Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities

Hiroaki Kaido
Boston University

Andres Santos
UC San Diego

The Model

Consider $X \sim P_0$ with $X \in \mathcal{X} \subseteq \mathbf{R}^{d_X}$, $\theta \in \Theta \subset \mathbf{R}^{d_\theta}$ and restriction:

$$F(\int m(x,\theta)dP_0(x)) \leq 0$$

where $m : \mathcal{X} \times \Theta \to \mathbf{R}^{d_m}$ and $F : \mathbf{R}^{d_m} \to \mathbf{R}^{d_F}$ are known functions.

Under Identification

- Semiparametric efficiency bound may exist.
- Possible to construct efficient estimator.

Partial Identification

- What does semiparametric efficiency bound mean?
- Is there an efficient estimator?

Basic Example

Suppose Y is unobserved, $Y_L \leq Y \leq Y_U$ almost surely and $\theta_0 = E[Y]$.

Identified set: $[E[Y_L], E[Y_U]]$

Natural Estimator: $[\bar{Y}_L, \bar{Y}_U]$

Efficiency

- Set-estimator is built from an efficient estimator of the boundary.
- Easy to characterize boundary as function of P_0.

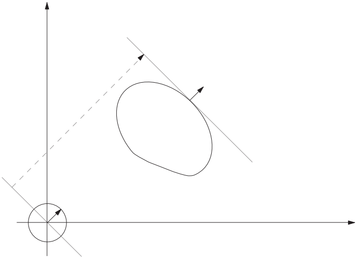
In General Model

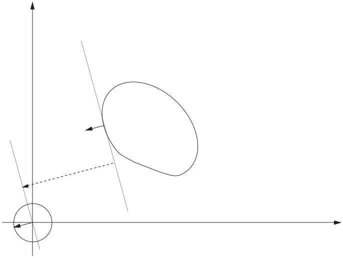
- Think of efficient estimation of the boundary of the set.

... but how do we characterize boundary as a function of P_0?

Under Convexity

Each boundary point can be identified by its supporting hyperplane:





Supporting Hyperplane

- Each boundary point is contained in a tangent hyperplane to set.
- **Intuition:** Tangent hyperplanes trace out the boundary of the set.

Goal: Construct an efficient estimator for the "support function".

Example 1

Manski & Tamer (2002): Let $Y \in \mathbf{R}$ be unobservable and satisfy:

$$Y = Z'\theta_0 + \epsilon ,$$

where $E[\epsilon|Z] = 0$ and $Z \in \mathbf{R}^{d_\theta}$ has discrete support $\mathcal{Z} \equiv \{z_1, \ldots z_K\}$.

If (Y_L, Y_U) are observable and $Y_L \leq Y \leq Y_U$ almost surely, then:

$$\Theta_0 = \{\theta \in \mathbf{R}^{d_\theta} : E[Y_L|Z] - Z'\theta \leq 0 \text{ and } Z'\theta - E[Y_U|Z] \leq 0\}$$

Comments

- Constraints are linear (convex) in θ, implies Θ_0 is convex.
- Challenging to think of $\partial\Theta_0$ as function of distribution of (Y_L, Y_U, Z).

Example 1

$$\Theta_0 = \{\theta \in \mathbf{R}^{d_\theta} : E[Y_L|Z] - Z'\theta \leq 0 \text{ and } Z'\theta - E[Y_U|Z] \leq 0\}$$

Role of m_L: Let $x \equiv (y_L, y_U, z)$, use $m_L(x)$ to pick moments we need.

$$\int m_L(x) dP_0(x) = \begin{pmatrix} (E[Y_L 1\{Z = z_1\}], \ldots, E[Y_L 1\{Z = z_K\}])' \\ (E[Y_U 1\{Z = z_1\}], \ldots, E[Y_U 1\{Z = z_K\}])' \\ (P(Z = z_1), \ldots, P(Z = z_K))' \end{pmatrix}$$

Role of F_L: Use F_L function to construct expressions we require:

$$F_L(\int m_L(x) dP_0(x)) = \begin{pmatrix} -(E[Y_L|Z = z_1], \ldots, E[Y_L|Z = z_K])' \\ (E[Y_U|Z = z_1], \ldots, E[Y_U|Z = z_K])' \end{pmatrix}$$

Combining: Set $A \equiv (-z_1, \ldots, -z_K, z_1, \ldots z_K)'$, $m(x, \theta) = (\theta' A', m_L(x)')'$,

$$F(\int m(x, \theta) dP_0(x)) = A\theta - F_L(\int m_L(x) dP_0(x))$$

Example 2

Pakes (2010): Agent chooses $Z \in \mathbf{R}^{d_Z}$ from $\mathcal{Z} \equiv \{z_1, \ldots, z_K\}$ to maximize:

$$E[\pi(Y, Z, \theta_0)|\mathcal{F}]$$

where Y is observable variable and $\mathcal{F}$ is the agent's information set.

A common specification is $\pi(Y, Z, \theta_0) = \psi(Y, Z) - Z'\theta_0$, which implies:

$$\Theta_0 = \{\theta \in \mathbf{R}^{d_\theta} : E[((\psi(Y, z_j) - \psi(Y, z_i)) - (z_j - z_i)'\theta)1\{Z^* = z_i\}] \leq 0\}$$

where Z^* is the agent's observed optimal decision.

Comments

- Identified set is convex and determined by linear inequalities in θ.
- Requires assumptions on agent's beliefs.
- ψ assumed known, though often separately estimated.

Example 3

Luttmer (1996): Under power utility and market frictions, modified Euler:

$$E[\frac{1}{1+\rho}Y^{-\gamma}Z - P] \leq 0$$

with Y future/present consumption, $Z \in \mathbf{R}^{dz}$ asset payoff and P prices.

Comments

- Constraints strictly convex if $Z \geq 0$ and $Z > 0$ with positive probability.
- Resulting identified set for (ρ, γ) is convex.
- Big Caveat: Our efficiency bound is for i.i.d. data.

General Outline

Preliminaries

- Background: support functions and efficiency.
- Linear constraints and regularity.

Efficiency

- Characterizing sources of irregularity.
- Semiparametric efficiency bound for support function.
- Show “plug-in” estimator is in fact efficient.

Confidence Regions

- Construct bootstrap procedure.
- Establish validity of confidence regions.

Literature Review

Partial Identification

Manski & Tamer (2002), Manski (2003), Imbens & Manski (2004), Chernozhukov, Hong & Tamer (2007), Romano & Shaikh (2008, 2010), Stoye (2009), Bontemps, Magnac & Maurin (2007), Beresteanu & Molinari (2008), Beresteanu, Molinari & Molchanov (2009), Kaido (2010).

Moment Inequalities

Andrews & Jia (2008), Menzel (2009), Rosen (2009), Andrews & Soares (2010), Bugni (2010), Canay (2010), Pakes (2010).

Efficiency and Regularity

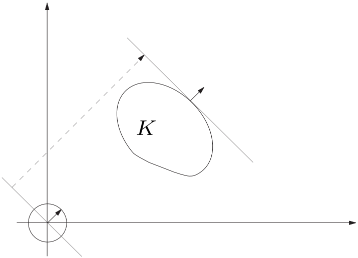
Chamberlain (1987, 1992), Brown & Newey (1998), Newey (2004), Ai & Chen (2009), Chen & Pouzo (2009), Hirano & Porter (2009), Chernozhukov, Lee & Rosen (2009), Song (2010).

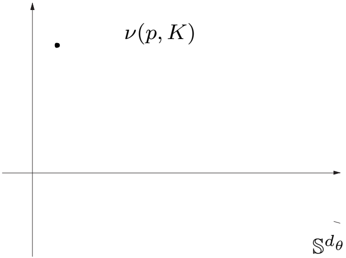
Support Function

Let $\mathbb{S}^{d_\theta} \equiv \{p \in \mathbf{R}^{d_\theta} : \|p\| = 1\}$ and K be a convex compact set.

