Empirical Welfare Maximization with Constraints

Liyang (Sophie) Sun, CEMFI.

ASSA 2023

Motivation

Welfare program eligibility criterion

- Given a limited budget, policymakers would adopt eligibility criterion
  - e.g. income thresholds to prioritize eligibilities for low-income population
- Ex. Medicaid is a U.S. public health insurance program intended for the low-income population
- Heated debate on how much to raise the income threshold for a Medicaid expansion:
  - Insurance reduces the price of health care and increases health care spending
  - Health insurance mitigates financial risk and provides health gain
  - Need causal estimates for the impact of Medicaid eligibility

Welfare program eligibility criterion

- Given a limited budget, policymakers would adopt eligibility criterion
  - e.g. income thresholds to prioritize eligibilities for low-income population
- Ex. Medicaid is a U.S. public health insurance program intended for the low-income population
- Heated debate on how much to raise the income threshold for a Medicaid expansion:
  - Insurance reduces the price of health care and increases health care spending
  - Health insurance mitigates financial risk and provides health gain
  - Need causal estimates for the impact of Medicaid eligibility

Empirical Example

- Great natural experiment: Oregon Health Insurance Experiment (OHIE)
  - Faced a budget constraint for its Medicaid program in 2008
  - Ran a lottery for fairness rather than ‘first come first serve’
  - Randomly assigned eligibility to apply for Medicaid
  - Measured health and health care spending after one year
- Finkelstein et al. (2012) found increase in health care spending but also improved subjective health (intent-to-treat estimates)
- Can we use the OHIE to answer the policy question:
  - How much should we raise the income threshold for a Medicaid expansion (potentially interacted with other characteristics)?

Empirical Example

- Great natural experiment: Oregon Health Insurance Experiment (OHIE)
  - Faced a budget constraint for its Medicaid program in 2008
  - Ran a lottery for fairness rather than 'first come first serve'
  - Randomly assigned eligibility to apply for Medicaid
  - Measured health and health care spending after one year
- Finkelstein et al. (2012) found increase in health care spending but also improved subjective health (intent-to-treat estimates)
- Can we use the OHIE to answer the policy question:
  - How much should we raise the income threshold for a Medicaid expansion (potentially interacted with other characteristics)?

Statistical rule

- Policymakers propose the eligibility criterion for a welfare program at the beginning of the fiscal year, aiming to balance budget and maximize welfare gain
- Translates to a constrained optimization problem in the target population
  - Choice variable is eligibility criterion, which can be more general than income thresholds
  - Both the objective function and the constraint need to be estimated (based on the RCT)
- Frontier approach proposes the Empirical Welfare Maximization (EWM) rule that accounts for estimation error in the objective function only
  - Kitagawa and Tetenov (2018); Athey and Wager (2020); among others
  - My contribution: theoretical implication of having to estimate the constraint in addition to the objective function, also see Donini et al. (2018), Sun, Du and Wager (2021)

Statistical rule

- Policymakers propose the eligibility criterion for a welfare program at the beginning of the fiscal year, aiming to balance budget and maximize welfare gain
- Translates to a constrained optimization problem in the target population
  - Choice variable is eligibility criterion, which can be more general than income thresholds
  - Both the objective function and the constraint need to be estimated (based on the RCT)
- Frontier approach proposes the Empirical Welfare Maximization (EWM) rule that accounts for estimation error in the objective function only
  - Kitagawa and Tetenov (2018); Athey and Wager (2020); among others
  - My contribution: theoretical implication of having to estimate the constraint in addition to the objective function, also see Donini et al. (2018), Sun, Du and Wager (2021)

Statistical rule

- Policymakers propose the eligibility criterion for a welfare program at the beginning of the fiscal year, aiming to balance budget and maximize welfare gain

- Translates to a constrained optimization problem in the target population
  - Choice variable is eligibility criterion, which can be more general than income thresholds
  - Both the objective function and the constraint need to be estimated (based on the RCT)

