ECMA31100 Introduction to Empirical Analysis II

Winter 2022, Week 8, Part 1: Discussion Session

Conroy Lau

The University of Chicago

March 2, 2022

Part 1 of Week 8

Topics for this week's TA session

- Due to popular request, I will focus on the following two topics:
  - Treatment choice
  - Difference-in-differences/event study

Part 1: Treatment choice

- I will focus on the main ideas of two recent papers
- Due to time constraint, I need to skip some technical details and the proofs
- I will focus on methodology and implementation so you can try to implement it

Motivation

- Kitagawa and Tetenov (2018) — empirical welfare maximization
- Mbakop and Tabord-Meehan (2021) — penalized welfare maximization

Idea

- A sample with different observable covariates
- We may not maximize welfare if all are treated
- There might also be some capacity constraints so that we cannot treat everyone
- Hence, we want to determine treatment allocation that maximize social welfare

Contents

Framework

Notations

- Binary treatment: $D \in \{0,1\}$
- Potential outcomes and covariates: $(Y(1), Y(0), X) \sim Q$ where $X \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$
- Observed outcome: $Y = Y(1)D + Y(0)(1-D)$
- Observed data: $\{Z_i\}_{i=1}^n = \{(Y_i, X_i, A_i)\}_{i=1}^n \sim P$, where $Z_i \in \mathcal{Z}$
- Set of all feasible treatment allocations: $\mathcal{G}$
- Conditional mean treatment response: $m_d(x) \equiv \mathbb{E}_Q[Y(d)|X=x]$ for $d \in \{0,1\}$
- Conditional average treatment effect: $\tau(x) \equiv m_1(x) - m_0(x)$
- Propensity score: $e(X) \equiv \mathbb{P}[D=1|X]$

Goal

- The planner wishes to max. utilitarian welfare $\mathbb{E}_Q\left[Y(1)\mathbb{1}[X \in G] + Y(0)\mathbb{1}[X \notin G]\right]$

Goal

Optimal treatment rule

- By the confoundedness assumption (i.e., $(Y(1), Y(0)) \perp\!\!\!\perp D|X$), we can write it as

$$\mathbb{E}_Q[Y(0)] + \underbrace{\mathbb{E}_P\left[\left(\frac{YD}{e(X)} - \frac{Y(1-D)}{1-e(X)}\right)\mathbb{1}[X \in G]\right]}_{\equiv W(G) = \mathbb{E}_P[\tau(X)\mathbb{1}[X \in G]]}$$

- We will focus on $W(G)$ because $\mathbb{E}_Q[Y(0)]$ does not depend on G
- The optimal treatment rule is $G^\star \in \arg\max_{G \in \mathcal{G}} W(G)$

First-best and second-best

- The **first-best** decision set is $G^\star_{\text{FB}} \equiv \{x \in \mathcal{X} : \tau(x) \geq 0\}$
- $W^\star_{\mathcal{G}} \equiv \sup_{G \in \mathcal{G}} W(G)$ is **second-best** if $W^\star_{\mathcal{G}} < W(G^\star_{\text{FB}})$

Any questions?

Contents

Empirical welfare maximization

Optimal treatment rule

$$G^\star \in \arg\max_{G \in \mathcal{G}} W(G) \tag{1}$$

Empirical welfare maximization (EWM)

- EWM maximizes a sample analog of (1), i.e.,

$$\widehat{G}_{\mathrm{EWM}} \in \arg\max_{G \in \mathcal{G}} W_n(G) \;\text{ where }\; W_n(G) \equiv \frac{1}{n}\sum_{i=1}^{n}\left[\underbrace{\left(\frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)}\right)}_{\equiv \tau_i} \mathbb{1}[X_i \in G]\right]$$

- This can be written as a mixed-integer linear program (MILP)

Assumptions and examples

Assumptions

- Unconfoundedness (UCF): $(Y_1, Y_0) \perp\!\!\!\perp D | X$
- Bounded outcomes (BO): There exists $M < \infty$ such that $Y \in [-\frac{M}{2}, \frac{M}{2}]$
- Strict overlap (SO): There exists $\kappa \in (0, \frac{1}{2})$ such that $e(x) \in [\kappa, 1-\kappa]$ for all $x \in \mathcal{X}$
- VC-class (VC): A class of decision sets $\mathcal{G}$ has finite VC-dim. $v < \infty$ and is countable

