Modeling Interference with Experiment Roll-out

Ari Boyarsky

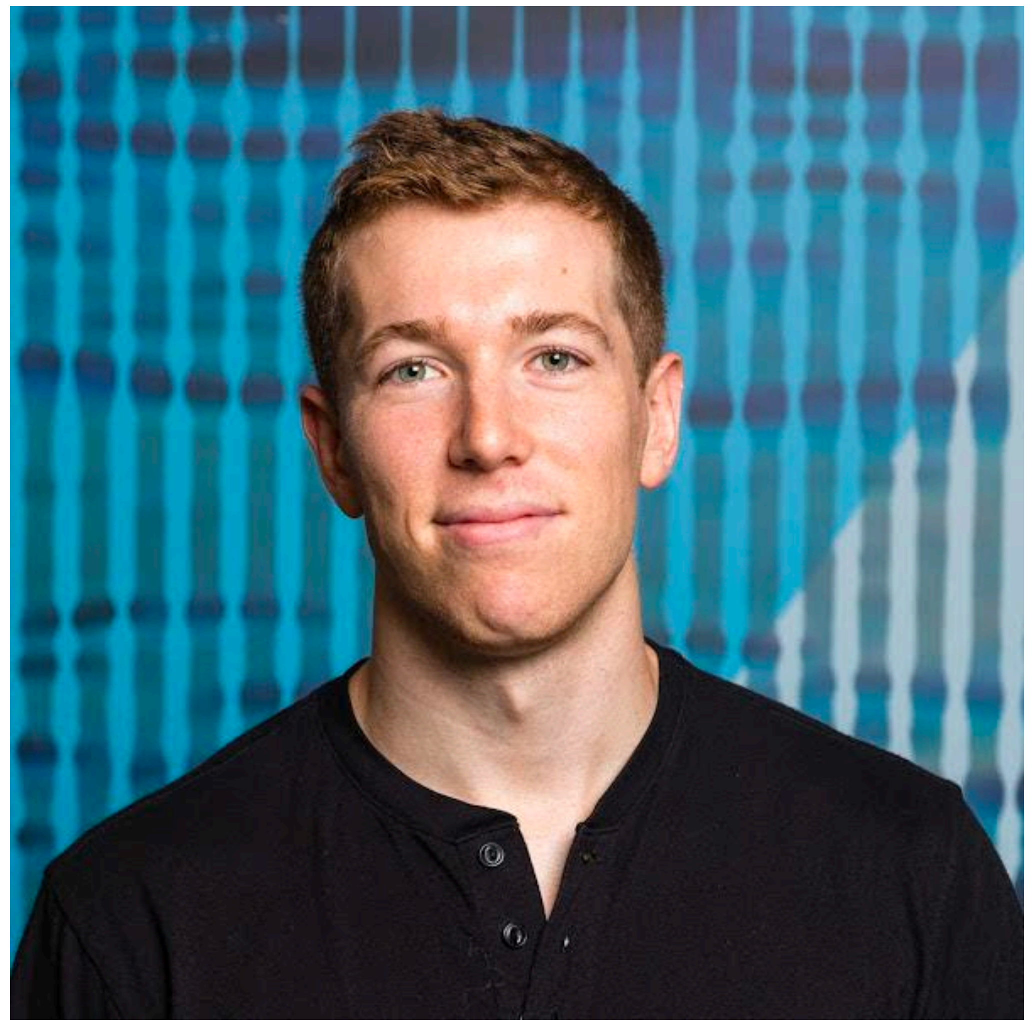
Columbia Business School

ariboyarsky.com

Hongseok Namkoong

Columbia Business School

hsnamkoong.github.io



Jean Pouget-Abadie

Google Research

jean.pouget-abadie.com

Outline

- Setting
- Motivating Example: Auctions
- Identification of Causal Effects
- Selecting between Outcome Models

The Problem of Interference

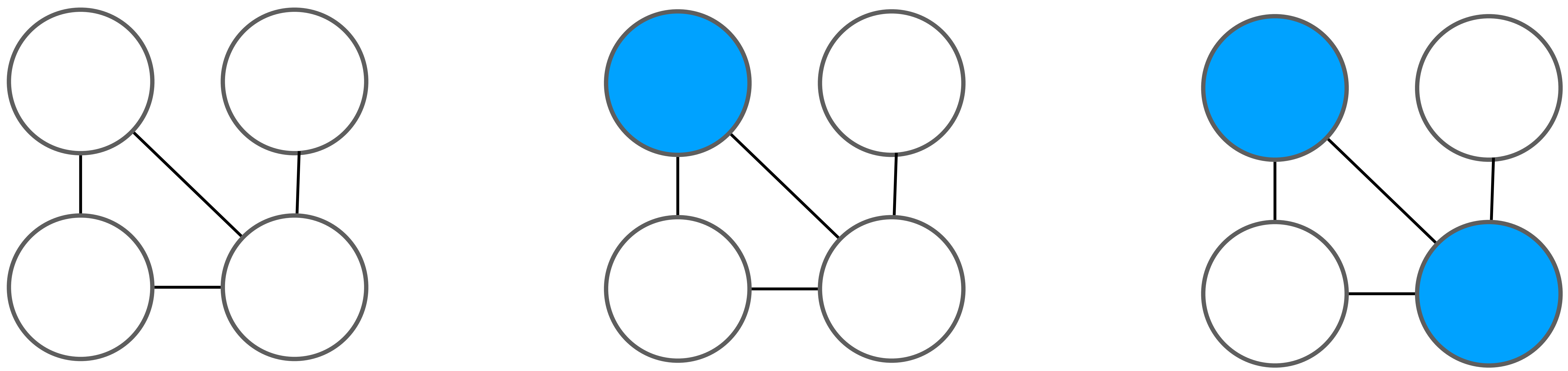
- Spillover effects between units violates a key implicit assumption: outcomes $Y_i(Z_i)$ depend only on unit i's treatment status

