# Adaptive Estimation of Intersection Bounds – a Classification Approach

Vira Semenova

University of California, Berkeley

## Introduction

Many causal and structural parameters are not point-identified and need to be bounded from above and below .

# Introduction

Many causal and structural parameters are not point-identified and need to be bounded from above and below .

Examples: Manski (1997), Heckman et al. (1997), Lee (2009), (e.g., Kalouptsidi et al. (2020)), etc.

Covariate set $\mathcal{X}$ maps to a class of bounds

- any covariate subset $\mathcal{X}' \subseteq \mathcal{X}$ maps to a pair of valid bounds
- the full set $\mathcal{X}$ maps to sharp (the tightest possible) bounds

## Introduction

Many causal and structural parameters are not point-identified and need to be bounded from above and below .

Examples: Manski (1997), Heckman et al. (1997), Lee (2009), (e.g., Kalouptsidi et al. (2020)), etc.

Covariate set $\mathcal{X}$ maps to a class of bounds

- any covariate subset $\mathcal{X}' \subseteq \mathcal{X}$ maps to a pair of valid bounds
- the full set $\mathcal{X}$ maps to sharp (the tightest possible) bounds

Sharp bounds are difficult to estimate, non-sharp bounds may not be very useful.

## Outline

1. Example
2. Debiased Inference
3. Linear Programming
4. Envelope Theorem

## Literature Review

1. **Envelope theorems. Stochastic programming** Shapiro (Annals of Statistics, 1989), Milgrom and Segal (2002)
2. **Bounds/partial identification**: (**identification**) Heckman (1976), Heckman (1979), Manski (1989), Manski (1990), Manski (1997), Heckman et al. (1997), Fan et al. (2017), Tetenov (2012), Kamat (2019) (**inference**) Fan and Park (2010, 2012), Chernozhukov et al. (2013), Kaido and Santos (2014), Kaido and White (2012), Kaido (2017) Kaido (2016), Kline and Tartari (2016), Abdulkadiroglu et al. (2020), Kaido et al. (2019), Hsieh et al. (2021), Fang et al. (2020)
3. **Policy Learning and Classification**: Tsybakov (2004), Qian and Murphy (2011), Kitagawa and Tetenov (2018), Athey and Wager (2021), Mbakop and Tabord-Meehan (2021), Sun (2021)
4. **Directional Differentiability** Fang and Santos (2018), Ponomarev (2022).
5. **Orthogonal/debiased machine learning**: Newey (1994), Belloni and Chernozhukov (2011), Chernozhukov et al. (2022), Belloni et al. (2017), Chernozhukov et al. (2018), Chiang et al. (2019), Sasaki and Ura (2020), Sasaki et al. (2020), Cha et al. (2022)

# (1): Example

## Example: setup

Notation

- $D = 1$ if subject wins a lottery
- $S(1) = 1$ employed if $D = 1$
- $S(0) = 1$ employed if $D = 0$
- $X$ pre-treatment (baseline) characteristics

Observed data: $(X, D, S)$ where $S = D \cdot S(1) + (1 - D)S(0)$.

## Example: setup

Notation

- $D = 1$ if subject wins a lottery
- $S(1) = 1$ employed if $D = 1$
- $S(0) = 1$ employed if $D = 0$
- $X$ pre-treatment (baseline) characteristics

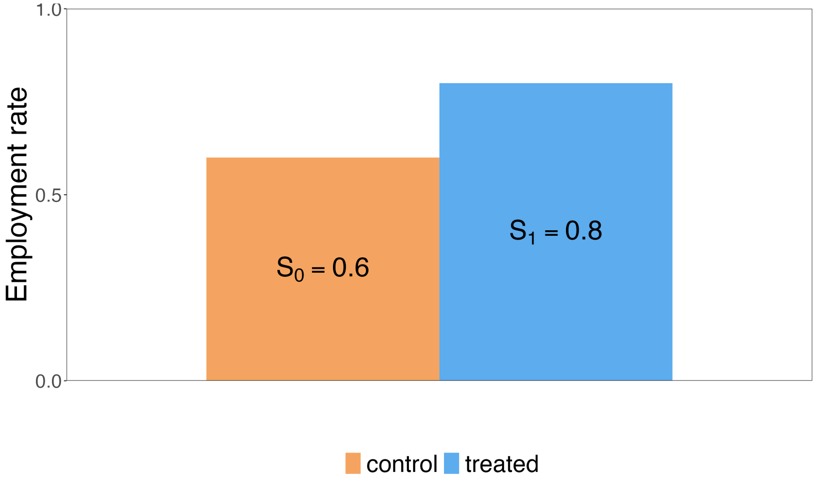
Observed data: $(X, D, S)$ where $S = D \cdot S(1) + (1 - D)S(0)$.

