Optimal Incentives under Moral Hazard: From Theory to Practice

George Georgiadis Michael Powell

Northwestern Kellogg

Motivation

- Imagine you have to design an employee performance pay plan.
- If you know all payoff-relevant parameters (*i.e.*, agent preferences, production function, etc), you can find optimal contract (in principle).
- Otherwise, agency theory gives us guiding principles (trade-offs, CS)

This paper: How to improve an existing PPP?

1. What information do you need?
2. And how should you use that information?

Motivation

- Imagine you have to design an employee performance pay plan.
- If you know all payoff-relevant parameters (*i.e.*, agent preferences, production function, etc), you can find optimal contract (in principle).
- Otherwise, agency theory gives us guiding principles (trade-offs, CS)

This paper: How to improve an existing PPP?

1. What information do you need?
2. And how should you use that information?

Motivation

- Imagine you have to design an employee performance pay plan.
- If you know all payoff-relevant parameters (*i.e.*, agent preferences, production function, etc), you can find optimal contract (in principle).
- Otherwise, agency theory gives us guiding principles (trade-offs, CS)

This paper: How to improve an existing PPP?

1. What information do you need?
2. And how should you use that information?

Preview

- *Framework:* Static agency model with a risk-averse agent
  - Principal knows *only* distribution of output following $w_A(\cdot)$ and $w_B(\cdot)$.
  - *Goal:* Find a new contract that raises profits as much as possible.

Key Lemma:

If the principal *takes a stance on* the agent's marginal utility for money, she can predict the distribution of output corresponding to *any* contract.

- Then, the principal can find an *optimal perturbation.*
- Application using real-effort experiment of DellaVigna and Pope ('17)
  1. *Predictions:* Use any pair of treatments to predict the other 5
  2. *Counterfactuals:* Estimate model and evaluate optimal perturbations

Preview

- *Framework:* Static agency model with a risk-averse agent
  - Principal knows *only* distribution of output following $w_A(\cdot)$ and $w_B(\cdot)$.
  - *Goal:* Find a new contract that raises profits as much as possible.

Key Lemma:

If the principal *takes a stance on* the agent's marginal utility for money, she can predict the distribution of output corresponding to *any* contract.

- Then, the principal can find an *optimal perturbation.*
- Application using real-effort experiment of DellaVigna and Pope ('17)
  1. *Predictions:* Use any pair of treatments to predict the other 5
  2. *Counterfactuals:* Estimate model and evaluate optimal perturbations

Preview

- *Framework:* Static agency model with a risk-averse agent
  - Principal knows *only* distribution of output following $w_A(\cdot)$ and $w_B(\cdot)$.
  - *Goal:* Find a new contract that raises profits as much as possible.

Key Lemma:

If the principal *takes a stance on* the agent's marginal utility for money, she can predict the distribution of output corresponding to *any* contract.

- Then, the principal can find an *optimal perturbation*.
- Application using real-effort experiment of DellaVigna and Pope ('17)
  1. *Predictions:* Use any pair of treatments to predict the other 5
  2. *Counterfactuals:* Estimate model and evaluate optimal perturbations

Preview

- *Framework:* Static agency model with a risk-averse agent
  - Principal knows *only* distribution of output following $w_A(\cdot)$ and $w_B(\cdot)$.
  - *Goal:* Find a new contract that raises profits as much as possible.

Key Lemma:

If the principal *takes a stance on* the agent's marginal utility for money, she can predict the distribution of output corresponding to *any* contract.

- Then, the principal can find an *optimal perturbation*.
- Application using real-effort experiment of DellaVigna and Pope ('17)
  1. *Predictions:* Use any pair of treatments to predict the other 5
  2. *Counterfactuals:* Estimate model and evaluate optimal perturbations

Related Literature

- Agency problems — Theory*:*
  - Mirrlees (1976), Holmström (1979), ...
  - Gibbons (1998), Murphy (1999), ...
- Agency problems — Empirics:
  - Lazear (2000), Shearer (2004), Bandiera et al. (2007, 2009), ...
  - Chiappori & Salanie (2002), Prendergast (2002), ...
- Sufficient statistics:
  - Monopoly pricing: Lerner (1934), Tirole (1988), ...
  - Optimal taxation: Saez (2001), Golosov et al. (2014), Chetty (2009), ..

Model

- Principal-agent model with the following timing:
  1. Principal offers a contract $w(\cdot)$.
  2. Agent observes $w(\cdot)$ and chooses effort $a(w) \in \mathbb{R}$.
  3. Output $x \sim f(\cdot|a(w))$ and payoffs are realized. (Normalize $\mathbb{E}[x|a] = a$.)
- *Preferences:*
  - Agent's utility: $\int v(w(x))f(x|a)dx - c(a)$
  - Principal's profit: $\pi(w) := ma(w) - \int w(x)f(x|a)dx$.
- *Information:*
  - Agent knows all payoff-relevant parameters
  - Principal knows (only) $f(\cdot|a(w_A))$, $f(\cdot|a(w_B))$, and

$$f_a(\cdot|a(w_A)) \simeq \frac{f(\cdot|a(w_B)) - f(\cdot|a(w_A))}{a(w_B) - a(w_A)}$$

Model

- Principal-agent model with the following timing:
  1. Principal offers a contract $w(\cdot)$.
  2. Agent observes $w(\cdot)$ and chooses effort $a(w) \in \mathbb{R}$.
  3. Output $x \sim f(\cdot|a(w))$ and payoffs are realized. (Normalize $\mathbb{E}[x|a] = a$.)
- *Preferences:*
  - Agent's utility: $\int v(w(x))f(x|a)dx - c(a)$
  - Principal's profit: $\pi(w) := ma(w) - \int w(x)f(x|a)dx$.
- *Information:*
  - Agent knows all payoff-relevant parameters
  - Principal knows (only) $f(\cdot|a(w_A))$, $f(\cdot|a(w_B))$, and

$$f_a(\cdot|a(w_A)) \simeq \frac{f(\cdot|a(w_B)) - f(\cdot|a(w_A))}{a(w_B) - a(w_A)}$$