- But potential outcomes often depend on the treatment status of **others**! $\longrightarrow$ What we really have is: $Y_i = Y_i(z_1, \ldots, z_N)$

- Many applications
  - Marketplaces, e.g. ride-sharing
  - Vaccine Trials

Roll-out Designs and Interference

- Roll-outs are a universal ‘release mechanism’ used by online platforms to guard against ‘faulty’ changes

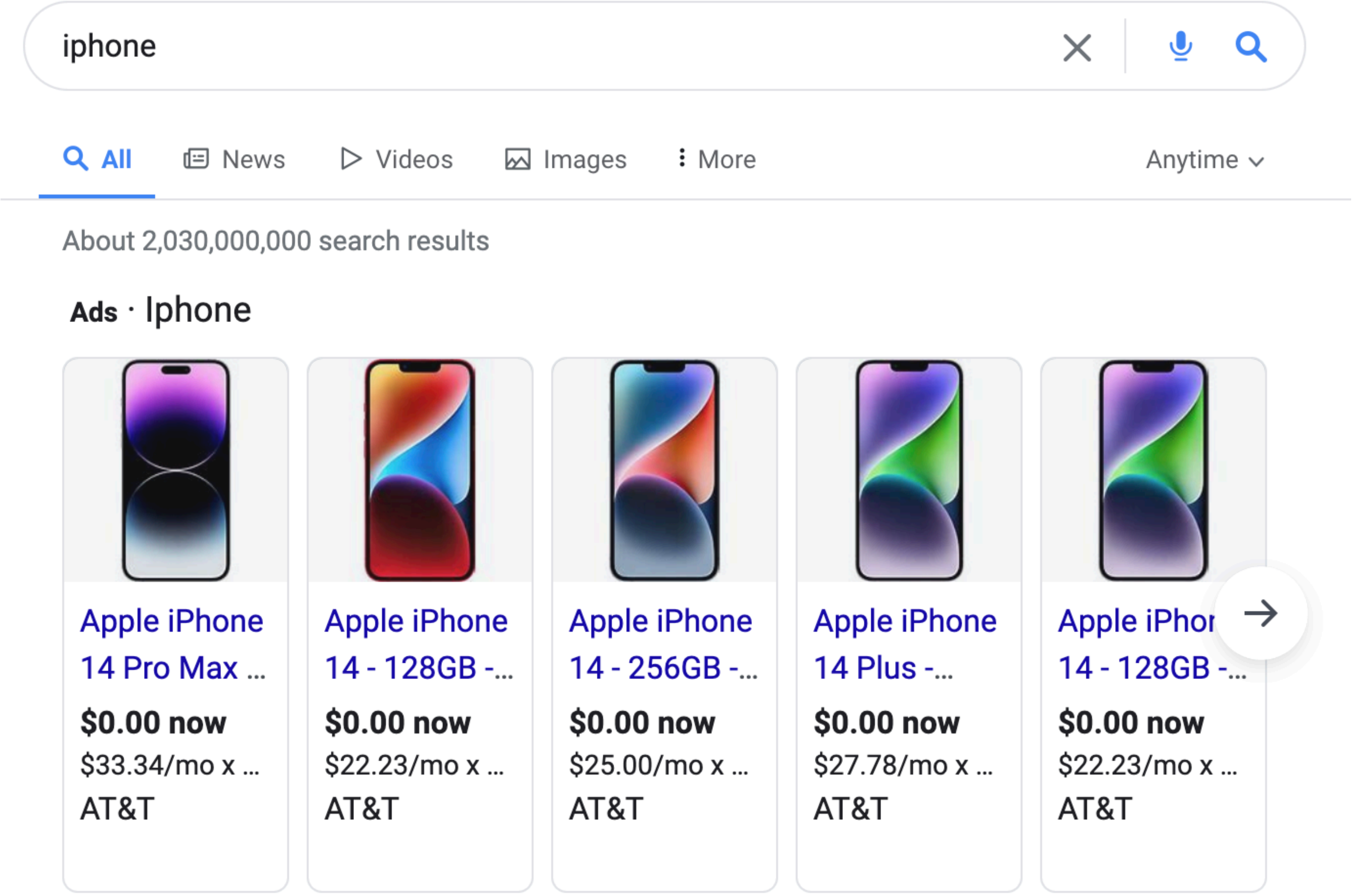


Period 1: 0% Treated Period 2: 25% Treated Period 3: 50% Treated

- See outcomes at multiple levels of treatment exposure
  - **If there is no interference roll-outs won’t change treatment effect**

