0           0
2             0            0           0
3             0            0           0
4             0            0           0
```

1) On average, having open surgery seems to be a bit worse



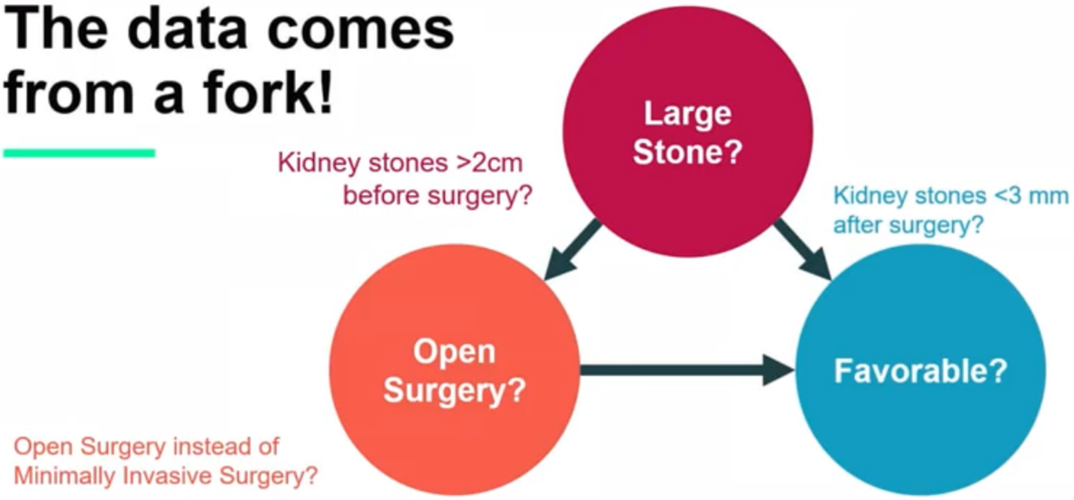
2) But when grouped by Largestone, it seems to be better for all groups?



How can this be possible?

↳ next slide

The data comes from a fork!



It is possible because we have a confounding variable.

Back door adjustment

$$ACE = E[y|do(X=1)] - E[y|do(X=0)]$$

Average Causal Effect

Assuming X, Y are binary, and Z is a valid adjustment set:

$$ACE = \sum_{x,z} p(x, y=1, z) \frac{(2x-1)}{p(x|z)}$$