| | | Control ($D = 0$) | |
|---|---|---|---|
| | | $S(0) = 1$ | $S(0) = 0$ |
| Treated ($D = 1$) | $S(1) = 1$ | always-takers ($\pi_{\text{AT}}$) | compliers ($\pi_{\text{comp}}$) |
| | $S(1) = 0$ | defiers ($\pi_{\text{defier}}$) | never-takers ($\pi_{\text{NT}}$) |

Target parameter is

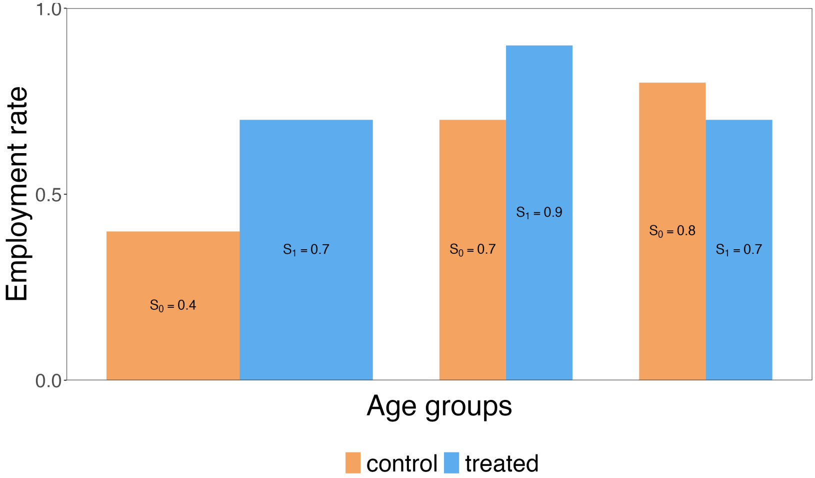
$$\bar{\pi} = (\pi_{\text{AT}}, \pi_{\text{C}}, \pi_{\text{D}}) \cdot (1, 0, 0) = \pi_{\text{AT}}$$

## Example: basic bound on $\pi_{AT}$




$$\pi_{AT} \leq \min(S_1, S_0) = \min(0.8, 0.6) = 0.6$$

## Example: tighter bound on $\pi_{AT}$




$$\mathbb{E} \min(s(0, X), s(1, X)) =$$

$$1/2 \min(0.4, 0.7) + 1/4 \min(0.7, 0.9) + 1/4 \min(0.8, 0.7) = 0.55$$

$$\boxed{\text{Jensen: } 0.55 < 0.6 = \min(S_1, S_0)}$$

# (2): Debiased Inference

## Example: age as a continuous variable

employment probability

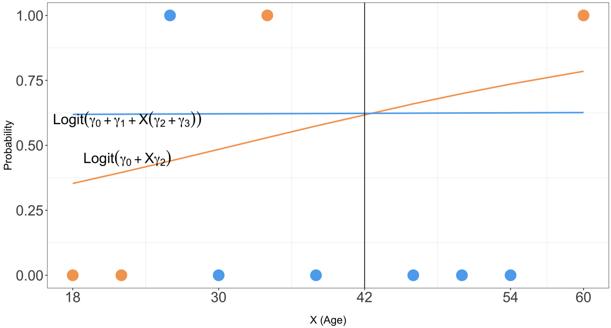
$$s(D, X) = \text{Logit}(\gamma_0 + D \cdot \gamma_1 + X \cdot \gamma_2 + D \cdot X \cdot \gamma_3)$$

