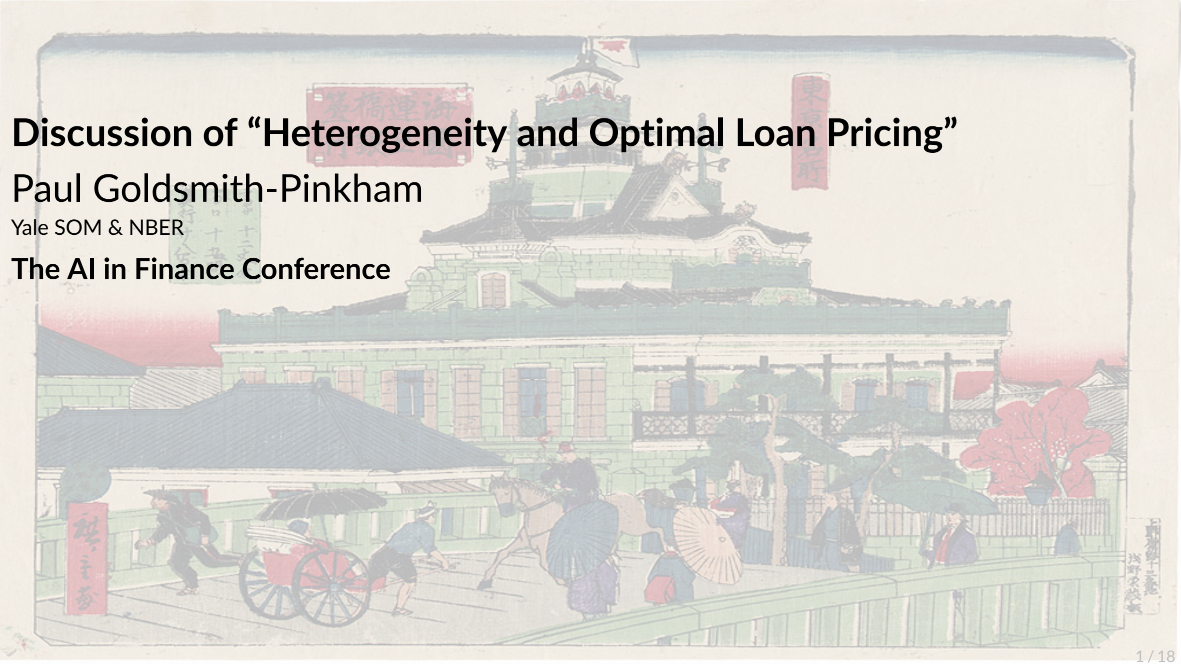


# Discussion of “Heterogeneity and Optimal Loan Pricing”

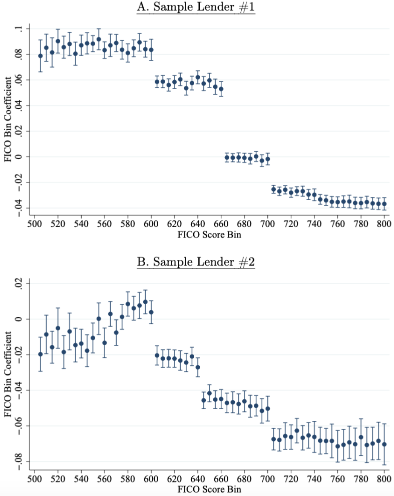
Paul Goldsmith-Pinkham
Yale SOM & NBER

**The AI in Finance Conference**

# Shifting interest rates shifts profits – a lovely paper

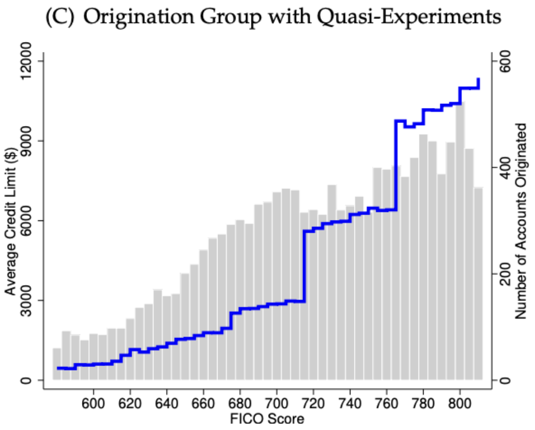
- Run experiment to randomize interest rates on loans
    - Motivated by fact that firms seem to have ”dumb” pricing strategies in lending
- Can firms do better?
    - Most deeply interesting question: *why* do firms do this?
- My goals today:
    1. Think about equilbirium consequences of shifting behavior
    2. Then, discuss some econometrics