The support function of K is then pointwise defined (on $\mathbb{S}^{d_\theta}$) by:

$$\nu(p, K) \equiv \sup_{k \in K} \langle p, k \rangle$$



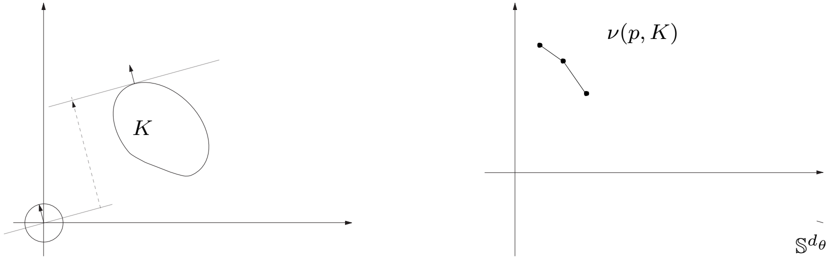


Support Function

Let $\mathbb{S}^{d_\theta} \equiv \{p \in \mathbf{R}^{d_\theta} : \|p\| = 1\}$ and K be a convex compact set.

The support function of K is then pointwise defined (on $\mathbb{S}^{d_\theta}$) by:

$$\nu(p, K) \equiv \sup_{k \in K} \langle p, k \rangle$$

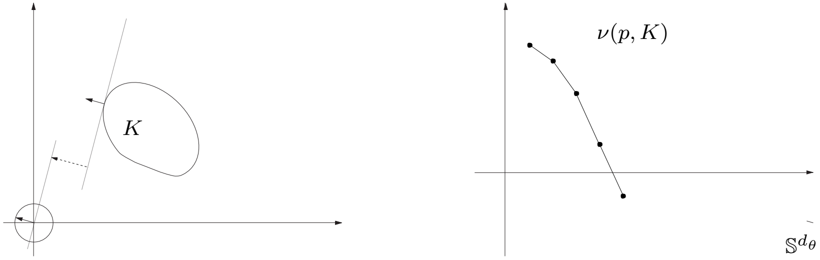


Support Function

Let $\mathbb{S}^{d_\theta} \equiv \{p \in \mathbf{R}^{d_\theta} : \|p\| = 1\}$ and K be a convex compact set.

The support function of K is then pointwise defined (on $\mathbb{S}^{d_\theta}$) by:

$$\nu(p, K) \equiv \sup_{k \in K} \langle p, k \rangle$$

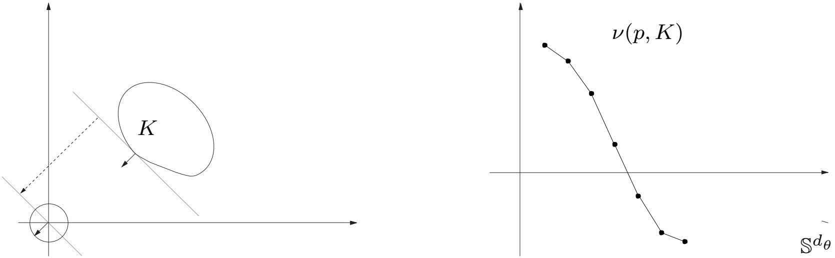


Support Function

Let $\mathbb{S}^{d_\theta} \equiv \{p \in \mathbf{R}^{d_\theta} : \|p\| = 1\}$ and K be a convex compact set.

The support function of K is then pointwise defined (on $\mathbb{S}^{d_\theta}$) by:

$$\nu(p, K) \equiv \sup_{k \in K} \langle p, k \rangle$$

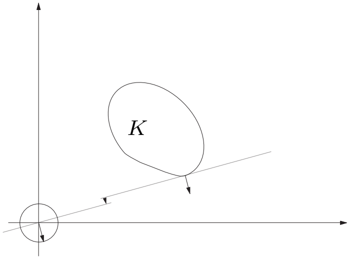


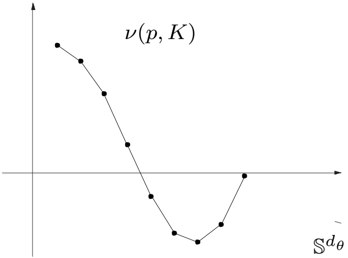
Support Function

Let $\mathbb{S}^{d_\theta} \equiv \{p \in \mathbf{R}^{d_\theta} : \|p\| = 1\}$ and K be a convex compact set.

The support function of K is then pointwise defined (on $\mathbb{S}^{d_\theta}$) by:

$$\nu(p, K) \equiv \sup_{k \in K} \langle p, k \rangle$$



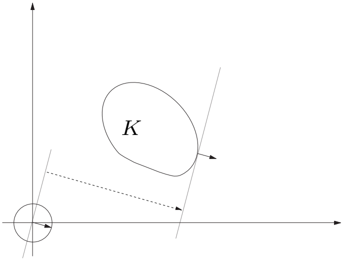


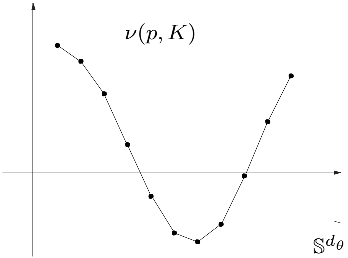
Support Function

Let $\mathbb{S}^{d_\theta} \equiv \{p \in \mathbf{R}^{d_\theta} : \|p\| = 1\}$ and K be a convex compact set.

The support function of K is then pointwise defined (on $\mathbb{S}^{d_\theta}$) by:

$$\nu(p, K) \equiv \sup_{k \in K} \langle p, k \rangle$$



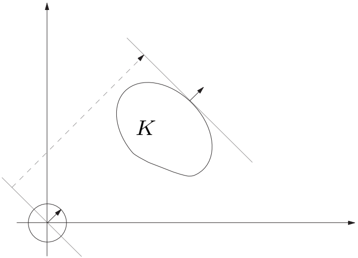


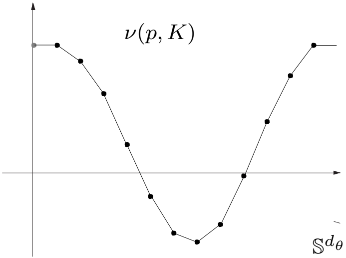
Support Function

Let $\mathbb{S}^{d_\theta} \equiv \{p \in \mathbf{R}^{d_\theta} : \|p\| = 1\}$ and K be a convex compact set.

The support function of K is then pointwise defined (on $\mathbb{S}^{d_\theta}$) by:

$$\nu(p, K) \equiv \sup_{k \in K} \langle p, k \rangle$$





Support Function

Let $\mathcal{C}(\mathbb{S}^{d_\theta})$ be the space of bounded continuous functions on $\mathbb{S}^{d_\theta}$.

$\Rightarrow$ Every convex compact K is associated with a unique function in $\mathcal{C}(\mathbb{S}^{d\theta})$.

Theorem (Hörmander) For any two convex compact, K_1 and K_2:

$$d_H(K_1, K_2) = \sup_{p \in \mathbb{S}^{d_\theta}} |\nu(p, K_1) - \nu(p, K_2)|$$

where $d_H(K_1, K_2)$ is the Hausdorff distance between the sets K_1 and K_2.

Norm Equality

- Relationship allows for inference; Beresteanu & Molinari (2008).
- Equipping $\mathcal{C}(\mathbb{S}^{d_\theta})$ with $\|\cdot\|_\infty$ implies embedding is isometric.

