Constrained Conditional Moment Restriction Models

Victor Chernozhukov	Whitney Newey	Andres Santos
MIT	MIT	UC San Diego

March 25, 2016

Shape Restrictions

Classic Work

- Test implications of consumer and producer theory.
- Exploit theoretically implied restrictions to sharpen estimation.
- Employ restrictions to establish nonparametric identification.

Recent Applications

- Complementarities in discrete games.
- Ramp-up and start-up costs in electricity production.
- Monotonicity of the pricing kernel.
- Literature on moment inequalities.

Goal: Device procedure for testing and imposing general shape restrictions.

Example 1: Demand

Quantity demanded Q_i given price P_i, income Y_i, and covariates W_i equals

$$Q_i = g_0(P_i, Y_i) + W_i'\gamma_0 + U_i$$

where g_0 and γ_0 are unknown function and vector, and $E[U_i|P_i, Y_i, W_i] = 0$.

Blundell et al. (2012): Impose Slutzky restriction for inference on g_0 – e.g.

$$H_0 : g_0(p_0, y_0) = c_0 \qquad H_1 : g_0(p_0, y_0) \neq c_0$$

Comments

- Exogeneity assumption can be relaxed given instrument.
- Mean independence can be replaced by quantile independence.
- Blundell et al. (2012) finds imposing Slutzky restriction important ...

  ... but asymptotics assume Slutzky is not binding.

Example 2: Monotonic RD

Outcome variable Y_i, treatment assigned when $R_i \geq 0$, and interested in

$$\tau_0 \equiv \lim_{r \downarrow 0} E[Y_i | R_i = r] - \lim_{r \uparrow 0} E[Y_i | R_i = r]$$

Impose monotonicity of regression function in neighborhood of τ_0 and test

$$H_0 : \tau_0 = 0 \qquad\qquad H_1 : \tau_0 \neq 0$$

Comments

- Relevant in Lee et al. (2004), Black et al. (2007).
- Obtain confidence region through test inversion.
- Sharp design can be extended to fuzzy or regression kink design.

Example 3: Complementarities

Agent's utility for bundle $a = (a_1, a_2) \in \{0,1\}^2$ of two goods $j \in \{1,2\}$ equals

$$U(a, Z_i, \epsilon_i) = \sum_{j=1}^{2} (W_i' \gamma_{0,j} + \epsilon_{i,j}) 1\{a_j = 1\} + \delta_0(Y_i) 1\{a_1 = 1, a_2 = 1\}$$

for δ_0 unknown function, (W_i, Y_i) covariates, $\epsilon = (\epsilon_1, \epsilon_2)$ normal distribution.

Consider test for whether goods $j \in \{1,2\}$ are always (in Y) substitutes

$H_0 : \delta_0(y) \leq 0$ for all y $\qquad$ $H_1 : \delta_0(y) > 0$ for some y

Comments

- Parametric approach in Gentzkow (2007) for print and online media.
- Applies to organizational design (Athey and Stern, 1998), and interactions in discrete games (De Paula and Tang, 2012).

Example 4: Hospital Referral

Ho and Pakes (2013) derive for two patients $j \in \{1, 2\}$ sent to hospital H_{ij}

$$E[\sum_{j=1}^{2}\{\gamma_0(P_{ij}(H_{ij}) - P_{ij}(H_{ij'})) + g_0(D_{ij}(H_{ij})) - g_0(D_{ij}(H_{ij'}))\}|Z_i] \leq 0$$

$P_{ij}(h)/D_{ij}(h)$ price/distance to hospital h, g_0 unknown increasing function.

Allow nonparametric monotonic g_0 while conducting inference on γ_0 – e.g.

$$H_0 : \gamma_0 = c_0 \qquad\qquad H_1 : \gamma_0 \neq c_0$$

Comments

- In general, moment inequalities with semiparametric specifications.
- Special case of moment equality restrictions with positivity constraint.

This Paper

Goal: Develop general tests that apply to previous examples.

Contributions

- Formalize common structure of shape restrictions.
- Models defined by finite conditional moment restrictions.
- Allow for potential partial identification.
- Analysis must be uniform in underlying distribution.

This Talk

- Focus on how to analyze shape restrictions.
- Model defined by single conditional moment restriction.
- Assume parameter is identified.
- Uniformity in the background.

General Outline

Formal Setup

- How do we think of shape restrictions in general terms?
- Introduce AM Spaces and their role in our problem.

Test Statistic

- Introduce the test statistic we study.
- Develop an asymptotic approximation to its distribution.

Bootstrap Approximation

- Develop bootstrap procedure to estimate asymptotic approximation.
- Establish bootstrap validity (uniformly in underlying distribution).

Related Literature

Shape Restrictions

Matzkin (1994), Hausman & Newey (1995), Lewbel (1995), Haag, Hoderlein & Pendakur (2007), Chetverikov (2012), Freyberger & Horowitz (2012), Beare & Schmidt (2014), Chetverikov & Wilhelm (2014), Armstrong (2015), Horowitz & Lee (2015).

Conditional Moment Models

Newey (1985), Chamberlain (1987, 1992), Ai & Chen (2003), Chen & Pouzo (2015), Hong (2011), Santos (2012), Tao (2014).

Partial Identification

Manski (2003), Andrews & Soares (2010), Chernozhukov, Lee, & Rosen (2013), Bugni, Canay, & Shi (2014).

The Model

The parameter of interest $\theta_0 \in \Theta$ is the unique solution to the restriction

$$E[\rho(X_i, \theta_0)|Z_i] = 0$$

for $X_i \in \mathbf{R}^{d_x}$, $Z_i \in \mathbf{R}^{d_z}$, and $\rho : \mathbf{R}^{d_x} \times \Theta \to \mathbf{R}$ is known function.