Figure 1: Examples of FICO-Based Discontinuities in Interest-Rate Policies



# Shifting interest rates shifts profits – a lovely paper

- Run experiment to randomize interest rates on loans
  - Motivated by fact that firms seem to have ”dumb” pricing strategies in lending
- Can firms do better?
  - Most deeply interesting question: *why* do firms do this?
- My goals today:
  1. Think about equilbirium consequences of shifting behavior
  2. Then, discuss some econometrics



# How are prices determined? First, in an abstract sense

- Recall that total profits are

$$qP(q) - C(q) = q(P(q) - AC(q))$$

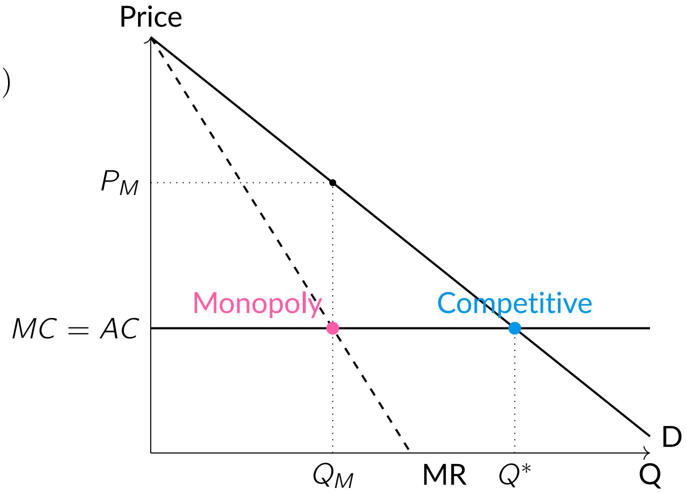
and maximization of profits depend on the market competition

- Under perfect competition:

$$P(q) = AC(q)$$

under monopoly:

$$P(q) = MC(q) \qquad \underbrace{-qP'(q)}_{\text{Marginal Consumer Surplus (MS(q))}}$$



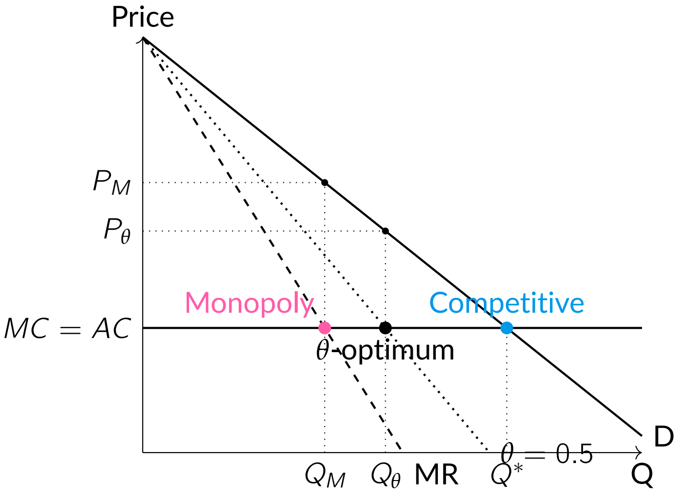
# How are prices determined?

- Mahoney and Weyl (2017) incorporate imperfect competition AND selection

- Imperfect competition as a linear mix:

$$P(q) = \theta\left[MS(q) + MC(q)\right] + (1-\theta)AC(q)$$

- Can be generated from different types of imperfect competition (Cournot, Bertrand)



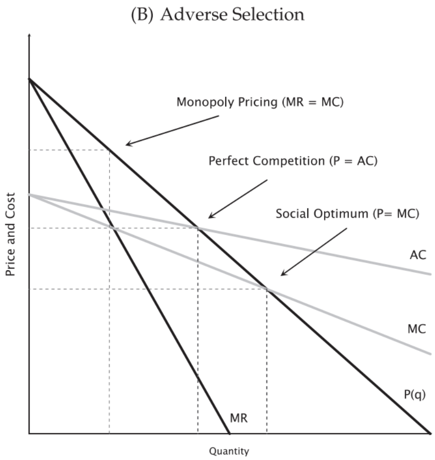
# How are prices determined?