regions of positive conditional ATE $s(1, X) - s(0, X)$

$$G := \{X : \quad s(1, X) - s(0, X) \geq 0\} = \{X : X \leq 42\}.$$

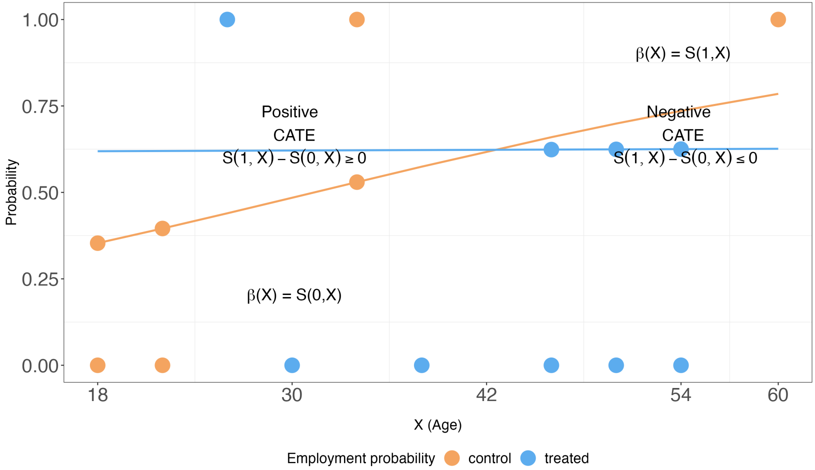
sharp bound is

$$\phi_0 = \mathbb{E} \min(s(0, X), s(1, X)) = 0.579$$



# Example: envelope regression

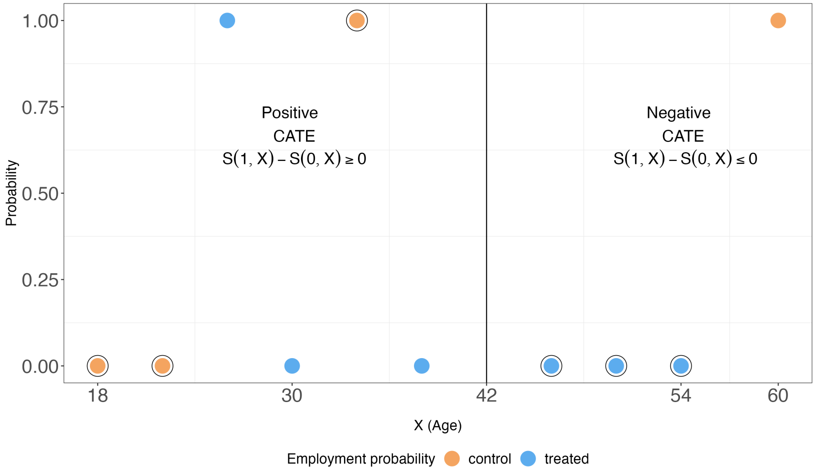
$$\widehat{\overline{\pi}}(\widehat{S}) := N^{-1} \sum_{i=1}^{N} \min(\widehat{S}(1, X_i), \widehat{S}(0, X_i))$$



## Example: envelope moment

region of positive CATE: $\widehat{G} := \{X : \widehat{s}(1, X) - \widehat{s}(0, X) \geq 0\}$

$$\widehat{\pi}(\widehat{G}) := (N^{-1} \sum_{i=1}^{N} 2S_i(1 - D_i)\{X_i \in \widehat{G}\} + 2S_iD_i\{X_i \in \widehat{G}^c\})$$



## Oracle property: result

**Assumptions.**

1. the covariate $X$ has bounded density (margin condition, Tsybakov (2004))
2. $\sup_{x\in\mathcal{X}} \sup_{t\in T} |\widehat{S}(t,x) - S_0(t,x)| = o_P(n^{-1/4})$

**Result.** The envelope moment $\widehat{\pi}(\widehat{G}) = \pi(\widehat{S})$ based on the plug-in estimator

$$\widehat{G} := \{X : \widehat{s}(1,X) - \widehat{s}(0,X) \geq 0\}$$

obeys oracle property

$$\sqrt{N}(\widehat{\pi}(\widehat{S}) - \widehat{\pi}(S_0)) \Rightarrow^P 0.$$

As a result, $\widehat{\pi}(\widehat{S})$ is asymptotically Gaussian with oracle variance

$$\sqrt{N}(\widehat{\pi}(\widehat{S}) - \pi_0) \Rightarrow N(0, V_\pi), \quad V_\pi = (2-\pi_0)\pi_0.$$

## Oracle property: discussion

- The result generalizes to infimum over infinite set $T$

$$\boxed{\phi_0 = \mathbb{E}_X \inf_{t \in T} s(t, X)}$$

## Oracle property: discussion

- The result generalizes to infimum over infinite set $T$

$$\boxed{\phi_0 = \mathbb{E}_X \inf_{t \in T} s(t, X)}$$

  - treatment effect CDF $\Pr(S(1) - S(0) \leq t)$ with $T = \mathrm{R}$
  - positive treatment effects $\mathbb{E}(S(1) - S(0))_+$ with $\cup_{t \in \mathrm{R}} T_t = \{1, 0\}$
  - best linear predictor of intersection bounds $\inf_{t \in T} s(t, X)$