Identified Set

$$\Theta_0(Q) \equiv \{\theta \in \Theta : F(\int m(x,\theta)dQ(x)) \leq 0\}$$

In turn, we can map $\Theta_0(Q)$ into its support function $p \mapsto \nu(p, \Theta_0(Q))$ by:

$$\nu(p, \Theta_0(Q)) = \sup_{\theta \in \Theta} \{\langle p, \theta \rangle \text{ s.t. } F(\int m(x,\theta)dQ(x)) \leq 0\}$$

Support Function

- Relatively simple dependence on Q (unlike $\partial\Theta_0(Q)$).
- $\Rightarrow$ With parametric model for P_0, could estimate by MLE.
- But note $\nu(\cdot, \Theta_0(Q)) \in \mathcal{C}(\mathbb{S}^{d_\theta})$ is an infinite dimensional parameter.

Efficiency in $\mathcal{C}(\mathbb{S}^{d_\theta})$

Finite Dimensional Setting

- Model $\mathbf{P}$, parameter $\rho : \mathbf{P} \to \mathbf{R}$, want to estimate $\rho(P_0)$ efficiently.
- Compute tangent space $\dot{\mathbf{P}}$, derivative $\dot{\rho}$, and project $\dot{\rho}$ onto $\dot{\mathbf{P}}$.

Problem: For us, $\rho(P_0) = \nu(\cdot, \Theta_0(P_0))$ which is in $\mathcal{C}(\mathbb{S}^{d_\theta})$.

Key: Tangent space remains the same, but differentiability changes...

Definition: For a model $\mathbf{P}$, parameter $\rho : \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ is pathwise weak differentiable at $P_0 \in \mathbf{P}$ if there is continuous linear operator $\dot{\rho} : \dot{\mathbf{P}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$:

$$\lim_{\eta \to 0} \Big| \int_{\mathbb{S}^{d_\theta}} \Big\{ \frac{\rho(P_\eta)(p) - \rho(P_0)(p)}{\eta} - \dot{\rho}(\dot{P}_0)(p) \Big\} dB(p) \Big| = 0$$

for any finite Borel measure B and submodel $\eta \mapsto P_\eta$ passing through P_0.

Convolution Theorem

Theorem (Háyek, LeCam) Under regularity conditions, if $\rho : \mathbf{P} \to \mathcal{C}(\mathbb{S}^{d_\theta})$ is pathwise weak differentiable at P_0 and $T_n \stackrel{L}{\to} \mathbb{G}$ is a regular estimator, then:

$$\mathbb{G} \stackrel{L}{=} \mathbb{G}_0 + \Delta_0$$

for a unique Gaussian process $\mathbb{G}_0$ and tight Borel r.v. Δ_0 with $\Delta_0 \perp \mathbb{G}_0$.

Comments

- Since parameter is in $\mathcal{C}(\mathbb{S}^{d_\theta})$, estimator $\{T_n\}$ converges in $\mathcal{C}(\mathbb{S}^{d_\theta})$.
- Gaussian process $\mathbb{G}_0$ does not depend on $\{T_n\}$, "noise term" Δ_0 does.

Intuition: Every regular estimator converges to $\mathbb{G}_0$ plus noise Δ_0 ...

$\Rightarrow$ An estimator is efficient if it converges in distribution to $\mathbb{G}_0$

Semiparametric Efficiency

Characterize Law of $\mathbb{G}_0$

- Finite dimensions: $\mathbb{G}_0$ is multivariate normal; report covariance matrix.
- In infinite dimensions ... find covariance kernel of $\mathbb{G}_0$.

Definition: The inverse information covariance functional for $\rho(P_0)$ is:

$$I^{-1}(p_1, p_2) \equiv \text{Cov}\{\mathbb{G}_0(p_1), \mathbb{G}_0(p_2)\}$$

Objectives

- Compute tangent space for $\mathbf{P}$ (must state assumptions on P_0).
- Establish $\rho(P_0) = \nu(\cdot, \Theta_0(P_0))$ is weakly pathwise differentiable.

$\Rightarrow$ First need to understand possible sources of irregularity...

Differentiability Problems

- Suppose $X = (X^{(1)}, X^{(2)})$, and $X \sim P_0$ with $E[X^{(1)}] > 0$ and $E[X^{(2)}] > 0$

$$F(\int m(x,\theta)dP_0(x)) = \begin{cases} \int (x^{(1)}\theta_1 + x^{(2)}\theta_2 - K)dP_0(x) \\ -\theta_2 \\ -\theta_1 \end{cases}$$

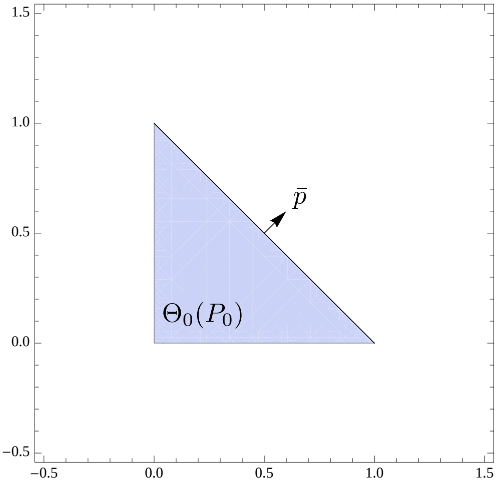
- Consider a submodel $\eta \mapsto P_\eta$ passing through P_0 and satisfying:

$$\int x^{(1)}dP_\eta(x) = E[X^{(1)}](1+\eta) \qquad \int x^{(2)}dP_\eta(x) = E[X^{(2)}]$$

Comments

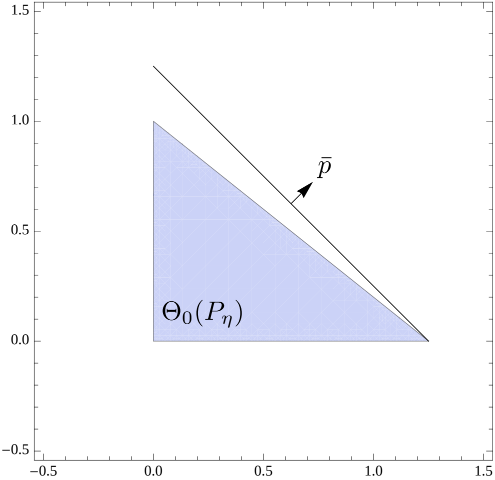
- Identified set is a triangle in positive orthant.
- What happens if we point $p \in \mathbb{S}^{d_\theta}$ at flat face of $\Theta_0(P_0)$?