Assumption (M)

- $\{X_i, Z_i\}_{i=1}^n$ is an i.i.d. sample distributed according to $P \in \mathbf{P}$.
- $\Theta \subseteq \mathbf{B}$ for Banach space $\mathbf{B}$ with norm $\|\cdot\|_{\mathbf{B}}$ (allow non/semi/parametric)
- The function $\rho : \mathbf{R}^{d_x} \times \Theta \to \mathbf{R}$ is differentiable in θ.

Results in paper

- Allow for θ_0 to be partially identified.
- Allow for $\rho(X_i, \theta)$ nondifferentiable in θ (e.g. quantile restrictions).

The Hypothesis

$$H_0 : \theta_0 \in R \qquad\qquad H_1 : \theta_0 \notin R$$

Goal: The set R must be general enough to include motivating examples ...

... and have enough structure for fruitful asymptotic analysis.

$$R \equiv \{\theta \in \mathbf{B} : \Upsilon_F(\theta) = 0 \text{ and } \Upsilon_G(\theta) \leq 0\}$$

Assumption (R)

- $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ for $\mathbf{F}$ a Banach space with norm $\|\cdot\|_{\mathbf{F}}$.
- $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ for $\mathbf{G}$ an AM space with order unit $\mathbf{1_G}$ and norm $\|\cdot\|_{\mathbf{G}}$.

AM Space

Basic Properties

- $\mathbf{G}$ has a partial order ($\Upsilon_G(\theta) \leq 0$ "makes sense").
- "$\leq$" and "$+$" interact as in $\mathbf{R}$ (e.g. $g_1 \geq g_2$ implies $g_1 + g_3 \geq g_2 + g_3$).
- Any pair $g_1, g_2 \in \mathbf{G}$ has a least upper bound $g_1 \vee g_2$.

Note: By above, can define an "absolute value" on $\mathbf{G}$ by $|g| \equiv g \vee (-g)$.

Order Unit

- **Definition:** $\mathbf{1_G}$ is an order unit means for any $g \in \mathbf{G}$ there is $\lambda \in \mathbf{R}$ so

$$|g| \leq \lambda \mathbf{1_G}$$

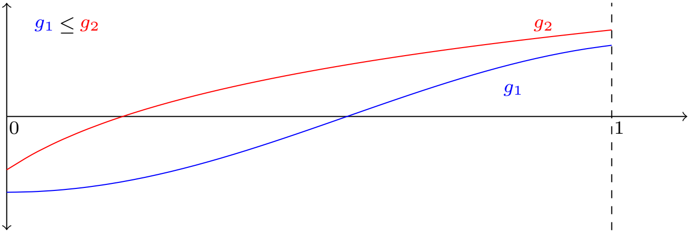
- **Intuition:** $\mathbf{1_G} \in \mathbf{G}$ can be made larger than any element by scaling.

AM Space (Example)

$$C([0,1]) \equiv \{g : [0,1] \to \mathbf{R} \text{ is continuous}\}$$

Properties

- Partial Order: "$g_1 \leq g_2$" iff $g_1(a) \leq g_2(a)$ for all $a \in [0,1]$.

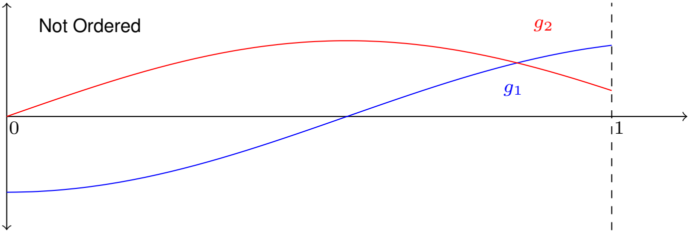


AM Space (Example)

$$C([0,1]) \equiv \{g : [0,1] \to \mathbf{R} \text{ is continuous}\}$$

Properties

- Partial Order: "$g_1 \leq g_2$" iff $g_1(a) \leq g_2(a)$ for all $a \in [0,1]$.

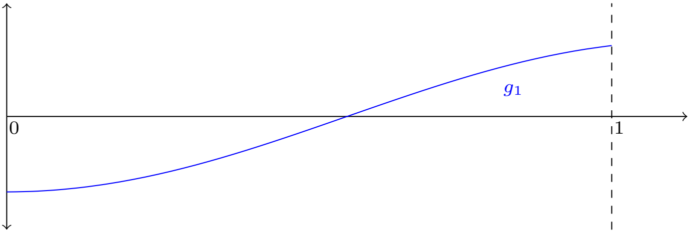


AM Space (Example)

$$C([0,1]) \equiv \{g : [0,1] \to \mathbf{R} \text{ is continuous}\}$$

Properties

- Partial Order: "$g_1 \leq g_2$" iff $g_1(a) \leq g_2(a)$ for all $a \in [0,1]$.
- Absolute Value: $|g|$ is the function $|g|(a) = |g(a)|$ for all $a \in [0,1]$.

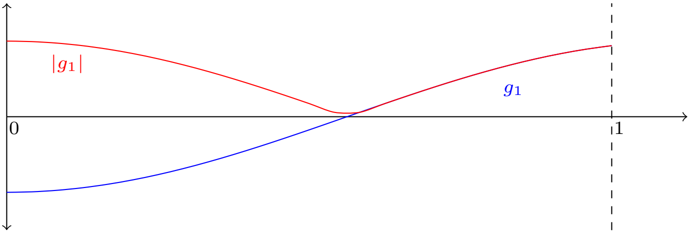


AM Space (Example)

$$C([0,1]) \equiv \{g : [0,1] \to \mathbf{R} \text{ is continuous}\}$$

Properties

- Partial Order: "$g_1 \leq g_2$" iff $g_1(a) \leq g_2(a)$ for all $a \in [0,1]$.
- Absolute Value: $|g|$ is the function $|g|(a) = |g(a)|$ for all $a \in [0,1]$.