Example: Second-Price Search Auctions

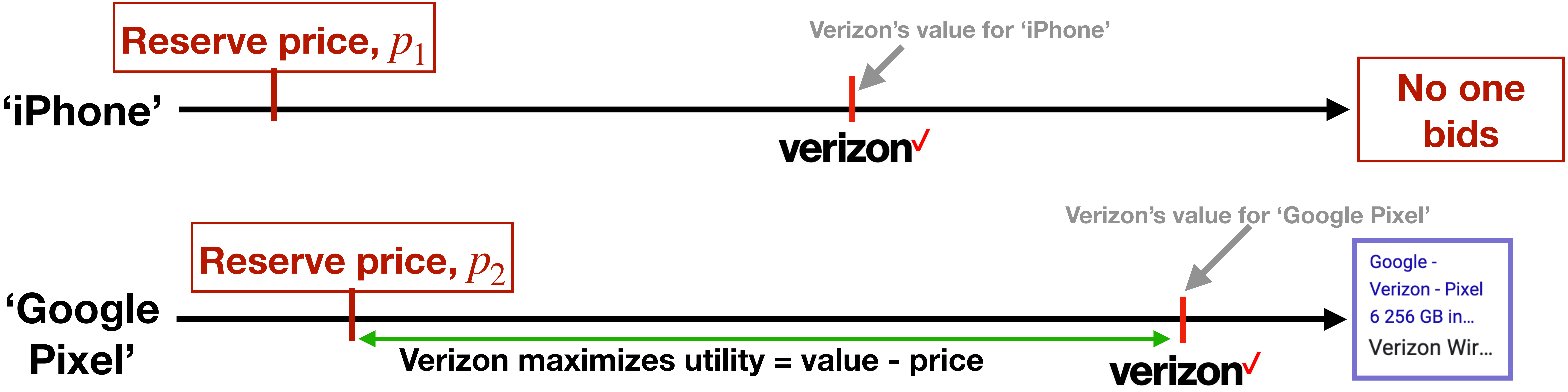
- **Units**: Keywords, e.g. ‘iPhone’
- **Treatment**: increase in reserve price
- **Outcome**: Total revenue from all auctions on keyword
- **Goal**: Choose reserve price policy that increases overall revenue
- **Challenge**: Interference



Status Quo

- Two keywords: ‘iPhone’ and ‘Google Pixel’
- 1 ad per keyword
- One bidder: Verizon *has budget for one auction*

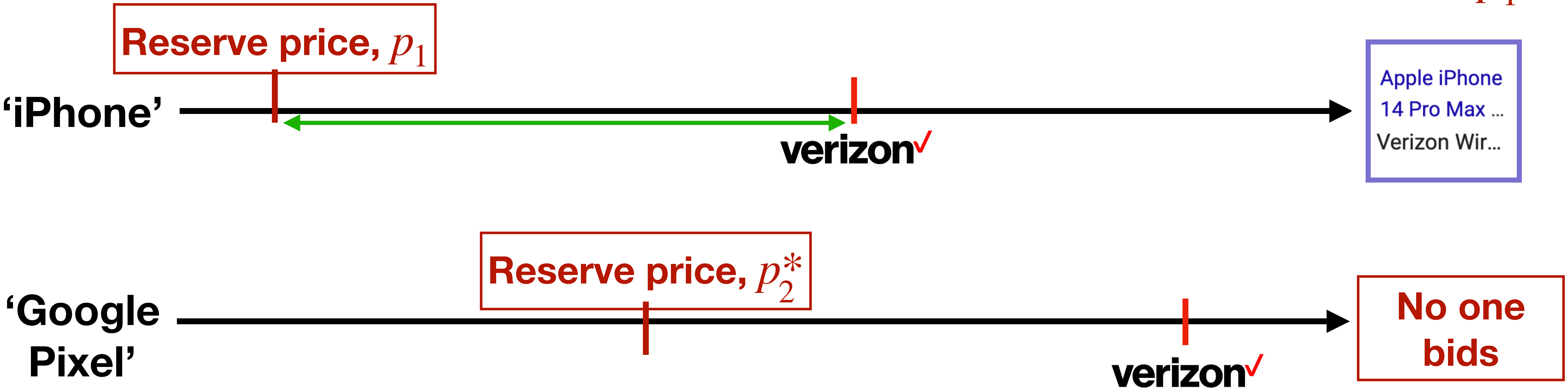
Total Revenue: p_2



Naive A/B Test

- **We treat only 'Google Pixel' keyword, use 'iPhone' as control**
- Budget constraint for Verizon binds
- Choose keyword that maximizes margin

