Eliminating Bias in Treatment Effect Estimates Arising from Adaptively Collected Data

János K. Divényi

Dec 20, 2019

Central European University, Emarsys

Outline

1. Introduction
2. Illustration of basic properties on a simple setup
3. Monte-Carlo simulation, main results

Introduction

Motivating example: online shop's pricing scheme

An online shop is considering to change its pricing scheme

Motivating example: online shop's pricing scheme

An online shop is considering to change its pricing scheme

- Innovation vs status quo
- Sequential arrival of subjects
- Experimenting is
  - cheap in terms of transactional costs
  - expensive in terms of opportunity costs

Motivating example: online shop's pricing scheme

An online shop is considering to change its pricing scheme

- Innovation vs status quo
- Sequential arrival of subjects
- Experimenting is
  - cheap in terms of transactional costs
  - expensive in terms of opportunity costs
- Two goals:
  1. profit
  2. treatment effect estimation

Standard solution: RCT & decide

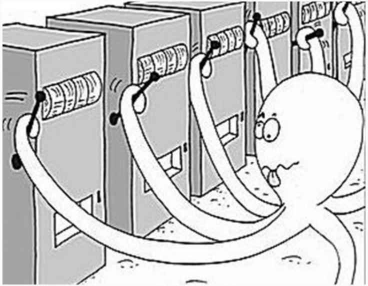
1. Conduct a Randomized Controlled Trial (aka AB test) on an experimental sample
2. Measure the treatment effect
3. Apply the treatment onwards if the effect is positive

Standard solution: RCT & decide

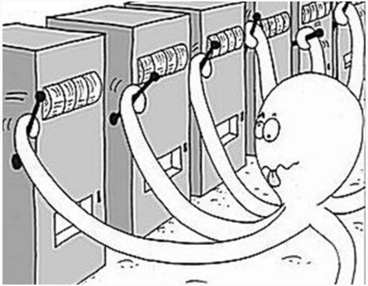
1. Conduct a Randomized Controlled Trial (aka AB test) on an experimental sample
2. Measure the treatment effect
3. Apply the treatment onwards if the effect is positive

Treatment choice literature: Manski (2004), Dehejia (2005), Kitagawa & Tetenov (2018)

Adaptive experiment: multi-armed bandits



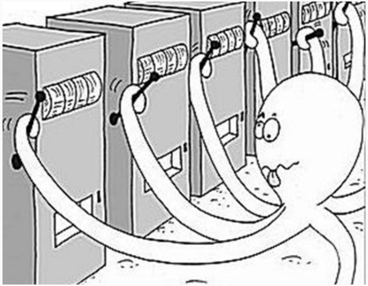
Adaptive experiment: multi-armed bandits



How to optimize which arm to pull in sequential decision-making?

Measure the rewards $\leftrightarrow$ Pull the highest

Adaptive experiment: multi-armed bandits



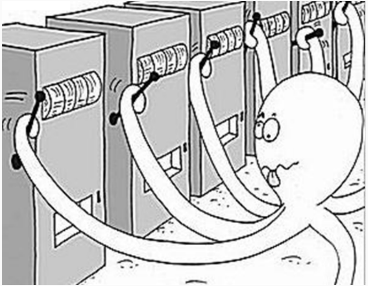
How to optimize which arm to pull in sequential decision-making?

Measure the rewards $\leftrightarrow$ Pull the highest

Exploration $\leftrightarrow$ Exploitation

Learning $\leftrightarrow$ Earning

Adaptive experiment: multi-armed bandits



How to optimize which arm to pull in sequential decision-making?

Measure the rewards $\leftrightarrow$ Pull the highest

Exploration $\leftrightarrow$ Exploitation

Learning $\leftrightarrow$ Earning

Bandit literature: Thompson (1933), Robbins (1952), Bubeck & Cesa-Bianchi (2012), Szepesvári & Lattimore (2019), Slivkins (2019)

Relation to literature

Sequential experiments Kasy & Sautmann (2019)

Dynamic treatment allocation Perchet et al. (2016), Kock & Thyrsgaard (2018)

Estimation considerations in bandits Villar, Bowden & Wason (2015), Nie et al. (2018), Dimakopoulou, Athey & Imbens (2019), Hadad et al. (2019)

My contribution

How to balance between estimation (RCT) and allocation (bandit) goal

- Systematic simulation exercise.
- There is a trade-off between welfare and estimation goals.
- IPW with limited propensity scores extends the set of choices.