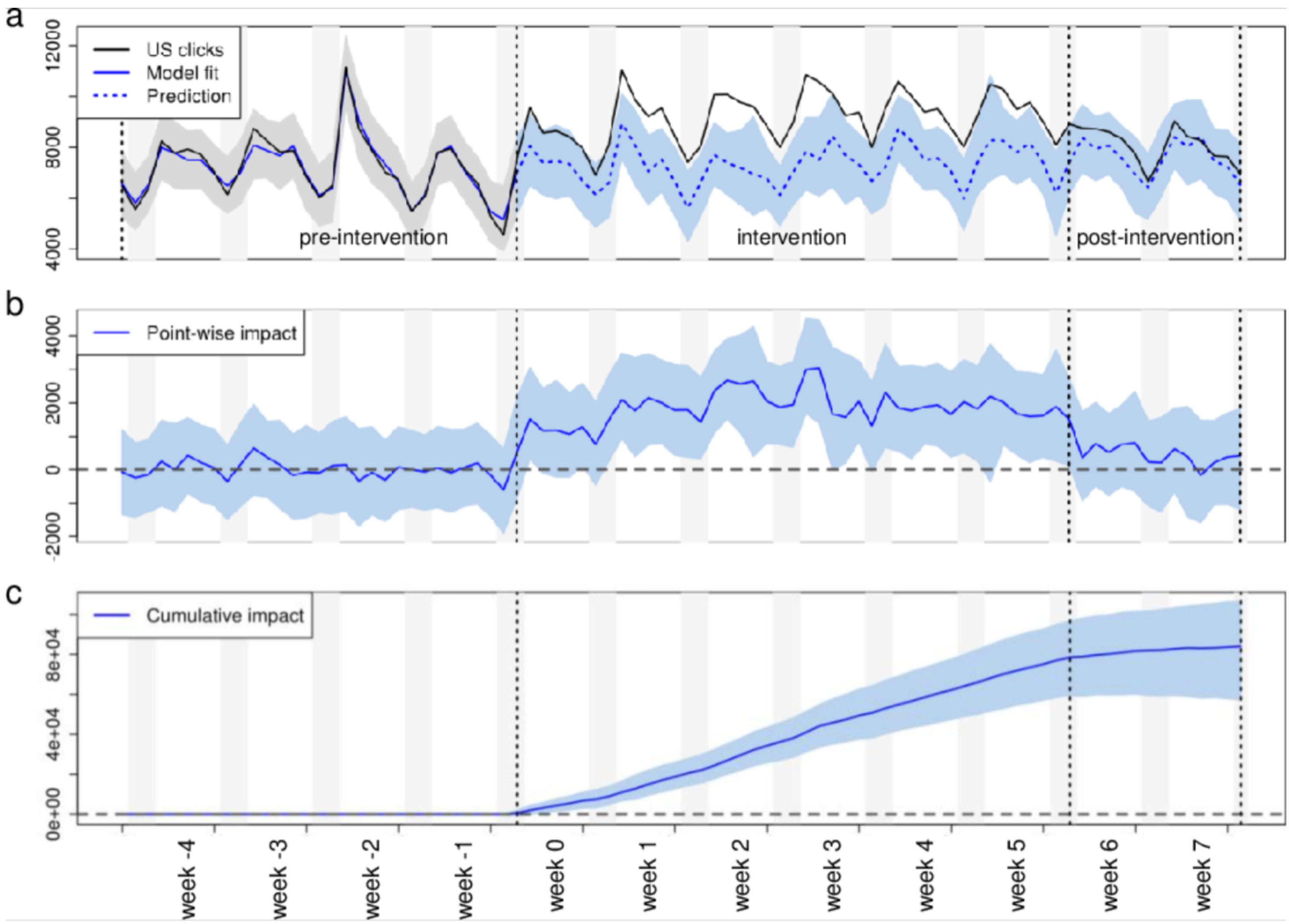
DoWhy adjusts correctly, given a graph!

```
In[5]: gml_graph=('graph [directed 1 ’
  ...:      'node [id 0 label "LargeStone?"]’
  ...:      'node [id 1 label "OpenSurgery?"]’
  ...:      'node [id 2 label "Favorable?"]’
  ...:      'edge [source 0 target 1]’
  ...:      'edge [source 0 target 2]’
  ...:      'edge [source 1 target 2]’
  ...:      ']')
```

```
In[6]: model = dw.CausalModel(
  ...:     data=df,
  ...:     treatment='OpenSurgery?’,
  ...:     outcome='Favorable?’,
  ...:     graph=gml_graph)
  ...:
  ...: identified_estimand = model.identify_effect(True)
  ...: estimate = model.estimate_effect(identified_estimand,