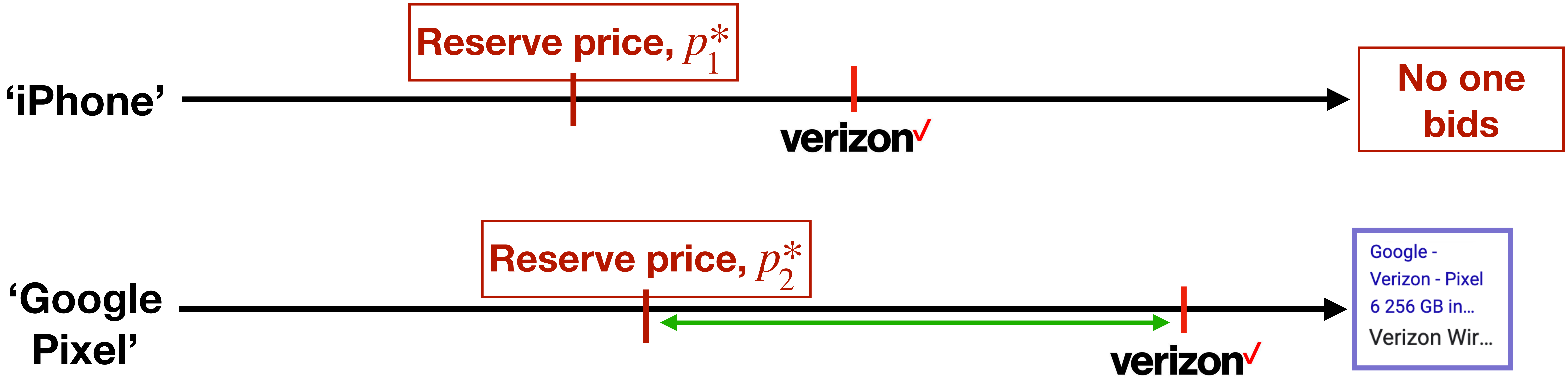
Total Revenue Decreases to: p_1



Counterfactual of Interest

- Two keywords: ‘iPhone’ and ‘Google Pixel’
- 1 ad per keyword
- ***Everyone is treated vs. no one***

Total Revenue Increases to: p_2^*



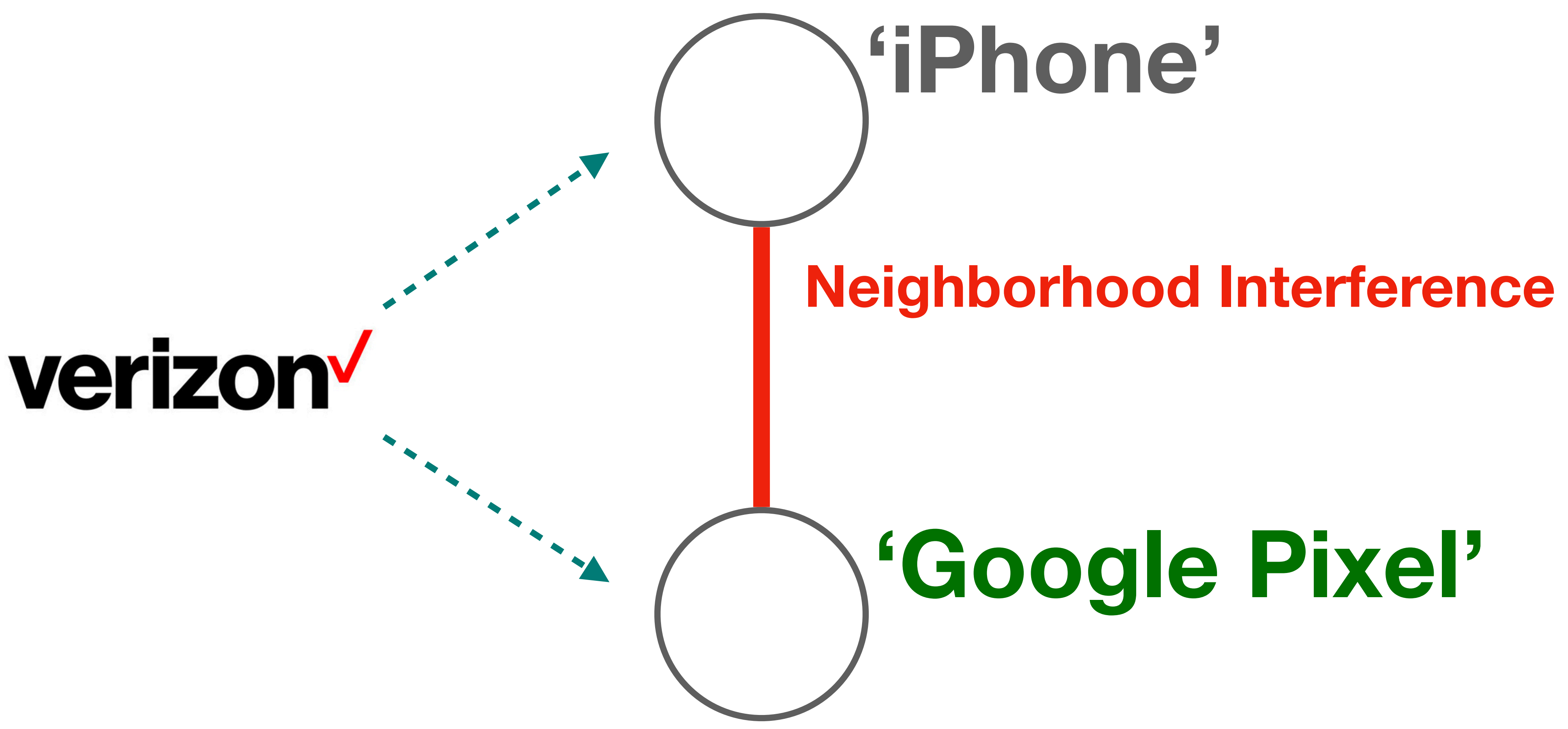
Models of Interference

- Modeling is one way to address interference
- Need to figure out whose treatment status matters for keyword i's outcome
  - e.g. keywords with similar advertisers
- **But determining if a model is correct is hard and often impossible**
  - How do we distinguish between good and bad models?