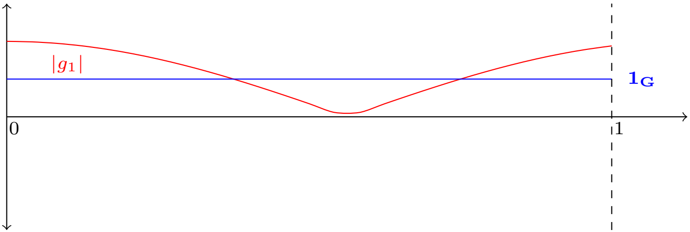


AM Space (Example)

$$C([0,1]) \equiv \{g : [0,1] \to \mathbf{R} \text{ is continuous}\}$$

Properties

- Partial Order: "$g_1 \leq g_2$" iff $g_1(a) \leq g_2(a)$ for all $a \in [0,1]$.
- Absolute Value: $|g|$ is the function $|g|(a) = |g(a)|$ for all $a \in [0,1]$.
- Order Unit: $\mathbf{1_G}$ is the constant function equal to one.

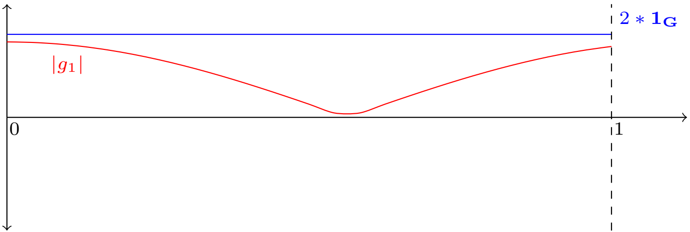


AM Space (Example)

$$C([0,1]) \equiv \{g : [0,1] \to \mathbf{R} \text{ is continuous}\}$$

Properties

- Partial Order: "$g_1 \leq g_2$" iff $g_1(a) \leq g_2(a)$ for all $a \in [0,1]$.
- Absolute Value: $|g|$ is the function $|g|(a) = |g(a)|$ for all $a \in [0,1]$.
- Order Unit: $\mathbf{1_G}$ is the constant function equal to one.



Running Example

For outcome variable $Y_i \in \mathbf{R}$, endogenous $W_i \in [0,1]$, instrument $Z_i \in \mathbf{R}$

$$Y_i = \theta_0(W_i) + \epsilon_i \qquad E[\epsilon_i | Z_i] = 0$$

Goal: Build a confidence region for $\theta_0(w_0)$ that imposes θ_0 is monotone.

$$R = \{\theta \in \mathbf{B} : \underbrace{\theta(w_0) - c_0 = 0}_{\Upsilon_F(\theta) = 0}, \text{ and } \underbrace{\theta'(w) \leq 0 \text{ for all } w}_{\Upsilon_G(\theta) \leq 0}\}$$

Procedure: Construct confidence region by test inverting (for different c_0)

$$H_0 : \theta_0 \in R \qquad H_1 : \theta_0 \notin R$$

Note: (i) $\Upsilon_F(\theta) = \theta(w_0) - c_0$, (ii) $\mathbf{F} = \mathbf{R}$, (iii) $\Upsilon_G(\theta) = \theta'$, (iv) $\mathbf{G} = C([0,1])$.

Test Statistic

By assumption, θ_0 is the unique element of Θ satisfying the restriction

$$E[\rho(X_i, \theta_0)|Z_i] = 0$$

For $\{q_j\}_{j=1}^{\infty}$ appropriate set of functions of Z_i, θ_0 is unique solution (in Θ) to

$$E[\rho(X_i, \theta_0)q_j(Z_i)] = 0 \text{ for all } j \qquad (\star)$$

Basic Idea

- If $\theta_0 \in R$, then there is a $\theta \in \Theta \cap R$ such that $(\star)$ holds.
- If $\theta_0 \notin R$, then there is no $\theta \in \Theta \cap R$ such that $(\star)$ holds.

$\Rightarrow$ Test whether $\theta_0 \in R$ by examining if $(\star)$ holds for some $\theta \in \Theta \cap R$

Test Statistic

Goal: Test if $\theta_0 \in R$, by examining if there is a $\theta \in \Theta \cap R$ such that

$$E[\rho(X_i, \theta_0) q_j(Z_i)] = 0 \text{ for all } j$$

Construct Statistic

- Replace population moments by sample moments.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta) q_j(Z_i)$$

Test Statistic

Goal: Test if $\theta_0 \in R$, by examining if there is a $\theta \in \Theta \cap R$ such that

$$E[\rho(X_i, \theta_0) q_j(Z_i)] = 0 \text{ for all } j$$

Construct Statistic

- Replace population moments by sample moments.
- Let $q^{k_n}(Z_i) = (q_1(Z_i), \ldots, q_{k_n}(Z_i))'$ and collect moments.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta) q^{k_n}(Z_i)$$

Test Statistic

Goal: Test if $\theta_0 \in R$, by examining if there is a $\theta \in \Theta \cap R$ such that

$$E[\rho(X_i, \theta_0) q_j(Z_i)] = 0 \text{ for all } j$$

Construct Statistic

- Replace population moments by sample moments.
- Let $q^{k_n}(Z_i) = (q_1(Z_i), \ldots, q_{k_n}(Z_i))'$ and collect moments.
- Search over $\theta \in \Theta \cap R$ to attempt to zero moments.

$$\inf_{\theta \in \Theta \cap R} \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta) q^{k_n}(Z_i) \|$$

Test Statistic

Goal: Test if $\theta_0 \in R$, by examining if there is a $\theta \in \Theta \cap R$ such that

$$E[\rho(X_i, \theta_0) q_j(Z_i)] = 0 \text{ for all } j$$

Construct Statistic

- Replace population moments by sample moments.
- Let $q^{k_n}(Z_i) = (q_1(Z_i), \ldots, q_{k_n}(Z_i))'$ and collect moments.
- Search over $\theta \in \Theta \cap R$ to attempt to zero moments.
- Replace Θ by approximating set Θ_n (e.g. splines, polynomials).

$$\inf_{\theta \in \Theta_n \cap R} \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta) q^{k_n}(Z_i) \|$$

Test Statistic

$$I_n(R) \equiv \inf_{\theta \in \Theta_n \cap R} \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta) q^{k_n}(Z_i) \|$$

Intuitively: J-test where parameter is restricted by the set R (sieve-GMM)

Comments

- We allow for weighting matrix, but suppress it here for simplicity.
- Norm need not be classic Euclidean norm.
- Class $\{q_j\}_{j=1}^{\infty}$ can change with n (e.g. B-Splines).
- Allowing for multiple moment restrictions is mostly notation.