Two examples of $\mathcal{G}$

1. Linear eligibility score: ($x_s \in \mathbb{R}^{d_s}$ is a subvector of x)

$$\mathcal{G}_{\text{LES}} \equiv \{\{x \in \mathbb{R}^{d_x} : \beta_0 + x_s' \beta_s \geq 0\} : (\beta_0, \beta_s') \in \mathbb{R}^{d_s+1}\}$$

2. Multiple linear index rules:

$$\mathcal{G}_{\text{MLIR}} \equiv \{\{x : x'\beta^1 \geq 0, \ldots, x'\beta^J \geq 0\} : \beta_1, \ldots, \beta_J \in B \subseteq \mathbb{R}^{d_x}\}$$

Capacity constraints

- Assume the % of target population that receive treatment cannot exceed $K \in (0, 1)$
- If we know the distribution of covariates P_X, then we can impose this on $\mathcal{G}$

Unknown distribution of covariates P_X

- If P_X is unknown, then $\widehat{G}$ may not satisfy the capacity constraint
  - Violation $\Longrightarrow$ allocate treatment to a fraction $\frac{K}{P_X(G)}$ of assigned receipents (**rationing**)
  - No violation $\Longrightarrow$ No rationing
- The **capacity-constrained welfare** is

$$W^K(G) \equiv \mathbb{E}_Q\left[\left[Y(1)\min\left\{1, \frac{K}{P_X(G)}\right\} + Y(0)\left(1 - \min\left\{1, \frac{K}{P_X(G)}\right\}\right)\right]\mathbb{1}[X \in G] + Y_0\mathbb{1}[X \notin G]\right]$$

- Then, maximize the sample analog of **capacity-constrained welfare gain**

$$\widehat{G}^K \equiv \arg\max_{G \in \mathcal{G}} V_n^K(G) \quad \text{where} \quad V_n^K(G) \equiv W_n^K(G) - W_n^K(\emptyset)$$

Any questions?

Contents

Penalized welfare maximization

Notations and assumption

- Let $\mathcal{G}$ be a "large" class that can be approximated by less complex classes:

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{G}_3 \subseteq \cdots \subseteq \mathcal{G}_k \subseteq \mathcal{G}_{k+1} \subseteq \cdots \subseteq \mathcal{G} \tag{2}$$

- Let $\widehat{G}_{n,k}$ be the EWM rule for class $\mathcal{G}_k$, i.e., $\widehat{G}_{n,k} \equiv \arg\max_{G \in \mathcal{G}_k} W_n(G)$
- We can write $\mathbb{E}_{P^n}[W^\star_{\mathcal{G}} - W(\widehat{G}_{n,k})] = \underbrace{\mathbb{E}_{P^n}[W^\star_{\mathcal{G}_k} - W(\widehat{G}_{n,k})]}_{\text{Estimation error}} + \underbrace{W^\star_{\mathcal{G}} - W^\star_{\mathcal{G}_k}}_{\text{Approximation error}}$

Example of sieves $\{\mathcal{G}_k\}_k$ for thresholds allocation in K-dimension

- $\mathcal{G} \equiv \{G \subset \mathcal{X} : G = \{x \in \mathcal{X} : s_k x_k \leq \bar{x}_k,\ k = 1, \ldots, K\}, \bar{x} \in \mathbb{R}^K, s \in \{-1, 1\}^K\}$
- For large K, VC($\mathcal{G}$) may be large relative to sample size
- The sieve sequence $\{\mathcal{G}_k\}_{k=1}^K$ where $\mathcal{G}_k$ uses $(k-1)$ out of K covariates

Penalized welfare maximization

Penalized welfare maximization (PWM)

$$\widehat{G}_n \equiv \widehat{G}_{n,\widehat{k}^\star} \equiv \arg\max_{G \in \mathcal{G}_{\widehat{k}^\star}} W_n(G)$$

where $\widehat{k}^\star \equiv \arg\max_k R_{n,k}(\widehat{G}_{n,k})$ and $R_{n,k}(G) \equiv W_n(G) - C_n(k) - \sqrt{\frac{t_k}{n}}$

- $C_n(k)$ is a penalty that measures the amount of "overfitting" from using $\widehat{G}_{n,k}$
- $\{t_k\}_{k=1}^{\infty}$ an increasing sequence and taken to be $t_k = k$ for simplicity