- Selection rotates cost curve (adverse downwards)

$$P(q) = \theta MS(q) + \sigma \underbrace{\left[\theta MC(q) + (1-\theta)AC(q)\right]}_{\text{Average Cost under Selection}} \quad (1)$$

$$+ (1-\sigma) \underbrace{AC(1)}_{\text{Average Cost for Population}} \quad (2)$$

- Correlation between those who *value* the product and those who are costly to serve



# General mapping to Aydin, Skrastins, Sraer (2025)

- How should we intrepret the counterfactual proposed in this paper?
  - Improvement on *cost function* – reduction in adverse selection
  - Improvement on *pricing* – increase in market power

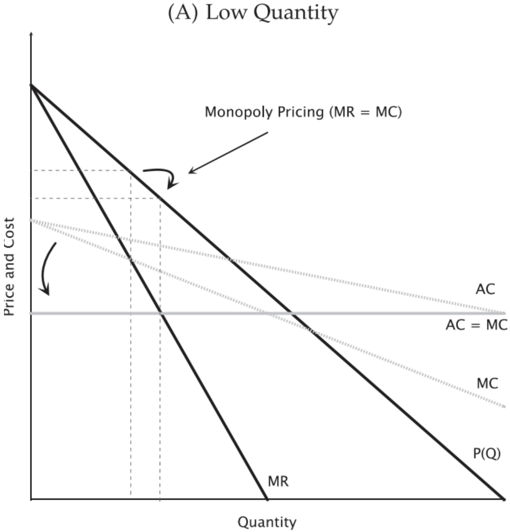
- Holding all else fixed, giving lenders the technology to differentiate on pricing encourages third-degree price discrimination and increases market power ($\theta$)

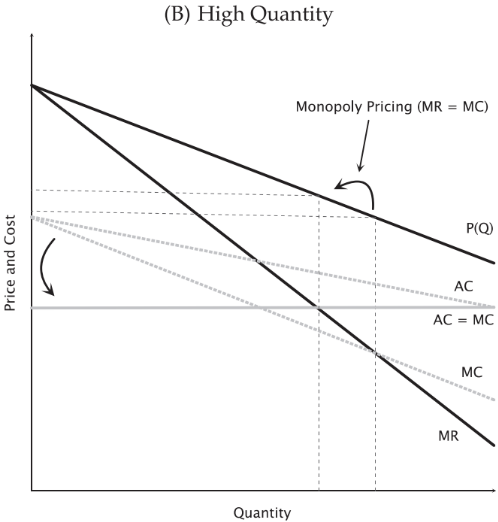
- Giving ability to differentiate on cost function reduces adverse selection ($\sigma$)

# General implications from Mahoney and Weyl (2017)

1. Market power increases producer surplus and decreases consumer surplus
2. With adverse selection, total social surplus falls with market power
3. Reducing adverse selection raises profits for all $\theta$
   - Better pricing in cost function with observables means higher profits
4. Reducing adverse selection under perfect competition increases consumer surplus **but** for $\theta < 1$, ambiguous effect on consumer surplus
   - Under monopoly, reducing adverse selection depends on the quantity level
   - At low quantities, reducing adverse selection raises consumer surplus
   - At higher quantities, reducing adverse selection reduces consumer surplus
   - Hence, the intermediate case of $\theta$ is also ambiguous

# Under monopoly, effect of reducing adverse selection depends on the quantity level

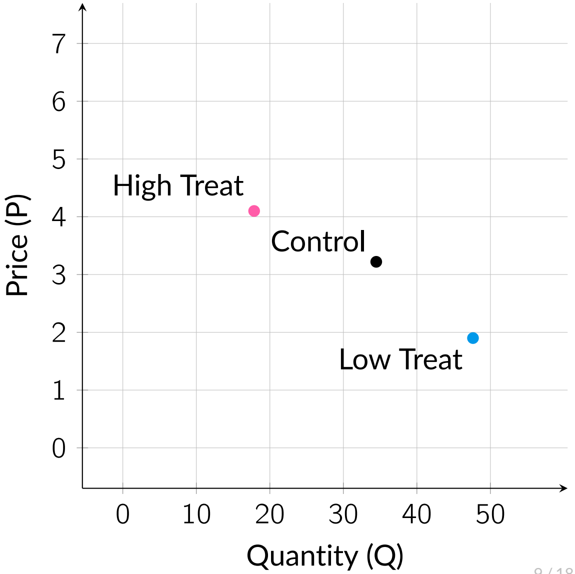


# In reality, is this what's going on?

- Run experiment to randomize interest rates on loans
  - Experiment traces out demand curve in aggregate
  - Caused by shifting the supply curve (cost curve) up and down

