Partial identification, distributional preferences, and the welfare ranking of policies

Maximilian Kasy

Department of Economics, UCLA

Two conflicting objectives in (micro)econometrics

1. Use only a priori justifiable assumptions (No functional forms!)
2. Evaluate the impact of counterfactual policies

- Relative weight given to these two is central in methodological debates.
- This paper: exploring the frontier in the tradeoff between the two objectives.
- Goal: Identification of the ranking of counterfactual policies based on models without functional form assumptions.

Questions:

1. How does the data distribution map into policy-rankings?
2. Under what conditions is the welfare ranking of policies fully / partially / not at all identified?

Setup considered:

- Allocation of binary treatment
- under partial identification of conditional average treatment effects
- with possibly restricted sets of feasible policies
- and general distributional preferences.

Answers depend on *interaction* of

1. the identified set for treatment effects,
2. the feasible policy set,
3. the objective function.

Contributions to literature

- To lit on partial identification of treatment effects; treatment choice (Manski (2003), Stoye (2011a)):
  partial identification of the welfare ranking of policies itself
- To lit on distributional decompositions
  (DiNardo et al. (1996), Firpo et al. (2009), Chernozhukov et al. (2009)):
  endogeneity of treatment; tractable bounds for effect of policies on statistics of the outcome distribution
- For practitioners:
  new objects of interest; simple calculation of these
  criteria for whether a given dataset is informative about the ranking of policies.

Further literature

- Optimal treatment assignment based on covariates:
  Manski (2004), Dehejia (2005), Bhattacharya and Dupas (2008), Hirano and Porter (2009), Chamberlain (2011),
- Relationship between policy sets and parameters of interest:
  Chetty (2009), Graham et al. (2008); Sen (1995)
- Debates about "causal" vs. "structural" approaches:
  Deaton (2009), Imbens (2010), Angrist and Pischke (2010), Nevo and Whinston (2010)
- Axiomatic decision theory:
  Knight (1921), Anscombe and Aumann (1963), Bewley (2002), Ryan (2009)
- Policy choice under ambiguity:
  Manski (2011), Stoye (2011b), Hansen and Sargent (2008)
- Robust statistics:
  Huber (1996)

Roadmap

1. Setup
2. Review of partial identification of average treatment effects
3. The identified welfare ranking of policies
4. Generalization to nonlinear objective functions
5. Relationship to axiomatic decision theory
6. Application to project STAR data
7. Outlook - Partial identification of optimal policy parameters in public finance models
8. Conclusion

Setup

- outcome of interest Y, generated by $Y = f(X, D, \epsilon)$
- treatment $D \in \{0, 1\}$, support of X, ϵ unrestricted
- potential outcomes $Y^d = f(X, d, \epsilon)$ for $d = 0, 1$
- conditional average treatment effects (ATE)

$$g(X) = E[Y^1 - Y^0 | X] \tag{1}$$

- counterfactual treatment assignment policies h:
  $P(D = 1|X) = h(X)$, $D \perp (Y^0, Y^1)|X$
- special case: deterministic policies $h(X) \in \{0, 1\} \Rightarrow D = h(X)$
- policy objective $\phi = \phi(f)$, where f is the distribution of Y
- special case considered first: $\phi = E[Y]$, $Y \in [0, 1]$

Potential applications: Assignment of

- income support programs to individuals, $Y =$ labor market outcomes
- indivisible capital goods to units of production, $Y =$ profits
- a medical treatment to patients, $Y =$ health outcomes
- students to integrated or segregated classes, $Y =$ rescaled test-scores

Limitations:

- discrete treatment
  (for identifiability)
- additively separable objective function
  (for expositional purposes; second part of talk generalizes)
- no informational / incentive compatibility constraints
  (excludes optimal taxation, ... - next project, see outlook)

Review of partial identification: instrumental variables (IV) c.f. Manski (2003)

Assumption (Instrumental variable setup)

The joint distribution of (X, Y, D, Z) *is observed, where* $D \in \{0, 1\}$, $Y \in [0, 1]$, $Y = D \cdot Y^1 + (1 - D) \cdot Y^0$ *for potential outcomes* Y^0, Y^1, *and* Z *is an instrumental variable satisfying*

$$Z \perp (Y^0, Y^1) | X. \tag{2}$$

Conditional exogeneity of Z, law of total probability $\Rightarrow$

$$\begin{aligned} g(X) = E[Y^1|X] - E[Y^0|X] \\ = \big(E[D|Z=z^1,X]\cdot E[Y^1|D=1,Z=z^1,X] \\ +E[1-D|Z=z^1,X]\cdot \mathbf{E[Y^1|D=0,Z=z^1,X]}\big) \\ -\big(E[1-D|Z=z^0,X]\cdot E[Y^0|D=0,Z=z^0,X] \\ +E[D|Z=z^0,X]\cdot \mathbf{E[Y^0|D=1,Z=z^0,X]}\big) \end{aligned} \quad (3)$$

- The data pin down all parts of this expression
- except for the counterfactual means $E[Y^1|D=0,Z=z^1,X]$, $E[Y^0|D=1,Z=z^0,X]$,
- which are bounded only by a priori restrictions on the support of Y.