Model

- Principal-agent model with the following timing:
  1. Principal offers a contract $w(\cdot)$.
  2. Agent observes $w(\cdot)$ and chooses effort $a(w) \in \mathbb{R}$.
  3. Output $x \sim f(\cdot|a(w))$ and payoffs are realized. (Normalize $\mathbb{E}[x|a] = a$.)
- *Preferences:*
  - Agent's utility: $\int v(w(x))f(x|a)dx - c(a)$
  - Principal's profit: $\pi(w) := ma(w) - \int w(x)f(x|a)dx$.
- *Information:*
  - Agent knows all payoff-relevant parameters
  - Principal knows (only) $f(\cdot|a(w_A))$, $f(\cdot|a(w_B))$, and

$$f_a(\cdot|a(w_A)) \simeq \frac{f(\cdot|a(w_B)) - f(\cdot|a(w_A))}{a(w_B) - a(w_A)}$$

The Canonical Principal-Agent Problem

- In the canonical formulation (Holmström, 1979), the principal solves

$$\max_{w(\cdot),\, a} \int [mx - w(x)]\, f(x|a)dx$$

$$\text{s.t.} \int v(w(x)) f(x|a)dx - c(a) \geq \underline{u} \qquad \text{(IR)}$$

$$a \in \arg\max_{\tilde{a}} \left\{ \int v(w(x)) f(x|\tilde{a})dx - c(\tilde{a}) \right\} \qquad \text{(IC)}$$

- To do so, she must know $v(\cdot)$, $\underline{u}$, $c(a)$, and $f(\cdot|a)$ for all a.
- In our setting, only knows $f(\cdot|a(w_i))$ for $i \in \{A, B\}$, and $f_a(\cdot|a(w_A))$
- *Notations:*

$$\widehat{a} := a(w_A) \;\; , \;\; \widehat{f} := f(\cdot|a(w_a)) \;\; , \text{ and } \widehat{f_a} := f_a(\cdot|a(w_a))$$

The Canonical Principal-Agent Problem

- In the canonical formulation (Holmström, 1979), the principal solves

$$\max_{w(\cdot),\, a} \int [mx - w(x)]\, f(x|a)dx$$

$$\text{s.t.} \int v(w(x))f(x|a)dx - c(a) \geq \underline{u} \tag{IR}$$

$$a \in \arg\max_{\tilde{a}} \left\{ \int v(w(x))f(x|\tilde{a})dx - c(\tilde{a}) \right\} \tag{IC}$$

- To do so, she must know $v(\cdot)$, $\underline{u}$, $c(a)$, and $f(\cdot|a)$ for all a.
- In our setting, only knows $f(\cdot|a(w_i))$ for $i \in \{A, B\}$, and $f_a(\cdot|a(w_A))$
- *Notations:*

$$\widehat{a} := a(w_A) \ , \ \widehat{f} := f(\cdot|a(w_a)) \ , \text{ and } \widehat{f_a} := f_a(\cdot|a(w_a))$$

The Canonical Principal-Agent Problem

- In the canonical formulation (Holmström, 1979), the principal solves

$$\max_{w(\cdot),\,a} \int [mx - w(x)]\, f(x|a)dx$$

$$\text{s.t.} \int v(w(x))f(x|a)dx - c(a) \geq \underline{u} \qquad \text{(IR)}$$

$$a \in \arg\max_{\tilde{a}} \left\{ \int v(w(x))f(x|\tilde{a})dx - c(\tilde{a}) \right\} \qquad \text{(IC)}$$

- To do so, she must know $v(\cdot)$, $\underline{u}$, $c(a)$, and $f(\cdot|a)$ for all a.
- In our setting, only knows $f(\cdot|a(w_i))$ for $i \in \{A, B\}$, and $f_a(\cdot|a(w_A))$
- *Notations:*

$$\widehat{a} := a(w_A) \ , \ \widehat{f} := f(\cdot|a(w_a)) \ , \text{and } \widehat{f_a} := f_a(\cdot|a(w_a))$$

Agent's Problem

- Assume optimal effort $a(w)$ satisfies the first-order condition

$$\int v(w(x)) f_a(x|a(w))dx = c'(a(w)) \tag{IC}$$

- Suppose $w(\cdot)$ is replaced by (some) contract $w(\cdot) + \theta t(\cdot)$, θ *small*.
- Define the directional (Gateaux) derivative

$$\mathcal{D}a(w,t) := \left.\frac{da(w+\theta t)}{d\theta}\right|_{\theta=0},$$

interpreted as the MC of a when w perturbed in the direction of $w + t$.

- Assume the principal knows

$$\mathcal{D}a(w_A, w_B - w_A) \simeq a(w_B) - a(w_A).$$

  - Implicitly assuming $\|w_B - w_A\| \simeq 0$ and $|a(w_B) - a(w_A)| \simeq 0$

Agent's Problem

- Assume optimal effort $a(w)$ satisfies the first-order condition

$$\int v(w(x)) f_a(x|a(w))dx = c'(a(w)) \tag{IC}$$

- Suppose $w(\cdot)$ is replaced by (some) contract $w(\cdot) + \theta t(\cdot)$, θ *small.*
- Define the directional (Gateaux) derivative

$$\mathcal{D}a(w,t) := \left. \frac{da(w+\theta t)}{d\theta} \right|_{\theta=0},$$

interpreted as the MC of a when w perturbed in the direction of $w+t$.

- Assume the principal knows

$$\mathcal{D}a(w_A, w_B - w_A) \simeq a(w_B) - a(w_A).$$

  - Implicitly assuming $\|w_B - w_A\| \simeq 0$ and $|a(w_B) - a(w_A)| \simeq 0$

Agent's Problem

- Assume optimal effort $a(w)$ satisfies the first-order condition

$$\int v(w(x)) f_a(x|a(w))dx = c'(a(w)) \tag{IC}$$

- Suppose $w(\cdot)$ is replaced by (some) contract $w(\cdot) + \theta t(\cdot)$, θ *small.*
- Define the directional (Gateaux) derivative

$$\mathcal{D}a(w,t) := \left.\frac{da(w+\theta t)}{d\theta}\right|_{\theta=0},$$

  interpreted as the MC of a when w perturbed in the direction of $w + t$.
- Assume the principal knows

$$\mathcal{D}a(w_A, w_B - w_A) \simeq a(w_B) - a(w_A).$$

  - Implicitly assuming $\|w_B - w_A\| \simeq 0$ and $|a(w_B) - a(w_A)| \simeq 0$

Principal's Problem

- If $w(\cdot)$ is replaced by (some) $w(\cdot) + \theta t(\cdot)$, then the principal's profit

$$\pi(w + \theta t) \simeq \pi(w) + \theta \mathcal{D}\pi(w, t),$$

where $\mathcal{D}\pi(w, t)$ is the derivative of $\pi(w)$ in direction of $w + t$, and

$$\mathcal{D}\pi(w, t) := \left.\frac{d\pi(w + \theta t)}{d\theta}\right|_{\theta=0} = \left(m - \int w f_a dx\right) \mathcal{D}a(w, t) - \int t f dx$$