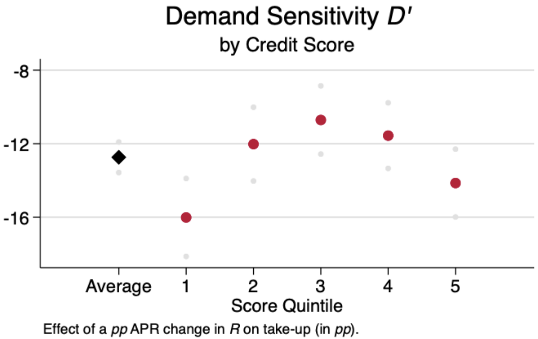
- However, key point is that the demand curves are combining different types of consumers



# In reality, is this what's going on?

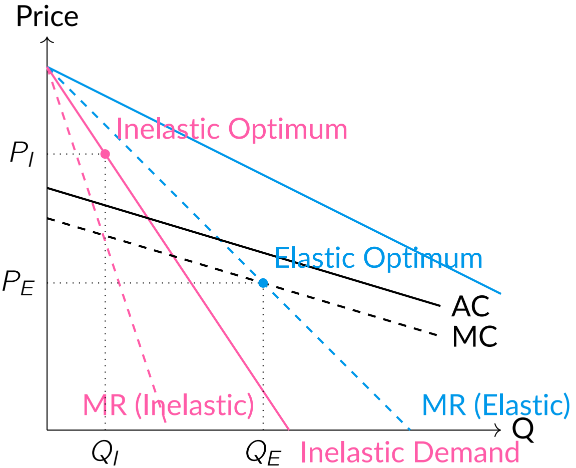
- Run experiment to randomize interest rates on loans
  - Experiment traces out demand curve in aggregate
  - Caused by shifting the supply curve (cost curve) up and down

- However, key point is that the demand curves are combining different types of consumers



# In reality, is this what's going on?

- Run experiment to randomize interest rates on loans
  - Experiment traces out demand curve in aggregate
  - Caused by shifting the supply curve (cost curve) up and down
- However, key point is that the demand curves are combining different types of consumers



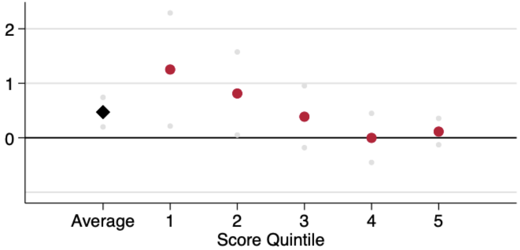
# What is the measure of costs capturing?

- However, the paper is also able to highlight how *costs* vary with changes in interest rates (unusual for IO!)

- Key point: the costs are only observed for *originated* loans
  - Typically an issue that this is combining extensive and intensive margin effects, (Lee (2009)) but not an issue here
  - Slope is steeper at low credit scores

- One implication: at higher credit scores, selection not an issue and market power dominates

Cost Sensitivity $C'$
by Credit Score



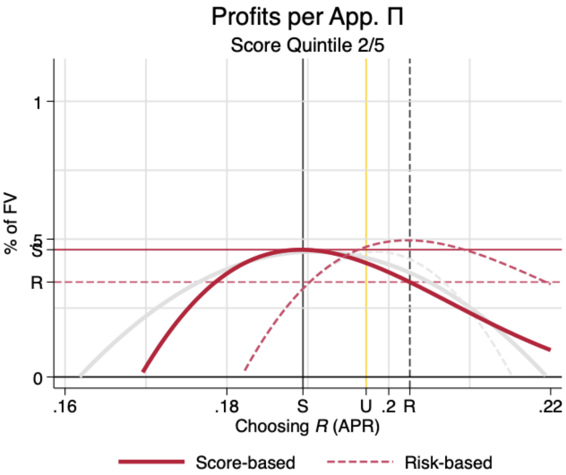

Effect of a *pp* APR change in *R* on cost of risk (as % of *FV*).