## Oracle property: discussion

- The result generalizes to infimum over infinite set $T$

$$\boxed{\phi_0 = \mathbb{E}_X \inf_{t \in T} s(t, X)}$$

  - treatment effect CDF $\Pr(S(1) - S(0) \leq t)$ with $T = \mathrm{R}$
  - positive treatment effects $\mathbb{E}(S(1) - S(0))_+$ with $\cup_{t \in \mathrm{R}} T_t = \{1, 0\}$
  - best linear predictor of intersection bounds $\inf_{t \in T} s(t, X)$

- Regularity of $\mathbb{E}_X \inf_{t \in T} s(t, X)$
  - Fang and Santos (2018): $\min_{t \in T}(s_t)$ not a regular parameter: bootstrap fails
  - this paper: $\mathbb{E}_X \inf_{t \in T} s(t, X)$ bootstrap applies

## Oracle property: discussion

- The result generalizes to infimum over infinite set $T$

$$\boxed{\phi_0 = \mathbb{E}_X \inf_{t \in T} s(t, X)}$$

  - treatment effect CDF $\Pr(S(1) - S(0) \leq t)$ with $T = \mathrm{R}$
  - positive treatment effects $\mathbb{E}(S(1) - S(0))_+$ with $\cup_{t \in \mathrm{R}} T_t = \{1, 0\}$
  - best linear predictor of intersection bounds $\inf_{t \in T} s(t, X)$
- Regularity of $\mathbb{E}_X \inf_{t \in T} s(t, X)$
  - Fang and Santos (2018): $\min_{t \in T}(s_t)$ not a regular parameter: bootstrap fails
  - this paper: $\mathbb{E}_X \inf_{t \in T} s(t, X)$ bootstrap applies
- Applications beyond bounds
  - welfare in statistical treatment choice (Kitagawa and Tetenov (2018))
  - Bayes (optimal) risk in classification literature

# (3): Linear Programming (LP)

## LP: setup

1. $\pi = (\pi_1, \pi_{-1}) \in \mathrm{R}^d$. basic upper bound on $\pi_1$ is

$$\bar{\pi} := \min_{\pi \in \mathrm{R}^d} -\pi_1 \text{ s. t. } A\pi = \mathbf{S} = \mathbb{E}[\mathcal{S}]$$

2. the group-specific constraint set is

$$A\pi = \mathbf{S}(x),$$

where the RHS is

$$\mathbf{S}(x) = \mathbb{E}[\mathcal{S} \mid X = x] \qquad \text{, A is known}$$

## LP: setup

1. $\pi = (\pi_1, \pi_{-1}) \in \mathrm{R}^d$. basic upper bound on $\pi_1$ is

$$\bar{\pi} := \min_{\pi \in \mathrm{R}^d} -\pi_1 \text{ s. t. } A\pi = \mathbf{S} = \mathbb{E}[\mathcal{S}]$$

2. the group-specific constraint set is

$$A\pi = \mathbf{S}(x),$$

where the RHS is

$$\mathbf{S}(x) = \mathbb{E}[\mathcal{S} \mid X = x] \qquad \text{, A is known}$$

3. the conditional bound is $\bar{\pi}(x)$. the target is the average conditional bound

$$\mathbb{E}_X \bar{\pi}(X)$$

## LP: setup

1. $\pi = (\pi_1, \pi_{-1}) \in \mathrm{R}^d$. basic upper bound on $\pi_1$ is

$$\bar{\pi} := \min_{\pi \in \mathrm{R}^d} -\pi_1 \text{ s. t. } A\pi = \mathbf{S} = \mathbb{E}[\mathcal{S}]$$

2. the group-specific constraint set is

$$A\pi = \mathbf{S}(x),$$

   where the RHS is

$$\mathbf{S}(x) = \mathbb{E}[\mathcal{S} \mid X = x] \qquad \text{, A is known}$$

3. the conditional bound is $\bar{\pi}(x)$. the target is the average conditional bound

$$\mathbb{E}_X \bar{\pi}(X)$$

4. by Jensen's inequality,

$$\mathbb{E}_X \bar{\pi}(X) \leq \bar{\pi}.$$

## LP: upper bound as an envelope of regression

1. Dual feasible set is data-free

$$\nu \in \mathrm{R}^r : A'\nu \geq (1, 0, \ldots, 0)'.$$

2. Dual set reduces to its vertices $\mathcal{T} = \{\nu_t\}$

$$\underbrace{\min_{\nu : A'\nu \geq e_1}}_{\text{infinite}} \nu' \mathbf{S}(x) = \underbrace{\min_{\nu_t \in \mathcal{T}}}_{\textit{finite}} \nu_t' \mathbf{S}(x)$$