Some examples of penalties

- Holdout penalty (focus of today)
- Rademacher penalty

Penalized welfare maximization — Holdout penalty

Sample splitting

- Let $S_n \equiv \{(Y_i, D_i, X_i)\}_{i=1}^n$ be the full sample, $\ell \in (0,1)$, $m \equiv n(1-\ell)$ and $r \equiv n - m$
- Construct $S_n^E \equiv \{(Y_i, D_i, X_i)\}_{i=1}^m$ (estimating) and $S_n^T \equiv \{(Y_i, D_i, X_i\}_{i=m+1}^n$ (testing)

Holdout penalty

- Let the empirical welfare for each rule $\widehat{G}_{\boldsymbol{m},k}$ on S_n^E and S_n^T be

$$W_m(\widehat{G}_{\boldsymbol{m},k}) = \frac{1}{\boldsymbol{m}} \sum_{i=1}^{\boldsymbol{m}} \tau_i \mathbb{1}[X_i \in \widehat{G}_{\boldsymbol{m},k}] \quad \text{and} \quad W_r(\widehat{G}_{\boldsymbol{m},k}) \equiv \frac{1}{r} \sum_{i=\boldsymbol{m}+1}^{n} \tau_i \mathbb{1}[X_i \in \widehat{G}_{\boldsymbol{m},k}]$$

- The holdout penalty is $C_{\boldsymbol{m}}(k) \equiv W_m(\widehat{G}_{\boldsymbol{m},k}) - W_r(\widehat{G}_{\boldsymbol{m},k})$
- PWM rule: $\widehat{G}_{\boldsymbol{m}} \equiv \arg\max_k \left[W_m(\widehat{G}_{\boldsymbol{m},k}) - C_{\boldsymbol{m}}(k) - \sqrt{\frac{k}{\boldsymbol{m}}} \right] = \arg\max_k \left[W_r(\widehat{G}_{\boldsymbol{m},k}) - \sqrt{\frac{k}{\boldsymbol{m}}} \right]$

Implementation

▸ MILP

Threshold allocations on d covariates

- Let $x \equiv (1, x^{(1)}, \ldots, x^{(d)})'$ be a $(d+1)$-vector
- Define $\mathcal{A} \equiv \{1, \ldots, d\}$ as the threshold dimension

MILP

▸ MLIR

- The formulation is similar to the implementation for multiple linear index rules
- We form a MILP that contains each of the thresholds $a \in \mathcal{A}$
- Each $\mathcal{G}_k$ uses $k-1$ covariates out of the d covariates
  - For $k = 1$, i.e., $\mathcal{G}_1$, we have either $\mathcal{G} = \emptyset$ or $\mathcal{G} = \mathcal{X}$
  - For $k = d+1$, i.e., $\mathcal{G}_d$, we use all available covariates

Any questions?

Contents

Estimated propensity score for EWM

- Propensity score is usually unknown

  $\Longrightarrow$ Kitagawa and Tetenov (2018) introduce **hybrid EWM** (with plug-in estimators)

m-hybrid rule

$$\widehat{G}_{\text{m-hybrid}} \in \arg\max_{G \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} \widehat{\tau}^{m}(X_i) \mathbb{1}[X_i \in G]$$

where $\widehat{\tau}^{m}(X_i) \equiv \widehat{m}_1(X_i) - \widehat{m}_0(X_i)$ and $\widehat{m}_d(x) \equiv \widehat{\mathbb{E}}[Y_d | X = x] = \widehat{\mathbb{E}}_n[Y | X = x, D = d]$ for $d \in \{0, 1\}$

e-hybrid rule

$$\widehat{G}_{\text{e-hybrid}} \in \arg\max_{G \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} \widehat{\tau}_i^{e} \mathbb{1}[X_i \in G]$$

where $\widehat{\tau}_i^{e} \equiv \left(\frac{Y_i D_i}{\widehat{e}(X_i)} - \frac{Y_i D_i}{1 - \widehat{e}(X_i)} \right) \mathbb{1}[\epsilon_n \leq \widehat{e}(X_i) \leq 1 - \epsilon_n]$ and $\epsilon_n \longrightarrow 0$ as $n \longrightarrow \infty$

Estimated propensity score for PWM

Unknown propensity score

- Same problem as in EWM with estimated propensity score $\widehat{e}(X_i)$
- We can define the hybrid EWM rules by using the sample analog and $\widehat{e}(X_i)$