First stage monotonic in $Z \Rightarrow$ bounds are tight for

$$z^1 = \underset{z}{\operatorname{argmax}}\; E[D|X,Z=z], \quad z^0 = \underset{z}{\operatorname{argmin}}\; E[1-D|X,Z=z].$$

Review of partial identification: panel data c.f. Chernozhukov et al. (2010)

Assumption (Panel data setup)

The joint distribution of (X, Y^T, D^T) *is observed, where* $D^T = (D_1, \ldots, D_T)$ *and* $Y^T = (Y_1, \ldots, Y_T)$*, and* $D_t \in \{0,1\}$*,* $Y_t \in [0,1]$*.* $Y_t = Y_t^1 \cdot D^t + Y_t^0 \cdot (1 - D_t)$ *for potential outcomes* Y_t^0, Y_t^1*. Potential outcomes satisfy the marginal stationarity condition*

$$(Y_t^0, Y_t^1)|X, D^T \sim Y_1^0, Y_1^1|X, D^T. \tag{4}$$

- Let $M_d = 1$ if there is a $t \leq T$ such that $D_t = d$, $M_d = 0$ else.
- If $M_d = 1$, choose t_d to be the smallest t such that $D_{t_d} = d$, and set $t_d = T + 1$ if $M_d = 0$.

Law of total probability $\Rightarrow$

$$
\begin{aligned}
g(X) = E[Y^1|X] - E[Y^0|X] \\
&= \big(E[M_1|X] \cdot E[Y^1|M_1 = 1, X] \\
&+ E[1 - M_1|X] \cdot \mathbf{E[Y^1|M_1 = 0, X]}\big) \\
&- \big(E[M_0|X] \cdot E[Y^0|M_0 = 1, X] \\
&+ E[1 - M_0|X] \cdot \mathbf{E[Y^0|M_0 = 0, X]}\big) \qquad (5)
\end{aligned}
$$

- The data pin down all parts of this expression (by marginal stationarity of potential outcomes $E[Y^d|M_d = 1, X] = E[Y_{t_d}|M_d = 1, X]$)
- except for the counterfactual means $E[Y^1|M_1 = 0, X]$, $E[Y^0|M_0 = 0, X]$,
- which are bounded only by a priori restrictions on the support of Y.

The welfare ranking of policies

- conditional average treatment effect $g(X) := E[Y^1 - Y^0 | X]$
- policy difference $h^{ab} = h^a - h^b$
- potential outcomes under either policy Y^a, Y^b
- difference in social welfare between h^a, h^b:

$$\begin{aligned} \phi^{ab} &= E[Y^a - Y^b] = E[(D^a - D^b)(Y^1 - Y^0)] \\ &= E[(h^a(X) - h^b(X))(Y^1 - Y^0)] \\ &= E[h^{ab}(X) g(X)] \end{aligned} \tag{6}$$

- h^a preferred to h^b if $\phi^{ab} > 0$

Geometry

- space of bounded measurable functions of X
- equipped with the inner product

$$\langle h, g \rangle := E[h(X)g(X)] \tag{7}$$

$\Rightarrow \phi^{ab} = \langle h^{ab}, g \rangle$

- set of policies

$$\mathscr{H} = \{h(.) : 0 \leq h(X) \leq 1\} \tag{8}$$

corresponding set of policy differences

$$d\mathscr{H} = \mathscr{H} - \mathscr{H} = \{h : \sup(|h|) \leq 1\}$$

- identified set for g: $\mathscr{G}$
  special case: rectangular sets

$$\mathscr{G} = \{g(.) : g(X) \in [\underline{g}(X), \overline{g}(X)]\} \tag{9}$$

Order relationships

Social welfare ranking of policies (complete order):

$$h^a \succ^g h^b :\Leftrightarrow \langle h^{ab}, g \rangle > 0$$
$$h^a \succeq^g h^b :\Leftrightarrow \langle h^{ab}, g \rangle \geq 0 \quad (10)$$

Identified welfare ranking of policies (partial order):

$$h^a \succ^{\mathscr{G}} h^b :\Leftrightarrow \langle h^{ab}, g \rangle > 0 \; \forall \; g \in \mathscr{G}$$
$$h^a \succeq^{\mathscr{G}} h^b :\Leftrightarrow \langle h^{ab}, g \rangle \geq 0 \; \forall \; g \in \mathscr{G} \quad (11)$$

We have

$$g \in \mathscr{G} \Rightarrow (h^a \succeq^{\mathscr{G}} h^b \Rightarrow h^a \succeq^g h^b). \quad (12)$$

