Overidentification in Regular Models

Xiaohong Chen
Yale University

Andres Santos
UC San Diego

March 9, 2016

Starting Point

$$Y = h_P(Z) + \epsilon$$

where h_P is nonparametric, and $E[\epsilon|W] = 0$ for an available instrument W.

Starting Point

$$Y = h_P(Z) + \epsilon$$

where h_P is nonparametric, and $E[\epsilon|W] = 0$ for an available instrument W.

Identification is well understood, is equivalent to a unique solution in h to

$$E[Y|W] = E[h(Z)|W]$$

Starting Point

$$Y = h_P(Z) + \epsilon$$

where h_P is nonparametric, and $E[\epsilon|W] = 0$ for an available instrument W.

Identification is well understood, is equivalent to a unique solution in h to

$$E[Y|W] = E[h(Z)|W]$$

Overidentification

- Original Question: Is this model overidentified or just identified?
- Broader Question: ◦ What do we actually mean by "overidentified"?
  ◦ How do we characterize "overidentification"?

First Definition?

Koopmans & Reiersøl, 1950.

- "This particular specification will be called observationally restrictive if the set of all distribution functions of observed variables generated by the structures is a proper subset of the set of all distribution functions."

First Definition?

Koopmans & Reiersøl, 1950.

- "This particular specification will be called observationally restrictive if the set of all distribution functions of observed variables generated by the structures is a proper subset of the set of all distribution functions."

- "A frequent case of an observationally restrictive specification is that where a parameter ... is restricted [by the structure] to a prescribed value. In this case, the specification in question has been called overidentifying."

First Definition?

Koopmans & Reiersøl, 1950.

- "This particular specification will be called observationally restrictive if the set of all distribution functions of observed variables generated by the structures is a proper subset of the set of all distribution functions."

- "A frequent case of an observationally restrictive specification is that where a parameter ... is restricted [by the structure] to a prescribed value. In this case, the specification in question has been called overidentifying."

Comments

- First definition: Discussed in context of potential refutability.
- But authors warn, not sufficient for testability (Romano, 2004).
- Second definition: Related to estimation and Hausman test.
- But "overidentifying" stronger than "observationally restrictive".

Is "overidentification" a useful concept?

Yes

- Word "overidentification" often associated with testability of the model.
  Examples: Chesher (2003), Matzkin (2003), Florens et. al. (2007).
- Word "overidentification" also associated with efficient estimation.
  Plug in Estimators: Newey (1994), Powell (1994).
  Two Stage: Newey & Powell (1999), Ackerberg et. al. (2014).

Is "overidentification" a useful concept?

Yes

- Word "overidentification" often associated with testability of the model.
  Examples: Chesher (2003), Matzkin (2003), Florens et. al. (2007).
- Word "overidentification" also associated with efficient estimation.
  Plug in Estimators: Newey (1994), Powell (1994).
  Two Stage: Newey & Powell (1999), Ackerberg et. al. (2014).

But

- What do we do in nonparametric IV model?
  $\Rightarrow$ Need precise definition of "overidentification" for general models.
  $\Rightarrow$ Should be intrinsically linked to testing and efficiency.

Overidentification in GMM

Let $\Theta \subseteq \mathbf{R}^{d_\beta}$, $X_i \in \mathbf{R}^{d_x}$, $\rho : \mathbf{R}^{d_x} \times \mathbf{R}^{d_\beta} \to \mathbf{R}^{d_\rho}$ with $d_\beta \leq d_\rho$ and suppose

$$\int \rho(X_i, \beta(P))dP = 0 \,.$$

Overidentification

- When is an overidentification test available? When $d_\rho > d_\beta$.
- When are efficiency considerations relevant? When $d_\rho > d_\beta$.
- Overidentification $\Longleftrightarrow d_\rho > d_\beta$.

Counting

- Counting intuition a widely used notion of overidentification.
- Stronger than "observationally restrictive".
- Not helpful in nonparametric instrumental variables.

Aim of Paper

The Literature

- The term overidentification is used in different ways in the literature.
- What is the precise definition that captures these ideas?

Aim of Paper

The Literature

- The term overidentification is used in different ways in the literature.
- What is the precise definition that captures these ideas?

Our Answer

- Introduce a simple condition we call local overidentification.
- Show it is equivalent to existence of more efficient estimators.
- Show it is equivalent to local testability of the model.

Implications

- Establish intrinsic link between efficiency and testability.
- Apply to conditional moment restrictions models.
- Apply to two stage and plug-in estimators.

Setup

We assume data consists of i.i.d. sample $\{X_i\}_{i=1}^n$, $X_i \in \mathbf{R}^{d_x}$, and $X_i \sim P$.

Setup

We assume data consists of i.i.d. sample $\{X_i\}_{i=1}^n$, $X_i \in \mathbf{R}^{d_x}$, and $X_i \sim P$.

Definition: A model $\mathbf{P}$ is a subset of the set of distributions on $\mathbf{R}^{d_x}$.

Setup

We assume data consists of i.i.d. sample $\{X_i\}_{i=1}^n$, $X_i \in \mathbf{R}^{d_x}$, and $X_i \sim P$.

Definition: A model $\mathbf{P}$ is a subset of the set of distributions on $\mathbf{R}^{d_x}$.

Definition: A "path" $t \mapsto P_{t,g}$ with $P_{t,g}$ a probability measure on $\mathbf{R}^{d_x}$ and

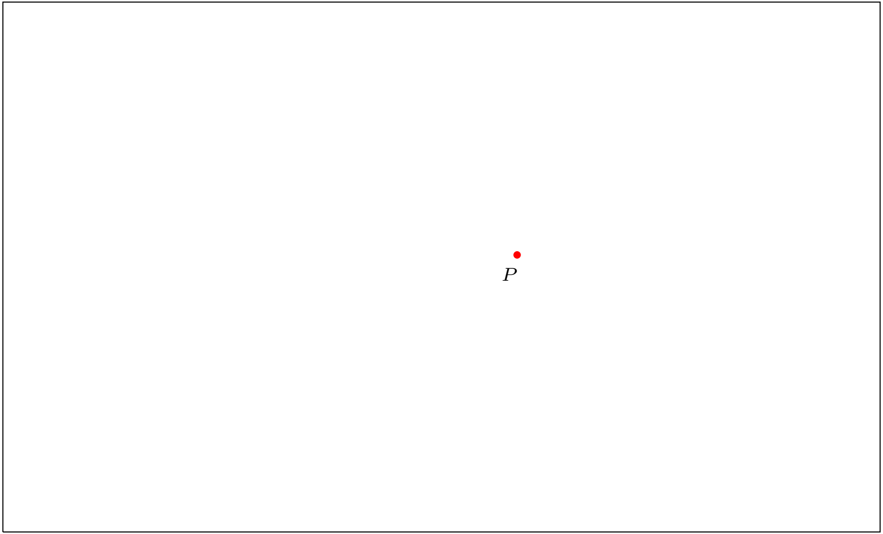
$$\lim_{t \to 0} \int [\frac{1}{t}(dP_{t,g}^{1/2} - dP^{1/2}) - \frac{1}{2} g dP^{1/2}]^2 = 0 \ .$$

The function $g : \mathbf{R}^{d_x} \to \mathbf{R}$ is referred to as the "score" of the path $t \mapsto P_{t,g}$.