Running Example

$$Y_i = \theta_0(W_i) + \epsilon_i \qquad\qquad E[\epsilon_i|Z_i] = 0$$

Goal: Build a confidence region for $\theta_0(w_0)$ that imposes θ_0 is monotone.

Example Specifics

- Specific structure of moment $\rho(X_i, \theta)$.

$$I_n(R) = \inf_{\theta \in \Theta_n \cap R} \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \theta(W_i)) q^{k_n}(Z_i) \|$$

Running Example

$$Y_i = \theta_0(W_i) + \epsilon_i \qquad E[\epsilon_i | Z_i] = 0$$

Goal: Build a confidence region for $\theta_0(w_0)$ that imposes θ_0 is monotone.

Example Specifics

- Specific structure of moment $\rho(X_i, \theta)$.
- Specific structure of restriction set R.

$$I_n(R) = \inf_{\theta \in \Theta_n} \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \theta(W_i)) q^{k_n}(Z_i) \|$$

s.t. (i) $\theta(w_0) = c_0,$ (ii) $\theta'(w) \leq 0$ for all $w \in [0, 1]$

Running Example

$$Y_i = \theta_0(W_i) + \epsilon_i \qquad\qquad E[\epsilon_i|Z_i] = 0$$

Goal: Build a confidence region for $\theta_0(w_0)$ that imposes θ_0 is monotone.

Example Specifics

- Specific structure of moment $\rho(X_i, \theta)$.
- Specific structure of restriction set R.
- Let $p^{j_n}(w) = (p_1(w), \ldots, p_{j_n}(w))$ and $\Theta_n = \{\theta = p^{j_n\prime}\beta$ some $\beta \in \mathbf{R}^{j_n}\}$.

$$I_n(R) = \inf_{\beta \in \mathbf{R}^{j_n}} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - p^{j_n}(W_i)'\beta) q^{k_n}(Z_i)\|$$

s.t. (i) $p^{j_n}(w_0)'\beta = c_0,$ (ii) $\nabla p^{j_n}(w)'\beta \leq 0$ for all $w \in [0, 1]$

Asymptotic Approximation

Goal: Approximate the finite sample distribution of our test statistic $I_n(R)$

$$I_n(R) \equiv \inf_{\theta \in \Theta_n \cap R} \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta) q^{k_n}(Z_i) \|$$

Local Space

- **Intuition:** The minimizer $\hat{\theta}_n$ of criterion close to θ_0 asymptotically.
- **Precisely:** Minimizer $\hat{\theta}_n$ asymptotically equal to $\theta_0 + \frac{h}{\sqrt{n}}$ with $\frac{h}{\sqrt{n}}$ in set

$$V_n(\theta_0, \ell_n) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B} : \underbrace{\theta_0 + \frac{h}{\sqrt{n}} \in \Theta_n \cap R}_{\hat{\theta}_n \text{ in } \Theta_n \cap R} \text{ and } \underbrace{\| \frac{h}{\sqrt{n}} \|_{\mathbf{B}} \leq \ell_n}_{\hat{\theta}_n - \theta_0 \text{ small}} \}$$

Key: Asymptotic distribution fundamentally affected by the set $V_n(\theta_0, \ell_n)$.

Asymptotic Approximation

Theorem: Under Assumptions M, R, and regularity conditions we obtain

$$I_n(R) = \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_n + E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]\| + o_p(1)$$

where $\mathbb{W}_n \in \mathbf{R}^{k_n}$ is a Gaussian r.v. and provided $\ell_n \downarrow 0$ appropriate rate.

Comments

- Intuitively, statistic equals distance between $\mathbb{W}_n$ and a set.
- Special case is J-test in Hansen (1982).
- Theorem is actually uniform in underlying distribution P of the data.

Running Example

$$Y_i = \theta_0(W_i) + \epsilon_i \qquad\qquad E[\epsilon_i|Z_i] = 0$$

Goal: Test null hypothesis $\theta_0(w_0) = c_0$ while imposing that θ_0 be monotone.

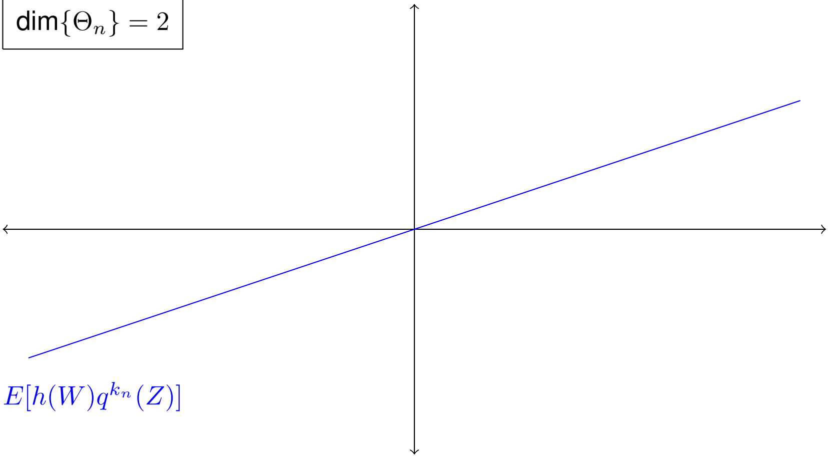
$$I_n(R) = \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_n + E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]\| + o_p(1)$$

$$= \inf_{h \in \Theta_n - \{\theta_0\}} \|\mathbb{W}_n - E[h(W_i) q^{k_n}(Z_i)]\|$$

$$\text{s.t.} \underbrace{\text{(i) } \theta_0(w_0) + \frac{h(w_0)}{\sqrt{n}} = c_0, \quad \text{(ii) } \theta_0' + \frac{h'}{\sqrt{n}} \leq 0,}_{\text{imposes that } \theta_0 + \frac{h}{\sqrt{n}} \in R} \quad \text{(iii) } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n$$