- Assume the principal's goal is to maximize $\mathcal{D}\pi(w_A, t)$ subject to $w_A + \theta t$ giving the agent at least as much utility as w_A.
- Using (IC), this (participation) constraint can be rewritten as

$$\int t v'(w_A) \widehat{f} dx \geq 0$$

- *Info Requirements:* $\mathcal{D}a(w_A, t)$ for all t & marg. utility function $v'(\cdot)$

Principal's Problem

- If $w(\cdot)$ is replaced by (some) $w(\cdot) + \theta t(\cdot)$, then the principal's profit

$$\pi(w + \theta t) \simeq \pi(w) + \theta \mathcal{D}\pi(w, t),$$

where $\mathcal{D}\pi(w, t)$ is the derivative of $\pi(w)$ in direction of $w + t$, and

$$\mathcal{D}\pi(w, t) := \left.\frac{d\pi(w + \theta t)}{d\theta}\right|_{\theta=0} = \left(m - \int wf_a dx\right)\mathcal{D}a(w, t) - \int tfdx$$

- Assume the principal's goal is to maximize $\mathcal{D}\pi(w_A, t)$ subject to $w_A + \theta t$ giving the agent at least as much utility as w_A.

- Using (IC), this (participation) constraint can be rewritten as

$$\int tv'(w_A)\widehat{f}dx \geq 0$$

- *Info Requirements:* $\mathcal{D}a(w_A, t)$ for all t & marg. utility function $v'(\cdot)$

Principal's Problem

- If $w(\cdot)$ is replaced by (some) $w(\cdot) + \theta t(\cdot)$, then the principal's profit

$$\pi(w + \theta t) \simeq \pi(w) + \theta \mathcal{D}\pi(w, t),$$

where $\mathcal{D}\pi(w, t)$ is the derivative of $\pi(w)$ in direction of $w + t$, and

$$\mathcal{D}\pi(w, t) := \left.\frac{d\pi(w + \theta t)}{d\theta}\right|_{\theta=0} = \left(m - \int w f_a dx\right) \mathcal{D}a(w, t) - \int t f dx$$

- Assume the principal's goal is to maximize $\mathcal{D}\pi(w_A, t)$ subject to $w_A + \theta t$ giving the agent at least as much utility as w_A.
- Using (IC), this (participation) constraint can be rewritten as

$$\int t v'(w_A) \widehat{f} dx \geq 0$$

- *Info Requirements:* $\mathcal{D}a(w_A, t)$ for all t & marg. utility function $v'(\cdot)$

Principal's Problem

- If $w(\cdot)$ is replaced by (some) $w(\cdot) + \theta t(\cdot)$, then the principal's profit

$$\pi(w + \theta t) \simeq \pi(w) + \theta \mathcal{D}\pi(w, t)\,,$$

where $\mathcal{D}\pi(w, t)$ is the derivative of $\pi(w)$ in direction of $w + t$, and

$$\mathcal{D}\pi(w, t) := \left.\frac{d\pi(w + \theta t)}{d\theta}\right|_{\theta=0} = \left(m - \int w f_a dx\right) \mathcal{D}a(w, t) - \int t f dx$$

- Assume the principal's goal is to maximize $\mathcal{D}\pi(w_A, t)$ subject to $w_A + \theta t$ giving the agent at least as much utility as w_A.
- Using (IC), this (participation) constraint can be rewritten as

$$\int t v'(w_A) \widehat{f} dx \geq 0$$

- *Info Requirements:* $\mathcal{D}a(w_A, t)$ for all t & marg. utility function $v'(\cdot)$

Simplifying the Informational Requirements

- Using (IC), we can write $\mathcal{D}a(w,t)$ in terms of primitives as

$$\mathcal{D}a(w,t) = \frac{\int t v'(w) f_a dx}{c''(a(w)) - \int v(w) f_{aa} dx}$$

Remark 1. For any (upper semi-continuous) t:

$$\mathcal{D}a(w_A,t) = \frac{\mathcal{D}a(w_A, w_B - w_A)}{\int (w_B - w_A) v'(w_A) \widehat{f}_a dx} \underbrace{\int t v'(w_A) \widehat{f}_a dx}_{\mathcal{D}M(w_A,t)}$$

- Perturbation leads to a change in the agent's marginal incentives, $\mathcal{D}M(w_A,t)$, which is predictable given v' and $\widehat{f}_a$. Locally,

$$\mathcal{D}a(w_A,t) = C \times \mathcal{D}M(w_A,t)\ , \text{ where } C = \frac{\mathcal{D}a(w_A, w_B - w_A)}{\mathcal{D}M(w_A, w_B - w_A)}.$$

- If the principal *takes a stance on* v', she can predict $\mathcal{D}a(w_A,t)$ $\forall t$.

Simplifying the Informational Requirements

- Using (IC), we can write $\mathcal{D}a(w,t)$ in terms of primitives as

$$\mathcal{D}a(w,t) = \frac{\int tv'(w) f_a dx}{c''(a(w)) - \int v(w) f_{aa} dx}$$

Remark 1. For any (upper semi-continuous) t:

$$\mathcal{D}a(w_A,t) = \frac{\mathcal{D}a(w_A, w_B - w_A)}{\int (w_B - w_A) v'(w_A) \widehat{f}_a dx} \underbrace{\int tv'(w_A) \widehat{f}_a dx}_{\mathcal{D}M(w_A,t)}$$

- Perturbation leads to a change in the agent's marginal incentives, $\mathcal{D}M(w_A,t)$, which is predictable given v' and $\widehat{f}_a$. Locally,

$$\mathcal{D}a(w_A,t) = C \times \mathcal{D}M(w_A,t) \text{ , where } C = \frac{\mathcal{D}a(w_A, w_B - w_A)}{\mathcal{D}M(w_A, w_B - w_A)}.$$

- If the principal *takes a stance on* v', she can predict $\mathcal{D}a(w_A,t)\ \forall t$.

