Improving Information from Manipulable Data

Alex Frankel Navin Kartik

July 2020

Allocation Problem

Designer uses data about an agent to assign her an allocation

Wants higher allocations for higher types

- Credit: Fair Isaac Corp maps credit behavior to credit score
  used to determine loan eligibility, interest rate, ...
  → Open/close accounts, adjust balances
- Web search: Google crawls web sites for keywords & metadata
  used to determine site's search rankings
  → SEO
- Product search: Amazon sees product reviews
  used to determine which products to highlight
  → Fake positive reviews

Given an allocation rule, agent will manipulate data to improve allocation

Manipulation changes inference of agent type from observables

Response to Manipulation

Allocation rule/policy $\rightarrow$ agent manipulation $\rightarrow$
inference of type from observables $\rightarrow$ allocation rule

- Fixed point policy: best response to itself
  - Rule is ex post optimal given data it induces
  - May achieve through adaptive process

- Optimal policy: commitment / Stackelberg solution
  - Maximizes designer's objective taking manipulation into account
  - Ex ante but (perhaps) not ex post optimal

Our interest:

1. How does optimal policy compare to fixed point?
2. What ex post distortions are introduced?

Fixed Point vs Optimal (commitment) policy

In our model:

1. How does optimal policy compare to fixed point?
   - Optimal policy is flatter than fixed point
     Less sensitive to manipulable data
2. What ex post distortions are introduced?
   - Commit to underutilize data
     Best response would be put more weight on data

Fixed Point vs Optimal (commitment) policy

Two interpretations of optimally flattening fixed point

- Designer with commitment power
  - Google search, Amazon product rankings, Government targeting
  - Positive perspective or prescriptive advice
- Allocation determined by competitive market
  - Use of credit scores (lending) or other test scores (college admissions)
  - Market settles on ex post optimal allocations
  - What intervention would improve accuracy of allocations?
    (Govt policy or collusion)

Related Literature

- Framework of "muddled information"
  - Prendergast & Topel 1996; Fischer & Verrecchia 2000; Benabou & Tirole 2006; Frankel & Kartik 2019
  - Ball 2020
  - Björkegren, Blumenstock & Knight 2020
- Related "flattening" to reduce manipulation in other contexts
  - Dynamic screening: Bonatti & Cisternas 2019
  - Finance: Bond & Goldstein 2015; Boleslavsky, Kelly & Taylor 2017
- Other mechanisms/contexts to improve info extraction
- CompSci / ML: classification algorithms with strategic responses

Background on Framework

Information Loss

In some models, fixed point policy yields full information,
so no need to distort

- When corresponding signaling game has separating eqm

Muddled information framework (FK 2019)

- Observer cares about agent's natural action η
  - Agent's action absent manipulation
- Agents also have heterogeneous gaming ability γ
  - Manipulation skill, private gain from improving allocation, willingness to cheat
- No single crossing: 2-dim type; 1-dim action
- When allocation rule rewards higher actions,
  high actions will muddle together high η with high γ

Muddled Information

Frankel & Kartik 2019

- Market information in a signaling equilibrium
  Analogous to fixed point in current paper

- Agent is the strategic actor
  - chooses x to maximize $V(\hat{\eta}(x), s) - C(x; \eta, \gamma)$
  - x is observable action, $\hat{\eta}$ is posterior mean,
    s is stakes / manipulation incentive
  - leading example: $s\hat{\eta}(x) - \frac{(x-\eta)^2}{\gamma}$

- Allocation implicit: agent's payoff depends on market belief

- Key result: higher stakes $\implies$ less eqm info (about natural action)
  - suitable general assumptions on $V(\cdot)$ and $C(\cdot)$
  - precise senses in which the result is true

Current paper **explicitly models allocation problem**;
How to use commitment to ↓ info loss and thereby ↑ alloc accuracy

Model

Designer's problem

- Agent(s) of type $(\eta, \gamma) \in \mathbb{R}^2$
- Designer wants to match allocation $y \in \mathbb{R}$ to natural action η:

$$\text{Utility} \equiv -(y-\eta)^2$$

- Allocation rule $Y(x)$, based on agent's observable $x \in \mathbb{R}$
- Agent chooses x based on (η, γ) and Y (details later)