## LP: upper bound as an envelope of regression

1. Dual feasible set is data-free

$$\nu \in \mathrm{R}^r : A'\nu \geq (1, 0, \dots, 0)'.$$

2. Dual set reduces to its vertices $\mathcal{T} = \{\nu_t\}$

$$\min_{\underbrace{\nu : A'\nu \geq e_1}_{\text{infinite}}} \nu'\mathbf{S}(x) = \min_{\underbrace{\nu_t \in \mathcal{T}}_{\textit{finite}}} \nu_t'\mathbf{S}(x)$$

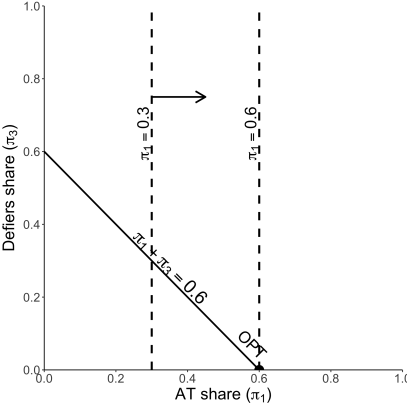
3. At the optimum, primal LP $=$ dual LP

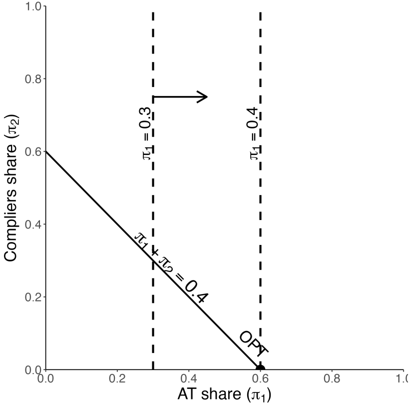
$$\bar{\pi}(x) = \inf_{\nu_t \in \mathcal{T}} \underbrace{\mathbf{S}(x)'\nu_t}_{} =: s(t, x) = \inf_{t \in \mathcal{T}} s(t, x).$$

Duality has been used in Kaido (2017), Fang et al. (2020), Hsieh et al. (2021) (JoE, 2021)

# Example: primal LP is a regression problem

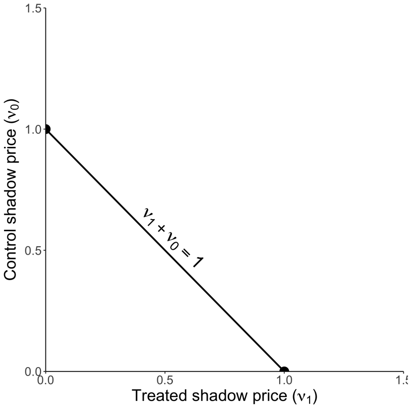
Example. $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$




$$\mathbf{S}(x_1) = \begin{pmatrix} 0.8 \\ 0.6 \end{pmatrix} \qquad \mathbf{S}(x_2) = \begin{pmatrix} 0.4 \\ 0.5 \end{pmatrix}$$

## Example: dual feasible set

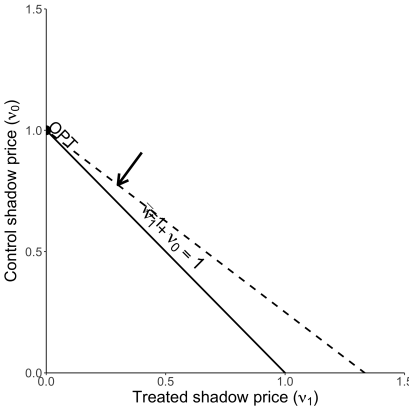



$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \nu, \quad A'\nu \geq (1, 0, 0)'$$

The vertex set $\mathcal{T} = \{(1,0), (0,1)\}$

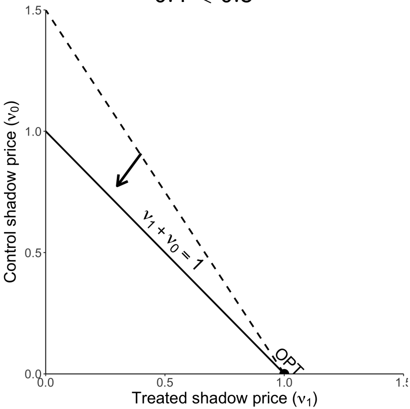
# Example: dual LP reduces to classification problem