Simplifying the Informational Requirements

- Using (IC), we can write $\mathcal{D}a(w,t)$ in terms of primitives as

$$\mathcal{D}a(w,t) = \frac{\int tv'(w) f_a dx}{c''(a(w)) - \int v(w) f_{aa} dx}$$

Remark 1. For any (upper semi-continuous) t:

$$\mathcal{D}a(w_A,t) = \frac{\mathcal{D}a(w_A, w_B - w_A)}{\int (w_B - w_A) v'(w_A) \widehat{f}_a dx} \underbrace{\int tv'(w_A) \widehat{f}_a dx}_{\mathcal{D}M(w_A,t)}$$

- Perturbation leads to a change in the agent's marginal incentives, $\mathcal{D}M(w_A,t)$, which is predictable given v' and $\widehat{f}_a$. Locally,

$$\mathcal{D}a(w_A,t) = C \times \mathcal{D}M(w_A,t) \text{ , where } C = \frac{\mathcal{D}a(w_A, w_B - w_A)}{\mathcal{D}M(w_A, w_B - w_A)}.$$

- If the principal *takes a stance on* v', she can predict $\mathcal{D}a(w_A,t)$ $\forall t$.

Principal's Problem (Cont'd)

- The principal solves

$$\max_{t \text{ u.s.c}} \mu \int t v'(w_A) \widehat{f}_a dx - \int t \widehat{f} dx$$

$$\text{s.t} \int t v'(w_A) \widehat{f} dx \geq 0$$

$$\int |t|^p \, dx \leq 1$$

where $p \in \{1, 2, ...\}$ normalizes the *length* of t.

- Problem is convex, so it can be solved using standard techniques.
  1. Necessary & sufficient condition for w_A to be optimal
  2. Opt. Perturbation: Replace w_A with $w \equiv w_A + \theta t$ for some $\theta > 0$ *small*

Principal's Problem (Cont'd)

- The principal solves

$$\max_{t \text{ u.s.c}} \mu \int t v'(w_A) \widehat{f}_a dx - \int t \widehat{f} dx$$
$$\text{s.t} \int t v'(w_A) \widehat{f} dx \geq 0$$
$$\int |t|^p dx \leq 1$$

where $p \in \{1, 2, ...\}$ normalizes the *length* of t.

- Problem is convex, so it can be solved using standard techniques.
    1. Necessary & sufficient condition for w_A to be optimal
    2. Opt. Perturbation: Replace w_A with $w \equiv w_A + \theta t$ for some $\theta > 0$ *small*

Principal's Problem (Cont'd)

- The principal solves

$$\max_{t \text{ u.s.c}} \mu \int t v'(w_A) \widehat{f}_a dx - \int t \widehat{f} dx$$
$$\text{s.t} \int t v'(w_A) \widehat{f} dx \geq 0$$
$$\int |t|^p \, dx \leq 1$$

where $p \in \{1, 2, ...\}$ normalizes the *length* of t.

- Problem is convex, so it can be solved using standard techniques.
  1. Necessary & sufficient condition for w_A to be optimal
  2. Opt. Perturbation: Replace w_A with $w \equiv w_A + \theta t$ for some $\theta > 0$ *small*

Principal's Problem (Cont'd)

- The principal solves

$$\max_{t \text{ u.s.c}} \mu \int t v'(w_A) \widehat{f}_a dx - \int t \widehat{f} dx$$
$$\text{s.t} \int t v'(w_A) \widehat{f} dx \geq 0$$
$$\int |t|^p \, dx \leq 1$$

where $p \in \{1, 2, \ldots\}$ normalizes the *length* of t.

- Problem is convex, so it can be solved using standard techniques.
  1. Necessary & sufficient condition for w_A to be optimal
  2. Opt. Perturbation: Replace w_A with $w \equiv w_A + \theta t$ for some $\theta > 0$ *small*

Non-Local Perturbations

- *Goal:* Develop algorithm for finding optimal *non-local* perturbations

A.1. For all a in some interval that contains $\widehat{a}$, $f_a(\cdot|a) \equiv \widehat{f_a}$

- Hence, the marginal incentive of effort corresponding to w,

$$M(w) = \int v(w)\widehat{f_a}dx$$

does not depend on a itself – *agent's FOC:* $M(w) = c'(a)$

A.2. For any w, effort and marginal incentives are related by

$$\log a(w) = \beta + \epsilon \log M(w),$$

where β and ϵ estimated using A-B test data and assumed $v'(\cdot)$

- Implicitly assuming the agent has isoelastic cost function.

Non-Local Perturbations

- *Goal:* Develop algorithm for finding optimal *non-local* perturbations

A.1. For all a in some interval that contains $\widehat{a}$, $f_a(\cdot|a) \equiv \widehat{f}_a$

- Hence, the marginal incentive of effort corresponding to w,

$$M(w) = \int v(w)\widehat{f}_a dx$$

does not depend on a itself – *agent's FOC:* $M(w) = c'(a)$

A.2. For any w, effort and marginal incentives are related by

$$\log a(w) = \beta + \epsilon \log M(w),$$

where β and ϵ estimated using A-B test data and assumed $v'(\cdot)$

- Implicitly assuming the agent has isoelastic cost function.

Non-Local Perturbations

- *Goal:* Develop algorithm for finding optimal *non-local* perturbations

A.1. For all a in some interval that contains $\widehat{a}$, $f_a(\cdot|a) \equiv \widehat{f}_a$

- Hence, the marginal incentive of effort corresponding to w,

$$M(w) = \int v(w)\widehat{f}_a dx$$

does not depend on a itself – *agent's FOC:* $M(w) = c'(a)$

A.2. For any w, effort and marginal incentives are related by

$$\log a(w) = \beta + \epsilon \log M(w),$$

where β and ϵ estimated using A-B test data and assumed $v'(\cdot)$

- Implicitly assuming the agent has isoelastic cost function.

Non-Local Perturbations

- *Goal:* Develop algorithm for finding optimal *non-local* perturbations

A.1. For all a in some interval that contains $\widehat{a}$, $f_a(\cdot|a) \equiv \widehat{f}_a$

- Hence, the marginal incentive of effort corresponding to w,

$$M(w) = \int v(w)\widehat{f}_a dx$$

does not depend on a itself – *agent's FOC:* $M(w) = c'(a)$

A.2. For any w, effort and marginal incentives are related by

$$\log a(w) = \beta + \epsilon \log M(w)\,,$$

where β and ϵ estimated using A-B test data and assumed $v'(\cdot)$

- Implicitly assuming the agent has isoelastic cost function.