Differentiability Problems



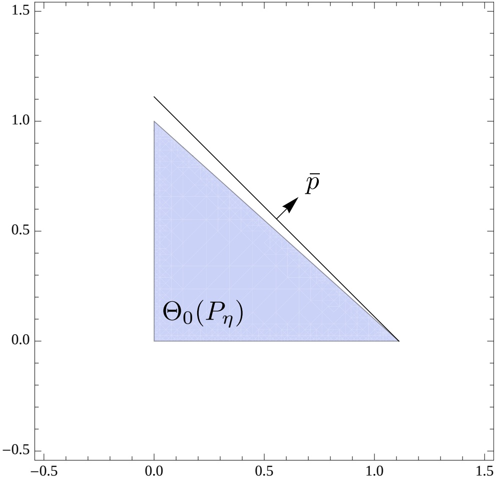
Identified Set

Differentiability Problems



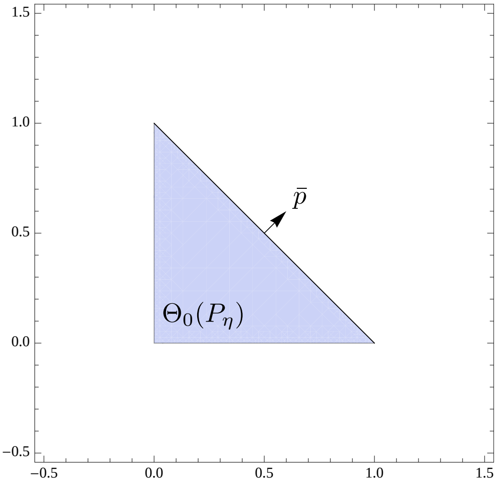
Identified Set at $\eta = -0.2$

Differentiability Problems



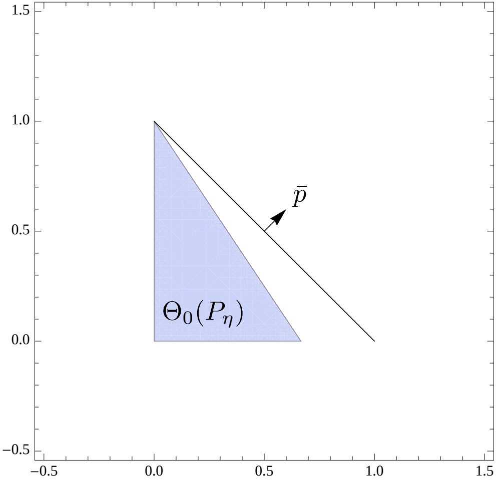
Identified Set at $\eta = -0.1$

Differentiability Problems



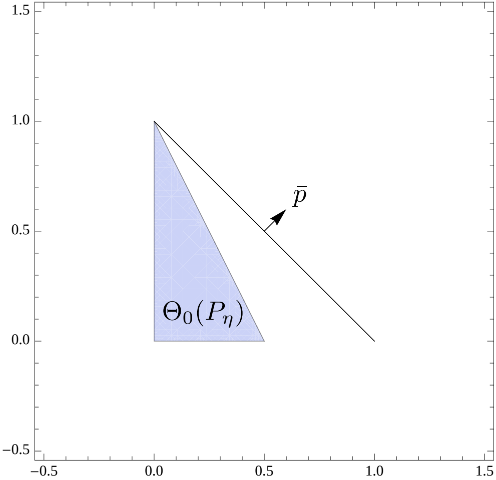
Identified Set at $\eta = 0$

Differentiability Problems



Identified Set at $\eta = 0.1$

Differentiability Problems



Identified Set at $\eta = 0.2$

Differentiability Problems

Formally: If $\bar{p} = \bar{s}/\|s\|$ for $\bar{s} = (E[X^{(1)}], E[X^{(2)}])$, then along $\eta \mapsto P_\eta$,

$$\nu(\bar{p}, \Theta_0(P_\eta)) = \begin{cases} \frac{K}{\|\bar{s}\|} & \text{if } \eta \geq 0 \\ \frac{K}{\|\bar{s}\|} \frac{E[X^{(1)}]}{(E[X^{(1)}]+\eta)} & \text{if } \eta < 0 \end{cases}$$

Implications

- When $d_\theta > 1$, slope of linear constraints should not depend on P_0.
- This is not a problem for strictly convex constraints.
- Not a problem in discussed examples.

Next Goal

- Restrict P_0, F, m so $\eta \mapsto \nu(\cdot, \Theta_0(P_\eta))$ is differentiable.
- Derive semiparametric efficiency bound and efficient estimator.

Model Details

Problem: Linear constraints may cause support function to be irregular.

Approach: Group constraints into linear and strictly convex ...

For $m(x,\theta)$: Let $m_S : \mathcal{X} \times \Theta \to \mathbf{R}^{d_{m_S}}$, $m_L : \mathcal{X} \to \mathbf{R}^{d_{m_L}}$, A a $d_{F_L} \times d_\theta$ matrix.

$$m(x,\theta) \equiv (m_S(x,\theta)', m_L(x)', \theta' A')'$$

For $F(v)$: Let $F_S : \mathbf{R}^{d_{m_S}} \to \mathbf{R}^{d_{F_S}}$ and $F_L : \mathbf{R}^{d_{m_L}} \to \mathbf{R}^{d_{F_L}}$, let:

$$F(\int m(x,\theta) dP_0(x)) = \begin{pmatrix} F_S(\int m_S(x,\theta) dP_0(x)) \\ A\theta - F_L(\int m_L(x) dP_0(x)) \end{pmatrix}$$

where $\theta \mapsto F_S^{(i)}(\int m_S(x,\theta) dP_0(x))$ is strictly convex for $1 \le i \le d_{F_S}$.

Key Assumptions

Assumptions (A)

(i) $\Theta_0(P_0)$ is contained in the interior of Θ (relative to $\mathbf{R}^{d_\theta}$).

(ii) There exists a $\theta_0 \in \Theta$ such that $F(E[m(X,\theta_0)]) < 0$.

(iii) $\theta \mapsto m(x,\theta)$ is differentiable (but not necessarily in x).

(iv) At each $\theta \in \partial\Theta_0(P_0)$ number of active constraints $\leq d_\theta$.

Discussion

- A(i) largely notation. May impose $\|\theta\|^2 \leq B$ or $\theta^{(i)} \geq C$ through F, m.
- A(ii) With moment equalities may lose convexity.
- A(iii) Allows discontinuous functions of x (e.g. $1\{X_i = x\}$).
- A(iv) analogue to intersection bounds (Hirano $\&$ Porter (2009)).

Key: A(ii)-A(iv) implied by linear independence requirement.

Lagrangian Representation

$$\nu(p, \Theta_0(P_0)) = \sup_{\theta \in \Theta} \{\langle p, \theta \rangle \text{ s.t. } F(\int m(x,\theta)dP_0(x)) \leq 0\}$$

$$= \sup_{\theta \in \Theta} \{\langle p, \theta \rangle + \lambda(p, P_0)' F(\int m(x,\theta)dP_0(x))\}$$

Intuition

- Each boundary point of $\Theta_0(P_0)$ is a maximizer for some $p \in \mathbb{S}^{d_\theta}$.
- $\lambda(p, P_0)$ reflects importance of constraints in keeping you inside $\Theta_0(P_0)$.

Reveals Dependence on P_η

$\Rightarrow$ Move along submodel $\eta \mapsto P_\eta$ $\Rightarrow$ Changes moment inequalities

$\Rightarrow$ Effect on set depends on constraint importance in shaping $\partial\Theta_0(P_\eta)$.

Efficiency Bound

Notation:

- $H(\theta) \equiv \nabla F(E[m(X,\theta)])$.
- $\Omega(\theta_1,\theta_2) \equiv E[(m(X,\theta_1) - E[m(X,\theta_1)])(m(X,\theta_2) - E[m(X,\theta_2)])']$.
- $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ such that $\theta^*(p) \in \arg\max_{\theta\in\Theta_0(P_0)}\langle p,\theta\rangle$ for all $p \in \mathbb{S}^{d_\theta}$.

Theorem: Under Assumption (A) and regularity conditions, we obtain:

$$I^{-1}(p_1,p_2) = \lambda(p_1,P_0)'H(\theta^*(p_1))\Omega(\theta^*(p_1),\theta^*(p_2))H(\theta^*(p_2))'\lambda(p_2,P_0)$$

In particular, efficiency bound for estimating $\nu(\bar{p},\Theta_0(P_0))$ at fixed $\bar{p}$ is:

$$\text{Var}\{\lambda(\bar{p},P_0)'\nabla F(E[m(X,\theta^*(\bar{p}))])m(X,\theta^*(\bar{p}))\}$$

i.e. “importance-weighted” linear combination of binding constraints.

Efficiency Bound

Notation:

- $H(\theta) \equiv \nabla F(E[m(X,\theta)])$.
- $\Omega(\theta_1,\theta_2) \equiv E[(m(X,\theta_1) - E[m(X,\theta_1)])(m(X,\theta_2) - E[m(X,\theta_2)])']$.
- $\theta^* : \mathbb{S}^{d_\theta} \to \Theta$ such that $\theta^*(p) \in \arg\max_{\theta\in\Theta_0(P_0)} \langle p,\theta\rangle$ for all $p \in \mathbb{S}^{d_\theta}$.