- Dual cone of $\mathscr{G}$: $\hat{\mathscr{G}} = \{h : \min_{g\in\mathscr{G}} \langle h, g\rangle \geq 0\}$
- Polar cone of $\mathscr{G}$: $\mathscr{G}^* = -\hat{\mathscr{G}} = \{h : \max_{g\in\mathscr{G}} \langle h, g\rangle \leq 0\}$
- Orthocomplement of g: $g^\perp = \{h : \langle h, g\rangle = 0\}$

Proposition (The maximal set of ordered policy pairs ▸ Sketch of proof)

Suppose

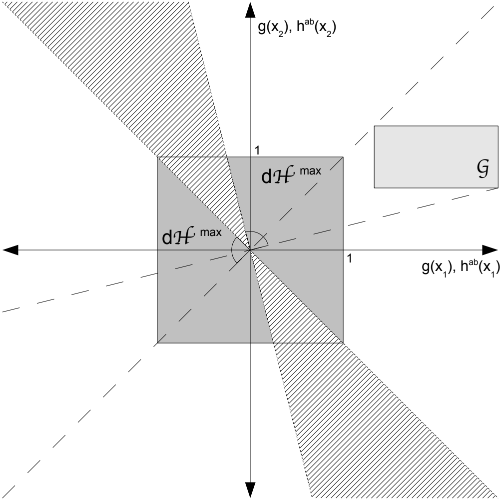
- *the identified set $\mathscr{G}$ is convex,*
- $0 \notin \overline{\mathscr{G}}$,
- $\operatorname{argmin}_{g\in\overline{\mathscr{G}}} ||g||$ *exists.*

Then:

- *$\mathscr{G}$ is uninformative about the ordering of h^a, h^b* $\Leftrightarrow$
- *neither $h^a \succeq^{\mathscr{G}} h^b$ nor $h^b \succeq^{\mathscr{G}} h^a$* $\Leftrightarrow$

$$h^{ab} \in d\mathscr{H} \setminus \left(\hat{\mathscr{G}} \cup \mathscr{G}^*\right) = d\mathscr{H} \cap \left(\bigcup_{g\in\mathscr{G}} g^\perp\right)^o \tag{13}$$

Illustration for the case $supp(X) = \{x_1, x_2\}$



Next:
Relationship between

1. feasible policy sets,
2. identification requirements.

In particular:
When is preference ordering on linearly restricted policy set

1. fully identified?
2. not identified at all?

Assumption (Affine restrictions on policy set)

The set of feasible policies is given by
$\mathscr{H}' = \{h \in \mathscr{H} : \langle h, c_i \rangle = C_i, i = 1 \ldots k\}$.

Proposition (Affine policy sets which are totally ordered by $\succeq^{\mathscr{G}}$)

Suppose

- $\mathscr{G}^o$ *is non-empty,*
- $\mathscr{H}' = \{h \in \mathscr{H} : \langle h, c_i \rangle = C_i, i = 1 \ldots k\}$.

Then:

If $\mathscr{H}'$ *is totally ordered by* $\succeq^{\mathscr{G}}$

$\Rightarrow$ $\mathscr{H}'$ *is at most one dimensional.*

▸ Sketch of proof

Proposition (Affine policy sets s.t. $\mathscr{G}$ is uninformative about $\succeq^{g}$)

Suppose

- $\mathscr{G}$ *is convex,*
- $\mathscr{G}^o$ *is non-empty,*
- $\mathscr{H}' = \{h \in \mathscr{H} : \langle h, c_i \rangle = C_i, i = 1 \dots k\}$.

Then:
There exist no $h^a \neq h^b \in \mathscr{H}'$ *such that* $h^a \succeq^{\mathscr{G}} h^b$
$\Leftrightarrow \sum_i \lambda_i c_i$ *is an element of* $\mathscr{G}^o$ *for some* $\lambda_i \in \mathbb{R}$.

▸ Sketch of proof

Nonlinear objective functions

Assume

- more general social welfare $\phi = \phi(f)$
- f: density of Y relative to the measure μ on $\mathcal{Y}$.
- Support of X: $\{x^0, \ldots, x^n\}$.
- The outcome distribution f^* of a status quo treatment assignment policy h^* is known. (e.g. $h^* = 0$)

Math review (1)

- $\mathscr{L}^p(\mathcal{Y},\mu)$: the set of measurable functions f of Y w. finite L^p norm

$$\|f\| := \left(\int |f|^p d\mu\right)^{1/p}.$$

- L^1 norm (on the space of densities $f(y)$) = the total variation norm.
- Dual space of a vector space: the set of continuous linear functionals on that space, w. the norm

$$\|\psi\| := \sup\{|\psi(f)| : \|f\| \leq 1\}.$$

- Dual space of L^p is L^q
  (for $1 \leq p < \infty$), where $1/p + 1/q = 1$:
  $\forall$ linear functionals ψ on L^p $\exists$ a function $IF \in L^q$, such that

$$\psi(f) = \int IF(y)f(y)d\mu(y).$$

- Dual space of L^1: L^∞, the set of bounded measurable functions.