$0.8 > 0.6$




$$\min 0.8\nu_1 + 0.6\nu_2 \text{ s. to } A'\nu \geq (1,0,0)'$$

$0.4 < 0.5$



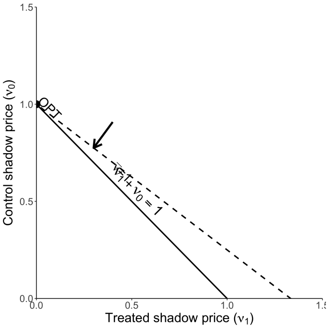

$$\min 0.4\nu_1 + 0.5\nu_2 \text{ s. to } A'\nu \geq (1,0,0)'$$

## Example: estimate reduces to weighted sample average

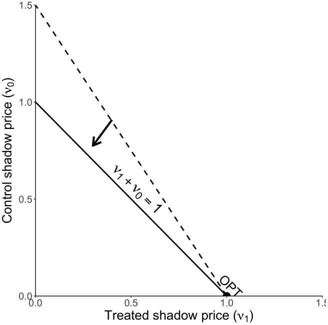
region of positive CATE:

$$\widehat{G} := \{x : \widehat{s}(1,x) - \widehat{s}(0,x) \geq 0\}$$

$x_1 \in \widehat{G}$




$x_2 \notin \widehat{G}$




The estimator is the sample average:

$$\widehat{\widehat{\pi}}(\widehat{G}) := (N^{-1} \textstyle\sum_{i=1}^{N} 2S_i(1-D_i)\{X_i \in \widehat{G}\} + 2S_iD_i\{X_i \in \widehat{G}^c\}).$$

## Example: main take-aways

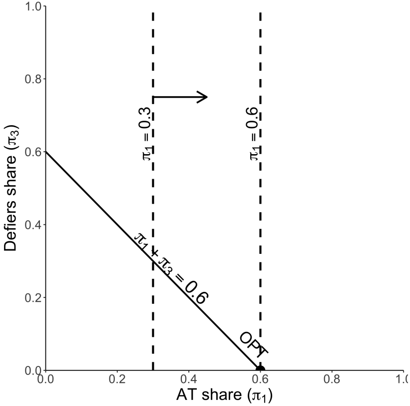
1. covariates tighten bounds
2. dual LP is a classification problem
3. dual LP is first-order insensitive to misclassification mistake
    - The dual vector $\bar{\nu}(X)$ is the Riesz representer function for the RHS function

# (3.b): Linear Programming (LP)

$A(x) \neq A$ depends on $x$

## General case: primal LP

$$\bar{\pi}(x) = \min_{\pi} -\pi_1 \ \text{ subject to } \ A(x)\cdot\pi = \mathbf{S}(x)$$




Example. $$A(x) = \begin{pmatrix} 1/0.8 & 1/0.8 & 0 \\ 1/0.6 & 0 & 1/0.6 \end{pmatrix}, \qquad \mathbf{S} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

## General case: from envelopes to saddle values

1. Define Lagrangian function

$$L(\pi, \nu, x) := -\pi_1 + \nu^\top (A(x)\pi - \mathbf{S}(x)).$$

## General case: from envelopes to saddle values

1. Define Lagrangian function

$$L(\pi, \nu, x) := -\pi_1 + \nu^\top (A(x)\pi - \mathbf{S}(x)).$$

2. The objective function is the *saddle value* of regression

$$\bar{\pi}(x) = \max_{\nu} \min_{\pi} L(\pi, \nu, x) = L(\pi^*(x), \nu^*(x), x)$$

3. $(\pi^*(x), \nu^*(x))$ is a saddle value of $L(\pi, \nu, x)$

4. Envelope moment is

$$g(W, \pi, \nu) := \textstyle\sum_{\nu,\pi} g_{\nu,\pi}(W) 1\{\pi = \pi^*(X), \nu = \nu^*(X)\},$$

where $g_{\nu,\pi}(W)$ is an unbiased signal for Lagrangian

$$\mathbb{E}[g_{\nu,\pi}(W) \mid X = x] = L(\pi, \nu, x).$$

## LP: oracle property for saddle moments

**Assumptions.**

1. the covariate $X$ has bounded density
2. $\sup_x \|\widehat{S}(x) - S_0(x)\| + \|\widehat{A}(x) - A_0(x)\| = o_P(n^{-1/4})$
3. new condition!: $(\widehat{\pi}, \widehat{\nu})$ must be a saddle-value, that is

$$L(x, \widehat{\pi}, \widehat{\nu}) = \max_\nu \min_\pi L(x, \widehat{\pi}, \widehat{\nu}) = \min_\pi \max_\nu L(x, \widehat{\nu}, \widehat{\pi})$$