Multiple Graphs lead to Multiple Models

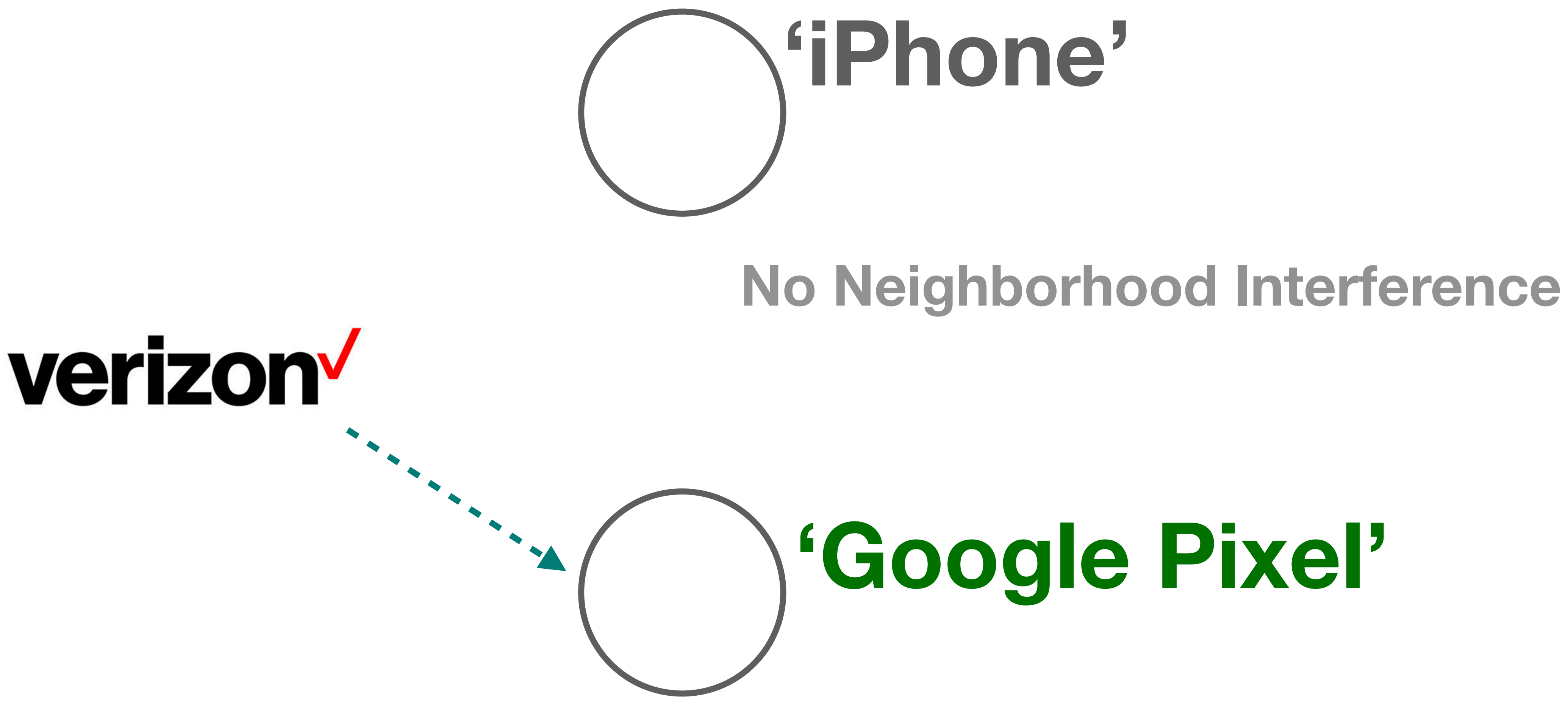
$$\text{Revenue for Keyword} \leftarrow f\left(\text{My Treatment}, \underbrace{\text{Neighbors Treated}}_{\text{Interference}}\right)$$

- How do we determine neighboring keywords?