- Frontier approach proposes the Empirical Welfare Maximization (EWM) rule that accounts for estimation error in the objective function only
  - Kitagawa and Tetenov (2018); Athey and Wager (2020); among others
  - My contribution: theoretical implication of having to estimate the constraint in addition to the objective function, also see Donini et al. (2018), Sun, Du and Wager (2021)

Constrained Optimization Problem

Setup

- Eligibility criterion is an indicator function $g(X)$
  - Equals 1 if individual with observable characteristic X is assigned with eligibility
  - 0 otherwise
- Welfare function of criterion g is its average outcome in the target population P:

$$\begin{aligned} W(g;P) &:= E_P[Y_0 \cdot (1 - g(X)) + Y_1 \cdot g(X)] \\ &:= E_P[\Gamma \cdot g(X)] + E_P[Y_0] \end{aligned}$$

- Gain $\Gamma = Y_1 - Y_0$ is the difference in individual's potential outcomes due to eligibility
  - Unknown because we do not observe both Y_1 and Y_0

Setup

- Eligibility criterion is an indicator function $g(X)$
  - Equals 1 if individual with observable characteristic X is assigned with eligibility
  - 0 otherwise

- Welfare function of criterion g is its average outcome in the target population P:

$$
\begin{aligned}
W(g;P) &:= E_P[Y_0 \cdot (1 - g(X)) + Y_1 \cdot g(X)] \\
&:= E_P[\Gamma \cdot g(X)] + E_P[Y_0]
\end{aligned}
$$

- Gain $\Gamma = Y_1 - Y_0$ is the difference in individual's potential outcomes due to eligibility
  - Unknown because we do not observe both Y_1 and Y_0

Setup

- Budget function of criterion g is its average cost in the target population P at the end of the fiscal year:

$$B(g;P) := E_P[0 \cdot (1 - g(X)) + R \cdot g(X)] = E_P[R \cdot g(X)]$$

- Cost R is the potential resource government will spend to provide eligibility to this individual
  - Unknown before an individual becomes eligible e.g., R can be zero if eligible individual decides not to enroll, or much larger due to price response
- Individual's (bounded) $\underbrace{(\Gamma, R, X)}_{(gain,\ cost,\ characteristic)}$ distributed according to P in the target population and in the RCT sample

†Notation: Welfare function $W(g;P) := E_P[\Gamma \cdot g(X)]$ where $g(X)$ is an eligibility criterion and $\Gamma := Y_1 - Y_0$ is individual's gain due to eligibility.

Setup

- Budget function of criterion g is its average cost in the target population P at the end of the fiscal year:

$$B(g; P) := E_P[0 \cdot (1 - g(X)) + R \cdot g(X)] = E_P[R \cdot g(X)]$$

- Cost R is the potential resource government will spend to provide eligibility to this individual
  - Unknown before an individual becomes eligible e.g., R can be zero if eligible individual decides not to enroll, or much larger due to price response
- Individual's (bounded) $\underbrace{(\Gamma, R, X)}_{(gain,\ cost,\ characteristic)}$ distributed according to P in the target population and in the RCT sample

†Notation: Welfare function $W(g; P) := E_P[\Gamma \cdot g(X)]$ where $g(X)$ is an eligibility criterion and $\Gamma := Y_1 - Y_0$ is individual's gain due to eligibility.

Constrained Optimization Problem

- The constrained optimization problem is

$$\max_{g \in \mathcal{G}, B(g;P) \leq k} W(g;P)$$

- $\mathcal{G}$ is the set of available criteria e.g., all possible income thresholds
- Budget constraint $B(g;P) \leq k$
- The set of feasible criteria depends on P
- The solution g_P^* is the feasible and welfare-efficient criterion

†Notation: Welfare function $W(g;P) \coloneqq E_P[\Gamma \cdot g(X)]$ and budget function $B(g;P) \coloneqq E_P[R \cdot g(X)]$ where $g(X)$ is an eligibility criterion, Γ is individual's gain and R is individual's cost to the gov't due to eligibility.