  ...:             method_name="backdoor.propensity_score_weighting")
  ...: f"The Average Causal Effect is {estimate.value:+.3%}"
Out[6]: 'The Average Causal Effect is +4.957%'
```

Google

Inferring the effect of an event using CausalImpact



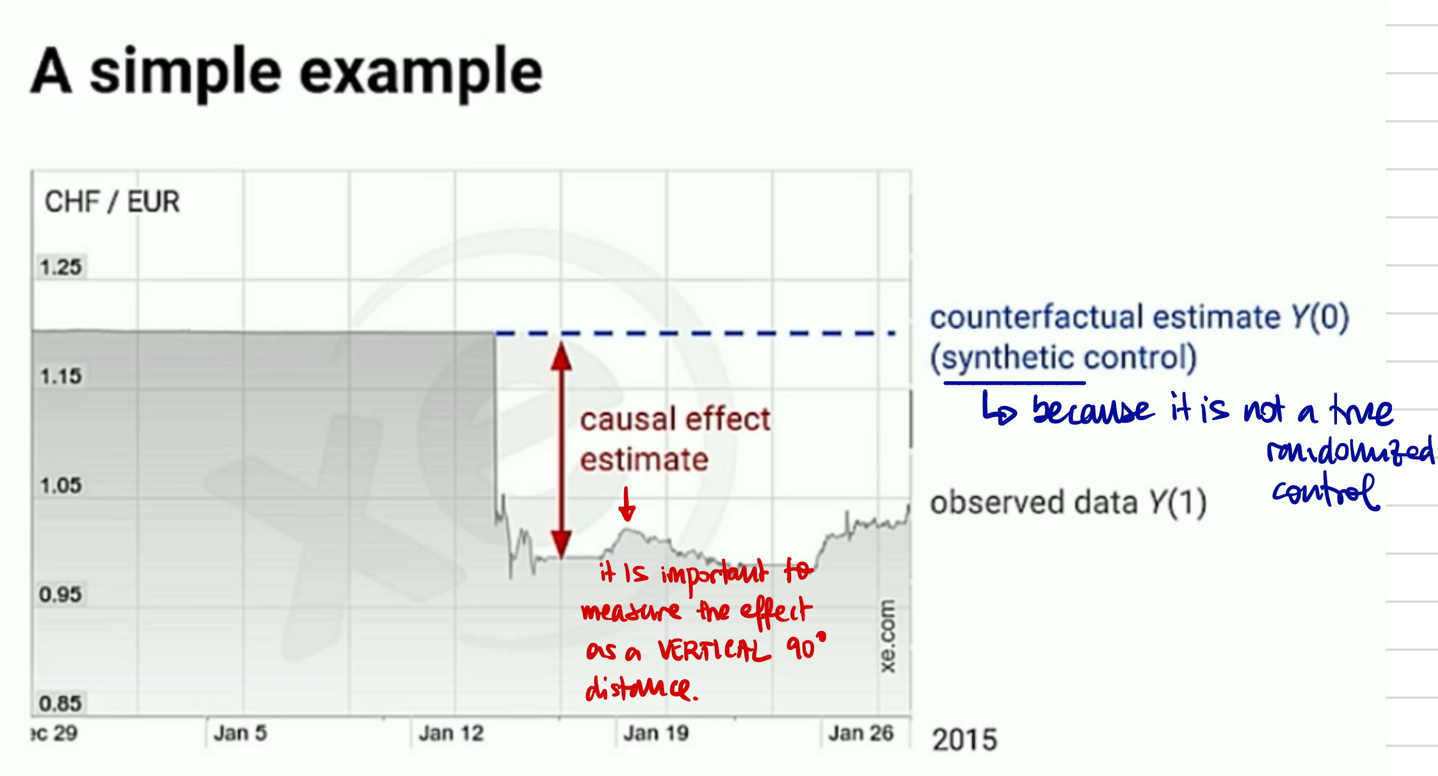
WHAT IS CAUSAL IMPACT?

- When no controlled randomized experiment can be designed and implemented, it would be very difficult to understand if a positive or negative effect of a change you made was true success, chance or an effect of seasonality.
- Causal Impact is an approach to estimating the effect of a designed intervention on a time series.

HOW DOES IT WORK?