Math review (2)

- ϕ is Fréchet differentiable at f^* for the norm $\|f\|$ if there exists a linear functional $\partial\phi/\partial f$, continuous with respect to the norm $\|f\|$, such that

$$\lim_{f \to f^*} \frac{\|(\phi(f) - \phi(f^*)) - \partial\phi/\partial f \cdot (f - f^*)\|}{\|f - f^*\|} = 0.$$

- ϕ L^p differentiable $\Rightarrow$
  $\exists$ dual representation of the linear functional $\partial\phi/\partial f$: $IF(y; f^*)$, the "influence function" of ϕ
- ϕ L^2 differentiable $\Rightarrow$
  $\mathrm{Var}(IF) < \infty$; ϕ is $\sqrt{n}$ estimable
- ϕ L^1 differentiable $\Rightarrow$
  IF is bounded, ϕ is a robust statistic

Lemma (Dual representations ‣ Sketch of proof)

Suppose ϕ is **$\mathbf{L^p}$ differentiable**.
Consider a family of policies $h(\theta)$, corresponding $f(h(\theta))$, $\phi(f(h(\theta)))$, denote $\check{f} = f(h(0))$.
$\Rightarrow$ there are functions $IF(y;\check{f})$, $g^f(y|x)$, and $g^\phi(x;\check{f})$, s.t.

$$\begin{aligned}
\phi_\theta &= \frac{\partial \phi}{\partial f} \cdot f_\theta = \int IF(y;\check{f}) f_\theta(y) d\mu(y) \\
f_\theta(y) &= \frac{\partial f}{\partial h} \cdot h_\theta = \langle h_\theta, g^f(y|.)\rangle \\
\phi_\theta &= \frac{\partial \phi}{\partial h} \cdot h_\theta = \langle h_\theta, g^\phi(.;\check{f})\rangle
\end{aligned} \tag{14}$$

Furthermore, $g^f(y|x) = f^1(y|x) - f^0(y|x)$, and

$$\begin{aligned}
g^\phi(x;\check{f}) &= \int IF(y;\check{f}) g^f(y|x) d\mu(y) \\
&= E[IF(Y^1;\check{f})|X=x] - E[IF(Y^0;\check{f})|X=x].
\end{aligned} \tag{15}$$

Back to partial identification of treatment effects

For an exogenous instrument Z,

$$
\begin{aligned}
g^f(y|x) = f^1(y|x) - f^0(y|x) \\
&= \Big(E[D|Z=z^1,X]\cdot f(y|D=1,z^1,x) \\
&\quad + E[1-D|Z=z^1,X]\cdot \mathbf{f^1(y|x,z^1,D=0)}\Big) \\
&\quad - \Big(E[1-D|Z=z^0,X]\cdot f(y|D=0,z^0,x) \\
&\quad + E[D|Z=z^0,X]\cdot \mathbf{f^0(y|x,z^0,D=1)}\Big)
\end{aligned} \tag{16}
$$

- The data pin down all parts of this expression
- except for the counterfactual densities
  $f^1(y|x,z^1,D=0)$,
  $f^0(y|x,z^0,D=1)$,
- which are only restricted to have their support on $\mathcal{Y}$.

Similarly under the panel data assumption:

$$\begin{aligned} g^f(y|x) = f^1(y|x) - f^0(y|x) \\ &= \big(E[M_1|X] \cdot f^1(y|M_1 = 1, x) \\ &+ E[1 - M_1|X] \cdot \mathbf{f^1(y|M_1 = 0, x)}\big) \\ &- \big(E[M_0|X] \cdot f^0(y|M_0 = 1, x) \\ &+ E[1 - M_0|X] \cdot \mathbf{f^0(y|M_0 = 0, x)}\big) \end{aligned} \tag{17}$$

- The data pin down all parts of this expression $(f^d(y|M_d = 1, X) = f(Y_{t_d}|M_d = 1, X))$
- except for the counterfactual densities $f^1(y|x, M_1 = 0)$, $f^0(y|x, M_0 = 0)$,
- which are again only restricted to have their support on $\mathcal{Y}$.

Thus, under either setup, the following is true:

Assumption

The identified set for g^f, $\mathscr{G}^f$, has the form

$$\mathscr{G}^f = \{g^f : g^f(.|x) = \tilde{g}^f(.|x) + \gamma^1(x) \cdot f^1(.|x, cf) - \gamma^0(x) \cdot f^0(.|x, cf)\},$$

where $\tilde{g}^f(.|x)$, $\gamma^1(x)$ and $\gamma^0(x)$ are known,
and $f^1(.|x, cf)$, $f^0(.|x, cf)$ are counterfactual outcome densities ranging over the set of probability densities relative to μ.

Recall:

$$g^\phi(x; f^*) = \int IF(y; f^*) g^f(y|x) d\mu(y)$$

$\Rightarrow$ identified set for g^f maps into identified set for $g^\phi(x; f^*)$.