Estimate the Welfare and Budget Functions

- Both $W(g;P)$ and $B(g;P)$ are unknown functions:

$$W(g;P) := E_P[\Gamma \cdot g(X)]$$
$$B(g;P) := E_P[R \cdot g(X)]$$

- Use sample-analog versions based on data

$$\widehat{W}_n(g) := \frac{1}{n}\sum_i \Gamma_i^* \cdot g(X_i), \quad \widehat{B}_n(g) := \frac{1}{n}\sum_i R_i^* \cdot g(X_i)$$

where Γ_i^* is the gain estimate for Γ and R_i^* is the cost estimate for R

†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \leq k} W(g;P)$ where $W(g;P)$ is the welfare function and $B(g;P)$ is the budget function

Gain and cost estimates (Γ_i^*, R_i^*)

- For an RCT where eligibility D_i is randomized by X_i and observed outcome Y_i and cost Z_i, we can use

$$\Gamma_i^* = \left(\frac{D_i}{p(X_i)} - \frac{1-D_i}{1-p(X_i)}\right) \cdot Y_i$$

$$R_i^* = \frac{D_i}{p(X_i)} \cdot Z_i$$

  for known propensity score $p(X_i) = Pr\{D_i = 1 \mid X_i\}$
- In general we can use doubly-robust score where we can use ML to estimate $p(X_i)$ formula
- Well-known estimates (in causal inference): unbiased, e.g. $E_P[\frac{1}{n}\sum_i \Gamma_i^*] = E_P[\Gamma]$, and asymptotically normal

†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \leq k} W(g;P)$ where $W(g;P)$ is the welfare function and $B(g;P)$ is the budget function

Uniform estimation quality

- Use sample-analogs from RCT data to estimate functions:

$$\widehat{W}_n(g) := \frac{1}{n}\sum_i \Gamma_i^* \cdot g(X_i) \text{ for } W(g;P) := E_P[\Gamma \cdot g(X)]$$

$$\widehat{B}_n(g) := \frac{1}{n}\sum_i R_i^* \cdot g(X_i) \text{ for } B(g;P) := E_P[R \cdot g(X)]$$

- For a fixed g, the sample-analogs are unbiased, e.g. $E_P[\widehat{W}_n(g)] = W(g;P)$, and asymptotically normal
- In fact, asymptotic normality holds for all criteria $g \in \mathcal{G}$ uniformly when $\mathcal{G}$ has a finite VC-dimension, and holds for a large class of P
- Allows us to consider the asymptotic behavior for statistical rules $\widehat{g}$ that are based on $\widehat{W}_n(g)$ and $\widehat{B}_n(g)$

†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \le k} W(g;P)$ where $W(g;P)$ is the welfare function and $B(g;P)$ is the budget function

Statistical Rules

What are some candidate rules?

- The direct extension to the existing EWM approach is a rule that selects the criterion

$$\widehat{g}_{\text{sample}} \in \arg\max_{g \in \mathcal{G}} \{\widehat{W}_n(g) \text{ subject to } \widehat{B}_n(g) \leq k\}.$$

- How should we compare $\widehat{g}_{\text{sample}}$ with some other data-based rule $\widehat{g}$?
- Need to define statistical properties that an ideal $\widehat{g}$ should satisfy
- Ideally the following probabilities should approach zero for any target population $P \in \mathcal{P}$ definitions
    1) likely to select feasible criteria i.e., small $Pr_{P^n}\{B(\widehat{g}; P) > k\}$
    2) likely to achieve close to (constrained) efficient welfare i.e., small $Pr_{P^n}\{W(\widehat{g}; P) - W(g_P^*; P) < -\epsilon\}$ for $\epsilon > 0$

†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \leq k} W(g; P)$ with solution g_P^* in the target population P; RCT data gives estimates $\widehat{W}_n(g)$ and $\widehat{B}_n(g)$; Statistical rule $\widehat{g}$ selects a criterion based on the RCT data of size n and $Pr_{P^n}(\cdot)$ denotes probability over repeated samples

What are some candidate rules?