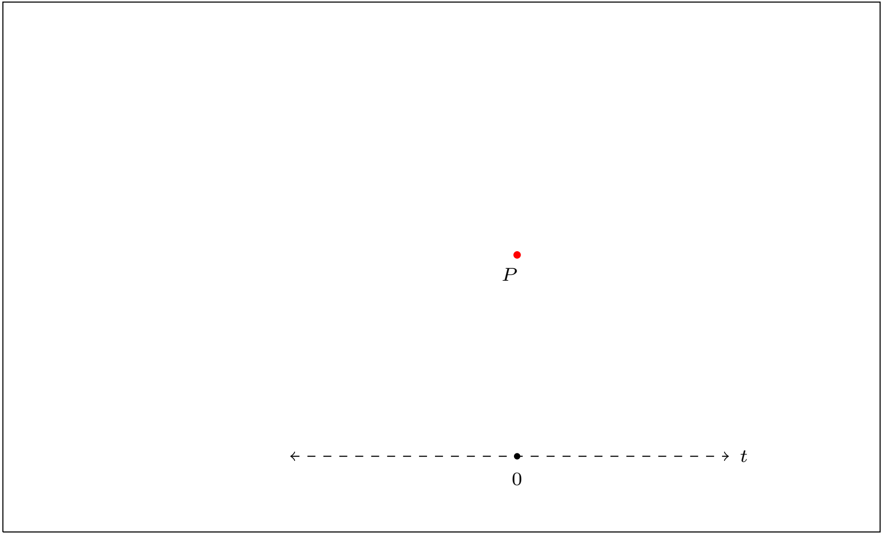
Comments

- $t \mapsto P_{t,g}$ is a smooth correctly specified likelihood ($P_{0,g} = P$).
- As usual, score g has mean zero and finite second moment.
- Only feature will matter to us is the score g.

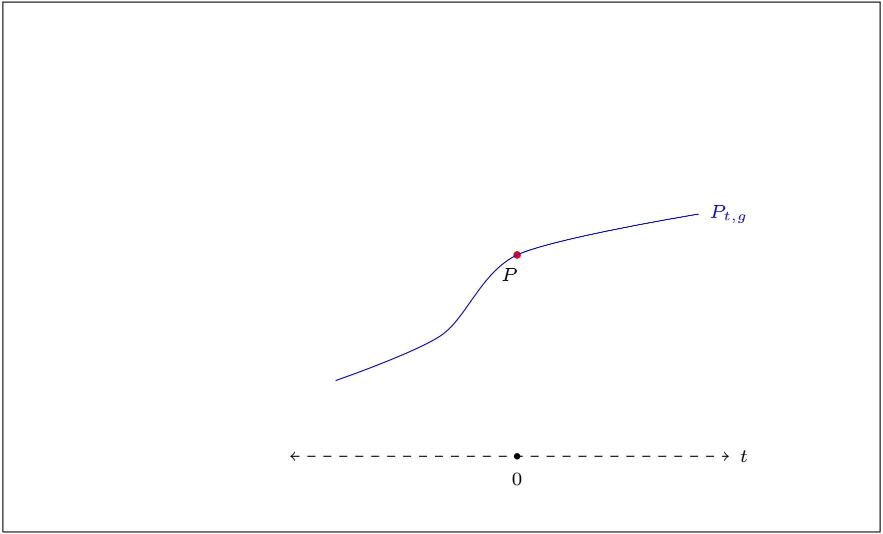
Space of Distributions



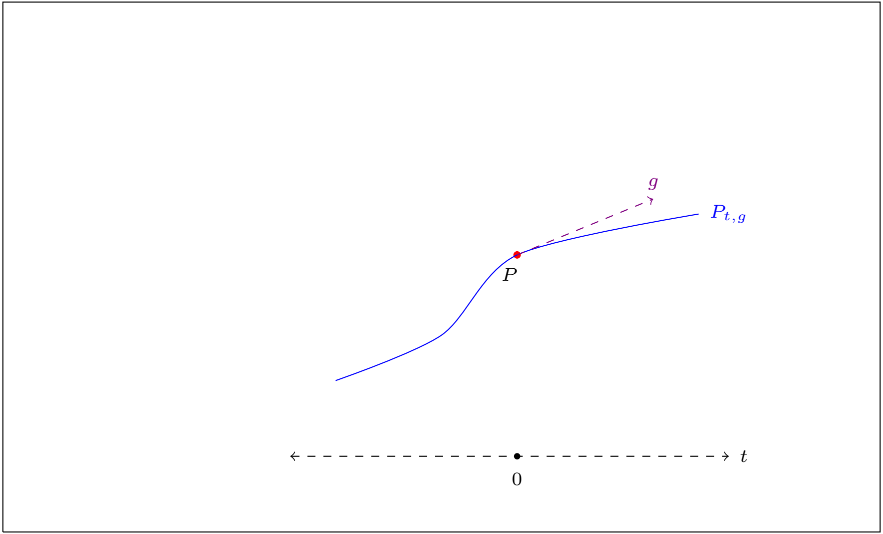
Space of Distributions



Space of Distributions



Space of Distributions



Tangent Space

$$L_0^2 \equiv \{g : \int g dP = 0 \text{ and } \int g^2 dP < \infty\}$$

Possible to show for any $g \in L_0^2$ we can find a path $t \mapsto P_{t,g}$ with score g.

Intuition: $\int g dP = 0$ is only restriction following from $P_{t,g}$ a measure.

Tangent Space

$$L_0^2 \equiv \{g : \int g dP = 0 \text{ and } \int g^2 dP < \infty\}$$

Possible to show for any $g \in L_0^2$ we can find a path $t \mapsto P_{t,g}$ with score g.

Intuition: $\int g dP = 0$ is only restriction following from $P_{t,g}$ a measure.

Question: What about information contained in model when $P \in \mathbf{P}$?

Tangent Space

$$L_0^2 \equiv \{g : \int g dP = 0 \text{ and } \int g^2 dP < \infty\}$$

Possible to show for any $g \in L_0^2$ we can find a path $t \mapsto P_{t,g}$ with score g.

Intuition: $\int g dP = 0$ is only restriction following from $P_{t,g}$ a measure.

Question: What about information contained in model when $P \in \mathbf{P}$?

- The tangent space is the set of scores that “agree” with the model $\mathbf{P}$

$$\bar{T}(P) \equiv \text{cl}\{g \in L_0^2 : g \text{ is score of some } t \mapsto P_{t,g} \in \mathbf{P}\}$$

Tangent Space

$$L_0^2 \equiv \{g : \int g dP = 0 \text{ and } \int g^2 dP < \infty\}$$

Possible to show for any $g \in L_0^2$ we can find a path $t \mapsto P_{t,g}$ with score g.

Intuition: $\int g dP = 0$ is only restriction following from $P_{t,g}$ a measure.

Question: What about information contained in model when $P \in \mathbf{P}$?

- The tangent space is the set of scores that "agree" with the model $\mathbf{P}$

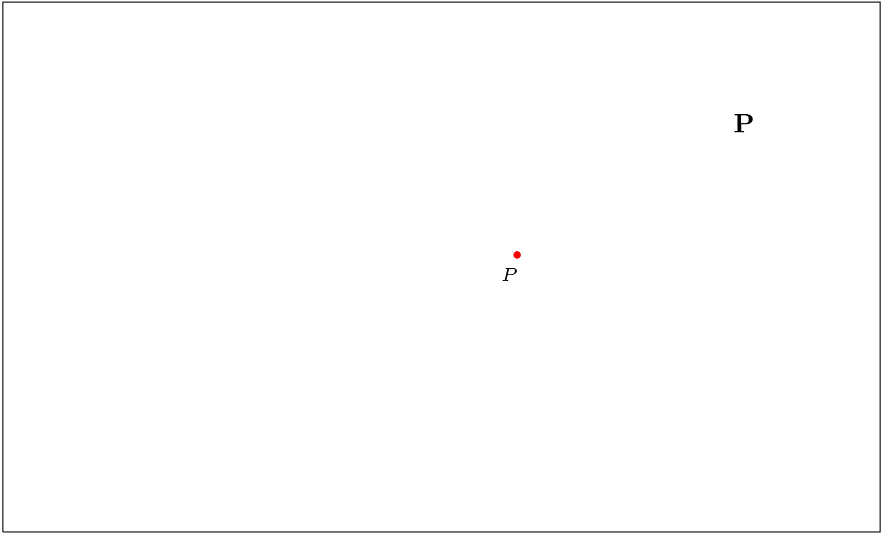
$$\bar{T}(P) \equiv \text{cl}\{g \in L_0^2 : g \text{ is score of some } t \mapsto P_{t,g} \in \mathbf{P}\}$$