**Result.** The saddle moment

$$\widehat{\phi}(\widehat{\nu}, \widehat{\pi}) := N^{-1} \sum_{i=1}^{N} \sum_{d=0}^{d=1} g_{\nu,\pi}(W_i) 1\{\nu = \widehat{\nu}(X_i), \pi = \widehat{\pi}(X_i)\}.$$

obeys oracle property

$$\sqrt{N}(\widehat{\phi}(\widehat{\nu}, \widehat{\pi}) - \widehat{\phi}(\nu_0, \pi_0)) \Rightarrow^P 0.$$

As a result, $\widehat{\phi}(\widehat{\nu}, \widehat{\pi})$ is asymptotically Gaussian with oracle variance

$$\sqrt{N}(\widehat{\phi}(\widehat{\nu}, \widehat{\pi}) - \phi_0) \Rightarrow N(0, V_\phi).$$

# (4): Envelope Theorem

## Key definitions and facts

(Envelope Theorem, Milgrom and Segal (2002)) The envelope function $V(\tau) := \inf_{t \in T} s(t, \tau)$ is differentiable. Its derivative

$$V'(\tau) = s_\tau(t^*(\tau), \tau), \qquad \tau \in [0, 1]$$

is calculated as if $t^*(\tau)$ was known.

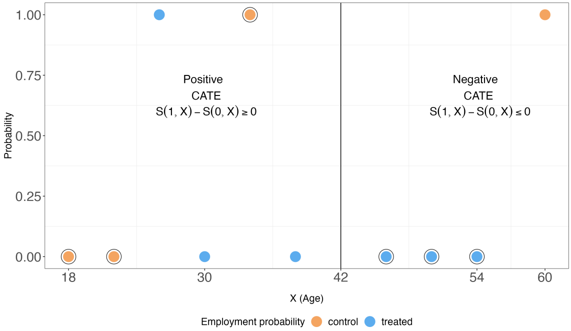
Let $\{P_\tau\}$ is a parametric submodel containing true distribution $P_0$. The target parameter $\phi(P_\tau), \tau \in [0, 1]$,

$$\frac{\partial \phi(P_\tau)}{\partial \tau} = \mathbb{E}\psi(W) S_\tau(W), \qquad \tau \in [0, 1]$$

where $\psi(W)$ is the **influence** function and $S_\tau(W)$ is the score. van der Vaart (1991)

## Influence function: $\phi(W) = \psi_1(W) + \psi_2(W)$

$$\phi(W) = \sum_{d=0}^{d=1} 2S\{D = d\}1\{s(d, X) = t^*(X)\} - \phi_0.$$



1. Envelope theorem $\Rightarrow$ no correction term (bounds, support function)
2. Implicit function theorem for DP operator

$$V(x;\gamma) = \zeta(x) + \beta \mathbb{E}_{\gamma}[V(x_{+}\gamma)|x]$$

   Differentiate the DP operator

$$\partial_\gamma V(x;\gamma_0) = \beta \partial_\gamma \int V(x_+;\gamma_0) f(x'|x;\gamma_0) dx_+ + \beta \mathbb{E} \partial_\gamma [V(x_+;\gamma_0)|x] dx_+$$

1. Envelope theorem $\Rightarrow$ no correction term (bounds, support function)
2. Implicit function theorem for DP operator

$$V(x;\gamma) = \zeta(x) + \beta\mathbb{E}_{\gamma}[V(x_{+}\gamma)|x]$$

Differentiate the DP operator

$$\partial_{\gamma} V(x;\gamma_0) = \beta\partial_{\gamma}\int V(x_{+};\gamma_0)f(x'|x;\gamma_0)dx_{+} + \beta\mathbb{E}\partial_{\gamma}[V(x_{+};\gamma_0)|x]dx_{+}$$

## Conclusion and Future Work

Proposed asymptotic theory for

- envelopes $\mathbb{E} \inf_{t \in T} s(t, X)$
- saddle-values $\mathbb{E} \max_{\nu} \inf_{\pi} L(\nu, \pi, X)$

Future Work

- quantify sharpness-complexity tradeoff

$$\arg\max_{\phi_0} \underbrace{\phi_0}_{\text{parameter}} + N^{-1/2} \underbrace{\sqrt{\phi_0(1-\phi_0)}}_{\text{std. error}} \text{ not sharp bound!}$$

- incorporate semi-supervised classifiers (use pre-treatment covariates to draw boundaries)

Abdulkadiroglu, A., Pathak, P. A., and Walters, C. R. (2020). Do parents value school effectiveness. *American Economic Review*, 110(5):1502–1539.