- Expected loss for designer:

$$\boxed{\text{Loss} \equiv \mathbb{E}[(Y(x)-\eta)^2]}$$

Nb: pure allocation/estimation problem

- Designer puts no weight on agent utility
- Effort is purely "gaming"

Designer's problem

- Agent(s) of type $(\eta, \gamma) \in \mathbb{R}^2$
- Designer wants to match allocation $y \in \mathbb{R}$ to natural action η:

$$\text{Utility} \equiv -(y-\eta)^2$$

- Allocation rule $Y(x)$, based on agent's observable $x \in \mathbb{R}$
- Agent chooses x based on (η, γ) and Y (details later)

- Expected loss for designer:

$$\boxed{\text{Loss} \equiv \mathbb{E}[(Y(x)-\eta)^2]}$$

Useful decomposition:

$$\text{Loss} = \underbrace{\mathbb{E}[(\mathbb{E}[\eta|x]-\eta)^2]}_{\text{Info loss from estimating } \eta \text{ from } x} + \underbrace{\mathbb{E}[(Y(x)-\mathbb{E}[\eta|x])^2]}_{\text{Misallocation loss given estimation}}$$

Linearity assumptions

We will focus on

- Linear allocation policies for designer:

$$\boxed{Y(x) = \beta x + \beta_0}$$

  - β is allocation sensitivity, strength of incentives

- Agent has a linear response function:

  Given policy (β, β_0), agent of type (η, γ) chooses

$$\boxed{x = \eta + m\beta\gamma}$$

  Parameter $m > 0$ captures manipulability of the data (or stakes)

  Such response is optimal if agent's utility is, e.g.,

$$y - \frac{(x-\eta)^2}{2m\gamma}$$

Summary of designer's problem

- Joint distribution over (η, γ)
  - Means μ_η, μ_γ; finite variances $\sigma_\eta^2, \sigma_\gamma^2 > 0$; correlation $\rho \in (-1, 1)$
  - $\rho \geq 0$ may be more salient, but $\rho < 0$ not unreasonable
  - Main ideas come through with $\rho = 0$

- Designer's optimum (β^*, β_0^*) minimizes expected quadratic loss:

$$\min_{\beta,\beta_0} \mathbb{E}\Big[\Big(\underbrace{\beta(\overbrace{\eta + m\beta\gamma}^{\text{agent's response } x}) + \beta_0}_{\text{allocation } Y(x)} - \eta\Big)^2\Big]$$

  - Simple model, but objective is quartic in β

Preliminaries

Linearly predicting type η from observable x

- Suppose Agent responds to allocation rule $Y(x) = \beta x + \beta_0$,
  then Designer gathers data on joint distr of (η, x)

- Let $\hat{\eta}_\beta(x)$ be the best linear predictor of η given x:

$$\hat{\eta}_\beta(x) = \hat{\beta}(\beta)x + \hat{\beta}_0(\beta),$$

where, following OLS, $\hat{\beta}(\beta) = \frac{\text{Cov}(x,\eta)}{\text{Var}(x)} = \frac{\sigma_\eta^2 + m\rho\sigma_\eta\sigma_\gamma\beta}{\sigma_\eta^2 + m^2\sigma_\gamma^2\beta^2 + 2m\rho\sigma_\eta\sigma_\gamma\beta}$

- Can rewrite designer's objective

$$\text{Loss} = \underbrace{\mathbb{E}[(\mathbb{E}[\eta|x] - \eta)^2]}_{\substack{\text{Info loss from} \\ \text{estimating } \eta \text{ from } x}} + \underbrace{\mathbb{E}[(Y(x) - \mathbb{E}[\eta|x])^2]}_{\substack{\text{Misallocation loss given} \\ \text{estimation}}}$$

Preliminaries

Linearly predicting type η from observable x

- Suppose Agent responds to allocation rule $Y(x) = \beta x + \beta_0$,
  then Designer gathers data on joint distr of (η, x)