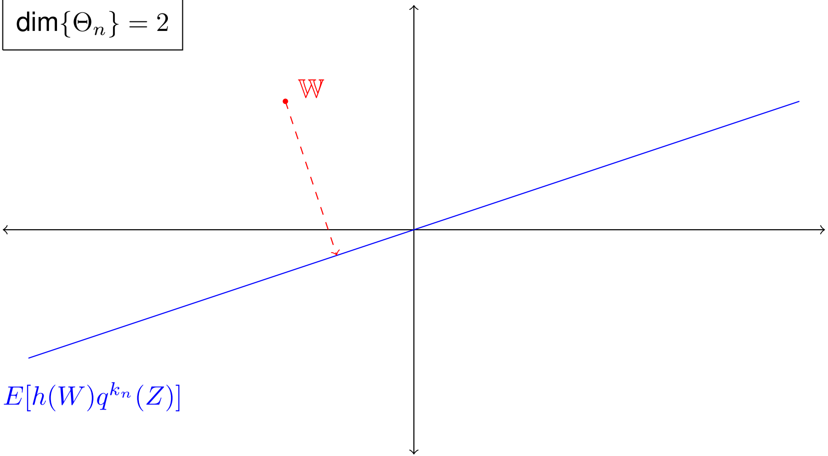
Case 1: No Monotonicity

$k_n = 2$
$\dim\{\Theta_n\} = 2$



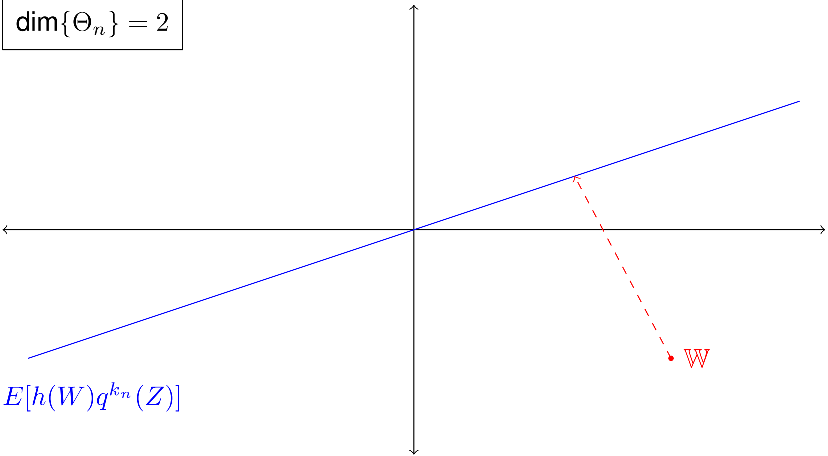
Case 1: No Monotonicity

$$k_n = 2$$
$$\text{dim}\{\Theta_n\} = 2$$



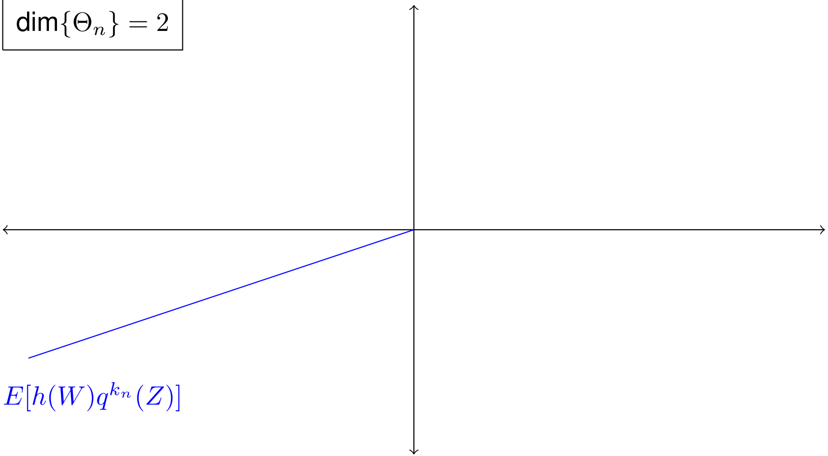
Case 1: No Monotonicity

$k_n = 2$
$\dim\{\Theta_n\} = 2$



Case 2: Monotonicity Binds

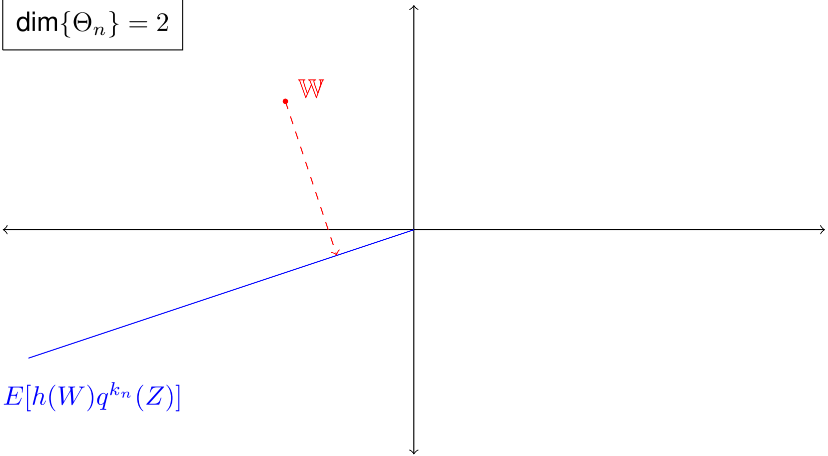
$$k_n = 2$$
$$\dim\{\Theta_n\} = 2$$



Case 2: Monotonicity Binds

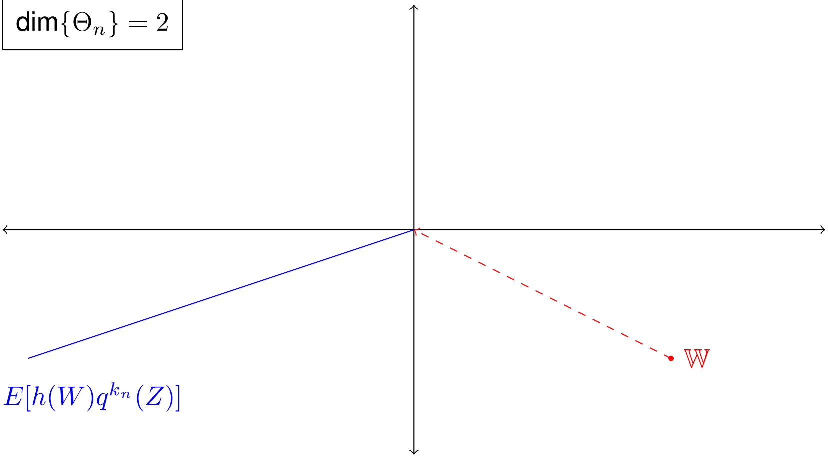
$k_n = 2$

$\text{dim}\{\Theta_n\} = 2$



Case 2: Monotonicity Binds

$$k_n = 2$$
$$\text{dim}\{\Theta_n\} = 2$$



Proof (Parametric Intuition)

$$I_n(R) = \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_n + E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]\| + o_p(1)$$

Step 1: Argue minimizer $\hat{\theta}_n$ equals $\theta_0 + \frac{h}{\sqrt{n}}$ with $\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)$ to obtain

$$I_n(R) = \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta_0 + \frac{h}{\sqrt{n}}) q^{k_n}(Z_i)\| + o_p(1)$$

Step 2: Conduct Taylor expansion around θ_0 to obtain for reminder $\mathcal{R}_n$ that

$$\inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \| \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta_0) q^{k_n}(Z_i)}_{\approx \mathbb{W}} + \underbrace{\frac{1}{n} \sum_{i=1}^{n} \nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)}_{\approx E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]} + \mathcal{R}_n \|$$

General Theorem

Challenge: In nonparametric problem, for $\mathcal{R}_n$ to disappear, either

- Model is linear in $\theta \Rightarrow \mathcal{R}_n = 0$ (running example).
- Rate of convergence for $\hat{\theta}_n$ is sufficiently fast (potentially unrealistic).

Theorem: Under Assumptions M, R, and regularity conditions we obtain

$$I_n(R) \leq \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_n + E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]\| + o_p(1)$$

where $\mathbb{W}_n \in \mathbf{R}^{k_n}$ is a Gaussian r.v. and provided $\ell_n \downarrow 0$ appropriate rate.

Comments

- Same asymptotic approximation, but potentially conservative.