Summary

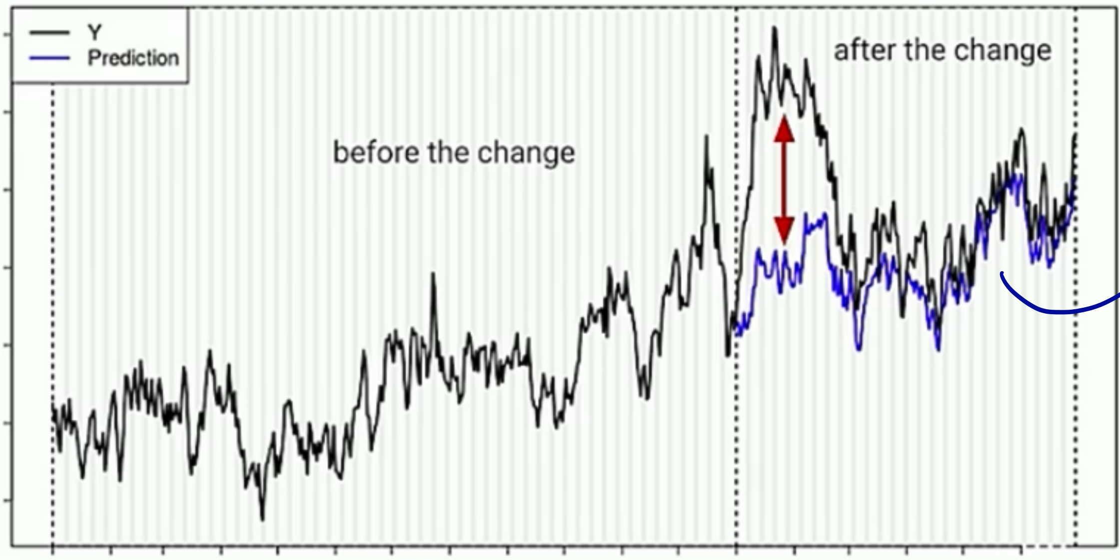
- 2 datasets
  - → A → control time-series (clicks in non-affected markets, clicks on versions of your site, etc)
  - → B → intervention time-series.
  - → Build a Bayesian Structural Time-Series model, which predicts the counterfactual:
    - → how would the response metric have evolved if the intervention had never existed?



In real life, clickstream data has a much higher volatility, and it will most certainly have trends and seasonality.

A harder example

however, the idea still remains the same, we want to predict the counterfactual.



Causal inference and potential outcomes

Unit Experiment	Treatment status T_i	Outcome under treatment $Y_i(1)$	Outcome under no treatment $Y_i(0)$	Covariates X_i
1	1	✓	estimate	✓

for example, a market.

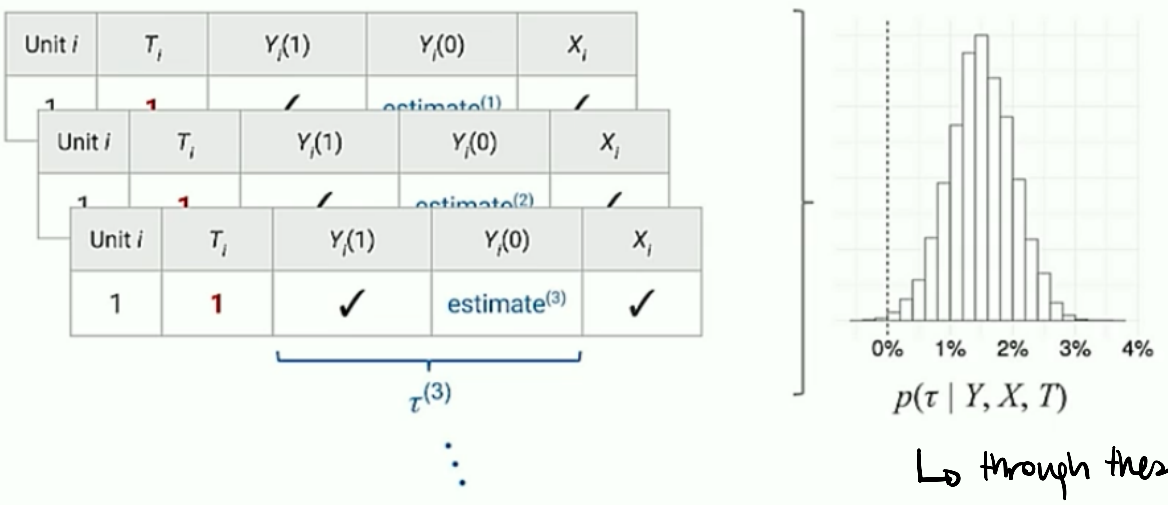
⇒ causal effect estimate

These are input features that:
1) Are not affected by the treatment when launched.
2) But that are predictive of the target before the treatment was launched.

For example:
- weather
- other markets that behave very similar to our treatment market?
- google trends, stock market, government indices.