- Let $\hat{\eta}_\beta(x)$ be the best linear predictor of η given x:

$$\hat{\eta}_\beta(x) = \hat{\beta}(\beta)x + \hat{\beta}_0(\beta),$$

where, following OLS, $\hat{\beta}(\beta) = \frac{\text{Cov}(x,\eta)}{\text{Var}(x)} = \frac{\sigma_\eta^2 + m\rho\sigma_\eta\sigma_\gamma\beta}{\sigma_\eta^2 + m^2\sigma_\gamma^2\beta^2 + 2m\rho\sigma_\eta\sigma_\gamma\beta}$

- Can rewrite designer's objective for linear policies

$$\text{Loss} = \underbrace{\mathbb{E}[(\hat{\eta}_\beta(x) - \eta)^2]}_{\substack{\text{Info loss from} \\ \text{linearly estimating } \eta \text{ from } x}} + \underbrace{\mathbb{E}[(Y(x) - \hat{\eta}_\beta(x))^2]}_{\substack{\text{Misallocation loss given} \\ \text{linear estimation}}}$$

  - Info loss $\propto 1 - R^2_{\eta x}$
  - For corr. $\rho \geq 0$, $\hat{\beta}(\beta)$ is $\downarrow$ on $\beta \geq 0$ $\quad (\because x = \eta + m\beta\gamma)$

Benchmarks

Benchmarks

Loss = Info loss from linear estimation + Misallocation loss given linear estimation

Constant policy: $Y(x) = 0 \cdot x + \beta_0$

- No manipulation, $x = \eta$
- Info loss is 0
- Misallocation loss may be very large

Naive policy: $Y(x) = 1 \cdot x + 0$

- Designer's b.r. to data generated by constant policy
  $Y(x) = \hat{\eta}_{\beta=0}(x) = \hat{\beta}(0)x + \hat{\beta}_0(0)$
- But after implementing this policy, agent's behavior changes
  Agent now responding to $\beta = 1$, not $\beta = 0$

Benchmarks

Loss = Info loss from linear estimation + Misallocation loss given linear estimation

Designer's b.r. if agent behaves as if policy is (β, β_0)

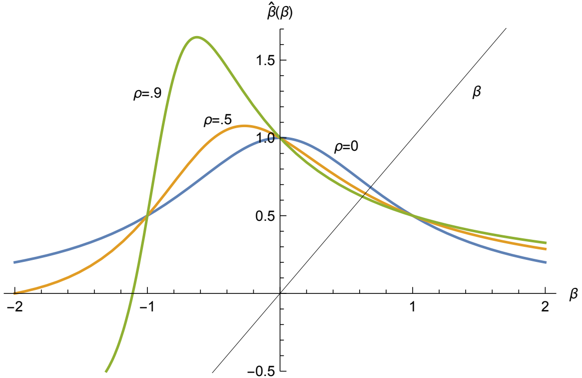
- Set $Y(x) = \hat{\eta}_\beta(x) = \hat{\beta}(\beta)x + \hat{\beta}_0(\beta)$
- Designer's optimum if agent's behavior were fixed

Fixed point policy: $Y(x) = \beta^{\text{fp}}x + \beta_0^{\text{fp}}$

- $\hat{\beta}_0(\beta^{\text{fp}}) = \beta_0^{\text{fp}}$ and $\hat{\beta}(\beta^{\text{fp}}) = \beta^{\text{fp}}$
- Simultaneous-move game's NE (under linearity restriction)
  - NE w/o restriction if (η, γ) is elliptically distr
- Misallocation loss given linear estimation = 0,
  allocations ex post optimal
- Info loss may be large