- We recover original theorem if rate of convergence is sufficiently fast.

Bootstrap Approximation

Goal: Build bootstrap estimate of the distribution for approximation to $I_n(R)$

$$U_n(R) \equiv \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_n + E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]\|$$

Standard Unknowns

- The law of the Gaussian r.v. $\mathbb{W}_n \in \mathbf{R}^{k_n}$.
- The derivative $E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]$ for $\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)$.

Challenging Unknown

- The local parameter space $V_n(\theta_0, \ell_n)$.

Standard Unknown I

$$\mathbb{W}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho(X_i, \theta_0) q^{k_n}(Z_i) + o_p(1)$$

Note: $\mathbb{W}_n$ has covariance matrix $\Sigma_n \equiv E[\rho^2(X_i, \theta_0) q^{k_n}(Z_i) q^{k_n}(Z_i)']$.

Multiplier Bootstrap

- Randomly drawn i.i.d. $\{\omega_i\}_{i=1}^n$ independent of data with $\omega_i \sim N(0,1)$

$$\hat{\mathbb{W}}_n \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i \{\rho(X_i, \hat{\theta}_n) q^{k_n}(Z_i) - \frac{1}{n} \sum_{j=1}^{n} \rho(X_j, \hat{\theta}_n) q^{k_n}(Z_j)\}$$

- Conditional on data, $\hat{\mathbb{W}}_n \sim N(0, \hat{\Sigma}_n)$ and $\hat{\Sigma}_n$ is sample analogue to Σ_n.

Standard Unknown II

$$E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]$$

For $\hat{\theta}_n$ the argmin found when computing full sample test statistic $I_n(R)$ let

$$\frac{1}{n}\sum_{i=1}^{n} \nabla_\theta \rho(X_i, \hat{\theta}_n)[h] q^{k_n}(Z_i)$$

Comments

- Estimator must be consistent uniformly on suitable set of h.
- Use numerical method when $\rho : \mathbf{R}^{d_x} \times \Theta \to \mathbf{R}$ is not differentiable.
- Numerical approach not linear in $h \Rightarrow$ Final statistic harder to compute.

Local Space

$$V_n(\theta_0, \ell_n) \equiv \{\frac{h}{\sqrt{n}} \in \mathbf{B} : \theta_0 + \frac{h}{\sqrt{n}} \in \Theta_n \cap R \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\}$$

Note: If we knew θ_0, then we would know $V_n(\theta_0, \ell_n)$ for any ℓ_n ...

$\Rightarrow$ Use local parameter space of $\hat{\theta}_n$ to "estimate" $V_n(\theta_0, \ell_n)$.