- The direct extension to the existing EWM approach is a rule that selects the criterion

$$\widehat{g}_{\text{sample}} \in \arg\max_{g \in \mathcal{G}} \{\widehat{W}_n(g) \text{ subject to } \widehat{B}_n(g) \leq k\}.$$

- How should we compare $\widehat{g}_{\text{sample}}$ with some other data-based rule $\widehat{g}$?
- Need to define statistical properties that an ideal $\widehat{g}$ should satisfy
- Ideally the following probabilities should approach zero for any target population $P \in \mathcal{P}$ definitions
    1) likely to select feasible criteria i.e., small $Pr_{P^n}\{B(\widehat{g}; P) > k\}$
    2) likely to achieve close to (constrained) efficient welfare i.e., small $Pr_{P^n}\{W(\widehat{g}; P) - W(g_P^*; P) < -\epsilon\}$ for $\epsilon > 0$

†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \leq k} \; W(g; P)$ with solution g_P^* in the target population P; RCT data gives estimates $\widehat{W}_n(g)$ and $\widehat{B}_n(g)$; Statistical rule $\widehat{g}$ selects a criterion based on the RCT data of size n and $Pr_{P^n}(\cdot)$ denotes probability over repeated samples

Sample-Analog Rule

- Recall the sample-analog rule

$$\widehat{g}_{\text{sample}} \in \arg\max_{g \in \mathcal{G}} \{\widehat{W}_n(g) \text{ subject to } \widehat{B}_n(g) \leq k\}.$$

- N.B. the sample-analog rule $\widehat{g}_{\text{sample}}$ compares $\widehat{B}_n(g)$ with k directly to assess whether $B(g; P) \leq k$
- The sample-analog budget $\widehat{B}_n(g)$ can always over- or under-estimate the budget $B(g; P)$

†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \leq k} W(g; P)$ with solution g_P^* in the target population P; RCT data gives estimates $\widehat{W}_n(g)$ and $\widehat{B}_n(g)$

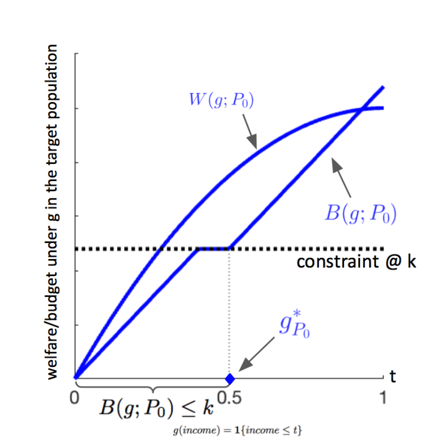
Sample-Analog Rule

- Consider a target population P_0 where the constraint binds exactly and $W(g^*_{P_0}; P_0)$ is separated from that of other non-binding criteria

  Assumption

- Real world example: "pay for themselves" programs suggested by Hendren and Sprung-Keyser (2020)
- The sample-analog rule $\widehat{g}_{\text{sample}}$ is neither asymptotically feasible nor asymptotically welfare efficient

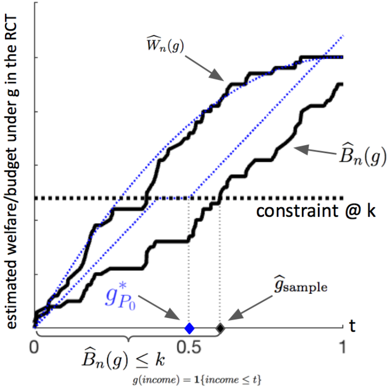
  Proposition



†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \leq k} W(g; P)$ with solution g^*_P in the target population P; The sample-analog rule is $\widehat{g}_{\text{sample}} \in \arg\max_{g \in \mathcal{G}} \{\widehat{W}_n(g)$ subject to $\widehat{B}_n(g) \leq k\}$