Towards an Optimal non-local Perturbation

Claim: Principal should solve

$$\max_{w(\cdot),\,\Delta a} \; m(\widehat{a}+\Delta a) - \int w(\widehat{f}+\Delta a \widehat{f}_a) \tag{P}$$

$$\text{s.t.} \int v(w)\widehat{f}_a = \left(\frac{\widehat{a}+\Delta a}{\widehat{a}}\right)^{1/\epsilon} \int v(w_A)\widehat{f}_a \tag{IC}$$

$$\int v(w)\left(\widehat{f}+\Delta a \widehat{f}_a\right) \geq \int v(w_A)\left(\widehat{f}+\Delta a \widehat{f}_a\right) \tag{IR}$$

- Suppose $a(w) = \widehat{a}+\Delta a$. Using a first-order approximation:

$$f(\cdot|\widehat{a}+\Delta a) \simeq \widehat{f}+\Delta a \widehat{f}_a \quad \text{and} \quad c(\widehat{a}+\Delta a) \simeq c(\widehat{a}) + \Delta a \int v(w_A)\widehat{f}_a$$

- It follows from $\log a(w) = \beta + \epsilon \log M(w)$ that w must satisfy (IC).
- Constraint that w gives at least as much utility as w_A:

$$\int v(w(x)) f(x|\widehat{a}+\Delta a) - c(\widehat{a}+\Delta a) \geq \int v(w_A)\widehat{f} - c(\widehat{a}) \implies \text{(IR)}$$

Towards an Optimal non-local Perturbation

Claim: Principal should solve

$$\max_{w(\cdot),\,\Delta a} \; m(\widehat{a}+\Delta a) - \int w(\widehat{f}+\Delta a \widehat{f}_a) \tag{P}$$

$$\text{s.t.} \int v(w)\widehat{f}_a = \left(\frac{\widehat{a}+\Delta a}{\widehat{a}}\right)^{1/\epsilon} \int v(w_A)\widehat{f}_a \tag{IC}$$

$$\int v(w)\left(\widehat{f}+\Delta a\widehat{f}_a\right) \geq \int v(w_A)\left(\widehat{f}+\Delta a\widehat{f}_a\right) \tag{IR}$$

- Suppose $a(w) = \widehat{a}+\Delta a$. Using a first-order approximation:

$$f(\cdot|\widehat{a}+\Delta a) \simeq \widehat{f}+\Delta a\widehat{f}_a \quad \text{and} \quad c(\widehat{a}+\Delta a) \simeq c(\widehat{a}) + \Delta a \int v(w_A)\widehat{f}_a$$

- It follows from $\log a(w) = \beta + \epsilon \log M(w)$ that w must satisfy (IC).
- Constraint that w gives at least as much utility as w_A:

$$\int v(w(x)) f(x|\widehat{a}+\Delta a) - c(\widehat{a}+\Delta a) \geq \int v(w_A)\widehat{f} - c(\widehat{a}) \implies \text{(IR)}$$

Towards an Optimal non-local Perturbation

Claim: Principal should solve

$$\max_{w(\cdot),\,\Delta a} m(\widehat{a}+\Delta a) - \int w(\widehat{f}+\Delta a\widehat{f_a}) \tag{P}$$

$$\text{s.t.} \int v(w)\widehat{f_a} = \left(\frac{\widehat{a}+\Delta a}{\widehat{a}}\right)^{1/\epsilon} \int v(w_A)\widehat{f_a} \tag{IC}$$

$$\int v(w)\left(\widehat{f}+\Delta a\widehat{f_a}\right) \geq \int v(w_A)\left(\widehat{f}+\Delta a\widehat{f_a}\right) \tag{IR}$$

- Suppose $a(w) = \widehat{a}+\Delta a$. Using a first-order approximation:

$$f(\cdot|\widehat{a}+\Delta a) \simeq \widehat{f}+\Delta a\widehat{f_a} \quad \text{and} \quad c(\widehat{a}+\Delta a) \simeq c(\widehat{a}) + \Delta a \int v(w_A)\widehat{f_a}$$

- It follows from $\log a(w) = \beta + \epsilon \log M(w)$ that w must satisfy (IC).
- Constraint that w gives at least as much utility as w_A:

$$\int v(w(x))f(x|\widehat{a}+\Delta a) - c(\widehat{a}+\Delta a) \geq \int v(w_A)\widehat{f} - c(\widehat{a}) \implies \text{(IR)}$$

Towards an Optimal non-local Perturbation

Claim: Principal should solve

$$\max_{w(\cdot),\,\Delta a} m(\widehat{a}+\Delta a) - \int w(\widehat{f}+\Delta a\widehat{f}_a) \tag{P}$$

$$\text{s.t.} \int v(w)\widehat{f}_a = \left(\frac{\widehat{a}+\Delta a}{\widehat{a}}\right)^{1/\epsilon} \int v(w_A)\widehat{f}_a \tag{IC}$$

$$\int v(w)\left(\widehat{f}+\Delta a\widehat{f}_a\right) \geq \int v(w_A)\left(\widehat{f}+\Delta a\widehat{f}_a\right) \tag{IR}$$

- Suppose $a(w) = \widehat{a}+\Delta a$. Using a first-order approximation:

$$f(\cdot|\widehat{a}+\Delta a) \simeq \widehat{f}+\Delta a\widehat{f}_a \quad \text{and} \quad c(\widehat{a}+\Delta a) \simeq c(\widehat{a}) + \Delta a \int v(w_A)\widehat{f}_a$$

- It follows from $\log a(w) = \beta + \epsilon \log M(w)$ that w must satisfy (IC).
- Constraint that w gives at least as much utility as w_A:

$$\int v(w(x))f(x|\widehat{a}+\Delta a) - c(\widehat{a}+\Delta a) \geq \int v(w_A)\widehat{f} - c(\widehat{a}) \implies \text{(IR)}$$

Solving for the Optimal non-local Perturbation

- *Stage 1:* For every Δa, solve

$$\widehat{\Pi}(\Delta a) = \max_{w(\cdot)} \; m(\widehat{a} + \Delta a) - \int w(\widehat{f} + \Delta a \widehat{f}_a)$$

$$\text{s.t.} \quad \int v(w)\widehat{f}_a = \left(\frac{\widehat{a} + \Delta a}{\widehat{a}}\right)^{1/\epsilon} \int v(w_A)\widehat{f}_a$$

$$\int v(w)\left(\widehat{f} + \Delta a \widehat{f}_a\right) \geq \int v(w_A)\left(\widehat{f} + \Delta a \widehat{f}_a\right)$$

- Optimization program is convex as long as $\widehat{f} + \Delta a \widehat{f}_a > 0$ for all x.
- *Stage 2:* Solve

$$\widehat{\Pi}^* = \max_{\Delta a} \widehat{\Pi}(\Delta a)$$

- *Info. requirements:* Must know $\widehat{f}$, $\widehat{f}_a$, and $v'(\cdot)$ (using $\int \widehat{f}_a = 0$)
- *Alternative:* Can approximate $v(w) \simeq v(w_A) + (w - w_A)v'(w_A)$ to make constraints linear in w—then stage 1 program is convex $\forall \Delta a$.