Illustration of basic properties

Formal setup

- Binary treatment W with constant effect $Y_i(1) = Y_i(0) + \tau$
- Fix population $\{Y_i(0)\}_{i=1}^n$, $Y(0) \ \sim \mathcal{N}(0, \sigma^2)$, arrival is random

Formal setup

- Binary treatment W with constant effect $Y_i(1) = Y_i(0) + \tau$
- Fix population $\{Y_i(0)\}_{i=1}^n$, $Y(0) \ \sim \mathcal{N}(0, \sigma^2)$, arrival is random
- Goals
  1. find treatment rule $\pi : \{1, ..., n\} \to \{0, 1\}$ that maximizes $\sum_i Y_i$
  2. estimate τ

Sequential arrival

1. A group of individuals $i \in B_j$ arrive, and they are assigned ($|B_j| = n_B$)
2. Outcomes $\{Y_i\}_{i \in B_j}$ are observed
3. A next group of individuals $i \in B_{j+1}$ arrive and steps 1-2 are repeated

Bandit approach: Adaptive data collection

Thompson sampling - an old heuristic suggested by Thompson (1933)

1. Assign the first batch with 50% probability to the treatment.
2. Derive beliefs for sampling means using normal density with calculated averages and their (known) standard deviations:

$$\mathcal{N}\left(\hat{\mu}_w^{(k)}, \frac{\sigma_w^{2(k)}}{n_w^{(k)}}\right)$$

3. Assign individuals to the treatment in the next batch by the probability that the treatment mean is higher than the control mean.

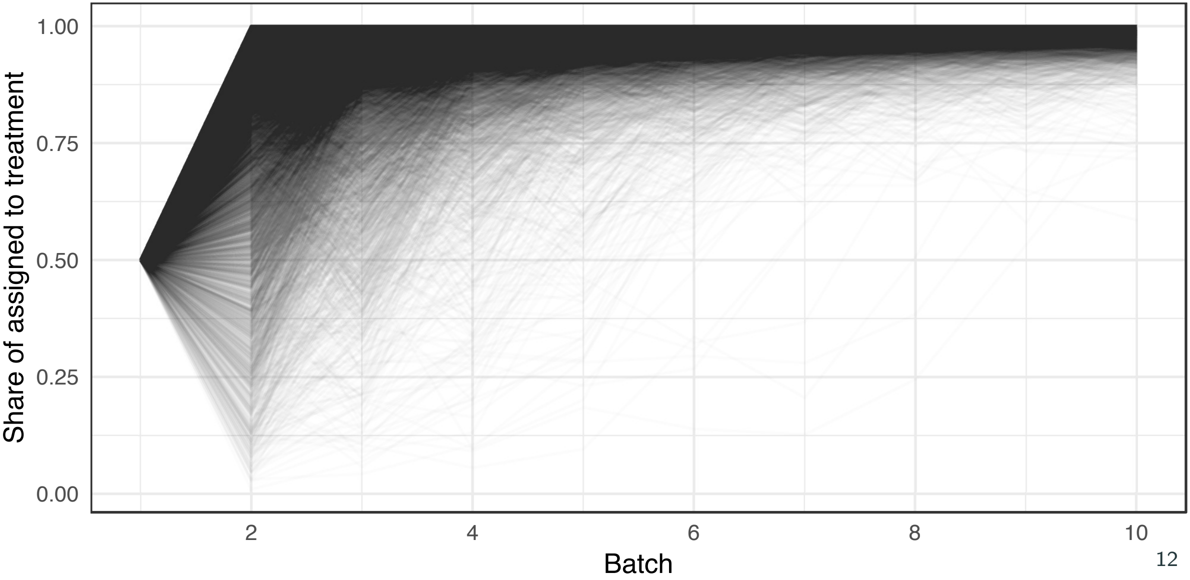
Parametrization

- $\tau = 1$ (normalization)
- $\sigma = 10$ (high signal-to-noise ratio)
- $n = 10,000$ in 10 batches ($n_B = 1,000$)