Proposition (Bounds on local policy effects and robustness [Sketch of proof])

Suppose ϕ *is* $\mathbf{L^p}$ **differentiable**. *Then*

$$\mathscr{G}^\phi(f^*) = \{g^\phi(;f^*): \ \underline{g}^\phi(x;f^*) \leq g^\phi(x;f^*) \leq \overline{g}^\phi(x;f^*)\},$$

where

$$\begin{aligned}\overline{g}^\phi(x;f^*) &= \int IF(y;f^*)\tilde{g}^f(y|x)d\mu(y) \\ &\quad + \gamma^1(x)\cdot \sup_{y\in\mathcal{Y}} IF(y;f^*) - \gamma^0(x)\cdot \inf_{y\in\mathcal{Y}} IF(y;f^*) \\ \underline{g}^\phi(x.;f^*) &= \int IF(y;f^*)\tilde{g}^f(y|x)d\mu(y) \\ &\quad + \gamma^1(x)\cdot \inf_{y\in\mathcal{Y}} IF(y;f^*) - \gamma^0(x)\cdot \sup_{y\in\mathcal{Y}} IF(y;f^*) \qquad (18)\end{aligned}$$

These bounds are finite if and only if ϕ *is* $\mathbf{L^1}$ **differentiable**, *i.e., iff the influence function IF is bounded on the support of* Y*.*

Ranking policies in a neighborhood of the status quo

Have studied welfare effect of local policy changes h_θ - what about policy changes from h to $h + h^{ab}$?
Lower bounds on welfare effects:

$$\underline{\Delta\phi}(h^{ab}; h) := \inf_{g^f \in \mathscr{G}^f} \Big(\phi(f^* + \langle h^{ab} + h - h^*, g^f \rangle) - \phi(f^* + \langle h - h^*, g^f \rangle) \Big)$$

$$\underline{d\phi}(h_\theta; h) := \inf_{g^f \in \mathscr{G}^f} \frac{\partial}{\partial\theta} \phi(f^* + \langle h(\theta) - h^*, g^f \rangle) = \inf_{g^\phi \in \mathscr{G}^\phi(h)} \langle h_\theta, g^\phi \rangle, \quad (19)$$

Theorem

Suppose ϕ is **continuously $\mathbf{L^1}$ differentiable***.*
Let h_θ be such that $\underline{d\phi}(h_\theta; h^*) > 0$. Then there exists a δ such that, for all h such that $\|h - h^*\| \leq \delta$ and all $0 < \gamma \leq \delta$,

$$\underline{\Delta\phi}(\gamma \cdot h_\theta; h) > 0.$$

Generalizing results from linear case

Assumption (Differentiable constraints)

The set of policies is given by $\mathscr{H}' = \left\{h \in [0,1]^{n+1} : C(h) = 0\right\}$, *where* $C : \mathbb{R}^{n+1} \to \mathbb{R}^k$, $k \leq n$, *is differentiable.*

$\Rightarrow$ tangent space at h^*:

$$T_{h^*}\mathscr{H}' = \left\{ h_\theta : \left\langle \frac{\partial C_i}{\partial h}(h^*), h_\theta \right\rangle = 0, i = 1 \ldots k \right\}$$

Proposition (Policy sets such that $T_{h^*}\mathscr{H}'$ is totally ordered / unordered)

Suppose ϕ is **$\mathbf{L^p}$ differentiable**, *$\mathscr{H}'$ is subject to a set of differentiable constraints, and $\mathscr{G}^{\phi^o} \neq \emptyset$.*

Then:
$T_{h^*}\mathscr{H}'$ is totally ordered by $\succeq^{\mathscr{G}^\phi}$
$\Rightarrow \mathscr{H}'$ is at most one dimensional, i.e., $k = n$.

Suppose furthermore $\mathscr{G}^\phi$ is convex.
Then:
There are no $h^a_\theta, h^b_\theta \in T_{h^*}\mathscr{H}'$ such that $h^a_\theta \succeq^{\mathscr{G}^\phi} h^b_\theta$
$\Leftrightarrow \sum_i \lambda_i \frac{\partial C_i}{\partial h}(h^*)$ is an element of $\mathscr{G}^{\phi^o}$ for some $\lambda_i \in \mathbb{R}$.

Theorem

Suppose ϕ *is* **continuously** L^1 **differentiable**, $\mathscr{H}'$ *is subject to a set of differentiable constraints,* $\mathscr{G}^\phi$ *has non-empty interior* ${\mathscr{G}^\phi}^o$, *and* $\mathscr{G}^\phi$ *is bounded.*

$\Rightarrow \exists$ *a neighborhood* N *of* h^* *in* $\mathscr{H}'$ *s.t., for all* $h \in N$*:*

(i) $T_{h^*}\mathscr{H}'$ *is totally unordered* $\Rightarrow$ N *is totally unordered.*

(ii) $T_{h^*}\mathscr{H}'$ *is partially ordered* $\Rightarrow$ N *is partially ordered.*