Solving for the Optimal non-local Perturbation

- *Stage 1:* For every Δa, solve

$$\widehat{\Pi}(\Delta a) = \max_{w(\cdot)} \; m(\widehat{a} + \Delta a) - \int w(\widehat{f} + \Delta a \widehat{f}_a)$$

$$\text{s.t.} \quad \int v(w)\widehat{f}_a = \left(\frac{\widehat{a} + \Delta a}{\widehat{a}}\right)^{1/\epsilon} \int v(w_A)\widehat{f}_a$$

$$\int v(w)\left(\widehat{f} + \Delta a \widehat{f}_a\right) \geq \int v(w_A)\left(\widehat{f} + \Delta a \widehat{f}_a\right)$$

- Optimization program is convex as long as $\widehat{f} + \Delta a \widehat{f}_a > 0$ for all x.
- *Stage 2:* Solve

$$\widehat{\Pi}^* = \max_{\Delta a} \widehat{\Pi}(\Delta a)$$

- *Info. requirements:* Must know $\widehat{f}$, $\widehat{f}_a$, and $v'(\cdot)$ (using $\int \widehat{f}_a = 0$)
- *Alternative:* Can approximate $v(w) \simeq v(w_A) + (w - w_A)v'(w_A)$ to make constraints linear in w—then stage 1 program is convex $\forall \Delta a$.

Solving for the Optimal non-local Perturbation

- *Stage 1:* For every Δa, solve

$$\widehat{\Pi}(\Delta a) = \max_{w(\cdot)} \; m(\widehat{a} + \Delta a) - \int w(\widehat{f} + \Delta a \widehat{f}_a)$$

$$\text{s.t.} \quad \int v(w)\widehat{f}_a = \left(\frac{\widehat{a} + \Delta a}{\widehat{a}}\right)^{1/\epsilon} \int v(w_A)\widehat{f}_a$$

$$\int v(w)\left(\widehat{f} + \Delta a \widehat{f}_a\right) \geq \int v(w_A)\left(\widehat{f} + \Delta a \widehat{f}_a\right)$$

- Optimization program is convex as long as $\widehat{f} + \Delta a \widehat{f}_a > 0$ for all x.
- *Stage 2:* Solve

$$\widehat{\Pi}^* = \max_{\Delta a} \widehat{\Pi}(\Delta a)$$

- *Info. requirements:* Must know $\widehat{f}$, $\widehat{f}_a$, and $v'(\cdot)$ (using $\int \widehat{f}_a = 0$)
- *Alternative:* Can approximate $v(w) \simeq v(w_A) + (w - w_A)v'(w_A)$ to make constraints linear in w—then stage 1 program is convex $\forall \Delta a$.

Solving for the Optimal non-local Perturbation

- *Stage 1:* For every Δa, solve

$$\widehat{\Pi}(\Delta a) = \max_{w(\cdot)} \; m(\widehat{a} + \Delta a) - \int w(\widehat{f} + \Delta a \widehat{f}_a)$$

$$\text{s.t.} \quad \int v(w)\widehat{f}_a = \left(\frac{\widehat{a} + \Delta a}{\widehat{a}}\right)^{1/\epsilon} \int v(w_A)\widehat{f}_a$$

$$\int v(w)\left(\widehat{f} + \Delta a \widehat{f}_a\right) \geq \int v(w_A)\left(\widehat{f} + \Delta a \widehat{f}_a\right)$$

- Optimization program is convex as long as $\widehat{f} + \Delta a \widehat{f}_a > 0$ for all x.
- *Stage 2:* Solve

$$\widehat{\Pi}^* = \max_{\Delta a} \widehat{\Pi}(\Delta a)$$

- *Info. requirements:* Must know $\widehat{f}$, $\widehat{f}_a$, and $v'(\cdot)$ (using $\int \widehat{f}_a = 0$)
- *Alternative:* Can approximate $v(w) \simeq v(w_A) + (w - w_A)v'(w_A)$ to make constraints linear in w—then stage 1 program is convex $\forall \Delta a$.

Extensions

1. *Bounded payments.* Assume that $w_A(x) + t(x) \in [\underline{w}, \overline{w}]$
   - New constraints are linear, so principal's problem remains convex.
2. *Heterogeneous abilities.* Assume that the principal offers a common contract to multiple agents who have heterogeneous effort costs.
   - Principal must classify the agents into types (ϕ), and estimate $\Pr\{\phi\}$, $\widehat{f}^{\phi}$, $\widehat{f}_a^{\phi}$, and $\mathcal{D}a^{\phi}(\widehat{w}, \widehat{t})$ for each ϕ.
   - Can induce selection by imposing participation for subset of types.
3. *Multidimensional effort.* Assume agent's effort $\mathbf{a} \in \mathbb{R}^N$ at cost $c(\mathbf{a})$
   - *e.g.*, effort towards quantity & quality, or selling different products.
   - Principal must have output data for $K \geq (N+3)/2$ contracts.

Extensions

4. *Parametric contract classes.* Assume the principal restricts attention to contracts of the form w_α, where α is a vector of parameters.
   - Find optimal perturbation direction $\mathbf{z}$. (*New contract*: $w_{\alpha+\theta\mathbf{z}}$)
   - Same informational requirements as general case.

5. *Other sources of incentives.* (Promotion, firing threat, prestige, etc)
   - Results hold verbatim if the agent's IC constraint can be written as

   $$\int v(w) f_a dx + I(a(w)) = c'(a(w)),$$

   where $I(a)$ denotes marginal benefit of effort due to *indirect incentives*.
   - *Key:* Additive separability and $I(\cdot)$ not directly dependent on w.

6. *Multiplicatively separable utility.* Agent's payoff $u(\omega, a) = v(\omega)c(a)$
   - *Example*: Agent's utility satisfies CARA.
   - Principal must take a stance on v (instead of v').