- The orthocomplement of $\bar{T}(P)$ are scores that "disagree" with $\mathbf{P}$

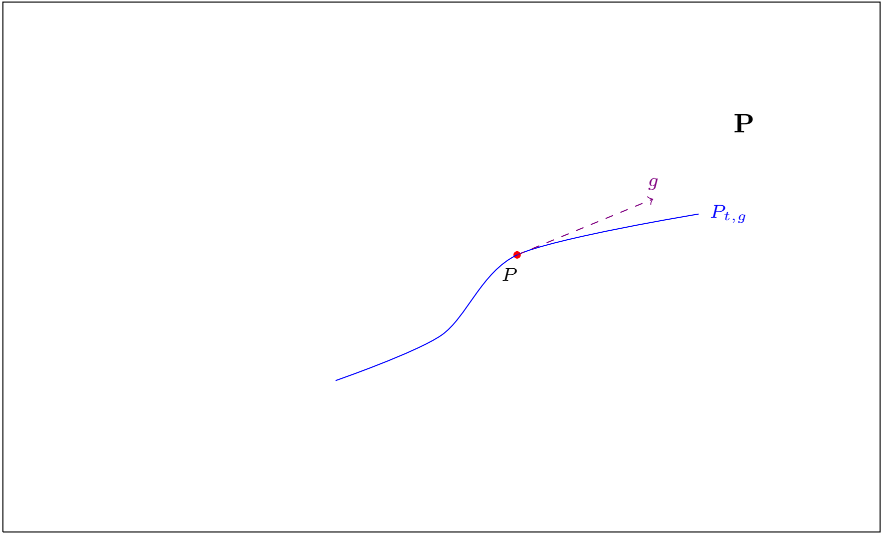
$$\bar{T}(P)^\perp \equiv \{g \in L_0^2 : \int g f dP = 0 \text{ for all } f \in \bar{T}(P)\}$$

Note: $\bar{T}(P)$ and $\bar{T}(P)^\perp$ decompose the set of all possible scores.