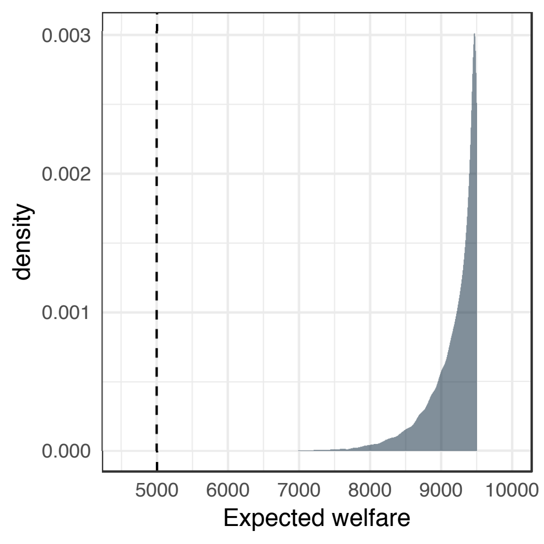
Parametrization

- $\tau = 1$ (normalization)
- $\sigma = 10$ (high signal-to-noise ratio)
- $n = 10,000$ in 10 batches ($n_B = 1,000$)
- I simulate assignment rules by randomizing the arrival, 20k runs

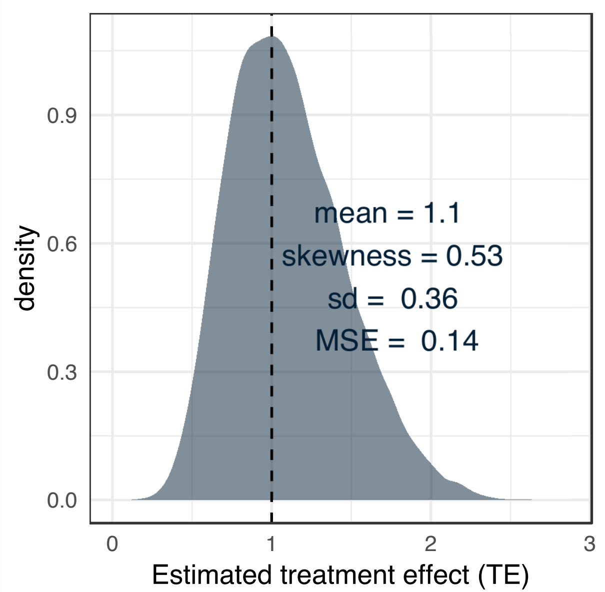
The bandits are adaptive (share of treated)...