PWM rule with hybrid holdout penalty

- Let $\widehat{e}^E(\cdot)$ be the estimated propensity score using S_n^E (similarly $\widehat{e}^T(\cdot)$ uses S_n^T)
- Define the empirical welfare on S_n^E and S_n^T as follows:

$$W_m^e(G) \equiv \frac{1}{\boldsymbol{m}} \sum_{i=1}^{\boldsymbol{m}} \widehat{\tau}_i^E \mathbb{1}[X_i \in G] \quad \text{where } \widehat{\tau}_i^E \equiv \left[\frac{Y_i D_i}{\widehat{e}^E(X_i)} - \frac{Y_i(1-D_i)}{1-\widehat{e}^E(X_i)} \right] \mathbb{1}[\epsilon_n \leq \widehat{e}^E(X_i) \leq 1 - \epsilon_i]$$

$$W_r^e(G) \equiv \frac{1}{r} \sum_{i=\boldsymbol{m}+1}^{n} \widehat{\tau}_i^E \mathbb{1}[X_i \in G] \quad \text{where } \widehat{\tau}_i^T \equiv \left[\frac{Y_i D_i}{\widehat{e}^T(X_i)} - \frac{Y_i(1-D_i)}{1-\widehat{e}^T(X_i)} \right] \mathbb{1}[\epsilon_n \leq \widehat{e}^T(X_i) \leq 1 - \epsilon_i]$$

- The hybrid holdout penalty is $C_m^e(k) \equiv W_m^e(\widehat{G}_{\boldsymbol{m},k}^e) - W_r^e(\widehat{G}_{\boldsymbol{m},k}^e)$

Any questions?

Contents

Empirical example

Background

- Experimental data from the national job training partnership act (JTPA) study
- Participants are randomized to receive job training services for 18 months
- Information used in this empirical exercise:
  - Pre-experiment: years of education and annual earnings
  - Post-experiment: annual earnings
- Number of observations in the sample used by the two papers: 9,223
- Used by Kitagawa and Tetenov (2018) and Mbakop and Tabord-Meehan (2021)

My replication

- Applies EWM and PWM to this sample
- I focus on the quadrant rule and the linear eligibility score

Data

Variables

- Denote Y_i as the post-experiment annual earnings
- Denote $X_{1,i}$ and $X_{2,i}$ as the pre-experiment years of education and annual earnings

Variable	Min.	Mean	Max.
Years of education ($X_{1,i}$)	7	11.61	18
Pre-experiment annual earnings ($X_{2,i}$)	0	3,232.56	63,000.00
Post-experiment annual earnings (Y_i)	0	16,093.20	155,760.00

Implementation

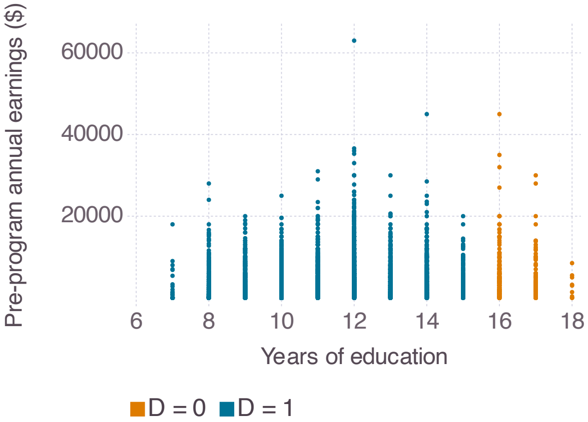
- Both papers take $e(X_i) = \frac{2}{3}$ as given
- Following Kitagawa and Tetenov (2018):
  - I demean the outcome variable, i.e., I use $Y_i^{\text{dm}} \equiv Y_i - \mathbb{E}_n[Y_i]$
  - I also normalize all the variables so they are bounded in $[-1, 1]$

Implementation

- I compute $\tau_i \equiv \frac{Y_i^{\text{dm}} D_i}{e(X_i)} - \frac{Y_i^{\text{dm}}(1-D_i)}{1-e(X_i)}$
- I follow the optimizer attributes used in Mbakop and Tabord-Meehan (2021)