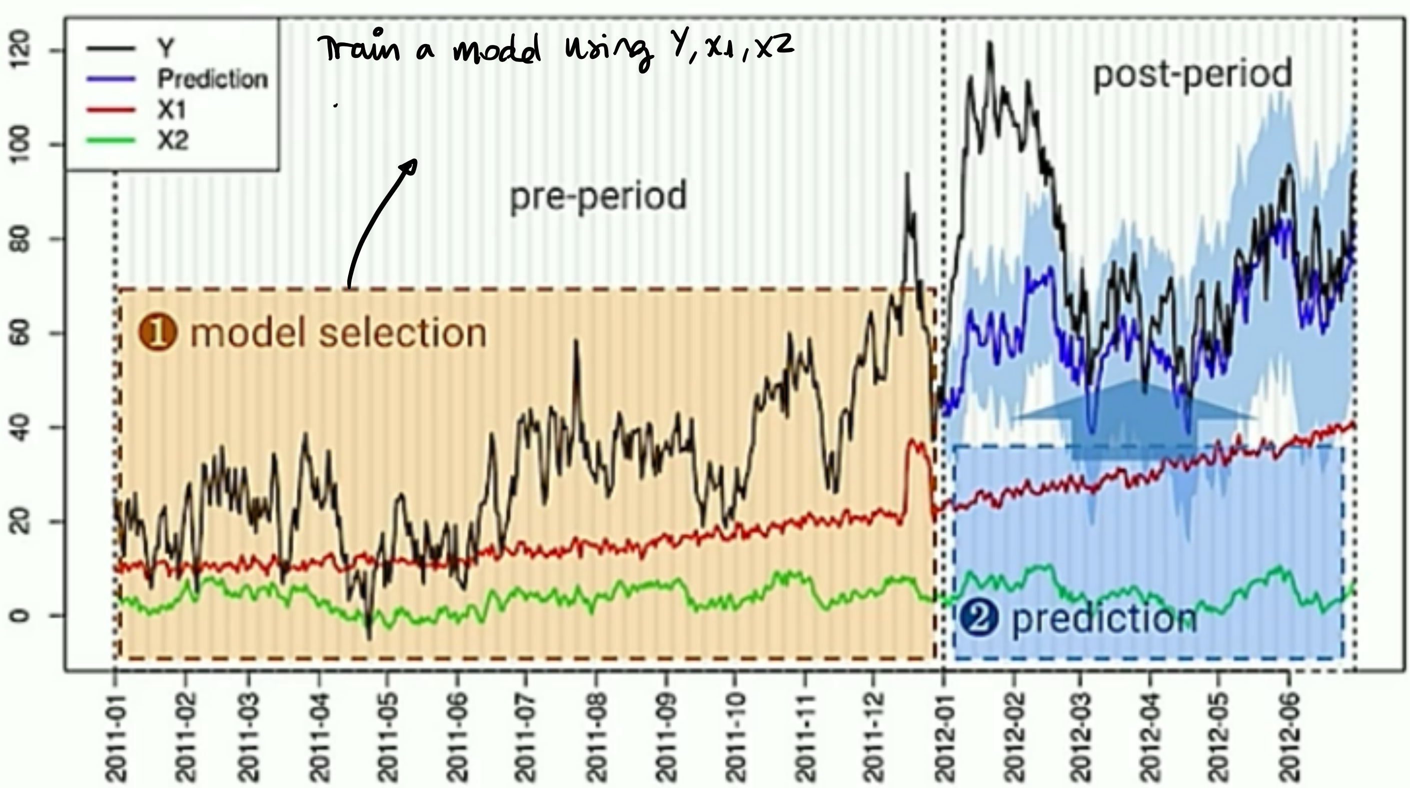
Inference

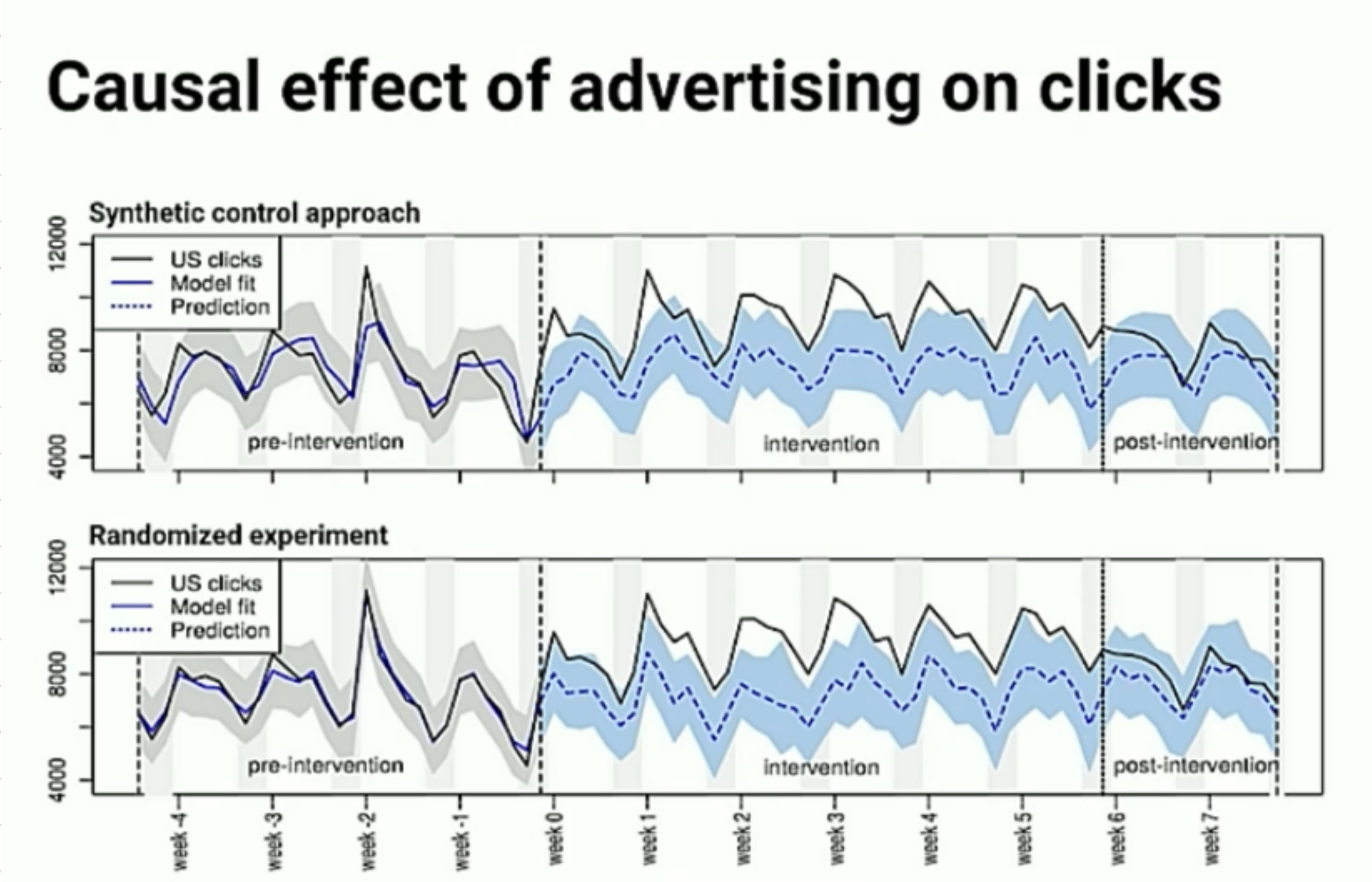
→ in order to estimate, we can use a Bayesian + MCMC simulations.



↳ through these simulations, we can get, not only an average point estimate, but also a credible interval.