(iii) $T_{h^*}\mathscr{H}'$ *is totally ordered* $\Rightarrow$ N *is totally ordered.*

An aside: Relationship to axiomatic decision theory

E.g. Anscombe and Aumann (1963), Bewley (2002), Ryan (2009)
Differences:

1. Space over which preferences are defined
    - axiomatic decision theory: acts
    - this paper: treatment assignment policies
2. Question of interest
    - axiomatic decision theory:
      restrictions on actual human behaviour, preferences
      $\Rightarrow$ characterizations of preferences (e.g., in terms of a set of priors)
    - this paper: "reverse question"
      identified set for conditional average treatment effects function
      $\Rightarrow$ derive preferences, behavior

Definition (Independence)

The relationship $\succ$ satisfies independence if, for all $h^a, h^b, h^c \in \mathscr{H}$, and all $\alpha \in (0,1)$, we have that $h^a \succ h^b$ if and only if

$$\alpha h^a + (1-\alpha)h^c \succ \alpha h^b + (1-\alpha)h^c.$$

adapting results from Ryan (2009) gives:

Proposition

A partial order $\succ$ on $\mathbb{R}^{\mathscr{X}}$ satisfies independence if and only if it can be represented as $\succ^{\mathscr{G}}$ for some convex set $\mathscr{G}$.

▸ Sketch of proof

Application to project STAR data

- 80 schools in Tennessee 1985-1986:
- Kindergarten students randomly assigned to small (13-17 students) / regular (22-25 students) classes within schools
- Students were supposed to remain in the same type class for 4 years
- Students entering school later were also randomly assigned
- Compliance was imperfect
- See Krueger (1999), Graham (2008)

This presentation: only point estimates of bounds, no inference!

This is an interesting application because:

- Large but imperfect compliance to experimental assignment
  $\Rightarrow$ non-trivial but informative bounds on treatment effects
- Heterogeneity in treatment effects
  $\Rightarrow$ reallocations s.t. budget constraint potentially welfare improving
- Potential for disagreement about objective function, identifying assumptions, budget constraint
  $\Rightarrow$ (identification of) policy ranking depends upon these

- Sample: students observed in grades 1 - 3
- Instrument $Z = 1$ for students assigned to a small class (upon first entering a project STAR school)
- realized treatment $D = 1$ for students in a small class (for all but at most 1 year during the study period)
- "poor": students receiving a free lunch
- Redistributive policies: Assigning all poor students to small classes, holding the average class size constant

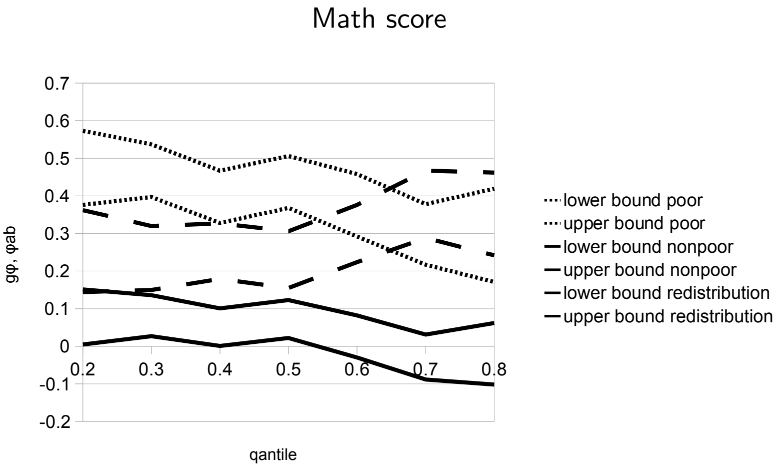
Table: THE JOINT DISTRIBUTION OF ASSIGNED AND REALIZED CLASS-SIZE

	D		Total
Z	0	1	
0	2,873	217	3,090
1	74	1,082	1,156
Total	2,947	1,299	4,246

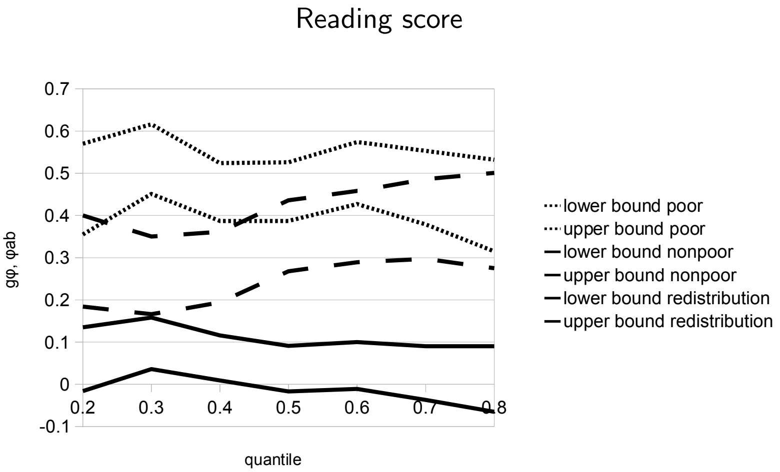
- Y: normalized average math scores in 3rd and 4th grade
- ϕ: quantiles of the test score distribution
- Following table: Bounds on $E[g^{\phi}|poor]$, $E[g^{\phi}|non-poor]$, $E[g^{\phi}]$, ϕ^{ab} for redistribution