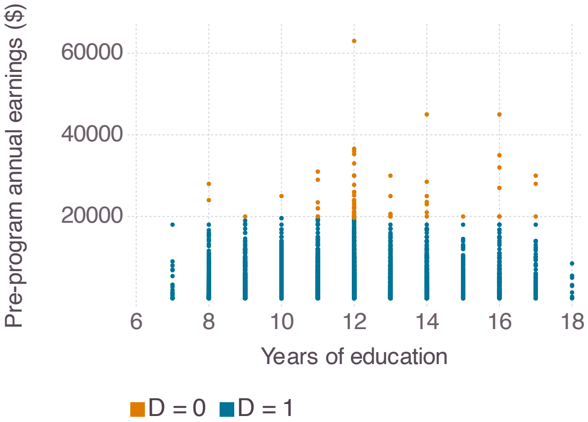
EWM with threshold rule on education

- $\mathcal{G}_{\text{edu}} \equiv \{\{(x_1) : \beta_0 + x_1\beta_1 \geq 0\}, \beta_0, \beta_1 \in [-1, 1]\}$
- Result: People with education ≤ 15



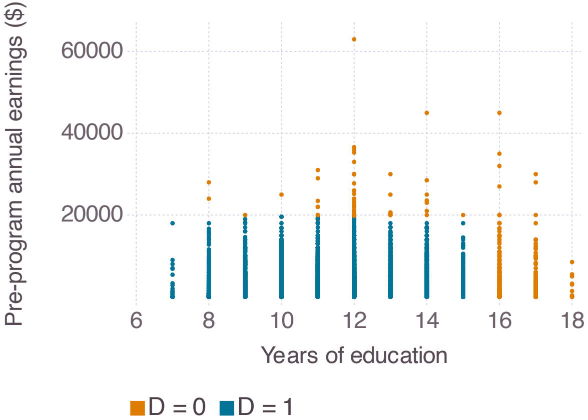
EWM with threshold rule on pre-experiment earnings

- $\mathcal{G}_{\text{edu}} \equiv \{\{(x_1) : \beta_0 + x_1\beta_1 \geq 0\}, \beta_0, \beta_1 \in [-1, 1]\}$
- Result: People with prior earnings $\leq$ 19,670 receive treatment



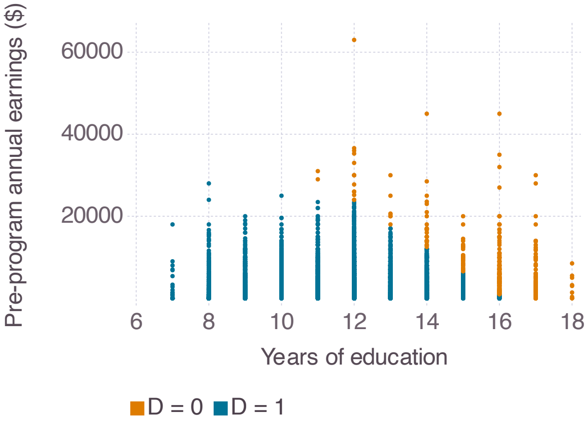
EWM with quadrant rule

- $\mathcal{G}_{QR} \equiv \{\{(x_1, x_2) : \beta_{0,1} + x_1\beta_{1,1} \geq 0, \beta_{0,2} + x_2\beta_{1,2} \geq 0\}, \beta_{0,0}, \beta_{1,1}, \beta_{0,2}, \beta_{1,2} \in [-1, 1]\}$
- Result: People with education $\leq$ 15 and prior earnings $\leq$ 19,670 receive treatment



EWM with linear eligibility score

- $\mathcal{G}_{\text{LES}} \equiv \{\{(x_1, x_2) : \beta_0 + x_1\beta_1 + x_2\beta_2 \geq 0\}, \beta_0, \beta_1, \beta_2 \in [-1, 1]\}$



References 1

KITAGAWA, T. AND A. TETENOV (2018): "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

MBAKOP, E. AND M. TABORD-MEEHAN (2021): "Model Selection for Treatment Choice: Penalized Welfare Maximization," *Econometrica*, 89, 825–848.