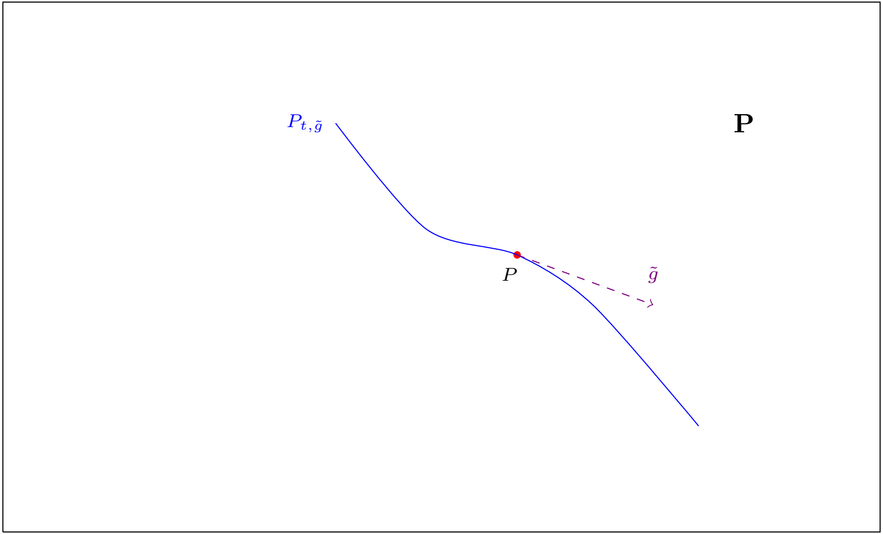
$\mathbf{P}$ = All Distributions



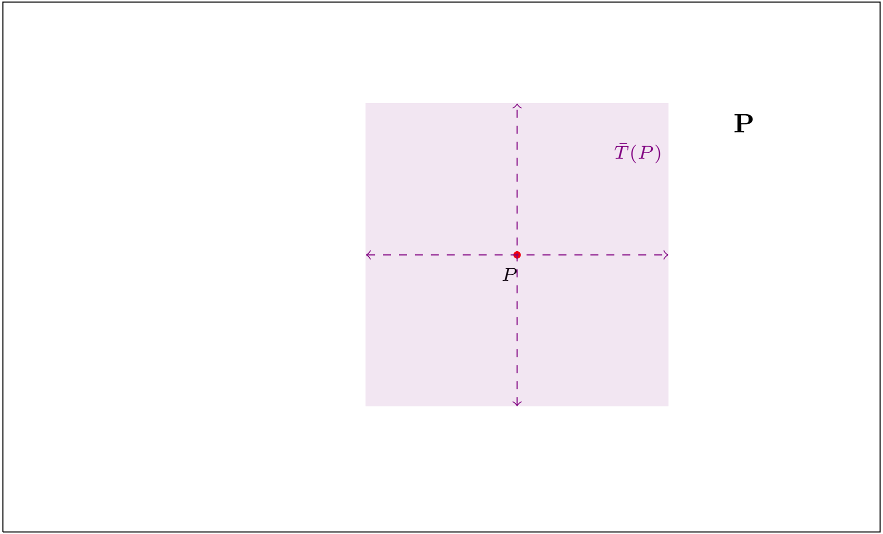
$\mathbf{P}$ = All Distributions



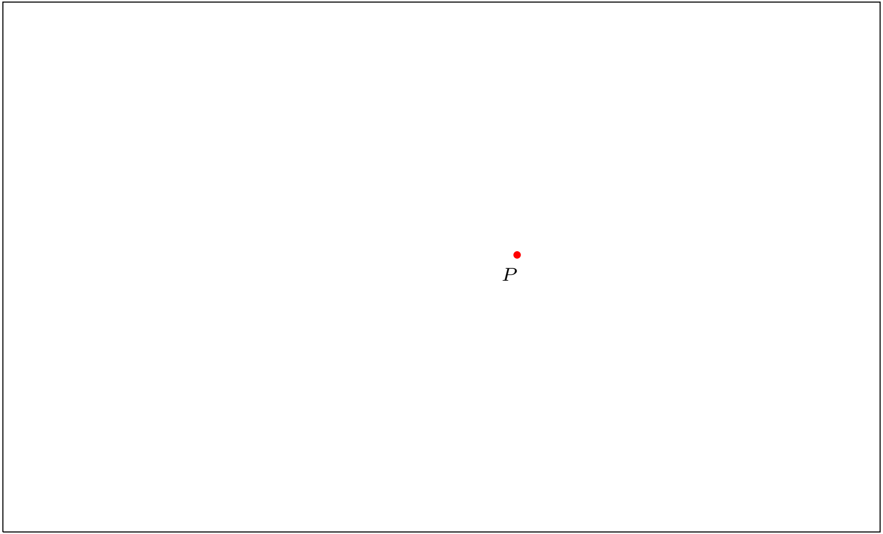
$\mathbf{P}$ = All Distributions



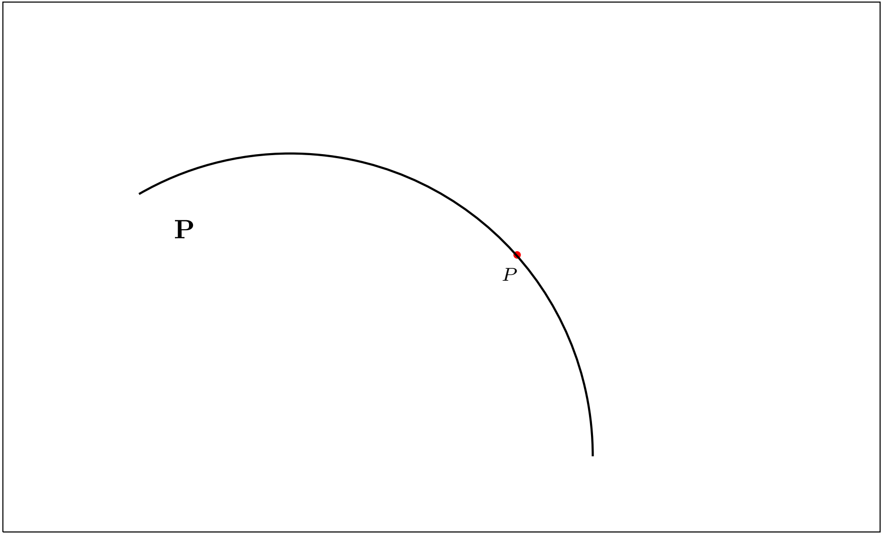
$\mathbf{P}$ = All Distributions



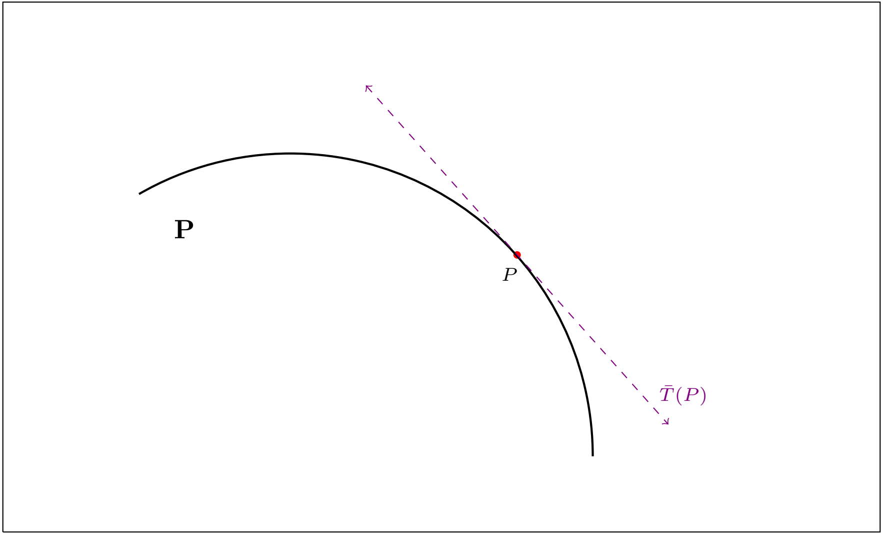
Informative P



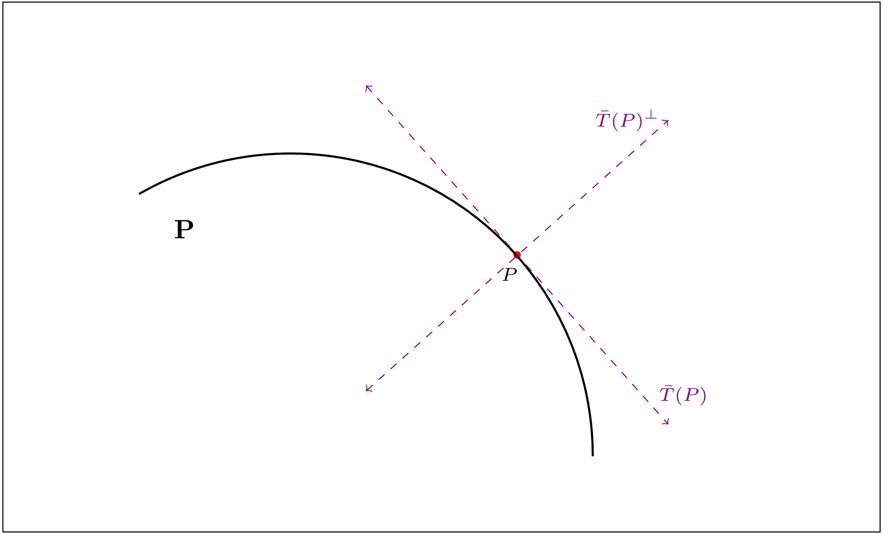
Informative $\mathbb{P}$



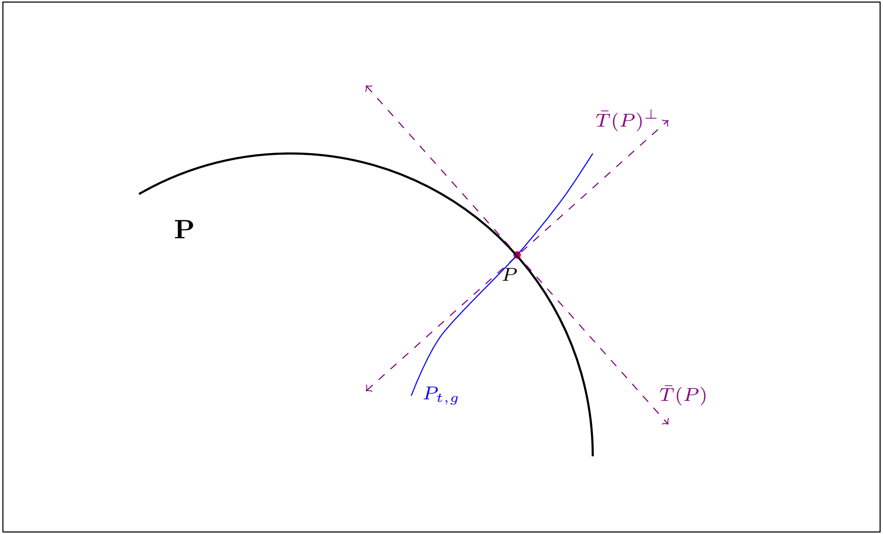
Informative $\mathbb{P}$



Informative $\mathbb{P}$



Informative $\mathbb{P}$



Local Overidentification

Assumption (R)

- $\{X_i\}_{i=1}^n$ is i.i.d. with $X_i \sim P_{1/\sqrt{n},g}$ for some path $P_{t,g}$ with $P_{0,g} = P \in \mathbf{P}$.
- $\bar{T}(P)$ is linear – i.e. if $g, f \in \bar{T}(P)$, $a, b \in \mathbf{R}$, then $ag + bf \in \bar{T}(P)$.

Local Overidentification

Assumption (R)

- $\{X_i\}_{i=1}^n$ is i.i.d. with $X_i \sim P_{1/\sqrt{n},g}$ for some path $P_{t,g}$ with $P_{0,g} = P \in \mathbf{P}$.
- $\bar{T}(P)$ is linear – i.e. if $g, f \in \bar{T}(P)$, $a, b \in \mathbf{R}$, then $ag + bf \in \bar{T}(P)$.

Main Definition

- If $\bar{T}(P) = L_0^2$ then we say P is locally just identified by $\mathbf{P}$.
- If $\bar{T}(P) \subsetneq L_0^2$ then we say P is locally overidentified by $\mathbf{P}$.