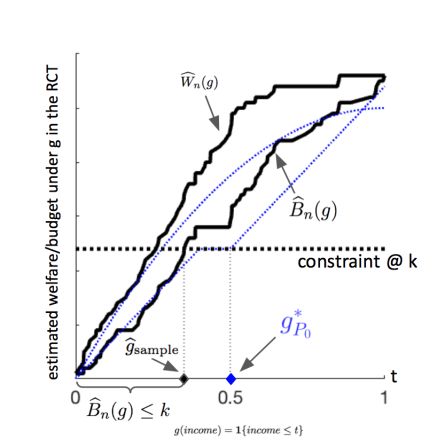
Asymptotic Infeasible Example

- Observe a particular draw of the data that gives under-estimated $\widehat{B}_n(g)$
- The sample-analog rule $\widehat{g}_{\text{sample}}$ selects an infeasible criterion
- Amount of under-estimation gets smaller, but there is always a chance



†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \leq k} \; W(g;P)$ with solution g^*_P in the target population P; The sample-analog rule is $\widehat{g}_{\text{sample}} \in \arg\max_{g \in \mathcal{G}} \{\widehat{W}_n(g)$ subject to $\widehat{B}_n(g) \leq k\}$

Asymptotic Inefficient Example



- Observe a particular draw of the data that gives over-estimated $\widehat{B}_n(g)$
- The sample-analog rule $\widehat{g}_{\text{sample}}$ selects a strictly inefficient criterion
- Amount of over-estimation gets smaller, but there is always a chance

†Notation: The population problem is $\max_{g\in\mathcal{G}, B(g;P)\leq k}\ W(g;P)$ with solution g^*_P in the target population P; The sample-analog rule is $\widehat{g}_{\text{sample}} \in \arg\max_{g\in\mathcal{G}}\{\widehat{W}_n(g)$ subject to $\widehat{B}_n(g) \leq k\}$

Theoretical results

- More generally, my paper shows an impossibility result: there exists no statistical rule that can approximate the solution to

$$\max_{g \in \mathcal{G}, B(g;P) \leq k} W(g;P)$$

uniformly well over $\mathcal{P}$ in terms of achieving both feasibility and efficient welfare even asymptotically theorem

- Intuition: when $\mathcal{P}$ is sufficiently rich, $\widehat{g}$ differ only marginally at nearby pairs of distributions. But g_P^*, the efficient and feasible criteria, can still look very different at nearby pairs of distributions (esp. corner solutions), causing uniformity problems in any $\widehat{g}$ visual

1) We can explicitly control the probability of violating the budget constraint with a mistake-controlling rule $\widehat{g}_{\text{mistake}}$
2) We can relax the budget constraint and solve a modified optimization problem with a trade-off rule $\widehat{g}_{\text{tradeoff}}$

†Notation: Statistical rule $\widehat{g}$ selects a criterion based on the RCT data of size n and $Pr_{P^n}(\cdot)$ denotes probability over repeated samples; Asymptotic feasibility requires $Pr_{P^n}\{B(\widehat{g};P) > k\} \to 0$ and asymptotic welfare efficiency requires $Pr_{P^n}\left\{W(\widehat{g};P) - W(g_P^*;P) < -\epsilon\right\} \to 0$ for any $\epsilon > 0$

Theoretical results

- More generally, my paper shows an impossibility result: there exists no statistical rule that can approximate the solution to

$$\max_{g \in \mathcal{G}, B(g;P) \leq k} W(g;P)$$

uniformly well over $\mathcal{P}$ in terms of achieving both feasibility and efficient welfare even asymptotically theorem

- Intuition: when $\mathcal{P}$ is sufficiently rich, $\widehat{g}$ differ only marginally at nearby pairs of distributions. But g^*_P, the efficient and feasible criteria, can still look very different at nearby pairs of distributions (esp. corner solutions), causing uniformity problems in any $\widehat{g}$ visual

1) We can explicitly control the probability of violating the budget constraint with a mistake-controlling rule $\widehat{g}_{\text{mistake}}$
2) We can relax the budget constraint and solve a modified optimization problem with a trade-off rule $\widehat{g}_{\text{tradeoff}}$

†Notation: Statistical rule $\widehat{g}$ selects a criterion based on the RCT data of size n and $Pr_{P^n}(\cdot)$ denotes probability over repeated samples; Asymptotic feasibility requires $Pr_{P^n}\{B(\widehat{g};P) > k\} \to 0$ and asymptotic welfare efficiency requires $Pr_{P^n}\left\{W(\widehat{g};P) - W(g^*_P;P) < -\epsilon\right\} \to 0$ for any $\epsilon > 0$