Contents

Mixed-integer linear programs

Linear programs

- Example:

$$\begin{aligned} \min_{x_1,x_2} \quad & 2x_1 + 3x_2 \\ \text{s.t.} \quad & x_1 - 2x_2 \leq 3 \\ & x_1, x_2 \geq 0 \end{aligned}$$

- Can be easily solved by many solvers
  - Free packages like `lpSolveAPI` in `R` or `GLPK` in `Julia`
  - Commercial softwares like `CPLEX` and `Gurobi` are available

Integer programs

- In the above example, we would need to impose the constraints where $x_1, x_2 \in \mathbb{Z}$
- These programs can also be solved by many solvers

Linear eligibility score (LES)

Objective function

- Recall that $\widehat{G}_{\text{EWM}} \in \arg\max_{G \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} \left[\left(\frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)} \right) \mathbb{1}[X_i \in G] \right]$
- $\mathcal{G}_{\text{LES}} \equiv \{\{x \in \mathbb{R}^{d_x} : \beta_0 + x_s' \beta_s \geq 0\} : (\beta_0, \beta_s') \in \mathbb{R}^{d_s+1}\}$ ($x_s \in \mathbb{R}^{d_s}$ is a subvector of x)
- Let $\tau_i \equiv \frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)}$ and represent $\mathbb{1}[X_i \in G]$ as z_i for each $i = 1, \ldots, n$
- Then, we can write the objective function as $\widehat{G}_{\text{EWM}} \in \arg\max_{G \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} \tau_i z_i$

Constraints

- We want $z_i = 1$ if $\beta_0 + x_s' \beta \geq 0$ and $z_i = 0$ otherwise for all $i = 1, \ldots, n$
- The variable z_i is binary by imposing $z_i \in \{0, 1\}$
- If $\beta_0 + x_s' \beta_s \geq 0$, we want $z_i = 1$, so impose $z_i > \frac{\beta_0 + x_s' \beta_s}{C_i}$ where $C_i > \sup_{\beta \in B} |X_i' \beta|$
- If $\beta_0 + x_s' \beta_s < 0$, we have $-1 \leq \frac{\beta_0 + x_s' \beta_s}{C_i} < 0$, so impose $z_i \leq 1 + \frac{\beta_0 + x_s' \beta_s}{C_i}$

Linear eligibility score (LES)

- $\mathcal{G}_{\text{LES}} \equiv \{\{x \in \mathbb{R}^{d_x} : \beta_0 + x_s'\beta_s \geq 0\} : (\beta_0, \beta_s') \in \mathbb{R}^{d_s+1}\}$ ($x_s \in \mathbb{R}^{d_s}$ is a subvector of x)

MILP

- Let $\tau_i \equiv \frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)}$, $X_i \equiv (1, x_s')'$, $\beta \equiv (\beta_0, \beta_s')' \in B \subseteq \mathbb{R}^{d_s+1}$, and $C_i > \sup_{\beta \in B} |X_i'\beta|$
- The MILP is:

$$\max_{\beta \in B,\, z_1,\ldots,z_n \in \mathbb{R}} \quad \sum_{i=1}^{n} \tau_i z_i$$

$$\text{s.t.} \quad \frac{X_i'\beta}{C_i} < z_i \leq 1 + \frac{X_i'\beta}{C_i} \qquad i = 1, \ldots, n \tag{3}$$

$$z_i \in \{0, 1\} \qquad i = 1, \ldots, n \tag{4}$$

Linear eligibility score (LES)

- $\mathcal{G}_{\text{LES}} \equiv \{\{x \in \mathbb{R}^{d_x} : \beta_0 + x_s'\beta_s \geq 0\} : (\beta_0, \beta_s') \in \mathbb{R}^{d_s+1}\}$ ($x_s \in \mathbb{R}^{d_s}$ is a subvector of x)

MILP

- Let $\tau_i \equiv \frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)}$, $X_i \equiv (1, x_s')'$, $\beta \equiv (\beta_0, \beta_s')' \in B \subseteq \mathbb{R}^{d_s+1}$, and $C_i > \sup_{\beta \in B} |X_i'\beta|$
- The MILP is:

$$
\begin{aligned}
\max_{\beta \in B,\, z_1,\ldots,z_n \in \mathbb{R}} \quad & \sum_{i=1}^{n} \tau_i z_i & & \\
\text{s.t.} \quad & \frac{X_i'\beta}{C_i} < z_i \leq 1 + \frac{X_i'\beta}{C_i} & i = 1, \ldots, n \qquad & (3) \\
& z_i \in \{0, 1\} & i = 1, \ldots, n \qquad & (4)
\end{aligned}
$$

  - Constraint (3) ensures that $z_i > 0$ if $X_i\beta \geq 0$

Linear eligibility score (LES)

- $\mathcal{G}_{\text{LES}} \equiv \{\{x \in \mathbb{R}^{d_x} : \beta_0 + x_s'\beta_s \geq 0\} : (\beta_0, \beta_s') \in \mathbb{R}^{d_s+1}\}$ ($x_s \in \mathbb{R}^{d_s}$ is a subvector of x)