Assumption (L)

- Parameter θ_0 in "interior" of Θ_n (i.e. only R matters).
- There is a linear map $\nabla\Upsilon_G(\theta_0) : \mathbf{B} \to \mathbf{G}$ and a neighborhood of θ_0 with

$$\|\Upsilon_G(\theta_1) - \Upsilon_G(\theta_0) - \nabla\Upsilon_G(\theta_0)[\theta_1 - \theta_0]\|_{\mathbf{G}} \leq K_g\|\theta_1 - \theta_0\|_{\mathbf{G}}^2$$

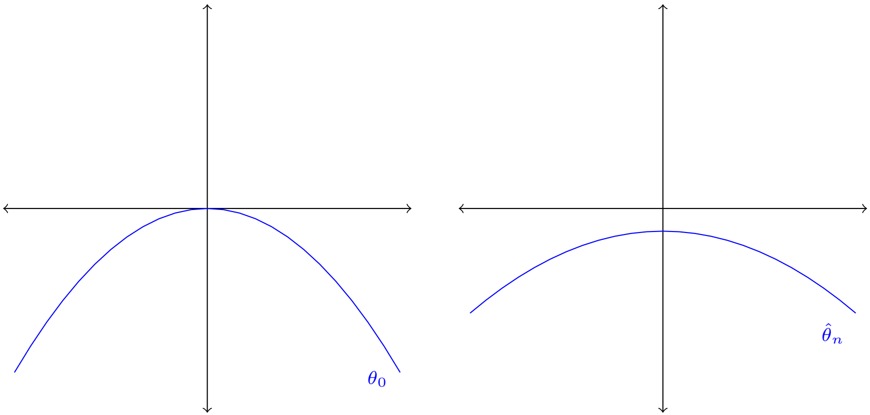
- There is a linear map $\nabla\Upsilon_F(\theta_0) : \mathbf{B} \to \mathbf{F}$ and a neighborhood of θ_0 with

$$\|\Upsilon_F(\theta_1) - \Upsilon_F(\theta_0) - \nabla\Upsilon_F(\theta_0)[\theta_1 - \theta_0]\|_{\mathbf{F}} \leq K_f\|\theta_1 - \theta_0\|_{\mathbf{F}}^2$$

Local Space: Inequalities

Example: For $\theta : \mathbf{R} \to \mathbf{R}$ suppose the only constraint is $\theta(w) \leq 0$ for all w.

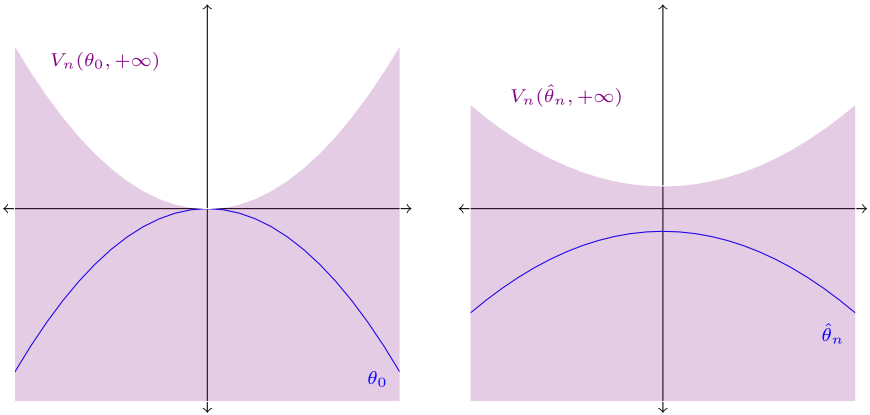
For size control we don't want to "overestimate" local parameter space ...



Local Space: Inequalities

Example: For $\theta : \mathbf{R} \to \mathbf{R}$ suppose the only constraint is $\theta(w) \leq 0$ for all w.

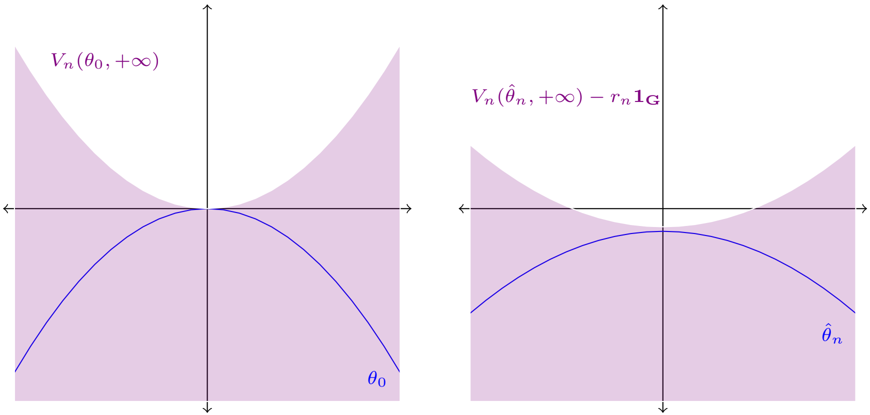
... which may happen due to estimation uncertainty!



Local Space: Inequalities

Example: For $\theta : \mathbf{R} \to \mathbf{R}$ suppose the only constraint is $\theta(w) \leq 0$ for all w.

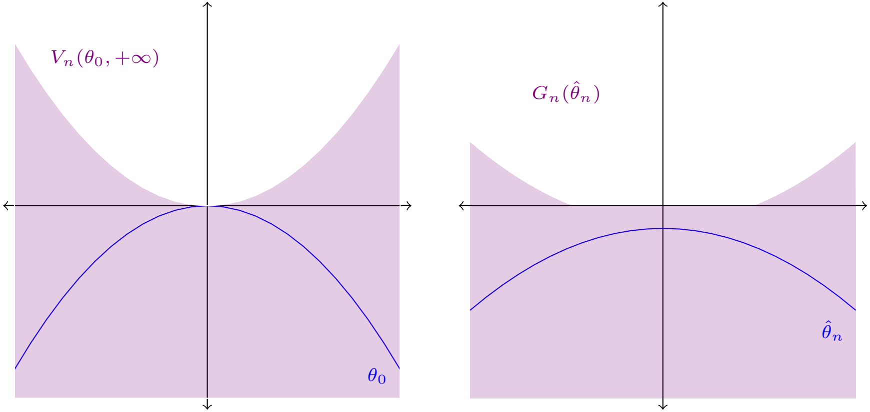
Fix: Adjust for estimation uncertainty in $\hat{\theta}_n$ ...