Designer best response $\hat{\beta}(\cdot)$ and fixed points

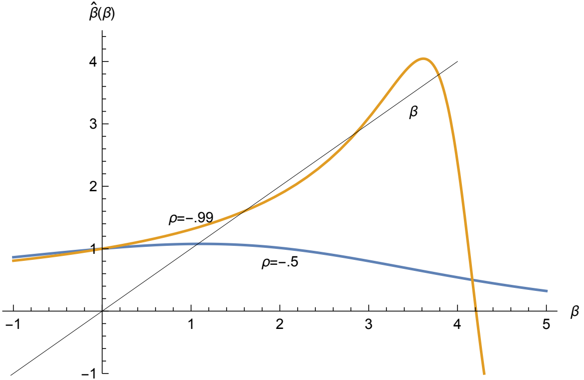
If (η, γ)'s corr. is $\boldsymbol{\rho \geq 0}$, then:



- For $\beta \geq 0$, best response sensitivity $\hat{\beta}(\beta)$ is positive and $\downarrow$
- Unique positive fixed point, and it is below naive b.r.: $\beta^{\mathrm{fp}} < 1$

Designer best response $\hat{\beta}(\cdot)$ and fixed points

If (η, γ)'s corr. is $\boldsymbol{\rho < 0}$, then:



- $\beta \gg 0 \implies$ higher x indicates lower $\eta \implies \hat{\beta}(\beta) < 0$
- $\hat{\beta}(\beta)$ can increase on $\beta \geq 0$
- Possible for fixed point sensitivity above naive: $\beta^{\text{fp}} > 1$
- Multiple positive fixed points possible

Main Result

Main Result

Designer chooses policy $Y(x) = \beta x + \beta_0$

Nb: Always at least one positive fixed point; just one if $\rho \geq 0$

Proposition

For the optimal policy's sensitivity β^*:

1. (Flattening.) $0 < \beta^* < \beta^{\text{fp}}$ for any $\beta^{\text{fp}} > 0$.
2. (Underutilize info.) $\hat{\beta}(\beta^*) > \beta^*$.

Commitment can yield large gains: $\exists$ params s.t.

$$L(\beta^{\text{fp}}) \simeq L(0) = \sigma_\eta^2, \text{ arbitrarily large}$$

$$L(\beta^*) \simeq 0, \text{ first best}$$

Main Result

Designer chooses policy $Y(x) = \beta x + \beta_0$

Nb: Always at least one positive fixed point; just one if $\rho \geq 0$

Proposition

For the optimal policy's sensitivity β^*:

1. (Flattening.) $0 < \beta^* < \beta^{\text{fp}}$ for any $\beta^{\text{fp}} > 0$.
2. (Underutilize info.) $\hat{\beta}(\beta^*) > \beta^*$.

Proof logic:

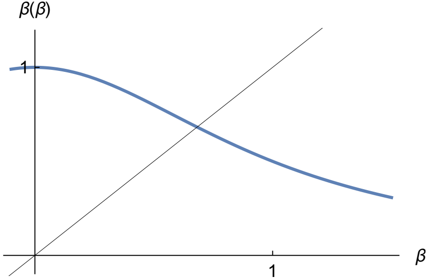
1. First order benefit of $\uparrow \beta$ from 0: constant policy not optimal
2. Lemma 1: First order benefit of $\downarrow \beta$ from any β^{fp}

   $\implies$ There is a local max in $(0, \beta^{\text{fp}})$
3. Show that such local max is global max

   (quartic polynomial)

Intuition for main result

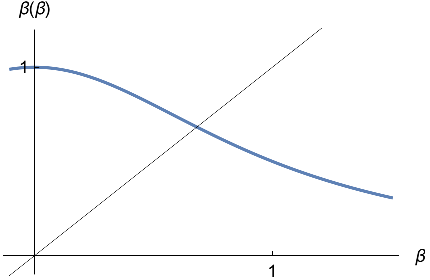
Loss = Info loss from linear estimation + Misallocation loss given linear estimation



- Misallocation loss is smaller when β close to b.r. $\hat{\beta}(\beta)$
- Info loss from estimation is smaller when β is smaller
  - Stronger incentives $\beta \implies$ more manipulation, less informative x
  - True for all $\beta > 0$ when $\rho \geq 0$, true for relevant range of β when $\rho < 0$

Intuition for main result

Loss = Info loss from linear estimation + Misallocation loss given linear estimation