Theorem: Under Assumption (A) and regularity conditions, we obtain:

$$I^{-1}(p_1,p_2) = \lambda(p_1,P_0)'E[m(X,\theta^*(p_1))m(X,\theta^*(p_2))']\lambda(p_2,P_0)$$

In particular, efficiency bound for estimating $\nu(\bar{p},\Theta_0(P_0))$ at fixed $\bar{p}$ is:

$$\text{Var}\{\lambda(\bar{p},P_0)'m(X,\theta^*(\bar{p}))\}$$

i.e. “importance-weighted” linear combination of binding constraints.

Proof Outline

Step 1: Establish restrictions on P_0 do not affect tangent space $\dot{\mathbf{P}}$.

Step 2: Show that in a neighborhood of P_0 (in the τ-topology) for all $p \in \mathbb{S}^{d_\theta}$:

$$\nu(p, \Theta_0(Q)) = \sup_{\theta \in \Theta} \{\langle p, \theta \rangle + \lambda(p, Q)' F(\int m(x, \theta) dQ(x))\}$$

Step 3: For $s_0(X)$ the score of $\eta \mapsto P_\eta$, show pointwise in $p \in \mathbb{S}^{d_\theta}$ that:

$$\frac{\partial}{\partial \eta} \nu(p, \Theta_0(P_\eta))\Big|_{\eta=0} = \lambda(p, P_0)' \nabla F(E[m(X, \theta^*(p))]) E[m(X, \theta^*(p)) s_0(X)]$$

Step 4: Extend result to obtain weak pathwise derivative $\dot{\rho} : \dot{\mathbf{P}} \to \mathcal{C}(\mathbb{S}^{d_\theta})$:

$$\dot{\rho}(s_0)(p) = \lambda(p, P_0)' \nabla F(E[m(X, \theta^*(p))]) E[m(X, \theta^*(p)) s_0(X)]$$

Proof Outline (Regularity)

In General: For $\Lambda(p, P_0)$ (set of multipliers), $\Xi(p, P_0)$ (set of maximizers):

$$\frac{\partial}{\partial \eta_+}\nu(p, \Theta_0(P_\eta)) = \max_{\theta^* \in \Xi(p,P_0)} \min_{\lambda \in \Lambda(p,P_0)} \lambda' \nabla F(E[m(X, \theta^*)]) E[m(X, \theta^*) s_0(X)]$$

$$\frac{\partial}{\partial \eta_-}\nu(p, \Theta_0(P_\eta)) = \min_{\theta^* \in \Xi(p,P_0)} \max_{\lambda \in \Lambda(p,P_0)} \lambda' \nabla F(E[m(X, \theta^*)]) E[m(X, \theta^*) s_0(X)]$$

Intuition

- Multiple Lagrange multipliers implies some constraint is redundant.
  $\Rightarrow$ Constraints are smooth in $\eta \mapsto P_\eta$ but relevant ones switch at P_0

- Multiple maximizers implies you are on a "flat face" of identified set.
  $\Rightarrow$ In constraint $A\theta - F_L(\int m_L(x) dP_\eta)$, θ^* does not enter derivative

Efficient Estimator

Intuition

- In simple example $[E[Y_L], E[Y_U]]$ we use plug-in estimator $[\bar{Y}_L, \bar{Y}_U]$.
- Tangent set $\dot{\mathbf{P}}$ not restricted ... model is not overidentified.

$\Rightarrow$ Expect "plug-in" estimator to be semiparametrically efficient

Define: $\hat{P}_n$ to be the empirical distribution ($\hat{P}_n(x) = \frac{1}{n}\sum_i 1\{X_i = x\}$) and:

$$\nu(p, \Theta_0(\hat{P}_n)) = \sup_{\theta \in \Theta} \{\langle p, \theta\rangle \text{ s.t. } F(\frac{1}{n}\sum_{i=1}^{n} m(X_i, \theta)) \leq 0\}$$

$$= \sup_{\theta \in \Theta} \{\langle p, \theta\rangle + \lambda(p, \hat{P}_n)' F(\frac{1}{n}\sum_{i=1}^{n} m(X_i, \theta))\}$$

Efficient Estimator

Assumption (B): $\{X_i\}_{i=1}^n$ is an i.i.d. sample with $X_i \sim P_0$.

Theorem: Under Assumptions (A), (B) and regularity conditions:

- Part A: $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ is a regular estimator for $\nu(\cdot, \Theta_0(P_0))$.
- Part B: Uniformly in $p \in \mathbb{S}^{d_\theta}$ we obtain the expansion:

$$\sqrt{n}(\nu(p, \Theta_0(\hat{P}_n)) - \nu(p, \Theta_0(P_0)))$$
$$= \lambda(p, P_0)' H(\theta^*(p)) \frac{1}{\sqrt{n}} \sum_{i=1}^n \{m(X_i, \theta^*(p)) - E[m(X, \theta^*(p))]\} + o_p(1)$$

- Part C: For $\mathbb{G}_0$ a mean zero tight Gaussian process on $\mathcal{C}(\mathbb{S}^{d_\theta})$:

$$\sqrt{n}(\nu(\cdot, \Theta_0(\hat{P}_n)) - \nu(\cdot, \Theta_0(P_0))) \xrightarrow{L} \mathbb{G}_0$$

where $\mathbb{G}_0$ satisfies Cov$\{\mathbb{G}_0(p_1), \mathbb{G}_0(p_2)\} = I^{-1}(p_1, p_2)$.

Efficient Estimator

$$\sqrt{n}(\nu(p,\Theta_0(\hat{P}_n)) - \nu(p,\Theta_0(P_0)))$$
$$= \lambda(p,P_0)'H(\theta^*(p))\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{m(X_i,\theta^*(p)) - E[m(X,\theta^*(p))]\} + o_p(1)$$

Lagrange Multipliers

- $\lambda(\cdot,Q)$ uniquely determined for all Q in a neighborhood of P_0.
- $\lambda(p,Q)$ is jointly continuous in (p,Q).
- Stochastic equicontinuity of the process is not obvious ...

... but Lagrange multipliers and complementary slackness conditions "smooth out" the process.

Back in Θ ...

Note: $\{\nu(\cdot, \Theta_0(\hat{P}_n))\}$ is identified with convex set $\hat{\Theta}_n = co\{\Theta_0(\hat{P}_n)\}$.

Theorem Let Assumptions (A), (B) and regularity conditions hold.

(i) $L : \mathbf{R}_+ \to \mathbf{R}_+$ is a nondecreasing continuous function

(ii) $L(0) = 0$ and $L(a) \leq Ma^\kappa$ for some M, κ and all $a \in \mathbf{R}_+$

If $\{K_n\}$ is a regular convex compact valued set estimator for $\Theta_0(P_0)$, then:

$$\liminf_{n\to\infty} E[L(\sqrt{n} d_H(K_n, \Theta_0(P_0)))] \geq \limsup_{n\to\infty} E[L(\sqrt{n} d_H(\hat{\Theta}_n, \Theta_0(P_0)))] = E[L(\|\mathbb{G}_0\|_\infty)]$$

Comments

- Lower bound holds without continuity of L, but attainment may not.
- Can be relaxed to $L(a) \leq M\exp(a\kappa)$ for limited values of κ.

Bootstrap

Problem: How do we obtain consistent bootstrap for the distribution of $\mathbb{G}_0$?

Approach: Perturb estimator of influence function ...

Definition: For random weights $\{W_i\}_{i=1}^n$ define G_n^* process pointwise by:

$$\lambda(p, \hat{P}_n)'\nabla F(\frac{1}{n}\sum_{i=1}^n m(X_i, \hat{\theta}(p)))\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\{m(X_i, \hat{\theta}(p)) - \frac{1}{n}\sum_{i=1}^n m(X_i, \hat{\theta}(p))\}$$

where $\hat{\theta} : \mathbb{S}^{d_\theta} \to \Theta$ satisfies $\hat{\theta}(p) \in \arg\max_{\theta \in \Theta_0(\hat{P}_n)} \langle p, \theta \rangle$ for all $p \in \mathbb{S}^{d_\theta}$.