Intuition

- P is just identified $\Leftrightarrow$ $\mathbf{P}$ locally consistent with any parametric model.
- P is overidentified $\Leftrightarrow$ $\mathbf{P}$ restricts possible parametric specifications.

Note: Reduces to traditional definition in GMM context.

Setup

$$H_0 : P \in \mathbf{P} \qquad\qquad H_1 : P \notin \mathbf{P}$$

Setup

$$H_0 : P \in \mathbf{P} \qquad\qquad H_1 : P \notin \mathbf{P}$$

We consider tests $\phi_n : \{X_i\}_{i=1}^n \to [0, 1]$ with well defined limiting local power

$$\lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} \equiv \pi(g)$$

for $X_i \sim P_{1/\sqrt{n},g}$ and with $P^n_{1/\sqrt{n},g} \equiv \bigotimes_{i=1}^n P_{1/\sqrt{n},g}$ the product measure.

Comments

- Note limiting power depends only on g – this is not an assumption.
- Mild conditions guarantee π exists when $\phi_n = 1\{T_n > c_{1-\alpha}\}$.

(Re)interpreting Test

Local size control demands that for any submodel $t \mapsto P_{t,g} \in \mathbf{P}$ we have

$$\pi(g) = \lim_{n \to \infty} \int \phi_n dP^n_{1/\sqrt{n},g} \leq \alpha$$

Equivalently, $\pi(g) \leq \alpha$ for any $g \in \bar{T}(P)$ – i.e. g "looks like" from submodel.

(Re)interpreting Test

Local size control demands that for any submodel $t \mapsto P_{t,g} \in \mathbf{P}$ we have

$$\pi(g) = \lim_{n \to \infty} \int \phi_n dP^n_{1/\sqrt{n},g} \leq \alpha$$

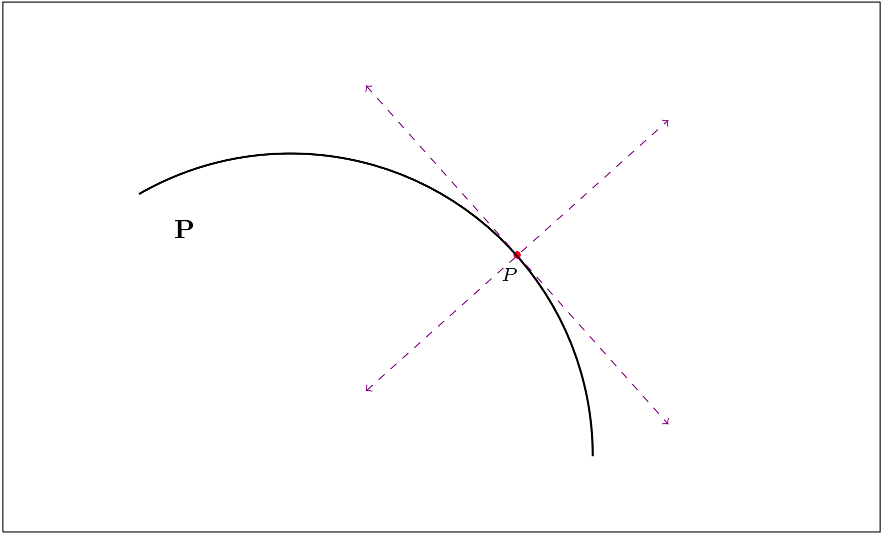
Equivalently, $\pi(g) \leq \alpha$ for any $g \in \bar{T}(P)$ – i.e. g "looks like" from submodel.

Notation

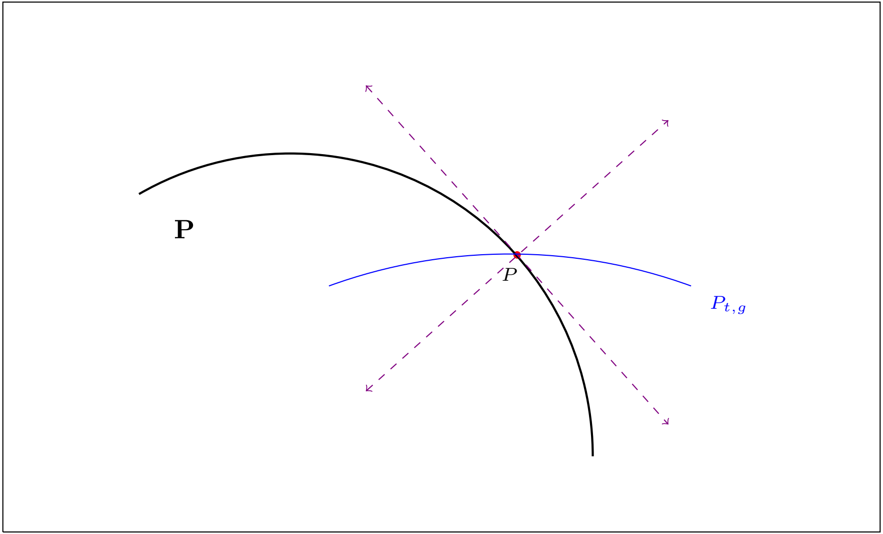
- Let $\Pi_T(g)$ denote projection of g into $\bar{T}(P)$ (in L^2_0).
- Let $\Pi_{T^\perp}(g)$ denote projection of g into $\bar{T}(P)^\perp$ (in L^2_0).

Note: For any $g \in L^2_0$ we have $g = \Pi_T(g) + \Pi_{T^\perp}(g)$.