...that results in higher total welfare

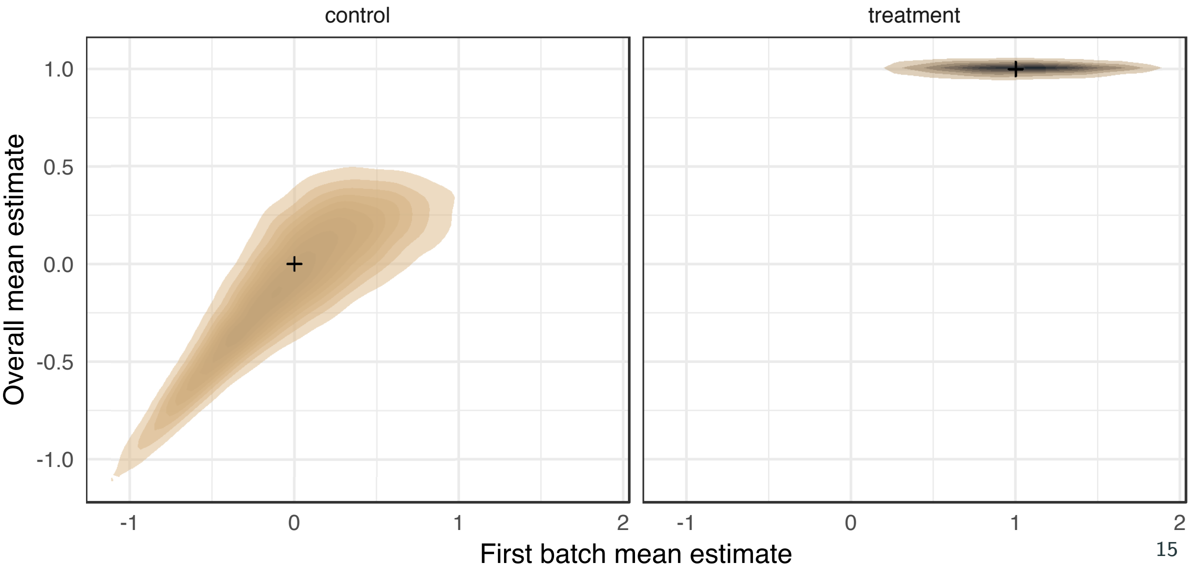


A price to pay: the implicit treatment effect estimator is biased



	Control	Treatment
Mean	-0.10	1.00
Skewness	-0.63	-1.49
Std. dev.	0.33	0.03
MSE	0.12	0.00

Intuition for the bias: asymmetric sampling due to adaptivity



Bias correction: Inverse Propensity Weighting (IPW)

- weight each observation by the probability of assigning them to the group they were actually assigned to (PS is estimated)

Bias correction: Inverse Propensity Weighting (IPW)

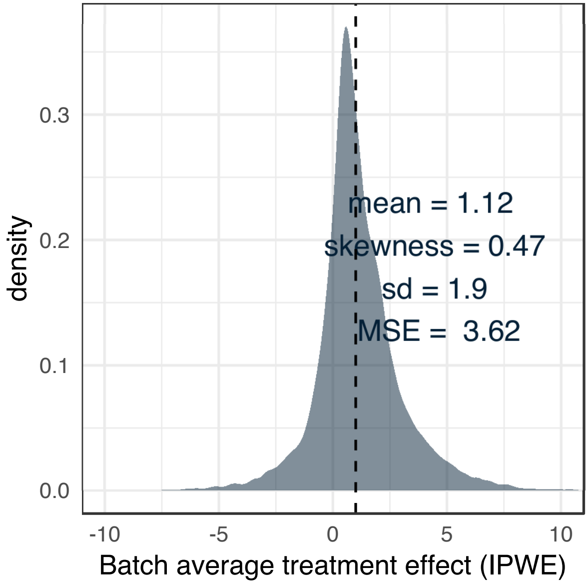
- weight each observation by the probability of assigning them to the group they were actually assigned to (PS is estimated)
- IPWE is the simple average of batch averages

Bias correction: Inverse Propensity Weighting (IPW)

- weight each observation by the probability of assigning them to the group they were actually assigned to (PS is estimated)
- IPWE is the simple average of batch averages

$$
\begin{aligned}
\hat{\tau}_{IPWE} &= \frac{1}{n}\left(\sum_{i=1}^{n}\frac{Y_i W_i}{p_i} - \sum_{i=1}^{n}\frac{Y_i\left(1-W_i\right)}{1-p_i}\right) \\
&= \frac{1}{n}\sum_{j=1}^{m}\left(\sum_{i\in B_j}\frac{Y_i W_i n_B}{\sum_{i\in B_j} W_i} - \sum_{i\in B_j}\frac{Y_i\left(1-W_i\right)n_B}{\sum_{i\in B_j}\left(1-W_i\right)}\right) \\
&= \frac{1}{m}\sum_{j=1}^{m}\underbrace{\frac{\sum_{i\in B_j} Y_i W_i}{\sum_{i\in B_j} W_i}}_{\substack{\text{batch treated}\\ \text{average}}} - \frac{1}{m}\sum_{j=1}^{m}\underbrace{\frac{\sum_{i\in B_j} Y_i(1-W_i)}{\sum_{i\in B_j}(1-W_i)}}_{\substack{\text{batch control}\\ \text{average}}}
\end{aligned}
$$