Why should this work?

- If $W_i \perp X_i$, expect to converge to efficient influence function.
- Law of G_n^* conditional on $\{X_i\}_{i=1}^n$ (but not $\{W_i\}_{i=1}^n$) consistent for $\mathbb{G}_0$.

Bootstrap

Assumption (C): $W \perp X$ with $E[W] = 0$, $E[W^2] = 1$ and $E[|W|^{2+\delta}] < \infty$.

Theorem If Assumptions (A), (B), (C) and regularity conditions hold, then:

$$G_n^* \stackrel{L^*}{\to} \mathbb{G}_0$$

(in prob.), where L^* denotes Law conditional on $\{X_i\}_{i=1}^n$ (but not $\{W_i\}_{i=1}^n$).

Comments:

- For example, $W \sim N(0,1)$ or W Rademacher.
- No need to recompute support function or estimate covariance kernel.

Critical Values

Let $\Psi_0 \subseteq \mathbb{S}^{d_\theta}$ and $\Upsilon : \mathbf{R} \to \mathbf{R}$. Critical values often are $1 - \alpha$ quantile of:

$$\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p))$$

Example 1: Let $\Psi_0 = \mathbb{S}^{d_\theta}$ and $\Upsilon(a) = |a|$. We need quantiles of:

$$\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p)) = \sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)|$$

Example 2: Let $\Psi_0 = \mathbb{S}^{d_\theta}$ and $\Upsilon(a) = |-a|_+$. We need quantiles of:

$$\sup_{p \in \Psi_0} \Upsilon(\mathbb{G}_0(p)) = \sup_{p \in \mathbb{S}^{d_\theta}} |-\mathbb{G}_0(p)|_+$$

Critical Values

Algorithm

Step 1: Compute $\nu(\cdot, \Theta_0(\hat{P}_n))$ to obtain $p \mapsto \lambda(p, \hat{P}_n)$ and $p \mapsto \hat{\theta}(p)$.

Step 2: Draw $\{W_i\}_{i=1}^n$ to construct G_n^*.

Step 3: Given Hausdorff consistent estimate $\hat{\Psi}_n$ for Ψ_0 define:

$$\hat{c}_{1-\alpha} \equiv \inf\{c : P(\sup_{p \in \hat{\Psi}_n} \Upsilon(G_n^*(p)) \leq c \,|\{X_i\}_{i=1}^n) \geq 1 - \alpha\}$$

Theorem: Under Assumptions (A), (B), (C) and regularity conditions:

$$\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}$$

One Sided Region

$$c^{(1)}_{1-\alpha} \equiv \inf\{c : P(\sup_{p\in\mathbb{S}^{d_\theta}} |-\mathbb{G}_0(p)|_+ \leq c) \geq 1-\alpha\}$$

Theorem: Under Assumptions (A), (B), (C) and regularity conditions:

$$\lim_{n\to\infty} P(\Theta_0(P_0) \subseteq \hat{\Theta}_n^{\hat{c}^{(1)}_{1-\alpha}/\sqrt{n}}) = 1-\alpha$$

Comments:

- Find $\hat{\Theta}_n^{\hat{c}^{(1)}_{1-\alpha}/\sqrt{n}}$ from its support function $\{\nu(\cdot,\Theta_0(\hat{P}_n)) + \hat{c}^{(1)}_{1-\alpha}/\sqrt{n}\}$.
- Test inversion of $H_0 : K \subseteq \Theta_0(P_0)$ using $T_n(K) \equiv \sqrt{n}\,\overrightarrow{d}_H(K,\hat{\Theta}_n)$.
- Duality first exploited in Beresteanu $\&$ Molinari (2008).

Two Sided Region

$$c_{1-\alpha}^{(2)} \equiv \inf\{c : P(\sup_{p \in \mathbb{S}^{d_\theta}} |\mathbb{G}_0(p)| \leq c) \geq 1-\alpha\}$$

Theorem: Under Assumptions (A), (B), (C) and regularity conditions:

$$\lim_{n\to\infty} P(\hat{\Theta}_n^{-\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}} \subseteq \Theta_0(P_0) \subseteq \hat{\Theta}_n^{\hat{c}_{1-\alpha}^{(2)}/\sqrt{n}}) = 1-\alpha$$

Comments:

- Provides uniform confidence interval for $\partial\Theta_0(P_0)$.
- Test inversion of $H_0 : K = \Theta_0(P_0)$ using $T_n(K) \equiv \sqrt{n} d_H(K, \hat{\Theta}_n)$.

Region for Parameter

$$\inf_{\theta \in \Theta_0(P_0)} \liminf_{n \to \infty} P(\theta \in \mathcal{P}_n) \geq 1 - \alpha$$

Standard Approach: Build $\mathcal{P}_n$ through test inversion of hypothesis:

$$H_0 : \theta \in \Theta_0(P_0) \qquad\qquad H_1 : \theta \notin \Theta_0(P_0)$$

Test Statistic: Use the efficient estimator to test this null hypothesis by:

$$H_n(\theta) \equiv \sqrt{n}\, \overrightarrow{d}_H(\{\theta\}, \hat{\Theta}_n)$$

Region for Parameter

Definition: Let $\mathfrak{M}(\theta)$ be set of maximizers of $p \mapsto \{\nu(p,\{\theta\}) - \nu(p,\Theta_0(P_0))\}$.

$$c_{1-\alpha}(\theta) \equiv \inf\{c : P(\sup_{p\in\mathfrak{M}(\theta)} |-\mathbb{G}_0(p)|_+ \leq c) \geq 1-\alpha\}$$

Note: Bootstrap with Hausdorff consistent estimate for $\mathfrak{M}(\theta)$ (Kaido 2010).

Theorem: Under Assumptions (A), (B), (C) and regularity conditions:

$$\inf_{\theta\in\Theta_0(P_0)} \liminf_{n\to\infty} P(\theta \in \hat{\mathcal{P}}_n) \geq 1-\alpha$$

where the confidence region is given by $\hat{\mathcal{P}}_n \equiv \{\theta \in \Theta : H_n(\theta) \leq \hat{c}_{1-\alpha}(\theta)\}$.

Local Power

Power Functions

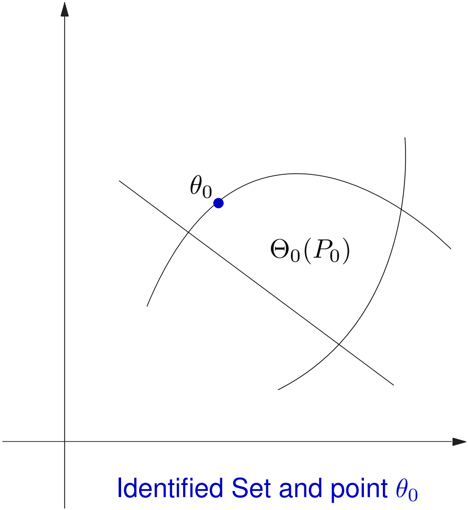
- Let $\pi_n(P_\eta;\theta_0)$ be probability test rejects $H_0:\theta_0\in\Theta_0(P_\eta)$ when $X\sim P_\eta$.
- Denote $\pi_n^*(P_\eta;\theta_0)$ for test that rejects when $H_n(\theta_0)>\hat{c}_{1-\alpha}(\theta_0)$.

Goal: Compare power functions along local parametric submodels.