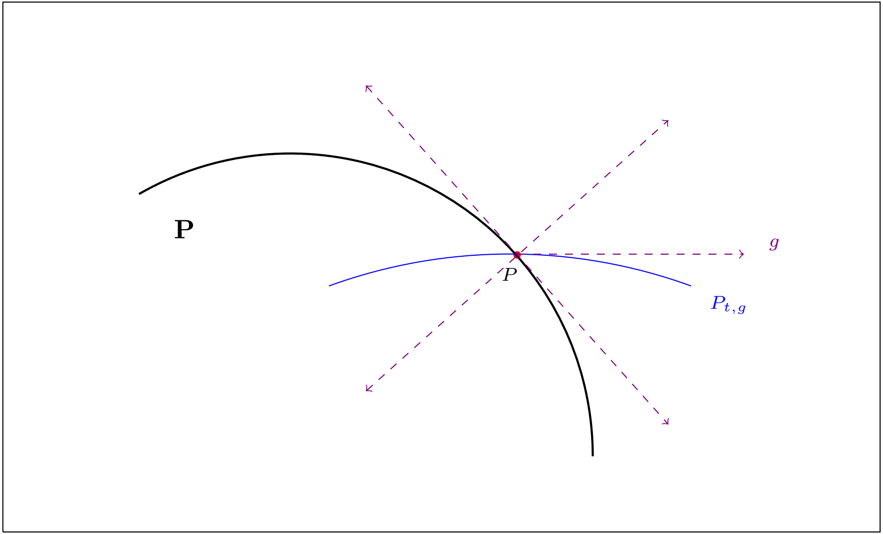
(Re)interpreting Test



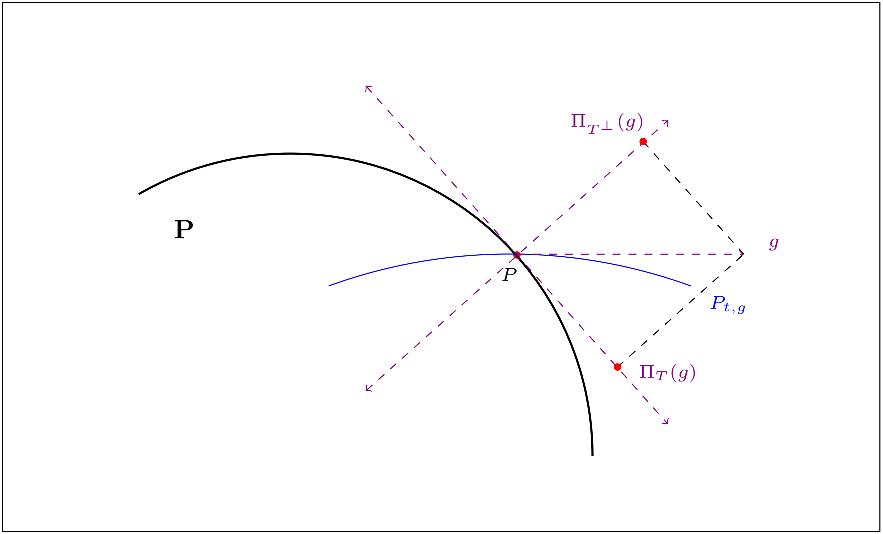
(Re)interpreting Test



(Re)interpreting Test



(Re)interpreting Test



(Re)interpreting Test

$$H_0 : \Pi_{T^\perp}(g) = 0 \qquad H_1 : \Pi_{T^\perp}(g) \neq 0$$

(Re)interpreting Test

$$H_0 : \Pi_{T^\perp}(g) = 0 \qquad\qquad H_1 : \Pi_{T^\perp}(g) \neq 0$$

Formally

- Let $\{\psi_k^T\}_{k=1}^{d_T}$ be orthonormal basis for $\bar{T}(P)$.
- Let $\{\psi_k^\perp\}_{k=1}^{d_{T^\perp}}$ be orthonormal basis for $\bar{T}(P)^\perp$.
- Let Q_g be distribution of $(Y, Z) \equiv (\{Y_k\}_{k=1}^{d_T}, \{Z_k\}_{k=1}^{d_{T^\perp}})$ on $\mathbf{R}^{d_T} \times \mathbf{R}^{d_{T^\perp}}$

$$Y_k \sim N(\int g\psi_k^T dP, 1) \text{ for } 1 \leq k \leq d_T$$

$$Z_k \sim N(\int g\psi_k^\perp dP, 1) \text{ for } 1 \leq k \leq d_{T^\perp}$$

Intuition: (Y, Z) is a "noisy" signal of the unknown score g.

(Re)interpreting Test

Theorem Let Assumption R hold, and ϕ_n satisfy for any $t \mapsto P_{t,g} \in \mathbf{P}$

$$\pi(g) \equiv \lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} \leq \alpha$$

(Re)interpreting Test

Theorem Let Assumption R hold, and ϕ_n satisfy for any $t \mapsto P_{t,g} \in \mathbf{P}$

$$\pi(g) \equiv \lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} \leq \alpha$$

Then there is a level α test $\phi : (Y, Z) \to [0, 1]$ of the null hypothesis

$$H_0 : \Pi_{T^\perp}(g) = 0 \qquad\qquad H_1 : \Pi_{T^\perp}(g) \neq 0$$

(Re)interpreting Test

Theorem Let Assumption R hold, and ϕ_n satisfy for any $t \mapsto P_{t,g} \in \mathbf{P}$

$$\pi(g) \equiv \lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} \leq \alpha$$

Then there is a level α test $\phi : (Y, Z) \to [0,1]$ of the null hypothesis

$$H_0 : \Pi_{T^\perp}(g) = 0 \qquad\qquad H_1 : \Pi_{T^\perp}(g) \neq 0$$

based on a single observation (Y, Z) such that for any path $t \mapsto P_{t,g}$

$$\pi(g) \equiv \lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} = \int \phi dQ_g$$

(Re)interpreting Test

Theorem Let Assumption R hold, and ϕ_n satisfy for any $t \mapsto P_{t,g} \in \mathbf{P}$

$$\pi(g) \equiv \lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} \leq \alpha$$

Then there is a level α test $\phi : (Y, Z) \to [0, 1]$ of the null hypothesis

$$H_0 : \Pi_{T^\perp}(g) = 0 \qquad\qquad H_1 : \Pi_{T^\perp}(g) \neq 0$$

based on a single observation (Y, Z) such that for any path $t \mapsto P_{t,g}$

$$\pi(g) \equiv \lim_{n\to\infty} \int \phi_n dP^n_{1/\sqrt{n},g} = \int \phi dQ_g$$

Comments

- Specification tests examine if g agrees with $\mathbf{P}$ based on signal (Y, Z).
- J-test corresponds to a Wald test on "signals" Z from $\bar{T}(P)^\perp$.

Overidentification and Testing

Implication: If P is locally just identified by $\mathbf{P}$, then $\mathbf{P}$ is locally untestable.

Overidentification and Testing

Implication: If P is locally just identified by $\mathbf{P}$, then $\mathbf{P}$ is locally untestable.

Key Question: If P is locally overidentified by $\mathbf{P}$, then is $\mathbf{P}$ locally testable?

Overidentification and Testing

Implication: If P is locally just identified by $\mathbf{P}$, then $\mathbf{P}$ is locally untestable.

Key Question: If P is locally overidentified by $\mathbf{P}$, then is $\mathbf{P}$ locally testable?

Assumption (B)

- There is a known subset $\mathcal{F} = \{f_k\}_{k=1}^{d_F} \subseteq \bar{T}(P)^{\perp} \subseteq L_0^2$.
- The set $\mathcal{F}$ satisfies $\sum_{k=1}^{d_F} f_k^2 dP < \infty$.

Note: Common tests implicitly estimate $\mathcal{F}$

Overidentification and Testing

Implication: If P is locally just identified by $\mathbf{P}$, then $\mathbf{P}$ is locally untestable.

Key Question: If P is locally overidentified by $\mathbf{P}$, then is $\mathbf{P}$ locally testable?

Assumption (B)

- There is a known subset $\mathcal{F} = \{f_k\}_{k=1}^{d_F} \subseteq \bar{T}(P)^{\perp} \subseteq L_0^2$.
- The set $\mathcal{F}$ satisfies $\sum_{k=1}^{d_F} f_k^2 dP < \infty$.

Note: Common tests implicitly estimate $\mathcal{F}$

$$\mathbb{G}_n \equiv \underbrace{(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_1(X_i), \ldots, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_{d_F}(X_i))'}$$

$\approx Z$ "signal" from $\bar{T}(P)^{\perp}$

Overidentification and Testing

Theorem (i) If Assumptions R and B hold, then for any path $t \mapsto P_{t,g} \in \mathbf{P}$

$$\lim_{n\to\infty} P^n_{1/\sqrt{n},g}(\|\mathbb{G}_n\| > c_{1-\alpha}) = \alpha$$

where $c_{1-\alpha}$ is $1-\alpha$ quantile of $\|\mathbb{G}_0\|$.