Athey, S. and Wager, S. (2021). Policy learning with observational data. *Econometrica*, 89:133–161.

Belloni, A. and Chernozhukov, V. (2011). $\ell_1$-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, 39(1):82–130.

Belloni, A., Chernozhukov, V., Fernandez-Val, I., and Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85:233–298.

Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814.

Bontemps, C., Magnac, T., and Maurin, E. (2012). Set identified linear models. *Econometrica*, 80:1129–1155.

Cha, J., Chiang, H. D., and Sasaki, Y. (2022). Inference in high-dimensional regression models without the exact or $l^p$ sparsity.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21:C1–C68.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2022). Locally Robust Semiparametric Estimation. *Econometrica*.

Chernozhukov, V., Lee, S., and Rosen, A. (2013). Intersection bounds: Estimation and inference. *Econometrica*, 81:667–737.

Chiang, H. D., Kato, K., Ma, Y., and Sasaki, Y. (2019). Multiway Cluster Robust Double/Debiased Machine Learning. *arXiv e-prints*, page arXiv:1909.03489.

Fan, Y., Guerre, E., and Zhu, D. (2017). Partial identification of functionals of the joint distribution of "potential outcomes".

Fan, Y. and Park, S. S. (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951.

Fan, Y. and Park, S. S. (2012). Confidence intervals for the quantile of treatment effects in randomized experiments. *Journal of Econometrics*, 167:330–344.

Fang, Z. and Santos, A. (2018). Inference on Directionally Differentiable Functions. *The Review of Economic Studies*, 86(1):377–412.

Fang, Z., Santos, A., Shaikh, A. M., and Torgovitsky, A. (2020). Inference for large-scale linear systems with known coefficients.

Heckman, J., Smith, J., and Clements, N. (1997). Making the most out of program evaluations and social experiments: accounting for heterogeneity in program impacts. *Review of Economic Studies*, 64:487–535.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4):475–492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Hsieh, Y.-W., Shi, X., and Shum, M. (2021). Inference on estimators defined by mathematical programming. *Journal of Econometrics*.

Ichimura, H. and Newey, W. K. (2022). The Influence Function of Semiparametric Estimators. *Quantitative Economics*, 13:29–61.

Kaido, H. (2016). A dual approach to inference for partially identified econometric models. *Journal of Econometrics*, 192:269–290.

Kaido, H. (2017). Asymptotically efficient estimation of weighted average derivatives with an interval censored variable. *Econometric Theory*, 33(5):1218–1241.

Kaido, H., Molinari, F., and Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*, 87(4):1397–1432.

Kaido, H. and Santos, A. (2014). Asymptotically efficient estimation of models defined by convex moment inequalities. *Econometrica*, 82(1):387–413.

Kaido, H. and White, H. (2012). Estimating misspecified moment inequality models. *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honour of Halbert L. White Jr.*

Kitagawa, T. and Tetenov, A. (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86:591–616.

Kline, P. and Tartari, M. (2016). Bounding the labor supply responses to a randomized welfare experiment: a revealed preference approach. *American Economic Review*, 106(4):972–1014.

Lee, D. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102.

Manski, C. (1997). Monotone treatment response. *Econometrica*, 65(6):1311–1334.

Manski, C. F. (1989). Anatomy of the selection problem. *The Journal of Human Resources*, 24(3):343–360.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323.

Mbakop, E. and Tabord-Meehan, M. (2021). Model selection for treatment choice: Penalized welfare maximization. *Econometrica*, 89:825–848.

Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70:583–601.

Newey, W. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, 62(6):245–271.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180 – 1210.

Sasaki, Y. and Ura, T. (2020). Estimation and inference for Policy Relevant Treatment Effects. *Journal of Econometrics*.

Sasaki, Y., Ura, T., and Zhang, Y. (2020). Unconditional quantile regression with high-dimensional data. *arXiv e-prints*, page arXiv:2007.13659.

Sun, L. (2021). Empirical welfare maximization with constraints.

Tetenov, A. (2012). Identification of positive treatment effects in randomized experiments with non-compliance.

Tsybakov, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, 32(1):135 – 166.