# Maximizing profits using experimental data

- The lack of variation in pricing suggests unsophisticated pricing – exactly how unsophisticated?
  - Directly estimate profits from the data + experiment

- Goal: define optimal $R(X)$ based on estimated $NPV(R,X)$, $AC(R,X)$ and $D(R,X)$ to maximize profits

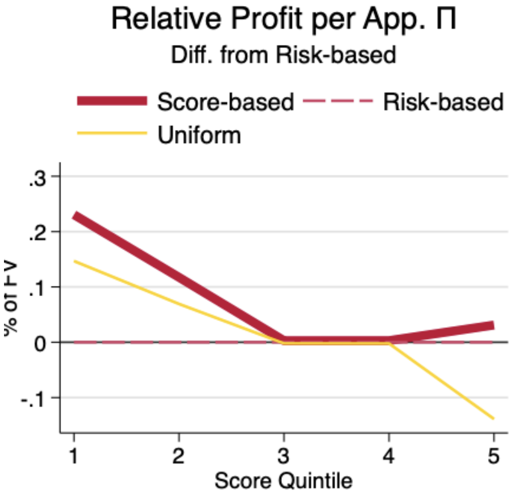
$$\Pi(\mathbf{R}) = \sum_X \Pi(R(X), X)$$
$$= \sum_X D(R(X), X) \times [NPV(R(X), X) - AC(R(X), X)]$$



# Maximizing profits using experimental data

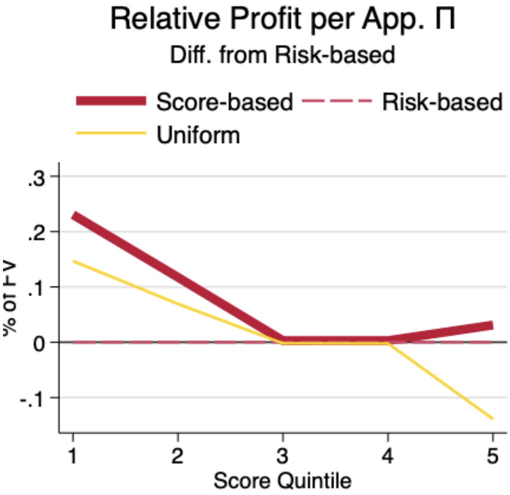
$$\Pi(\mathbf{R}) = \sum_X \Pi(R(X), X)$$
$$= \sum_X D(R(X), X) \times$$
$$[NPV(R(X), X) - AC(R(X), X)]$$

- With parameterized functions, this is a straightforward numerical optimization problem



# Maximizing profits using experimental data

- Evidence suggests that banks could do much better than they currently do for pricing decision, *just* from thinking more carefully about pricing with credit scores
  - Million dollar question is *why* banks don't (and haven't) priced more aggressively here
- Can we use more inputs beyond credit scores?

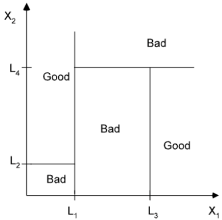


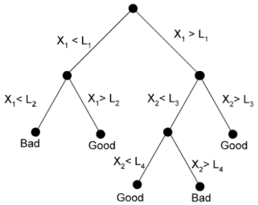
# So can we do more with machine learning?

- Simplest problem is to just bin borrowers by credit score and estimate changes in outcomes across treatment groups
    - These are CATEs: conditional average treatment effects

- With infinite data, we could do this with arbitrary number of variables, but in reality, curse of dimensionality bites
    - Prime location for machine learning
    - Significant recent advancements in adopting ML techniques to estimate CATE effects