IPW can even exacerbate the bias



	Control	Treatment
Mean	-0.12	1.00
Skewness	-0.44	-1.82
Std. dev.	1.89	0.04
MSE	3.59	0.00

Straight-forward strategies with unbiased treatment effect

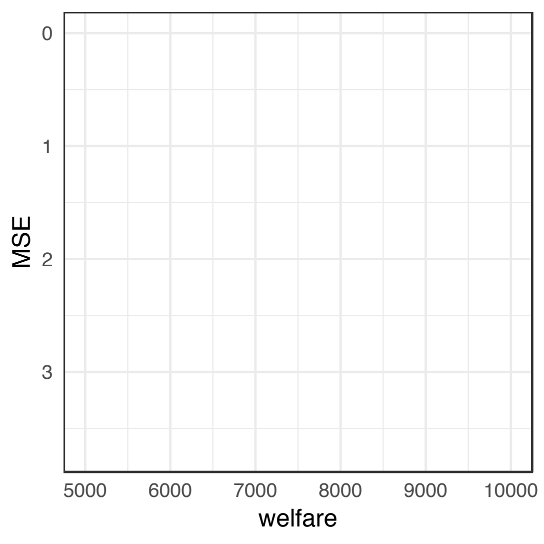
- Use only the first batch of bandit for estimation: FBTE

Straight-forward strategies with unbiased treatment effect

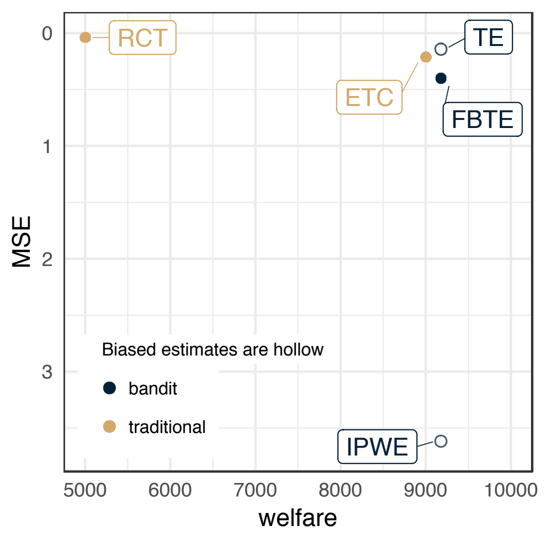
- Use only the first batch of bandit for estimation: FBTE
- RCT & decide = "explore-then-commit": ETC