Bid Amounts

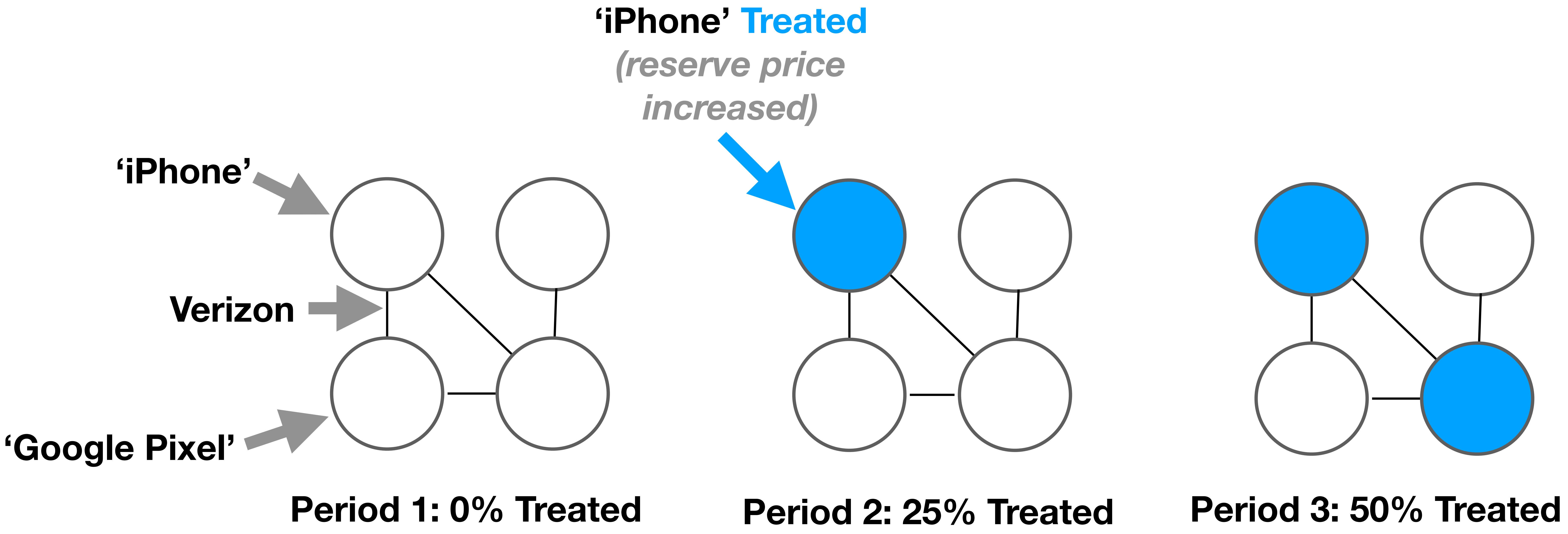
vs.



Winning Bids

Roll-out Designs and Interference

- To choose between interference models we need to observe different levels of treatment exposure
  - **Roll-outs induce temporal variation in treatment exposure — exploit for identification**



Identification and Estimation

- Roll-outs allow us to *identify* the **total treatment effect:** everyone treated vs. no-one:

$$TTE := \frac{1}{n}\sum_{i=1}^{n} Y_i(\overrightarrow{1}) - Y_i(\overrightarrow{0})$$

- Need interference to induce sufficient temporal variation into untreated units

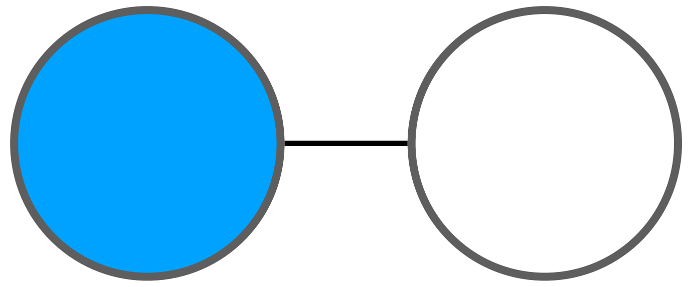
$$\textbf{Keyword Revenue} \leftarrow \underbrace{\tau}_{\text{'Direct' Effect}} \times \textbf{ Treated? } + \underbrace{\eta}_{\text{Interference Effect}} \times Average\left(\textbf{Neighbors Treated}\right)$$