Objective ϕ	poor students	non-poor students	all students	effect of redistribution
Assuming only instrument exogeneity				
0.3rd quantile	[0.185, 0.537]	[-0.030, 0.320]	[0.045, 0.396]	[-0.047, 0.199]
0.5th quantile	[0.174, 0.506]	[-0.025, 0.306]	[0.045, 0.376]	[-0.046, 0.186]
0.7th quantile	[-0.025, 0.378]	[0.065, 0.467]	[0.033, 0.435]	[-0.173, 0.110]
Assuming instrument exogeneity and monotonicity				
0.3rd quantile	[0.397, 0.537]	[0.150, 0.320]	[0.235, 0.393]	[0.027, 0.136]
0.5th quantile	[0.368, 0.506]	[0.155, 0.306]	[0.228, 0.374]	[0.022, 0.123]
0.7th quantile	[0.217, 0.378]	[0.288, 0.467]	[0.261, 0.432]	[-0.088, 0.031]

The same as a picture:



Now for reading scores:



Summary of empirical findings: Role of

- Identifying assumptions:
  - only instrument exogeneity:
    unidentified policy ranking
  - instrument exogeneity and monotonicity of outcomes:
    partially identified policy ranking
- Objective function:
  - Lower quantiles:
    redistributing to poor unambiguously positive
  - Top quantiles:
    redistributing to poor ambiguous effect
- Feasible policies:
  - Redistributing to poor:
    ambiguous for top
  - Decreasing class size for all:
    unambiguously positive (w.out assuming monotonicity!)

Outlook

Next project: **"Partial Identification of optimal policy parameters in public finance models"**, such as

- Optimal income taxation
  (Mirrlees (1971), Saez (2001))
- Optimal unemployment insurance
  (Baily (1978), Chetty (2006))

Common features (see Chetty (2009)):

- weighted utilitarian SWF
- Envelope arguments yield simple FOCs for optimal policy
- These are expressible in terms of a single or few response functions

Empirical implementation?

- Need to estimate continuous response functions
  (e.g., unemployment benefits $\Rightarrow$ unemployment rate;
  marginal tax rate $\Rightarrow$ tax base)
- Existing work: functional form assumptions
  (e.g., Saez (2001): constant elasticity extrapolation)
- I will consider nonparametric setups
  (e.g., monotonic treatment response and instrument exogeneity)
- $\Rightarrow$ identified sets for optimal policy parameters
  (e.g., lower bound on optimal top tax rate)

Goals:

- General characterization of these identified sets
- Inference
- Reanalysis of existing work, dropping functional form assumptions
- Possibly: Conceptualizing a feedback process of policy updating and new data; convergence to optimum?

Conclusion

- Goal of this paper: Exploring the frontier in the trade-off between
  - recognition of the limits of our knowledge,
  - and the necessity to give informed policy recommendations.
- In particular:
  1. How does the data distribution map into policy-rankings?
  2. Under what conditions is the welfare ranking of policies fully / partially / not at all identified?
- Depends on *interaction* of
  1. identified set,
  2. feasible policy set,
  3. objective function.

Thanks for your time!

Sketch of proof:

- $h^a \succeq^{\mathscr{G}} h^b \Leftrightarrow h^{ab} \in \hat{\mathscr{G}}$; the first claim follows immediately
- $h^a \succ^{\mathscr{G}} h^b$ or $h^b \succ^{\mathscr{G}} h^a$ iff $h^{ab} \notin \left(\bigcup_{g \in \mathscr{G}} g^{\perp}\right)$:
  - $\mathscr{G}$ convex $\Rightarrow$ connected
    $\Rightarrow \langle h^{ab}, \mathscr{G} \rangle$ connected
    $\Rightarrow \langle h^{ab}, \mathscr{G} \rangle$ contains both positive and non-positive values only if it contains 0
  - there is a $g \in \mathscr{G}$ such that $\langle h^{ab}, g \rangle = 0$
    $\Leftrightarrow$ there is a $g \in \mathscr{G}$ such that $h^{ab} \in g^{\perp}$
- The equality now follows from topological arguments, requiring existence of a separating hyperplane between 0 and $\mathscr{G}$.

Back

Proposition (Affine policy sets which are totally ordered by $\succeq^{\mathscr{G}}$)

Suppose

- $\mathscr{G}^o$ *is non-empty,*
- $\mathscr{H}' = \{h \in \mathscr{H} : \langle h, c_i \rangle = C_i, i = 1 \ldots k\}$.