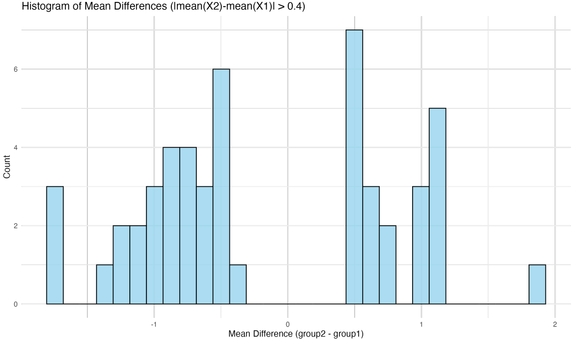
- This paper focuses on *tree* methods



# Why do we have to make things so complicated?

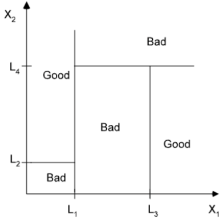
- First innovation in this space has been the use of "honest" methods
  - Consider $Y_i = \varepsilon_i$. $X$ is binary. $E(X_i) = 0.5$.
  - If we choose to split into $\tau(x) = E(Y_i|X_i x)$ only when the difference is large enough (which is what ML tree methods do), then we will get biased estimates of the difference (selection on noise)
  - Exacerbated with many covariates and low obs

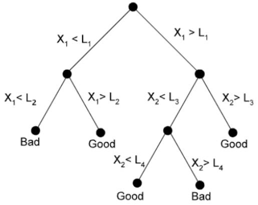


# Second innovation is to focus on different objective

- Athey and Wager (2019) propose *maximizing heterogeneity*, rather than minimizing error
    - This does a much better job of estimating CATEs

- This paper adds the profit maximization on top of this problem
    - Specifically, they maximize profits penalized by standard error (in original problem too?)



# Some comments on this approach

- Adding the profit maximization on top is a great addition
- I am not sure it's structural per se – it adds a linear combination of IV estimates on top of the standard GRF approach
  - Closest analogy is simulated method of moments
- How different would it look to just estimate the CATEs and then use the original optimization problem?
- How much is this relying on the extrapolation between points?
  - We really only have three points estimated on each curve

Table 12: Algorithm: Summary

*Training* Half
for Partitioning

- For each bootstrap $b \in \{1, \dots, B\}$:
  - Create candidate partition $\omega_b$
  - Estimate parameters $\hat{\theta}(\omega_b)$
  - Calculate objective $\mathcal{O}(\hat{\theta}(\omega_b))$
  - Repeat until objective maximizing partition $\omega_b^*$ identified

*Testing* Half
for *Honest* Estimation of Parameters and Objective

- For each bootstrap $b \in \{1, \dots, B\}$ and partition $\omega_b^*$:
  - Estimate sensitivities $\hat{\theta}_b^{\text{Honest}}(\omega_b^*)$
- Calculate $\bar{\theta} = \frac{1}{B}\sum_1^B \hat{\theta}_b^{\text{Honest}}$, which yields partition $\bar{\omega}$.
- Calculate objective $\mathcal{O}(\bar{\theta}(\bar{\omega}))$

| | Causal Forest | *Structural* Causal Forest |
|---|---|---|
| Objective $\mathcal{O}$ | $var(\hat{\theta})$ | $\Pi^*(\hat{\theta}(\omega_b))$ |

*Notes.*

# Some comments on this approach

- The paper avoids the best parts of ML by aggregating up the variables
  - Why do ML if you're going to reduce the dimensionality anyway?
- Part of the reason is to preserve interpretability, but this is a false goal
  - Rashomon effect
- What about using Chernozhukov et al. (2024) to estimate the CATEs directly, and then just calculate profits? How does that compare?
  - Then you can consider GATES as well

Table 14: Algorithm: Observables—Raw

| | | |
|---|---|---|
| *Liquidity* | Application | Day of month |
| | Application | Days since statement |
| | Application | Days since last application |
| | Application | Ratio of amount to max ever |
| | Utilization | Overdraft |
| | Utilization | Credit card |
| | Count | Loans (3m) |
| | Count | Applications (6m) |
| | Count | Checking withdrawals (3m) |
| *Experience* | Count | Accounts (ever) |
| | Count | Accounts (closed) |
| | First open | Credit card |
| | First open | Any account |

# Some comments on this approach

- The paper avoids the best parts of ML by aggregating up the variables
  - Why do ML if you're going to reduce the dimensionality anyway?
- Part of the reason is to preserve interpretability, but this is a false goal
  - Rashomon effect
- What about using Chernozhukov et al. (2024) to estimate the CATEs directly, and then just calculate profits? How does that compare?
  - Then you can consider GATES as well

$\hat{Y}$ vs $\hat{\beta}$