$$+ \textbf{ Noise}$$

Identification and Estimation

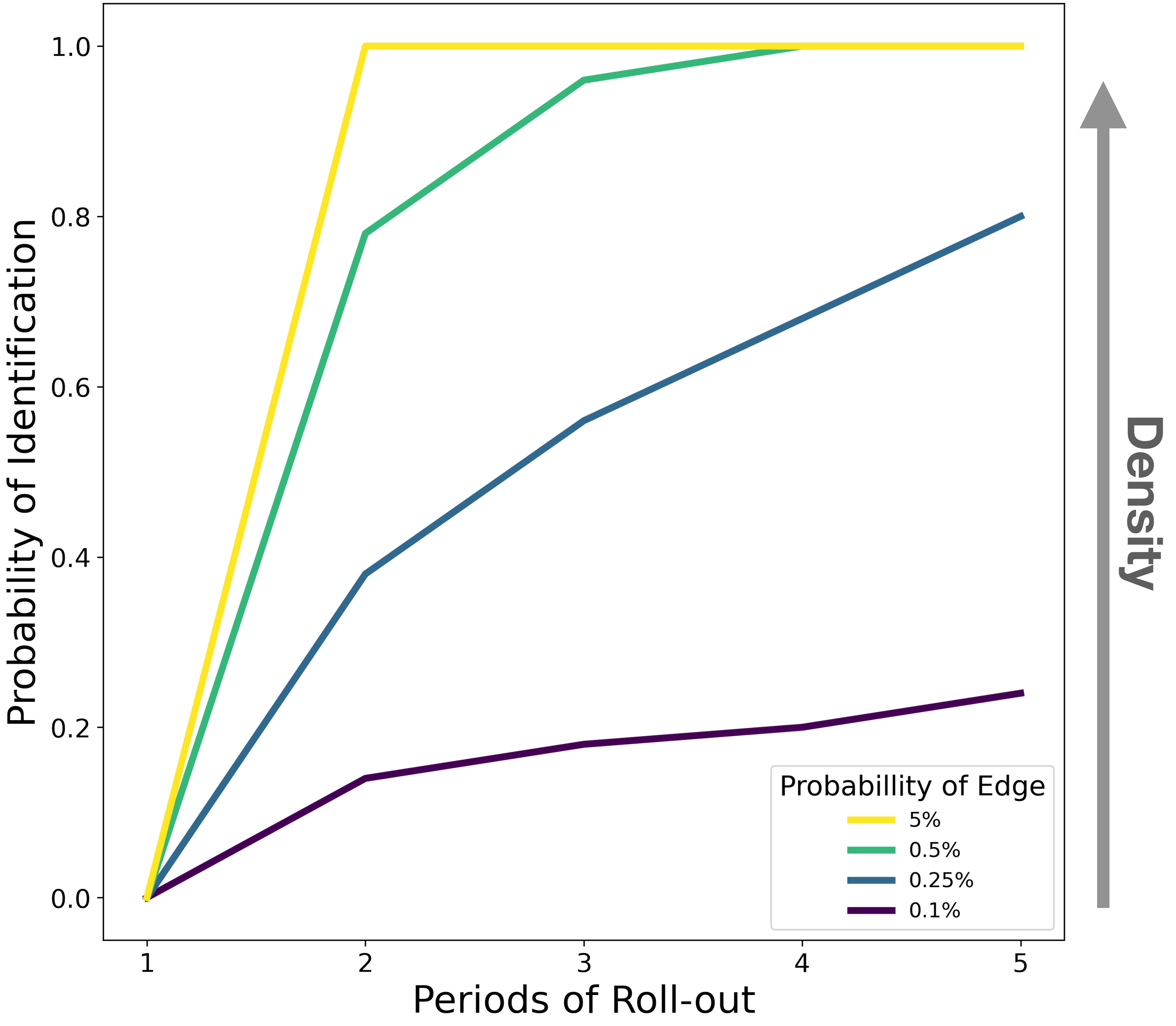
Possible Model: **Revenue for Keyword** $\leftarrow$ $\tau \times$ **Treated?**
$+\ \eta \times \mathit{Average}($**Neighbors Treated**$)\ +$ **Noise**

- **Theorem** (Identification):
  - ☑ $T > 1$ roll-out
  - ☑ *At* least one untreated unit is connected to at least one treated unit under selected network structure



➡Then can identify the total treatment effect (everyone vs. no-one treated)