Welfare-estimation trade-off

Welfare-estimation trade-off

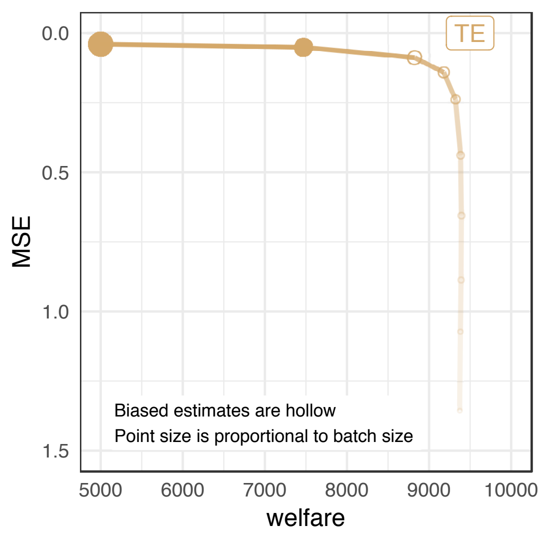


Welfare-estimation trade-off

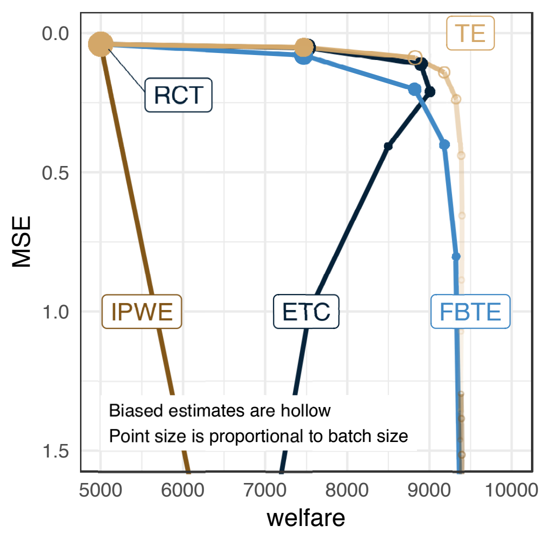


Welfare-estimation trade-off - varying batch size

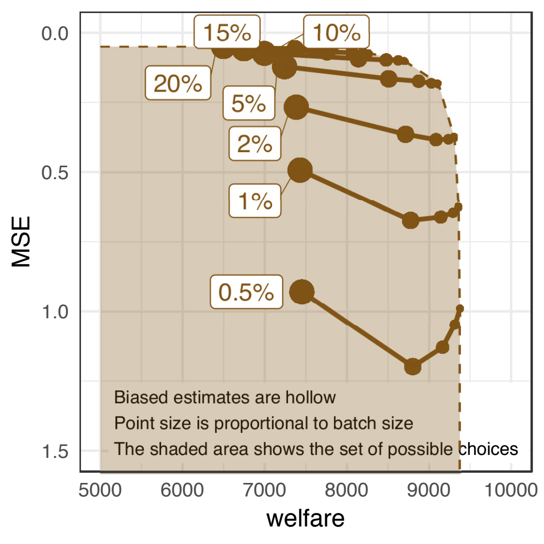
Welfare-estimation trade-off - varying batch size



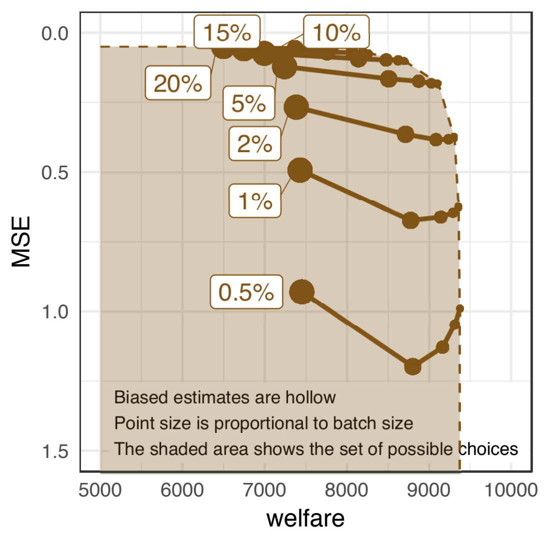
Welfare-estimation trade-off - varying batch size