Dataset

- *Goal:* Illustrate application & evaluate methodology
- Dataset from DellaVigna and Pope (2017)
- Real-effort experiment on M-Turk: Subjects press a-b keys for 10 min
- 7 treatments with different monetary incentives:

Contract (in ¢)	Mean effort	N
$w_1(x) = 100$	1521	540
$w_2(x) = 100 + 0.001x$	1883	538
$w_3(x) = 100 + 0.01x$	2029	558
$w_4(x) = 100 + 0.04x$	2132	566
$w_5(x) = 100 + 0.10x$	2175	538
$w_6(x) = 100 + 40\,\mathbb{I}_{\{x \geq 2000\}}$	2136	545
$w_7(x) = 100 + 80\,\mathbb{I}_{\{x \geq 2000\}}$	2188	532

- Each subject participates in a single treatment, once.

Two Exercises

- Assume subjects are identical, and make assumptions about v' and m

I. Given data for any two treatments, predict effort & profits for others.

  - Test predictions of two models:

$$\log a(w) = \beta + \epsilon \log M(w)$$
$$a(w) = \beta_0 + \beta_1 M(w)$$

  where $M(w) = \int v(w)\widehat{f_a}$, and constants are estimated using A-B test.

  - *Sensitivity analysis:* Prediction accuracy vs. assumptions about v'

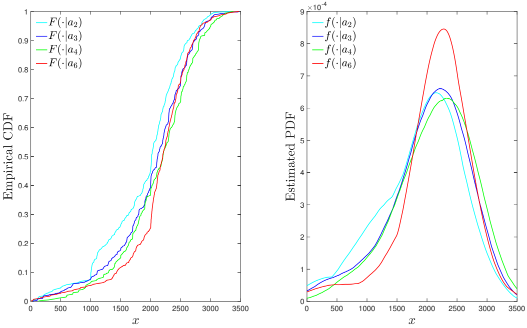
II. Counterfactuals:

  1. Use all seven treatments to estimate the parameters of the model
  2. Characterize optimally perturbed contract
  3. Compare projected profits to those of w_A and optimal contract

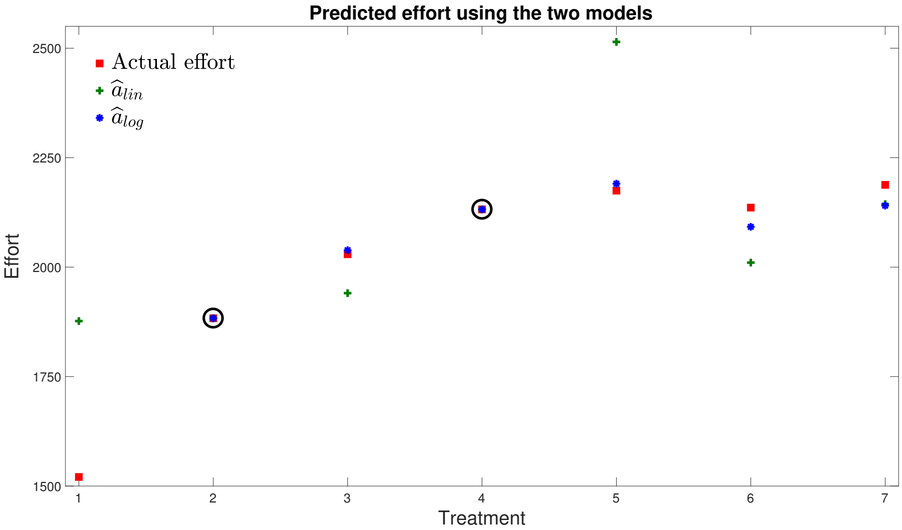
Step 1

1. Assume subjects have CRRA utility — specifically, $v'(\omega) = \omega^{-0.3}$
2. Normalize $a(w_i) = (\text{Mean effort})_i$.
3. Given A-B test, estimate $f(\cdot|a(w_i))$ for $i \in \{A, B\}$, and compute

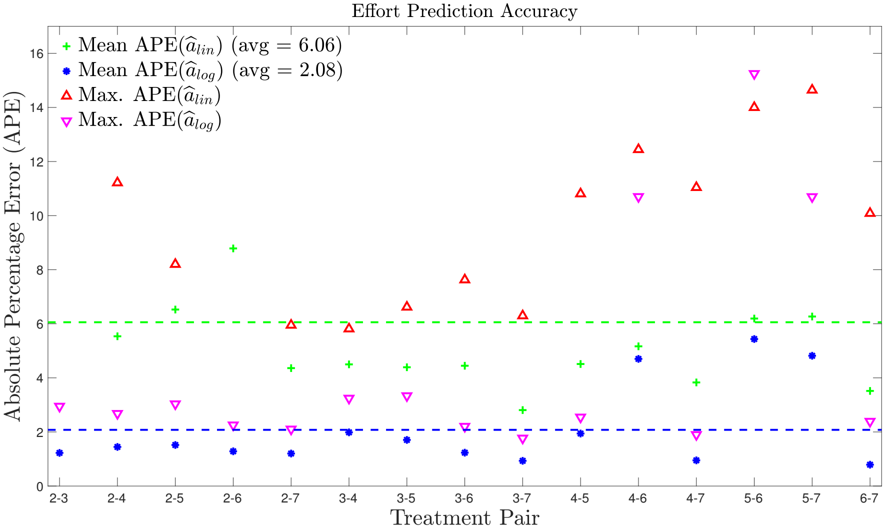
$$\widehat{f}_a(x) = \frac{f(x|a(w_B)) - f(x|a(w_A))}{a(w_B) - a(w_A)}$$



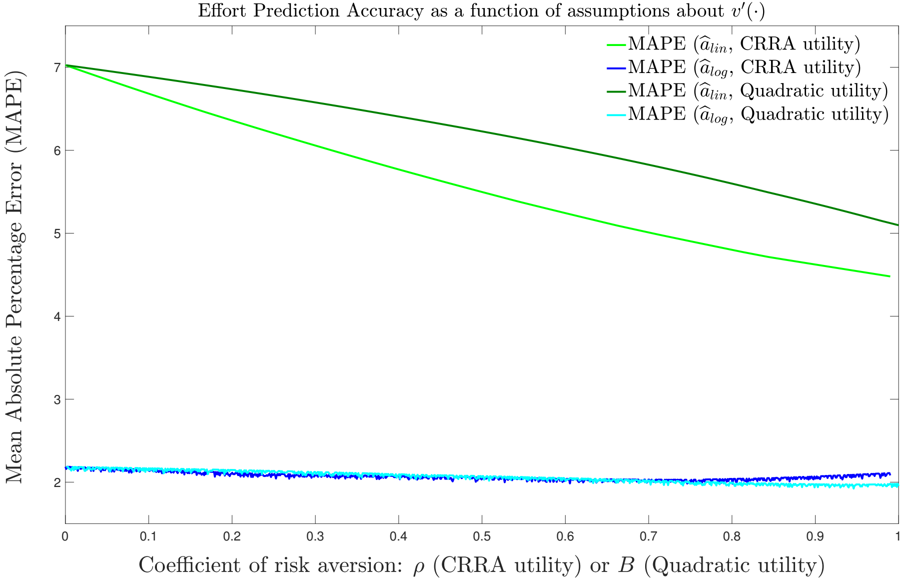
Exercise 1(a): Effort Predictions given Treatments 2 and 4