Definition: For $\theta_0\in\partial\Theta_0(P_0)$ let $\mathbf{H}(\theta_0)$ be set of submodels with:

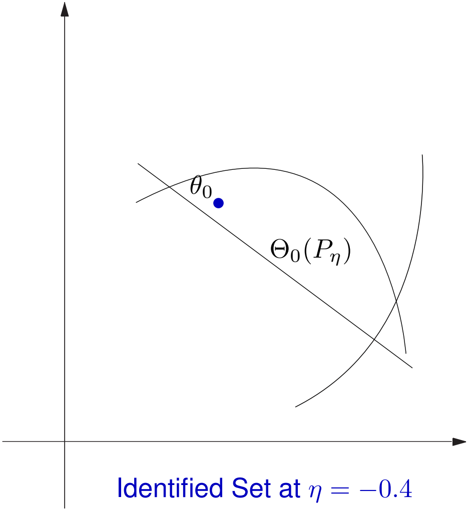
1. If $\eta\le 0$ then $\theta_0\in\Theta_0(P_\eta)$.
2. If $\eta>0$ then $\theta_0\notin\Theta_0(P_\eta)$.

Local Path



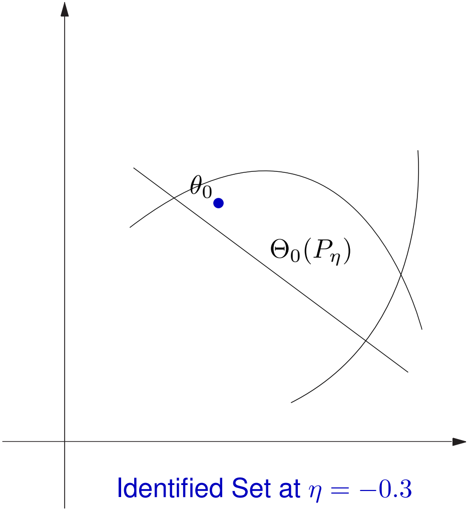
Identified Set and point θ_0

Local Path



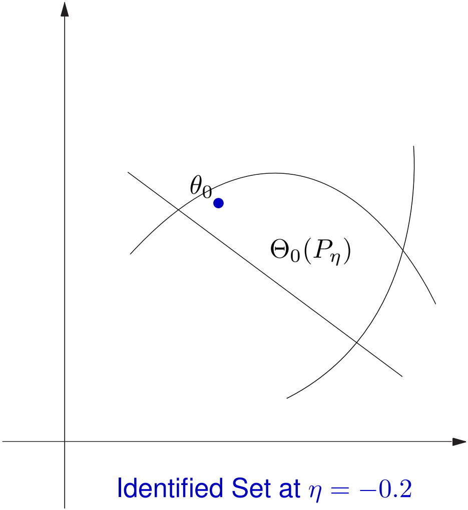
Identified Set at $\eta = -0.4$

Local Path



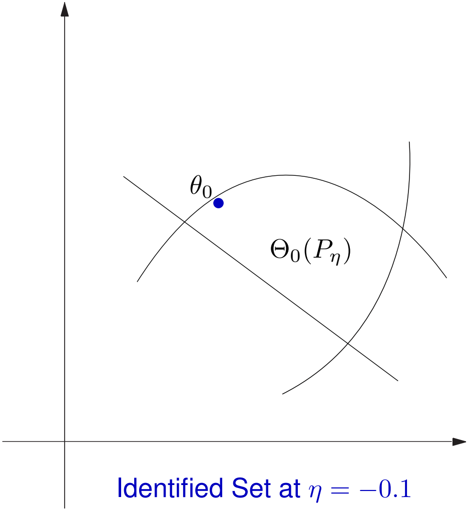
Identified Set at $\eta = -0.3$

Local Path



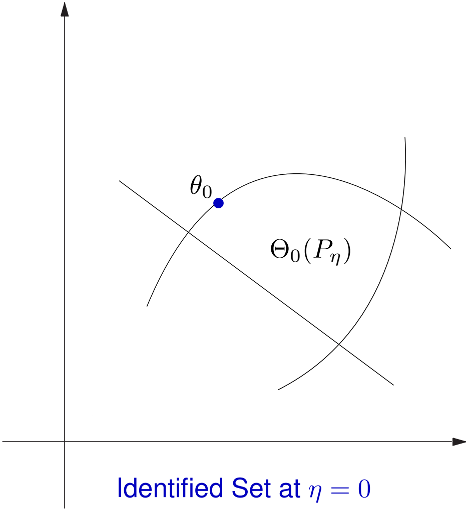
Identified Set at $\eta = -0.2$

Local Path



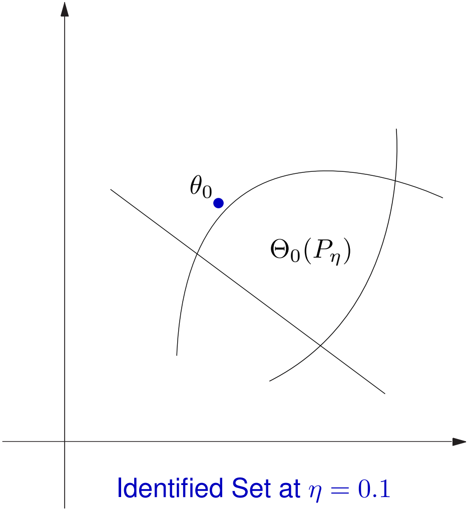
Identified Set at $\eta = -0.1$

Local Path



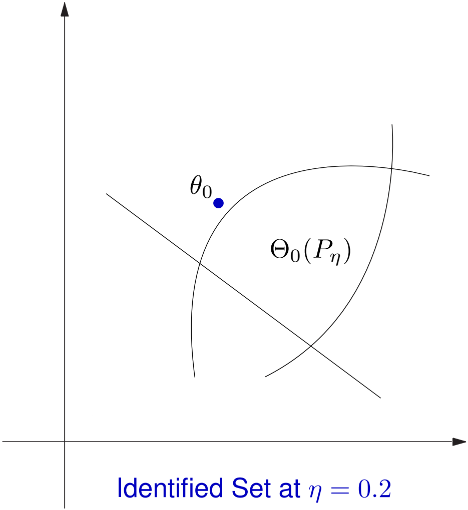
Identified Set at $\eta = 0$

Local Path



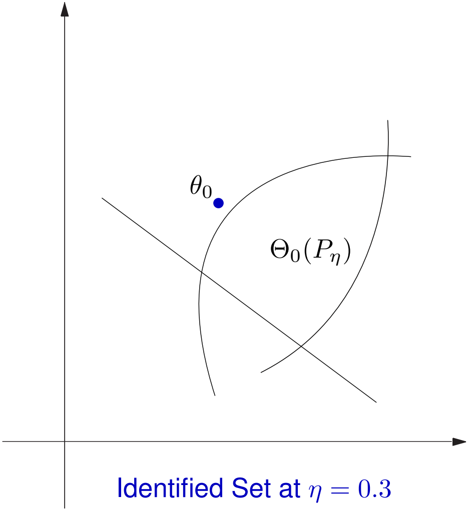
Identified Set at $\eta = 0.1$

Local Path



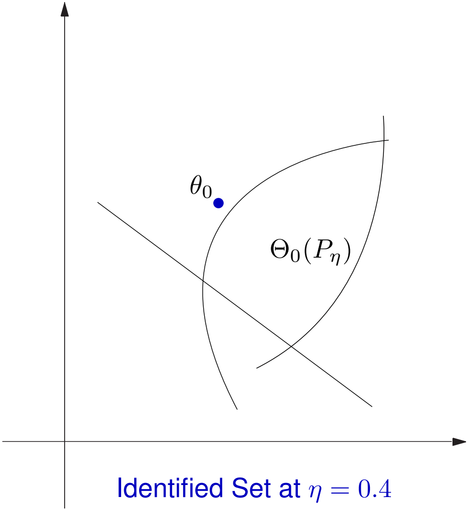
Identified Set at $\eta = 0.2$

Local Path



Identified Set at $\eta = 0.3$

Local Path



Identified Set at $\eta = 0.4$

Local Power

Theorem: Let Assumptions (A), (B), (C) and regularity conditions hold and

$$\limsup_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}};\theta_0) \leq \alpha$$

for every $P_\eta \in \mathbf{H}(\theta_0)$ and $\eta < 0$. If $\mathfrak{M}(\theta_0) = \{p_0\}$, then for any $P_\eta \in \mathbf{H}(\theta_0)$

$$\limsup_{n\to\infty} \pi_n(P_{\eta/\sqrt{n}};\theta_0) \leq \lim_{n\to\infty} \pi_n^*(P_{\eta/\sqrt{n}};\theta_0) = 1 - \Phi\Big(z_{1-\alpha} - \eta\frac{E[\tilde{l}(X)s_0(X)]}{\sqrt{E[\mathbb{G}_0^2(p_0)]}}\Big)$$

where $s_0(x)$ is the score of $\eta \mapsto P_\eta$ and $\tilde{l}(x) = -\lambda(p_0, P_0)'H(\theta_0)m(x,\theta_0)$.