MILP

- Let $\tau_i \equiv \frac{Y_i D_i}{e(X_i)} - \frac{Y_i(1-D_i)}{1-e(X_i)}$, $X_i \equiv (1, x_s')'$, $\beta \equiv (\beta_0, \beta_s')' \in B \subseteq \mathbb{R}^{d_s+1}$, and $C_i > \sup_{\beta \in B} |X_i'\beta|$
- The MILP is:

$$\max_{\beta \in B,\, z_1,\ldots,z_n \in \mathbb{R}} \quad \sum_{i=1}^{n} \tau_i z_i$$

$$\text{s.t.} \quad \frac{X_i'\beta}{C_i} < z_i \leq 1 + \frac{X_i'\beta}{C_i} \qquad i = 1, \ldots, n \tag{3}$$

$$z_i \in \{0, 1\} \qquad i = 1, \ldots, n \tag{4}$$

  - Constraint (3) ensures that $z_i > 0$ if $X_i\beta \geq 0$
  - Constraint (4) ensures that z is binary

Multiple linear index rules (MLIR)

- $\mathcal{G}_{\text{MLIR}} \equiv \{\{x : x'\beta^1 \geq 0, \ldots, x'\beta^J \geq 0\} : \beta_1, \ldots, \beta_J \in B \subseteq \mathbb{R}^{d_x}\}$

MILP

Also in PWM

- The MILP is:

$$\max_{\beta^1,\ldots,\beta \in B,\ \{\{z_i^j\}_{j=1}^J, z_i^\star\}_{i=1}^n \in \mathbb{R}} \sum_{i=1}^n \tau_i z_i^\star$$

$$\text{s.t.} \quad \frac{X_i'\beta^j}{C_i} < z_i^j \leq 1 + \frac{X_i'\beta^j}{C_i} \qquad i = 1, \ldots, n \text{ and } j = 1, \ldots, J \quad (5)$$

$$1 - J + \sum_{j=1}^J z_i^j \leq z_i^\star \leq J^{-1} \sum_{j=1}^J z_i^j \qquad i = 1, \ldots, n \quad (6)$$

$$z_i^1, \ldots, z_i^J, z_i^\star \in \{0, 1\} \qquad i = 1, \ldots, n \quad (7)$$

Multiple linear index rules (MLIR)

- $\mathcal{G}_{\text{MLIR}} \equiv \{\{x : x'\beta^1 \geq 0, \ldots, x'\beta^J \geq 0\} : \beta_1, \ldots, \beta_J \in B \subseteq \mathbb{R}^{d_x}\}$

MILP

Also in PWM

- The MILP is:

$$\max_{\beta^1,\ldots,\beta \in B,\ \{\{z_i^j\}_{j=1}^J, z_i^\star\}_{i=1}^n \in \mathbb{R}} \sum_{i=1}^n \tau_i z_i^\star$$

$$\text{s.t.} \quad \frac{X_i'\beta^j}{C_i} < z_i^j \leq 1 + \frac{X_i'\beta^j}{C_i} \qquad i = 1, \ldots, n \text{ and } j = 1, \ldots, J \quad (5)$$

$$1 - J + \sum_{j=1}^J z_i^j \leq z_i^\star \leq J^{-1} \sum_{j=1}^J z_i^j \qquad i = 1, \ldots, n \quad (6)$$

$$z_i^1, \ldots, z_i^J, z_i^\star \in \{0, 1\} \qquad i = 1, \ldots, n \quad (7)$$

  - Constraint (5) ensures that $z_i^j > 0$ if $X_i\beta^j \geq 0$ for each threshold j

Multiple linear index rules (MLIR)

- $\mathcal{G}_{\text{MLIR}} \equiv \{\{x : x'\beta^1 \geq 0, \ldots, x'\beta^J \geq 0\} : \beta_1, \ldots, \beta_J \in B \subseteq \mathbb{R}^{d_x}\}$