Overidentification and Testing

Theorem (i) If Assumptions R and B hold, then for any path $t \mapsto P_{t,g} \in \mathbf{P}$

$$\lim_{n\to\infty} P^n_{1/\sqrt{n},g}(\|\mathbb{G}_n\| > c_{1-\alpha}) = \alpha$$

where $c_{1-\alpha}$ is $1-\alpha$ quantile of $\|\mathbb{G}_0\|$. (ii) If instead $t \mapsto P_{t,g}$ satisfies

$$\liminf_{n\to\infty} \inf_{Q\in\mathbf{P}} n \int [dQ^{1/2} - dP^{1/2}_{1/\sqrt{n},g}]^2 > 0$$

and in addition $\mathcal{F}$ is such that cl$\{$lin$\{\mathcal{F}\}\} = \bar{T}(P)^{\perp}$, then it also follows that

$$\liminf_{n\to\infty} P^n_{1/\sqrt{n},g}(\|\mathbb{G}_n\| > c_{1-\alpha}) > \alpha$$

Overidentification and Testing

Theorem (i) If Assumptions R and B hold, then for any path $t \mapsto P_{t,g} \in \mathbf{P}$

$$\lim_{n\to\infty} P^n_{1/\sqrt{n},g}(\|\mathbb{G}_n\| > c_{1-\alpha}) = \alpha$$

where $c_{1-\alpha}$ is $1-\alpha$ quantile of $\|\mathbb{G}_0\|$. (ii) If instead $t \mapsto P_{t,g}$ satisfies

$$\liminf_{n\to\infty} \inf_{Q\in\mathbf{P}} n \int [dQ^{1/2} - dP^{1/2}_{1/\sqrt{n},g}]^2 > 0$$

and in addition $\mathcal{F}$ is such that $\text{cl}\{\text{lin}\{\mathcal{F}\}\} = \bar{T}(P)^{\perp}$, then it also follows that

$$\liminf_{n\to\infty} P^n_{1/\sqrt{n},g}(\|\mathbb{G}_n\| > c_{1-\alpha}) > \alpha$$

In Words: If P is locally overidentified by $\mathbf{P}$, then $\mathbf{P}$ is locally testable.

$\Rightarrow$ $\mathbf{P}$ is locally testable if and only if P is locally overidentified.

1. Local Overidentification

2. Specification Testing

3. **Estimation**

4. Conditional Moment Models

Preview

Question: What are implications for estimation of local overidentification?

GMM Intuition: Weighting matrix not important under just identification.

Preview

Question: What are implications for estimation of local overidentification?

GMM Intuition: Weighting matrix not important under just identification.

In General

- Regular estimators asymptotically equivalent under just identification.
- Asymptotically distinct estimators exist under over identification.

Comments

- Finite dimensional case follows from role of $\bar{T}(P)$ (Newey, 1990).
- We require generalization to infinite dimensional for Hausman test.
- Will show "abstract" test can implemented through Hausman test.

Multiple Estimators

Assumption (E)

- $\theta : \mathbf{P} \to \mathbf{B}$ is a known map with $\mathbf{B}$ a Banach space.
- There exists an asymptotically linear regular estimator $\hat{\theta}_n$ of $\theta(P)$.

Note: Restriction on parameter $\theta(P)$, not on model $\mathbf{P}$.

Multiple Estimators

Assumption (E)

- $\theta : \mathbf{P} \to \mathbf{B}$ is a known map with $\mathbf{B}$ a Banach space.
- There exists an asymptotically linear regular estimator $\hat{\theta}_n$ of $\theta(P)$.

Note: Restriction on parameter $\theta(P)$, not on model $\mathbf{P}$.

Theorem Let Assumptions R and E hold.

(i) If P is locally just identified and $\tilde{\theta}_n$ is regular and asymptotically linear

$$\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\} = o_p(1) \quad \text{(in } \mathbf{B}\text{)}$$

Multiple Estimators

Assumption (E)

- $\theta : \mathbf{P} \to \mathbf{B}$ is a known map with $\mathbf{B}$ a Banach space.
- There exists an asymptotically linear regular estimator $\hat{\theta}_n$ of $\theta(P)$.

Note: Restriction on parameter $\theta(P)$, not on model $\mathbf{P}$.

Theorem Let Assumptions R and E hold.

(i) If P is locally just identified and $\tilde{\theta}_n$ is regular and asymptotically linear

$$\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\} = o_p(1) \quad \text{(in } \mathbf{B}\text{)}$$

(ii) If P is locally overidentified, there is regular asymptotically linear $\tilde{\theta}_n$

$$\sqrt{n}\{\hat{\theta}_n - \tilde{\theta}_n\} \xrightarrow{L} \Delta \neq 0 \quad \text{(in } \mathbf{B}\text{)}$$

Multiple Estimators

In Words: P is locally overidentified if and only if efficiency matters.

Multiple Estimators

In Words: P is locally overidentified if and only if efficiency matters.

P is locally overidentified if and only if $\mathbf{P}$ is locally testable.

Multiple Estimators

In Words: P is locally overidentified if and only if efficiency matters.

P is locally overidentified if and only if $\mathbf{P}$ is locally testable.

Efficiency matters if and only if $\mathbf{P}$ is locally testable.

Multiple Estimators

In Words: P is locally overidentified if and only if efficiency matters.

P is locally overidentified if and only if $\mathbf{P}$ is locally testable.

Efficiency matters if and only if $\mathbf{P}$ is locally testable.

Implications

- GMM link between efficiency and testing not coincidental.
- Efficiency literature can be exploited to determine when $\mathbf{P}$ is testable.
- $\Rightarrow$ Semiparametric models are locally testable (Ai $\&$ Chen, 2003).

1. Local Overidentification
2. Specification Testing
3. Estimation
4. Conditional Moment Models

Setup

Question: So is the nonparametric IV model locally overidentified?

Goal: Use results from efficiency literature to answer (Ai & Chen, 2012).

Setup

Question: So is the nonparametric IV model locally overidentified?

Goal: Use results from efficiency literature to answer (Ai $\&$ Chen, 2012).

For $X = (Y, Z, W)$ with $Y \in \mathbf{R}^{d_y}$, $Z \in \mathbf{R}^{d_z}$, $W \in \mathbf{R}^{d_w}$, and $\rho : \mathbf{R}^{d_y+1} \to \mathbf{R}$

$$E[\rho(Y, h_P(Z))|W] = 0$$

for some unknown function $h_P : \mathbf{R}^{d_z} \to \mathbf{R}$ satisfying $\int h_P^2 dP < \infty$.

Examples

- (NPIV) $E[Y - h_P(Z)|W] = 0$ maps to $\rho(Y, h(Z)) = Y - h(Z)$.
- (NPQIV) $P(Y \leq h_P(Z)|W) = \tau$ maps to $\rho(Y, h(Z)) = 1\{Y \leq h(Z)\} - \tau$.

Setup

$$m(W,h) \equiv E[\rho(Y,h(Z))|W]$$

Setup

$$m(W,h) \equiv E[\rho(Y,h(Z))|W]$$

Think of $m(W,\cdot)$ as map $m : L^2_Z \to L^2_W$ and assume differentiability in that

$$\nabla m(W,h_P)[s] \equiv \frac{\partial}{\partial \tau} m(W,h_P+\tau s)\Big|_{\tau=0}$$

Setup

$$m(W,h) \equiv E[\rho(Y,h(Z))|W]$$

Think of $m(W,\cdot)$ as map $m : L^2_Z \to L^2_W$ and assume differentiability in that

$$\nabla m(W,h_P)[s] \equiv \frac{\partial}{\partial \tau} m(W,h_P+\tau s)\Big|_{\tau=0}$$

Consider derivative as a map $\nabla m(W,h_P) : L^2_Z \to L^2_W$ and denote its range

$$\mathcal{R} \equiv \{f \in L^2_W : f = \nabla m(W,h_P)[s] \text{ for some } s \in L^2_Z\}$$

Setup

NPIV Example

- Here, $m(W, h) = E[Y - h(Z)|W]$.
- Which implies $\nabla m(W, h_P)[s] = -E[s(Z)|W]$, and therefore

$$\mathcal{R} = \{f \in L^2_W : f(W) = E[s(Z)|W] \text{ for some } s \in L^2_Z\}$$

Setup

NPIV Example

- Here, $m(W,h) = E[Y - h(Z)|W]$.
- Which implies $\nabla m(W,h_P)[s] = -E[s(Z)|W]$, and therefore

$$\mathcal{R} = \{f \in L^2_W : f(W) = E[s(Z)|W] \text{ for some } s \in L^2_Z\}$$