Comments:

- Applies to θ_0 not at kink of boundaries.
- $P_\eta \in \mathbf{H}(\theta_0)$ if and only if $E[\tilde{l}(X)s_0(X)] > 0$.
- Weak "size control" requirement ... locality of semiparametric efficiency.

Subvectors

Suppose $\Theta = \Theta_1 \times \Theta_2$ with $\Theta_1 \subset \mathbf{R}^{d_{\theta_1}}$, $\Theta_2 \subset \mathbf{R}^{d_{\theta_2}}$ and $\theta = (\theta_1, \theta_2)$.

$$\Theta_{0,M}(P_0) \equiv \{\theta_1 \in \Theta_1 : (\theta_1, \theta_2) \in \Theta_0(P_0) \text{ for some } \theta_2 \in \Theta_2\}$$

Key: For $p_1 \in \mathbb{S}^{d_{\theta_1}}$ and $(p_1, p_2) = p \in \mathbb{S}^{d_{\theta_1}+d_{\theta_2}}$, it follows that:

$$\begin{aligned}\nu(p_1, \Theta_{0,M}(P_0)) &= \sup_{\theta_1 \in \Theta_{0,M}(P_0)} \langle p_1, \theta_1 \rangle \\ &= \sup_{(\theta_1,\theta_2) \in \Theta_0(P_0)} \{\langle p_1, \theta_1 \rangle + \langle 0, \theta_2 \rangle\} = \nu((p_1, 0), \Theta_0(P_0))\end{aligned}$$

$\Rightarrow$ The efficient estimator for $\nu(\cdot, \Theta_{0,M}(P_0))$ is just $\nu((\cdot, 0), \Theta_0(\hat{P}_n))$.

$\Rightarrow$ All our results apply to the identified set $\Theta_{0,M}$ as well.

Regression with Interval Outcome

- Let $\epsilon_i \sim N(0,1)$ and Y_i (unobservable) be generated according to:

$$Y_i = Z_i'\theta_0 + \epsilon_i$$

where $Z_i = (1, Z_{i,2})$, $Z_{i,2}$ uniform on K equally spaced points in $[-5, 5]$.

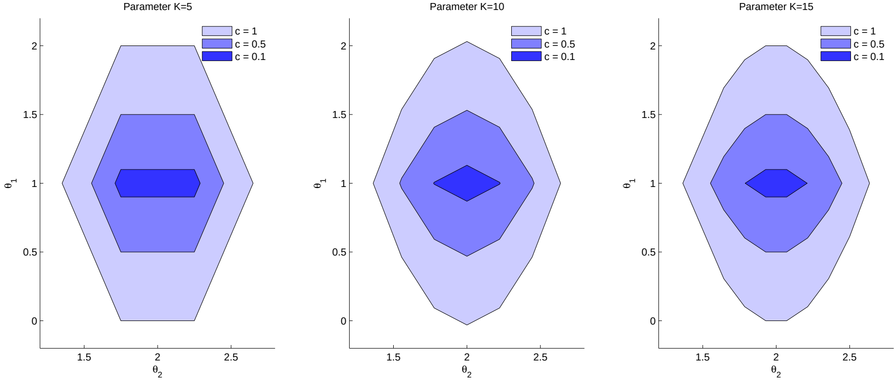
- For $V_i \sim U[0, 0.2]$ independent of (Y_i, Z_i) create $Y_{L_i} \leq Y \leq Y_{U,i}$ by:

$$Y_{L,i} = Y_i - C + V_i Z_i^2$$
$$Y_{U,i} = Y_i + C + V_i Z_i^2$$

Design Parameters:

- C controls the diameter of the identified set (identification at $C = 0$).
- K controls severity of "intersection bounds" problem.

Figure: Identified Set as a Function of C and K



Expected Hausdorff Distance

Table: Average $d_H(\hat{\Theta}_n, \Theta_0(P_0))$

		C		
	K	0.1	0.5	1
$n = 200$	5	0.153	0.150	0.151
	10	0.275	0.250	0.250
	15	0.514	0.361	0.359
$n = 500$	5	0.094	0.094	0.094
	10	0.177	0.155	0.155
	15	0.360	0.218	0.219
$n = 1,000$	5	0.066	0.066	0.066
	10	0.130	0.109	0.109
	15	0.201	0.154	0.154

Expected Inner Hausdorff Distance

Table: Average $\vec{d}_H(\hat{\Theta}_n, \Theta_0(P_0))$

	K	C 0.1	0.5	1
$n = 200$	5	0.137	0.144	0.144
	10	0.106	0.105	0.108
	15	0.308	0.045	0.047
$n = 500$	5	0.088	0.090	0.090
	10	0.085	0.093	0.094
	15	0.043	0.051	0.054
$n = 1,000$	5	0.063	0.064	0.064
	10	0.075	0.081	0.081
	15	0.061	0.055	0.053

Expected Outer Hausdorff Distance

Table: Average $\vec{d}_H(\Theta_0(P_0), \hat{\Theta}_n)$

		C		
	K	0.1	0.5	1
$n = 200$	5	0.150	0.145	0.145
	10	0.273	0.249	0.250
	15	0.296	0.359	0.359
$n = 500$	5	0.092	0.090	0.090
	10	0.185	0.154	0.154
	15	0.360	0.218	0.219
$n = 1,000$	5	0.064	0.064	0.064
	10	0.130	0.108	0.108
	15	0.185	0.154	0.154

One Sided Confidence Interval

Table: Nominal Level 0.95

		C		
	K	0.1	0.5	1
$n = 200$	5	0.946	0.944	0.942
	10	0.988	0.915	0.900
	15	0.995	0.896	0.824
$n = 500$	5	0.953	0.941	0.941
	10	0.958	0.912	0.910
	15	0.896	0.886	0.877
$n = 1,000$	5	0.962	0.950	0.950
	10	0.926	0.921	0.921
	15	0.978	0.893	0.891

Two Sided Confidence Interval

Table: Nominal Level 0.95

	K	C = 0.1	0.5	1
$n = 200$	5	0.958	0.945	0.943
	10	0.996	0.953	0.936
	15	0.909	0.948	0.886
$n = 500$	5	0.964	0.942	0.942
	10	0.983	0.953	0.951
	15	0.949	0.938	0.931
$n = 1,000$	5	0.971	0.949	0.949
	10	0.963	0.952	0.952
	15	0.987	0.892	0.939

Conclusion

Semiparametric Efficiency

- Proposed a notion of semiparametric efficiency.
- Characterized sources of irregularity in higher dimensions.
- Derived the semiparametric efficiency bound.

Efficient Estimation

- Showed “plug-in” estimator is efficient.
- Obtained consistent bootstrap procedure.
- Employed efficient estimator to construct confidence regions.

Challenges

- Efficiency is local concept, often more uniformity is desired.
- Sensitivity to “intersection bound” problems in higher dimensions?
- Different use of efficient estimator (Imbens & Manski (2004)).
- Other efficient estimators may “behave” better ... Sieve MLE? EL?