Example

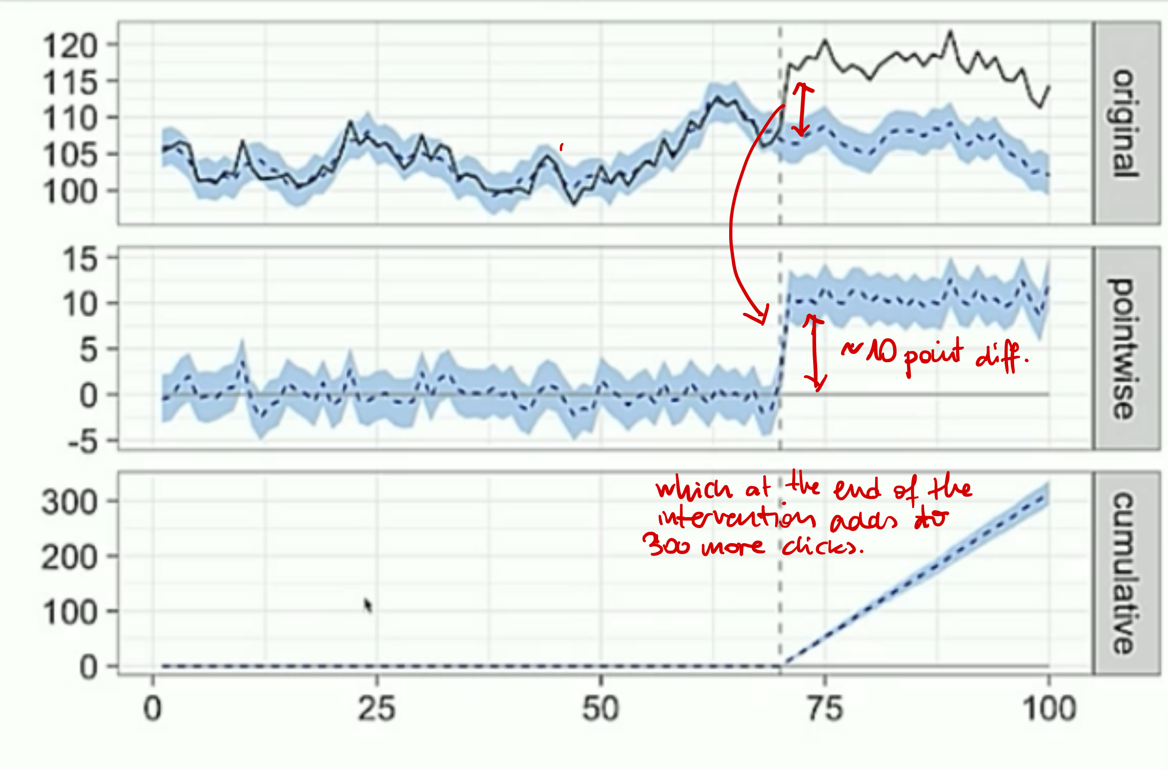




1. Randomized experiment chart.
   - Ideally, this would be the way of running the experiment
   - 50% of US $\underbrace{\text{have received treatment}}_{\text{black line}}$, $\underbrace{\text{50\% hasn't.}}_{\text{blue line.}}$
   - You can then measure, the delta in number of clicks between both groups during the intervention.

2. Sythetic control approach
   - You can see how the results are similar, even when we don't have access to a pseudo-controlled group.

How can we present results?




```
Posterior inference {CausalImpact}

                          Average          Cumulative
Actual                    117              3511
Prediction (s.d.)         107 (0.37)       3196 (11.15)
95% CI                    [106, 107]       [3175, 3219]

Absolute effect (s.d.)    11 (0.37)        315 (11.15)
95% CI                    [9.7, 11]        [292.5, 337]

Relative effect (s.d.)    9.9% (0.35%)     9.9% (0.35%)
95% CI                    [9.1%, 11%]      [9.1%, 11%]

Posterior tail-area probability p:   0.001
Posterior prob. of a causal effect:  99.9%

For more details, type: summary(impact, "report")
```

BEST PRACTICES

1) How can we avoid finding spurious correlations when looking for X?

↳ we can backtest the analysis → take a period before the treatment and treat it as a test set → is the result that there is no effect?

2) Do not rely on only 1 X predictor.
↳ you are at the mercy of this one predictor
↳ what happens if this predictor has many spikes? can you mitigate spiked effects with another timeseries?