How Likely is Identification



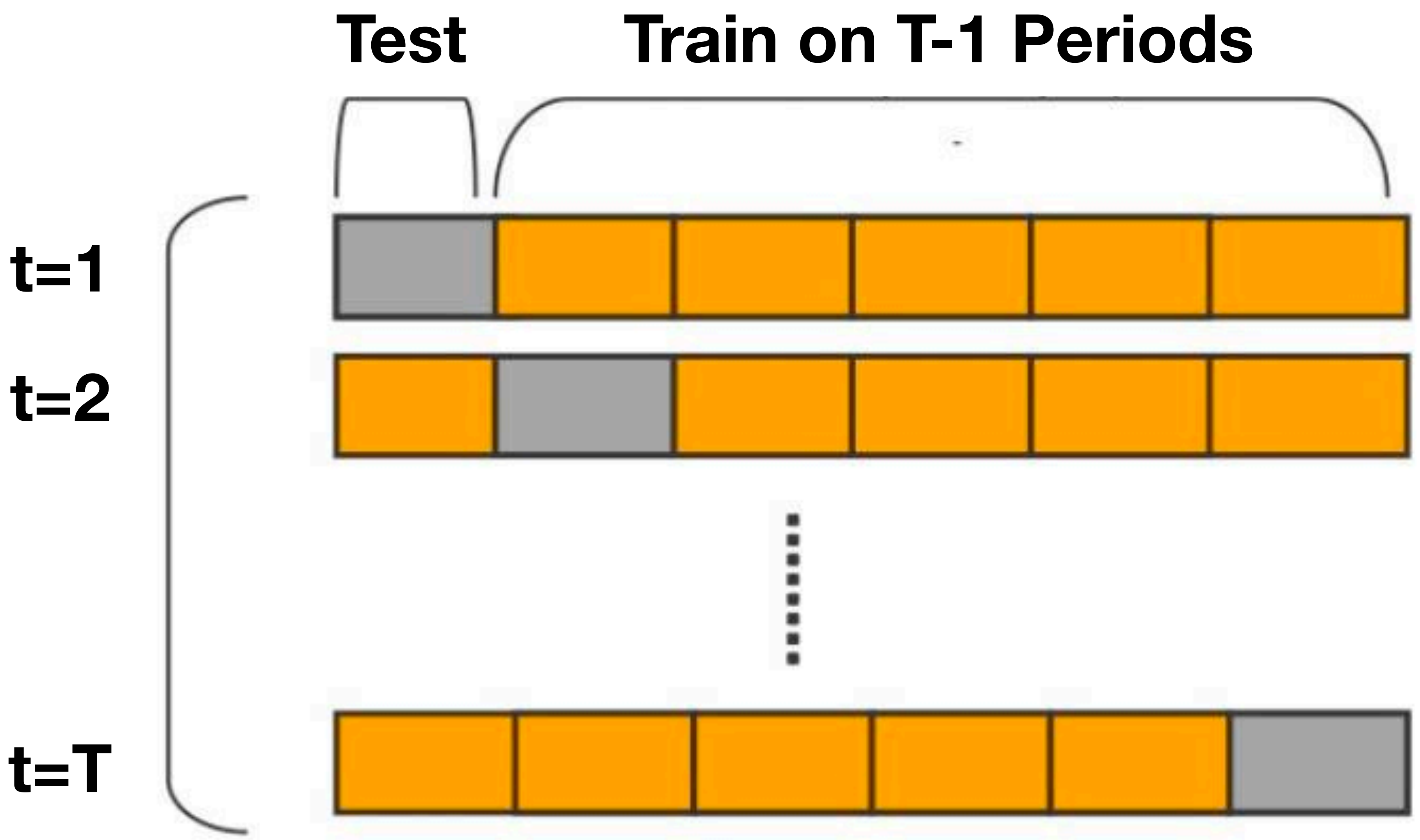
Key Takeaway

Roll-outs induce variation that helps identify parameters

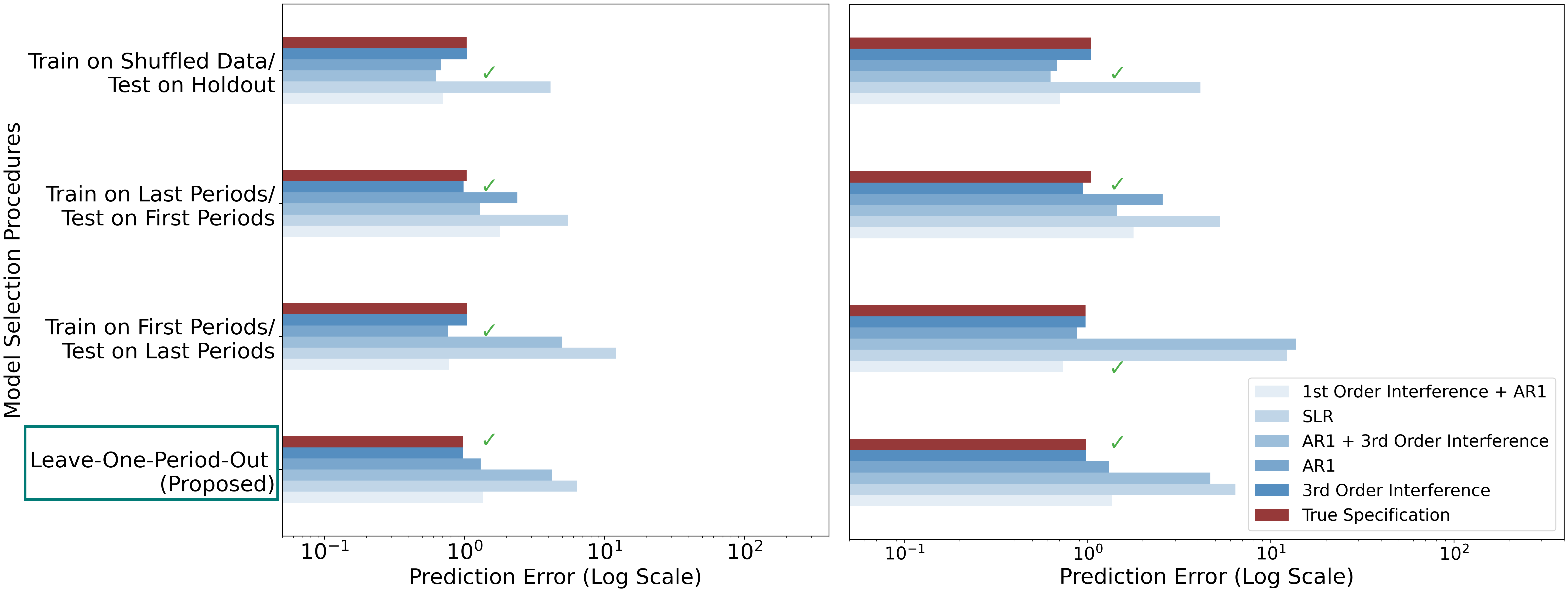
Model Selection Procedure

“Leave One Period Out (LOPO)”

- **Key Intuition:** Each period outcomes are under differing treatment exposures — exploit this variation in every period
- “Correct” outcome model must extrapolate to each period’s treatment exposure



Comparing Model Selection Procedures



Erdos-Renyi Graph with p=0.01 **Complete Graph**

Roll-outs provide us a mechanism to select between outcome models