Then:
If $\mathscr{H}'$ *is totally ordered by* $\succeq^{\mathscr{G}}$
$\Rightarrow$ $\mathscr{H}'$ *is at most one dimensional.*

Sketch of proof:

- Suppose $\dim(\mathscr{H}') > 1 \Rightarrow$ choose $h^1, h^2, h^3 \in \mathscr{H}'$, such that $h^1 - h^3$ and $h^2 - h^3$ are linearly independent; choose $g \in \mathscr{G}^o$.
- Define $h^* = \langle h^2 - h^3, g\rangle (h^1 - h^3) - \langle h^1 - h^3, g\rangle (h^2 - h^3)$.
  $\Rightarrow \langle h^*, g\rangle = 0$; $h^* \neq 0$
- Choose h^4, h^5 in $\mathscr{H}'$, such that $h^4 - h^5 = const. \cdot h^*$.

Back

Proposition (Affine policy sets s.t. $\mathscr{G}$ is uninformative about $\succeq^g$)

Suppose

- $\mathscr{G}$ *is convex,*
- $\mathscr{G}^o$ *is non-empty,*
- $\mathscr{H}' = \{h \in \mathscr{H} : \langle h, c_i \rangle = C_i, i = 1 \ldots k\}$.

Then:
There exist no $h^a \neq h^b \in \mathscr{H}'$ *such that* $h^a \succeq^{\mathscr{G}} h^b$
$\Leftrightarrow \sum_i \lambda_i c_i$ *is an element of* $\mathscr{G}^o$ *for some* $\lambda_i \in \mathbb{R}$.

Sketch of proof: (for case $k = 1$)

- Suppose $\lambda c = g \in \mathscr{G}^o \Rightarrow \langle h^a - h^b, g \rangle = 0$ for all $h^a, h^b \in \mathscr{H}'$
- Suppose $\lambda c \notin \mathscr{G}^o$ for any $\lambda \Rightarrow \exists$ a separating hyperplane $\tilde{h}^\perp$ s.t.

$$\sup_{\lambda \in \mathbb{R}} \langle \tilde{h}, \lambda c \rangle \leq \inf_{g \in \mathscr{G}} \langle \tilde{h}, g \rangle.$$

- $\Rightarrow \langle \tilde{h}, c \rangle = 0,\ 0 \leq \inf_{g \in \mathscr{G}} \langle \tilde{h}, g \rangle$.
- Choose $h^a, h^b \in \mathscr{H}'$ such that $h^{ab} = const. \cdot \tilde{h} \Rightarrow h^a \succeq^{\mathscr{G}} h^b$

Back

Sketch of proof:

- Existence of IF: differentiability, dual of L^p is isomorphic to L^q
- $f(y;\theta) - \check{f}(y) = \langle h(\theta) - h(0), g^f(y|.)\rangle$ by construction
  $\Rightarrow$ differentiate
- ϕ as a differentiable mapping from h to $\mathbb{R}$
  $\Rightarrow$ existence of g^ϕ by Riesz representation theorem.
- Combining these expressions:

$$\langle h_\theta, g^\phi(.;\check{f})\rangle = \int IF(y;\check{f})\langle h_\theta, g^f(y|.)\rangle d\mu(y).$$

Exchanging the order of integration, w.r.t. x and y:

$$\langle h_\theta, g^\phi(.;\check{f})\rangle = \langle h_\theta, \int IF(y;\check{f}) g^f(y|.) d\mu(y)\rangle.$$

Since this holds for all h_θ, the last claim follows.

Back

Sketch of proof: By lemma 1,

$$
\begin{aligned}
g^{\phi}(x; f^*) &= \int IF(y; f^*) g^f(y|x) d\mu(y) = \int IF(y; f^*) g^f(y|x) d\mu(y) \\
&= \int IF(y; f^*) \left(\tilde{g}^f(.|x) + \gamma^1(x) \cdot f^1(.|x, cf) - \gamma^0(x) \cdot f^0(.|x, cf) \right) d\mu(y) \\
&= \int IF(y; f^*) \tilde{g}^f(.|x) d\mu(y) \\
&\quad + \gamma^1(x) \cdot E[IF(Y^1; f^*)|x, cf] - \gamma^0(x) \cdot E[IF(Y^0; f^*)|x, cf], \quad (20)
\end{aligned}
$$

Sharp bounds on the conditional expectations are given by

$$
\inf_{y \in \mathcal{Y}} IF(y; f^*), \quad \sup_{y \in \mathcal{Y}} IF(y; f^*).
$$

▸ Back

Proposition

A partial order $\succ$ on $\mathbb{R}^{\mathscr{X}}$ satisfies independence if and only if it can be represented as $\succ^{\mathscr{G}}$ for some convex set $\mathscr{G}$.

Sketch of proof:

- Independence $\Rightarrow$ upper contour sets are translations of a convex cone
- Dual cone theorem: The dual cone of the dual cone of a convex cone is the original cone.
- Thus: Take as $\mathscr{G}$ any set which spans the dual cone of the upper contour set of 0.

Back