At $\beta = \beta^{\text{fp}}$, misallocation loss is minimized

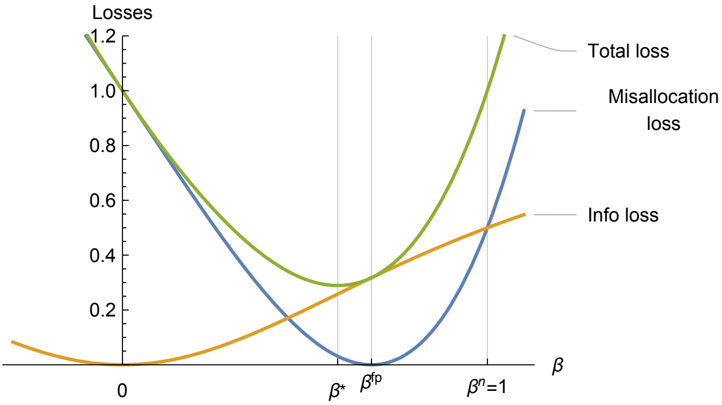
Slightly reducing sensitivity β yields

- First order benefit from $\downarrow$ info loss
- Second order harm from $\uparrow$ misallocation loss

(Analogously for $\uparrow \beta$ from 0, because there info loss minimized.)

Intuition for main result

Loss = Info loss from linear estimation + Misallocation loss given linear estimation



(In general, Loss not convex or even quasiconvex on $\mathbb{R}$.)

Some comparative statics

Recall $x = \eta + m\beta\gamma$

Let $k \equiv m\sigma_\gamma/\sigma_\eta$ describe susceptibility to manipulation

Proposition

1. As $k \to \infty$, $\beta^* \to 0$; As $k \to 0$, $\beta^* \to 1$;
   When $\rho \geq 0$, $\beta^* \downarrow$ in k.

2. When $\rho = 0$, $\beta^*/\beta^{\text{fp}} \downarrow$ in k;
   $\beta^*/\beta^{\text{fp}} \to 1$ as $k \to 0$ and $\beta^*/\beta^{\text{fp}} \to \sqrt[3]{1/2} \simeq .79$ as $k \to \infty$.

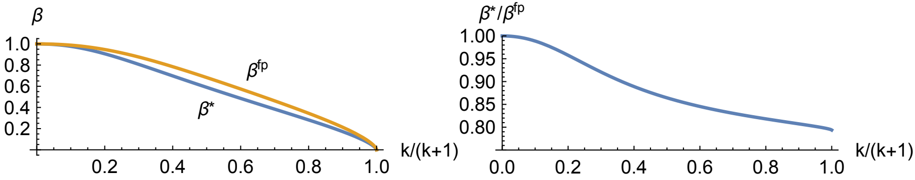


Figure with $\rho = 0$.

Conclusion

Discussion

- Can nonlinear allocation rules do better?
  - Typically yes
  - Linear rules are simple, easier to verify/commit to
  - Comparable to linear fixed points, which exist for elliptical distrs and to naive, which is linear
- If designer wants to reduce manipulation costs, $\downarrow \beta^*$
- If manipulation is productive effort, $\uparrow \beta^*$
- Crucial asymmetry in agent behavior $x = \eta + m\beta\gamma$
  - E.g., agent chooses effort (cost) e to generate data $x = \eta + \sqrt{\gamma}\sqrt{e}$
    Is effort a substitute or complement to the trait designer's values?
  - If designer wants to match allocation to γ, logic flips
    $\rightarrow$ For $\rho \geq 0$, $\beta^* > \beta^{\text{fp}}$ for any β^{fp}
  - If designer wants to match $(1-w)\eta + w\gamma$,
    $\rightarrow$ For $\rho = 0$, $\text{sign}[\beta^* - \beta^{\text{fp}}] = \text{sign}[w - w^*]$

Discussion

- Our model: info loss driven by heterogeneous response to incentives

  Does flattening fixed point extend to other sources of info loss?

  - Appendix: simple model of info loss driven by bounded action space

- More research: counterparts to “flattening” / “underutilizing information” in general allocation problems

Thank you!