#1 Mistake-Controlling Rule

- We can explicitly control the probability of violating the budget constraint with a mistake-controlling rule $\widehat{g}_{\text{mistake}}$

$$\widehat{g}_{\text{mistake}} \in \arg\max_{g \in \hat{\mathcal{G}}_\alpha \subset \mathcal{G}} \widehat{W}_n(g)$$

$$\hat{\mathcal{G}}_\alpha = \left\{ g : g \in \mathcal{G} \text{ and } \widehat{B}_n(g) \leq k + \frac{\widehat{\Sigma}^B(g,g)^{1/2}}{\sqrt{n}} \cdot c_\alpha \right\}$$

- $c_\alpha < 0$, a critical value that can be calculated based on bootstrap for a pre-specified significance level, e.g. $\alpha = 5\%$
- $\widehat{\Sigma}^B(\cdot,\cdot)$, a consistent estimator for the asymptotic variance of $\widehat{B}_n(\cdot)$
- The chance $\hat{\mathcal{G}}_\alpha$ contains any infeasible criterion is less than α asymptotically

slackness

#2 Trade-Off Rule

- We can relax the budget constraint and solve a modified optimization problem

$$\widehat{g}_{\text{tradeoff}} \in \arg\max_{g \in \mathcal{G}} \left\{ \widehat{W}_n(g) - \overline{\lambda} \cdot \max\{\widehat{B}_n(g) - k, 0\} \right\}$$

  for a trade-off coefficient $\overline{\lambda}$ that bounds the marginal gain from relaxing the constraint
- Ex. $\bar{\lambda}$ converts self-reported health to 0.6 of a statistical life year (VSLY) for the application of Medicaid (Finkelstein, Hendren and Luttmer, 2019)
- Approximates the solution to $\max_{g \in \mathcal{G}} \left\{ W(g;P) - \overline{\lambda} \cdot \max\{B(g;P) - k, 0\} \right\}$ instead of the original problem, which achieves at least the constrained efficient welfare and violates the budget no larger than $\frac{W(\widetilde{g}_P;P) - W(g_P^*;P)}{\overline{\lambda}}$

Application

Medicaid Expansion

- A more flexible Medicaid expansion criterion that varies income thresholds by number of children:

$$\mathcal{G} = \left\{ g(x) = \begin{cases} \mathbf{1}\{\text{income} \leq \beta_1\}, & \text{number of children} = 0 \\ \mathbf{1}\{\text{income} \leq \beta_2\}, & \text{number of children} = 1 \\ \mathbf{1}\{\text{income} \leq \beta_3\}, & \text{number of children} \geq 2 \end{cases} \right\}$$

  for characteristics $X = (\text{income}, \text{ number of children})$ and $\beta_j \geq \mathbf{0}$.
- Health gain Γ: increase in the probability of reporting "excellent/very good/good" on self-reported health (as opposed to "poor/fair") after receiving Medicaid eligibility
- Cost to Medicaid R: health care expenditure that needs to be reimbursed by Medicaid, in excess of previous year's level ($6,000)
- Budget threshold is normalized $k = 0$
- Both (Γ, R) are unknown and estimated by the OHIE

OHIE Sample Summary Statistics

- If make everyone eligible, overall health gain but costlier than pervious year

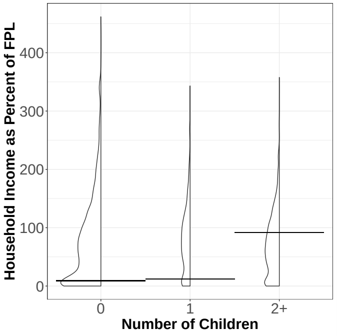
	Sample size	Sample mean of Γ_i^*	Sample mean of R_i^*
Overall	10,135	3.9%	\$358