MILP

▸ Also in PWM

- The MILP is:

$$\max_{\beta^1,\ldots,\beta \in B,\ \{\{z_i^j\}_{j=1}^J, z_i^\star\}_{i=1}^n \in \mathbb{R}} \sum_{i=1}^n \tau_i z_i^\star$$

$$\text{s.t.} \quad \frac{X_i'\beta^j}{C_i} < z_i^j \leq 1 + \frac{X_i'\beta^j}{C_i} \qquad i = 1, \ldots, n \text{ and } j = 1, \ldots, J \quad (5)$$

$$1 - J + \sum_{j=1}^J z_i^j \leq z_i^\star \leq J^{-1} \sum_{j=1}^J z_i^j \qquad i = 1, \ldots, n \quad (6)$$

$$z_i^1, \ldots, z_i^J, z_i^\star \in \{0, 1\} \qquad i = 1, \ldots, n \quad (7)$$

- Constraint (5) ensures that $z_i^j > 0$ if $X_i\beta^j \geq 0$ for each threshold j
- Constraint (6) ensures that $z_j^\star > 0$ if $X_i\beta^j \geq 0$ for all thresholds

Multiple linear index rules (MLIR)

- $\mathcal{G}_{\text{MLIR}} \equiv \{\{x : x'\beta^1 \geq 0, \ldots, x'\beta^J \geq 0\} : \beta_1, \ldots, \beta_J \in B \subseteq \mathbb{R}^{d_x}\}$

MILP

▸ Also in PWM

- The MILP is:

$$\max_{\beta^1,\ldots,\beta \in B,\ \{\{z_i^j\}_{j=1}^J, z_i^\star\}_{i=1}^n \in \mathbb{R}} \sum_{i=1}^n \tau_i z_i^\star$$

$$\text{s.t.} \quad \frac{X_i'\beta^j}{C_i} < z_i^j \leq 1 + \frac{X_i'\beta^j}{C_i} \qquad i = 1, \ldots, n \text{ and } j = 1, \ldots, J \quad (5)$$

$$1 - J + \sum_{j=1}^J z_i^j \leq z_i^\star \leq J^{-1} \sum_{j=1}^J z_i^j \qquad i = 1, \ldots, n \quad (6)$$

$$z_i^1, \ldots, z_i^J, z_i^\star \in \{0, 1\} \qquad i = 1, \ldots, n \quad (7)$$

  - Constraint (5) ensures that $z_i^j > 0$ if $X_i\beta^j \geq 0$ for each threshold j
  - Constraint (6) ensures that $z_j^\star > 0$ if $X_i\beta^j \geq 0$ for all thresholds
  - Constraint (7) ensures that z_i^j and $z_j^\star$ are all binary

Contents

Threshold allocations on d covariates

Back

- Let $x \equiv (1, x^{(1)}, \ldots, x^{(d)})'$ be a $(d+1)$-vector and $\mathcal{A} \equiv \{1, \ldots, d\}$ be the threshold dimension
- Define the threshold β_k on $x^{(k)}$ to be a length $(d+1)$ vector, such that:
  - First component in $[-1, 1]$, the $(k+1)$-th component in $\{-1, 0, 1\}$ and others are all 0

MILP

$$
\begin{aligned}
\max_{\{\beta_a\}_{a\in A}, \{z_1^a, \ldots, z_n^a\}_{a\in A}, z_1^\star, \ldots, z_n^\star} \quad & \sum_{i=1}^n \tau_i z_i^\star \\
\text{s.t.} \quad & \frac{x_i'\beta_a}{c} < z_i^a \le 1 + \frac{x_i'\beta_a}{c} && i = 1, \ldots, n \text{ and } a \in \mathcal{A} \\
& 1 - |A| + \textstyle\sum_{a\in\mathcal{A}} z_i^a \le z_i^\star \le |A|^{-1} \textstyle\sum_{a\in\mathcal{A}} z_i^a && i = 1, \ldots, n \\
& \beta_a^{(1)} \in [-1, 1] && a \in \mathcal{A} \\
& \beta_a^{(j)} = 0 && j > 1,\ j \neq a+1,\ \text{and } a \in \mathcal{A} \\
& \textstyle\sum_{a\in\mathcal{A}} e_a = k \\
& -e_a \le \beta_a^{(1)} \le e_a \text{ and } \beta_a^{(a+1)} = y_{a,1} - y_{a,2} && a \in \mathcal{A} \\
& y_{a,1} + y_{a,2} = e_a && a \in \mathcal{A} \\
& \{z_i^a\}_{i=1}^n, e_a, y_{a,1}, y_{a,2} \in \{0, 1\} \text{ for } a \in \mathcal{A} \text{ and } z_i^\star \in \{0, 1\} \text{ for } i = 1, \ldots, n
\end{aligned}
$$