NPQIV Example

- Here, $m(W,h) = P(Y \leq h(Z)|W) - \tau$.
- Which implies $\nabla m(W,h_P)[s] = E[f_{Y|Z,W}(h_P(Z)|Z,W)s(Z)|W]$, and

$$\mathcal{R} = \{f \in L^2_W : f(W) = E[f_{Y|Z,W}(h_P(Z)|Z,W)s(Z)|W] \text{ for some } s \in L^2_Z\}$$

Local Overidentification

Theorem Under regularity conditions, P is locally just identified if and only if

$$\bar{\mathcal{R}} = L^2_W$$

Local Overidentification

Theorem Under regularity conditions, P is locally just identified if and only if

$$\bar{\mathcal{R}} = L^2_W$$

Relation to GMM

$$E[\rho(X, \beta(P))] = 0 \qquad\qquad E[\rho(Y, h_P(Z))|W] = 0$$

Local Overidentification

Theorem Under regularity conditions, P is locally just identified if and only if

$$\bar{\mathcal{R}} = L^2_W$$

Relation to GMM

$$E[\rho(X, \beta(P))] = 0 \qquad\qquad E[\rho(Y, h_P(Z))|W] = 0$$

$$\underbrace{\nabla_\beta E[\rho(X, \beta(P))]} \qquad\qquad \underbrace{\nabla E[\rho(Y, h_P(Z))|W]}$$

Maps: $\mathbf{R}^{d_\beta}$ to $\mathbf{R}^{d_\rho}$ $\qquad\qquad$ L^2_Z to L^2_W

Local Overidentification

Theorem Under regularity conditions, P is locally just identified if and only if

$$\bar{\mathcal{R}} = L^2_W$$

Relation to GMM

	$E[\rho(X, \beta(P))] = 0$	$E[\rho(Y, h_P(Z))\|W] = 0$
	$\underbrace{\nabla_\beta E[\rho(X, \beta(P))]}$	$\underbrace{\nabla E[\rho(Y, h_P(Z))\|W]}$
Maps:	$\mathbf{R}^{d_\beta}$ to $\mathbf{R}^{d_\rho}$	L^2_Z to L^2_W
Just ID:	Map is onto $(d_\beta = d_\rho)$	Map is “onto” $(\bar{\mathcal{R}} = L^2_W)$

Local Overidentification

Theorem Under regularity conditions, P is locally just identified if and only if

$$\bar{\mathcal{R}} = L^2_W$$

Relation to GMM

$$E[\rho(X, \beta(P))] = 0 \qquad\qquad E[\rho(Y, h_P(Z))|W] = 0$$

$$\underbrace{\nabla_\beta E[\rho(X, \beta(P))]} \qquad\qquad \underbrace{\nabla E[\rho(Y, h_P(Z))|W]}$$

Maps:	$\mathbf{R}^{d_\beta}$ to $\mathbf{R}^{d_\rho}$	L^2_Z to L^2_W
Just ID:	Map is onto $(d_\beta = d_\rho)$	Map is “onto” $(\bar{\mathcal{R}} = L^2_W)$
Over ID:	Not onto $(d_\beta < d_\rho)$	Not onto $(\bar{\mathcal{R}} \subsetneq L^2_W)$

Local Overidentification

Alternative Characterization

$$\bar{\mathcal{R}} = L^2_W \text{ if and only if } \{0\} = \{s \in L^2_W : \nabla m(W, h_P)^*[s] = 0\}$$

where $\nabla m(W, h_P) : L^2_W \to L^2_Z$ is the adjoint of $\nabla m(W, h_P) : L^2_Z \to L^2_W$.

Local Overidentification

Alternative Characterization

$$\bar{\mathcal{R}} = L^2_W \text{ if and only if } \{0\} = \{s \in L^2_W : \nabla m(W, h_P)^*[s] = 0\}$$

where $\nabla m(W, h_P) : L^2_W \to L^2_Z$ is the adjoint of $\nabla m(W, h_P) : L^2_Z \to L^2_W$.

NPIV Example

- Here, $\nabla m(W, h_P)^*[s] = -E[s(W)|Z]$.
- Therefore, in NPIV P is locally just identified by $\mathbf{P}$ if and only if

$$E[s(W)|Z] = 0 \text{ implies } s(W) = 0 \text{ for all } s \in L^2_W$$

Local Overidentification

Alternative Characterization

$$\bar{\mathcal{R}} = L^2_W \text{ if and only if } \{0\} = \{s \in L^2_W : \nabla m(W, h_P)^*[s] = 0\}$$

where $\nabla m(W, h_P) : L^2_W \to L^2_Z$ is the adjoint of $\nabla m(W, h_P) : L^2_Z \to L^2_W$.

NPIV Example

- Here, $\nabla m(W, h_P)^*[s] = -E[s(W)|Z]$.
- Therefore, in NPIV P is locally just identified by $\mathbf{P}$ if and only if

$$E[s(W)|Z] = 0 \text{ implies } s(W) = 0 \text{ for all } s \in L^2_W$$

NPQIV Example

- Similarly, in NPQIV P is locally just identified by $\mathbf{P}$ if and only if

$$E[f_{Y|Z,W}(h_P(Z)|Z, W)s(W)|Z] = 0 \text{ implies } s(W) = 0 \text{ for all } s \in L^2_W$$

Discussion

Exogenous Models

- Suppose $W = (Z, V)$ and that $\rho(W, h(Z))$ is differentiable in h.
- Then, it follows that P is locally just identified if and only if

$$P(E[V] = V) = 1 \text{ and } P(E[\nabla_h \rho(Y, h_P(Z))|Z] \neq 0) = 1$$

Discussion

Exogenous Models

- Suppose $W = (Z, V)$ and that $\rho(W, h(Z))$ is differentiable in h.
- Then, it follows that P is locally just identified if and only if

$$P(E[V] = V) = 1 \text{ and } P(E[\nabla_h \rho(Y, h_P(Z))|Z] \neq 0) = 1$$

Efficiency Implications

- Parameters θ such as average derivatives, consumer surplus, etc.

  $\Rightarrow$ Plug-in typically efficient under "exogeneity" (Newey, 1994).

  $\Rightarrow$ Plug-in need not be efficient under "endogeneity" (Ai $\&$ Chen, 2012).

Discussion

Exogenous Models

- Suppose $W = (Z, V)$ and that $\rho(W, h(Z))$ is differentiable in h.
- Then, it follows that P is locally just identified if and only if

$$P(E[V] = V) = 1 \text{ and } P(E[\nabla_h \rho(Y, h_P(Z))|Z] \neq 0) = 1$$

Efficiency Implications

- Parameters θ such as average derivatives, consumer surplus, etc.

  $\Rightarrow$ Plug-in typically efficient under "exogeneity" (Newey, 1994).

  $\Rightarrow$ Plug-in need not be efficient under "endogeneity" (Ai $\&$ Chen, 2012).

- Two stage estimation problems $\theta(P) = \arg\max_\theta E[q(X, \theta, h_P(Z))]$.

  $\Rightarrow$ Two stage efficient under "exogeneity" (Ackerberg et al. 2014).

  $\Rightarrow$ Two stage estimation need not be efficient under "endogeneity".

Conclusion

Local Overidentification

- Is model locally consistent with any parametric model?
- Abstracts from counting “parameters” and “restrictions”.
- Partial generalization to nonregular models.

Refutability and Efficiency

- Intrinsic link between refutability and efficiency.
- Generalizes connection present in GMM to regular models.

Conditional Moment Models

- Simple characterization by exploiting efficiency literature.
- Implications for plug-in and two stage estimation.