- Heterogeneous health gain and cost by parental status

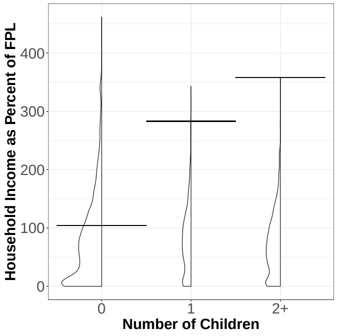
Number of children	Sample size	Sample mean of Γ_i^*	Sample mean of R_i^*
0	5,758	3.1%	\$651
1	1,736	10.3%	\$348
$\geq$2	2,641	1.5%	-\$275

Selected Eligibility Criterion

- $\widehat{g}_{\text{mistake}}$ solves $\max_{g \in \hat{\mathcal{G}}_\alpha} \widehat{W}_n(g)$ for a significance level $\alpha = 5\%$



- $\widehat{g}_{\text{tradeoff}}$ solves $\max_{g \in \mathcal{G}} \{\widehat{W}_n(g) - \bar{\lambda} \cdot \max\{\widehat{B}_n(g) - k, 0\}\}$ for $\bar{\lambda}$ converting self-reported health into value of a statistical life year (VSLY)



Notes: The horizontal bars mark the income thresholds selected by each rule, below which individuals are assigned Medicaid eligibility.

Conclusion

Conclusion

- I build on the EWM literature to allow for a constraint
- Tackle the challenge where the set of feasible criteria needs to be estimated
  - Key results have theoretical implications for many practical settings
- I propose two new rules and apply them to select new income thresholds for U.S. public health insurance
  - Mistake-controlling rule
  - Trade-off rule
- Thank you!

Conclusion

- I build on the EWM literature to allow for a constraint
- Tackle the challenge where the set of feasible criteria needs to be estimated
  - Key results have theoretical implications for many practical settings
- I propose two new rules and apply them to select new income thresholds for U.S. public health insurance
  - Mistake-controlling rule
  - Trade-off rule
- Thank you!

Doubly-Robust Score

- Suppose eligibility (D_i) is random after adjusting for X_i. Define the conditional expectation function (CEF) of a random variable U_i as $\gamma^U = E[U_i \mid X_i, D_i]$.
- For Y_i the observed outcome, define the doubly-robust score e.g. Chernozhukov et al. (2018)

$$\Gamma_i^* = \widehat{\gamma}^Y(X_i, 1) - \widehat{\gamma}^Y(X_i, 0) + \widehat{\alpha}(X_i, D_i) \cdot \left(Y_i - \widehat{\gamma}^Y(X_i, D_i)\right)$$

where $\widehat{\alpha}(X_i, D_i) = \frac{D_i}{\widehat{p}(X_i)} - \frac{1-D_i}{1-\widehat{p}(X_i)}$ is the estimated Horvitz-Thompson weight

- For Z_i the observed cost,

$$R_i^* = \widehat{\gamma}^Z(X_i, 1) + \frac{D_i}{\widehat{p}(X_i)} \cdot \left(Z_i - \widehat{\gamma}^Z(X_i, D_i)\right)$$

return

Uniform Version

Definition 1.

A statistical rule $\widehat{g}$ is *uniformly asymptotically feasible* over the class of distributions $\mathcal{P}$ if

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} Pr_{P^n}\{B(\widehat{g};P) > k\} = 0.$$

Definition 2.

A statistical rule $\widehat{g}$ is *uniformly asymptotically optimal* over the class of distributions $\mathcal{P}$ if for any $\epsilon > 0$

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} Pr_{P^n}\left\{W(\widehat{g};P) - W(g_P^*;P) < -\epsilon\right\} = 0.$$

return

Impossibility Result

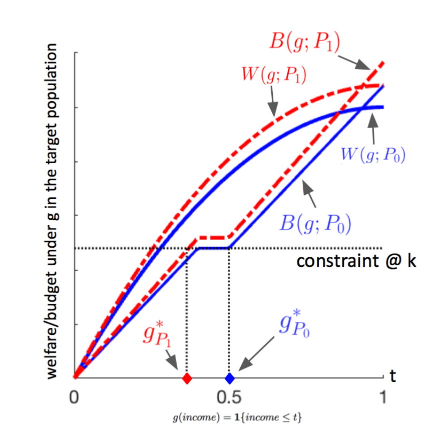
Theorem 1.