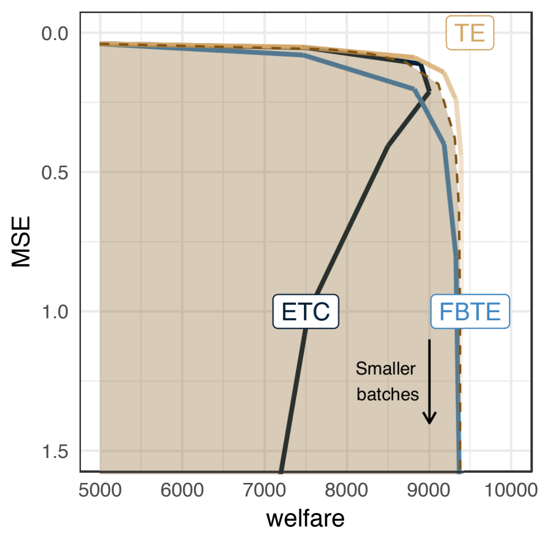


Solution: limited (capped) IPW

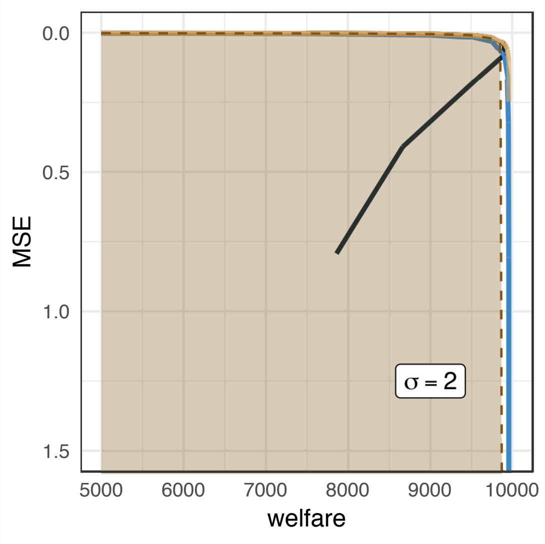


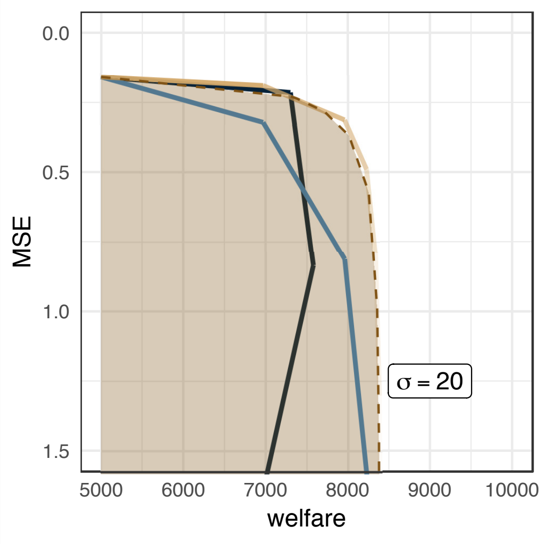
Solution: limited (capped) IPW



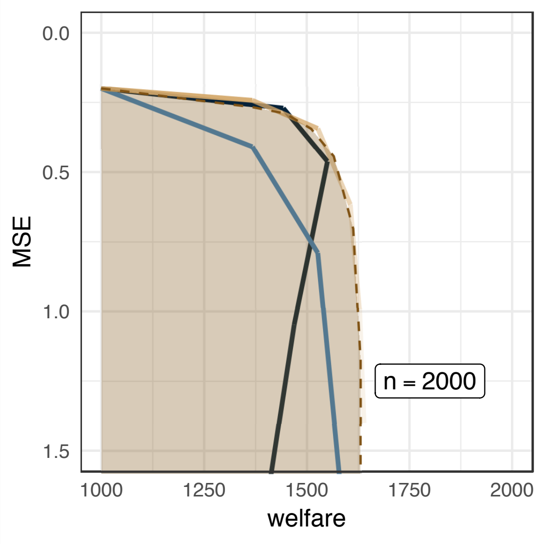


Different σ

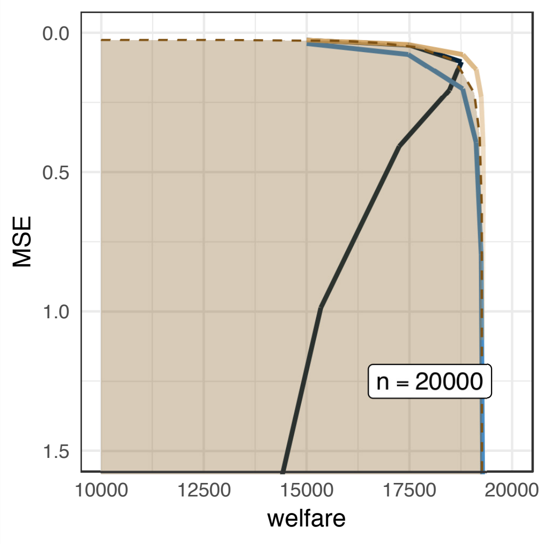




Different n



0.0
0.5
1.0
1.5
MSE
n = 2000
1000
1250
1500
1750
2000
welfare



0.0
0.5
1.0
1.5
MSE
n = 20000
10000
12500
15000
17500
20000
welfare

Conclusion

Summary of results

- there is a trade-off between welfare and estimation goals
- "quick" bandit and IPW with limited propensity scores extends the set of choices
- especially if the problem is hard (large noise-to-signal ratio)

Thank you for your attention