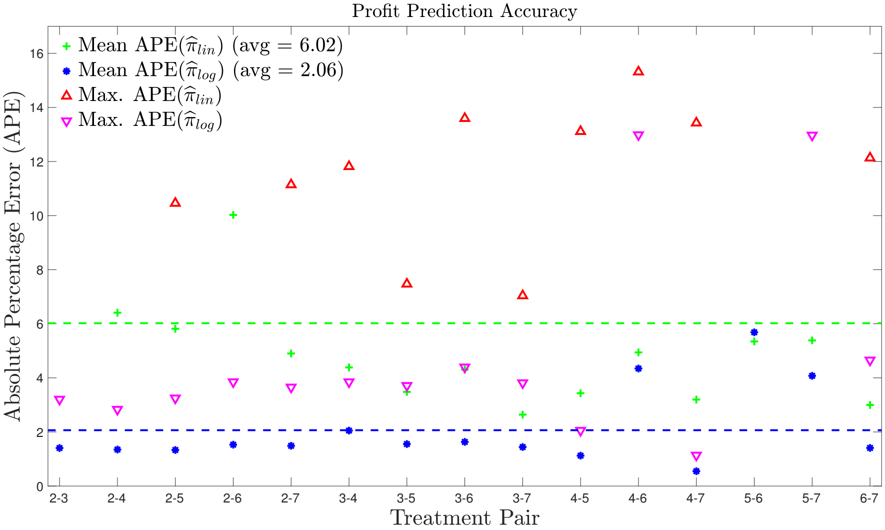
Exercise 1(b): Effort Prediction Accuracy



Exercise 1(c): Sensitivity Analysis



Exercise 1(d): Profit Prediction Accuracy



Estimate Model

1. Use estimates of $\{f(\cdot|a(w_i))\}_i$ to fit $f(\cdot|a)$ for all a using linear interpolation (thus assuming $f_a(x|a)$ is piece-wise linear in a)

2. Assume agent has CRRA utility and isoelastic costs; *i.e.*,

$$v(\omega) = \frac{\omega^{1-\rho}}{1-\rho} \quad \text{and} \quad c(a) = \frac{c_0}{p+1} a^{p+1},$$

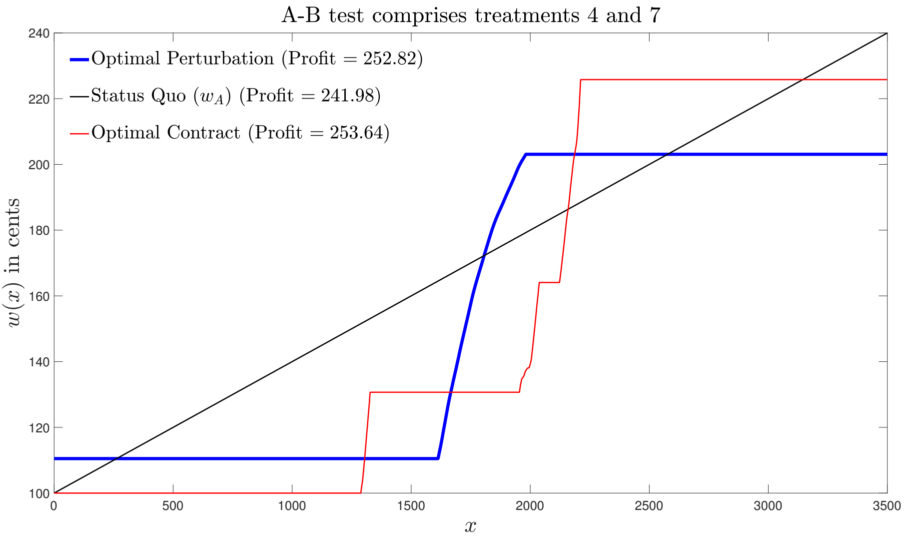
   and given w, he chooses his effort $a(w)$ such that

$$\int v(w) f_a(\cdot|a(w))dx + I = c^p(a(w)).$$

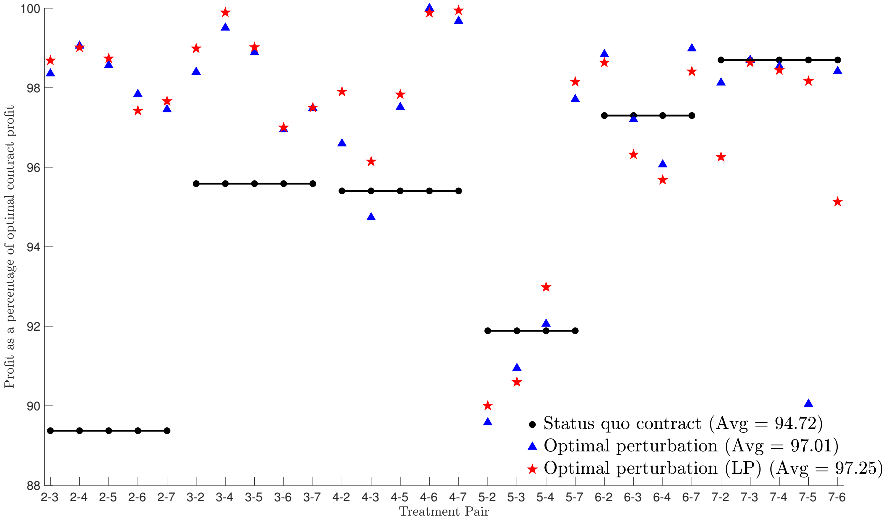
   Then, we estimate the unknown coefficients.

3. Assign value to principal's marginal profit — specifically, $m = 0.2$

Exercise 2(a): Optimal Perturbation



Exercise 2(b): Profits relative to Optimal Contract



Summary & Future Work

- Framework for using agency theory to address an empirical question.
  - How to improve an existing performance pay plan?
  - What information do you need to do so?
- Other questions:
  - Optimal experimentation (ratchet effects, behavioral constraints)?
  - Extend to other settings (non-monetary instruments, dynamics)?