Under some regularity conditions on the class of distributions $\mathcal{P}$, no statistical rule $\widehat{g}$ can be both asymptotically feasible and asymptotically welfare efficient uniformly over all $P \in \mathcal{P}$

- When $\mathcal{P}$ is sufficiently rich, $\widehat{g}$ differ only marginally at nearby pairs of distributions assumptions
- But g_P^*, the optimal and feasible criteria, can still look very different (esp. corner solutions), causing uniformity problems in any $\widehat{g}$

†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \leq k} \ W(g;P)$ with solution g_P^* in the target population P; Statistical rule $\widehat{g}$ selects a criterion based on the RCT data of size n; Asymptotic feasibility requires $Pr_{P^n}\{B(\widehat{g};P) > k\} \to 0$ and asymptotic welfare efficiency requires $Pr_{P^n}\left\{W(\widehat{g};P) - W(g_P^*;P) < -\epsilon\right\} \to 0$ for any $\epsilon > 0$

Visual Explanation for the Impossibility Result



- Consider P_0 (blue) and P_1 (red) that belong to $\mathcal{P}$
- Induce similar data and any $\widehat{g}$ selects similar criteria under P_0 and P_1
- But P_0 and P_1 have distinct feasible sets
- If $\widehat{g}$ is required to be asymptotically feasible and asymptotically welfare efficient under P_0, then $\widehat{g}$ cannot be asymptotically feasible under P_1

return

†Notation: The population problem is $\max_{g \in \mathcal{G}, B(g;P) \leq k} \; W(g;P)$ with solution g^*_P in the target population P; Statistical rule $\widehat{g}$ selects a criterion based on the RCT data of size n and $Pr_{P^n}(\cdot)$ denotes probability over repeated samples; Asymptotic feasibility requires $Pr_{P^n}\{B(\widehat{g};P) > k\} \to 0$ and asymptotic welfare efficiency requires $Pr_{P^n}\left\{W(\widehat{g};P) - W(g^*_P;P) < -\epsilon\right\} \to 0$ for any $\epsilon > 0$

Regularity Conditions

Assumption 1.

Contiguity. There exists a distribution $P_0 \in \mathcal{P}$ *under which a non-empty set of eligibility criteria satisfies the constraint exactly* $\mathcal{G}_0 = \{g : B(g; P_0) = k\}$*. Furthermore, the class of distributions* $\mathcal{P}$ *includes a sequence of data distributions* $\{P_{h_n}\}$ *contiguous to* P_0*, under which for all* $g \in \mathcal{G}_0$*, there exists some* $C > 0$ *such that*

$$\sqrt{n} \cdot (B(g; P_{h_n}) - k) > C.$$

Assumption 2.

Binding constraint. Under the data distribution P_0*, the constraint is satisfied exactly at the constrained optimum i.e.* $B(g^*_{P_0}; P_0) = k$*.*

return

Regularity Conditions

Assumption 3.

$W(g^*_{P_0}; P_0)$ is separated from other feasible criteria with different $B(g; P_0)$. Under the data distribution P_0, $\exists \epsilon > 0$ such that for any feasible criterion g, whenever

$$\left| B(g; P_0) - B(g^*_{P_0}; P_0) \right| > 0,$$

we have

$$W(g^*_{P_0}; P_0) - W(g; P_0) > \epsilon.$$

return

Feasibility and welfare efficiency under slackness

Theorem 2.

If $\alpha_n \to 0$ *as* $n \to \infty$*, then the mistake-controlling rule* $\widehat{g}_{mistake}$ *is asymptotically feasible uniformly over* $P \in \mathcal{P}$*.*
For a slack population $B(g_P^*; P) < k$*, if* $\alpha_n \to 0$ *at a rate such that* $c_{\alpha_n} = o(n^{1/2})$*,* $\widehat{g}_{mistake}$ *is asymptotically welfare efficient and asymptotically feasible*

- For example, can take $c_{\alpha_n} = -\log n$ return