Local Space: Inequalities

Example: For $\theta : \mathbf{R} \to \mathbf{R}$ suppose the only constraint is $\theta(w) \leq 0$ for all w.

... and incorporate additional information on constraint.



Local Space: Inequalities

$$G_n(\hat{\theta}_n) \equiv \{\frac{h}{\sqrt{n}} : \Upsilon_G(\hat{\theta}_n + \frac{h}{\sqrt{n}}) \leq (\Upsilon_G(\hat{\theta}_n) - K_g r_n \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \mathbf{1_G}) \vee (-r_n \mathbf{1_G})\}$$

where $r_n \downarrow 0$ at a rate slower than the uniform (in P) $\hat{\theta}_n$ rate of convergence.

Comments

- Main instance in which AM structure and order unit exploited.
- Presence of K_g necessary for nonlinear constraints Υ_G.
- Related to generalized moment selection (Andres $\&$ Soares 2010).

Running Example

$$Y_i = \theta_0(W_i) + \epsilon_i \qquad\qquad E[\epsilon_i | Z_i] = 0$$

Goal: Test null hypothesis $\theta_0(w_0) = c_0$ while imposing that θ_0 be monotone.

Here: $\Upsilon_G : \mathbf{B} \to \mathbf{G}$ where $\Upsilon_G(\theta) = \theta'$ (linear) and $\mathbf{G} = C([0,1])$.

$$G_n(\hat\theta_n) \equiv \{\frac{h}{\sqrt{n}} : \underbrace{\Upsilon_G(\hat\theta_n + \frac{h}{\sqrt{n}})}_{\hat\theta_n' + \frac{h'}{\sqrt{n}}} \leq (\underbrace{\Upsilon_G(\hat\theta_n)}_{\hat\theta_n'} - \underbrace{K_g r_n \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}}}_{0} \mathbf{1_G}) \vee (-r_n \mathbf{1_G})\}$$

$$\Rightarrow G_n(\hat\theta_n) \equiv \{\frac{h}{\sqrt{n}} : \frac{h'(w)}{\sqrt{n}} \leq \max\{0, -r_n - \hat\theta_n'(w)\} \text{ for all } w \in [0,1]\}$$

Local Space: Equalities

$$\{\frac{h}{\sqrt{n}} : \Upsilon_F(\theta_0 + \frac{h}{\sqrt{n}}) = 0\}$$

Special Case

- Suppose the constraint $\Upsilon_F : \mathbf{B} \to \mathbf{F}$ is linear.
- Since under the null hypothesis $\Upsilon_F(\theta_0) = 0$, linearity implies that

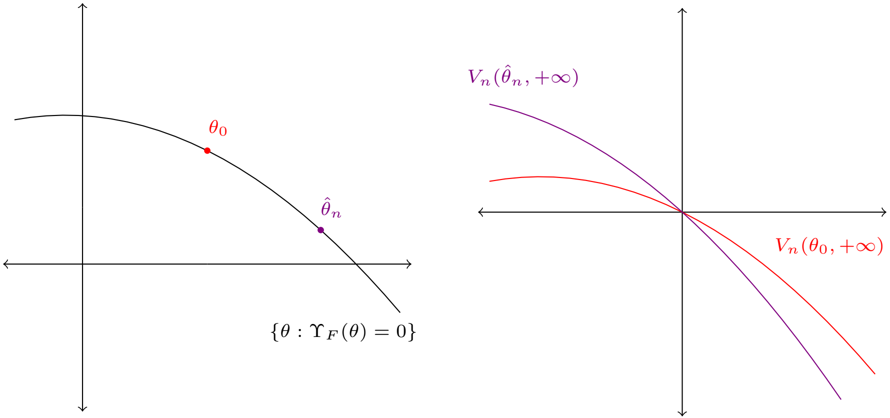
$$\{\frac{h}{\sqrt{n}} : \Upsilon_F(\theta_0 + \frac{h}{\sqrt{n}}) = 0\} = \{\frac{h}{\sqrt{n}} : \Upsilon_F(\frac{h}{\sqrt{n}}) = 0\}$$

$\Rightarrow$ Under linearity, the impact of Υ_F on local parameter space is known!

Local Space: Equalities

Example: Suppose $\mathbf{B} = \mathbf{R}^2$ and $\mathbf{F} = \mathbf{R}$, and no inequality constraints.

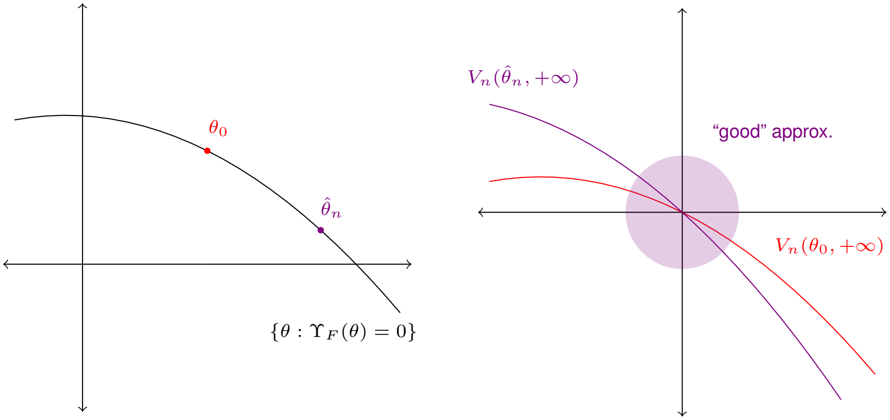
When constraints are nonlinear, local parameter space can be different ...



Local Space: Equalities

Example: Suppose $\mathbf{B} = \mathbf{R}^2$ and $\mathbf{F} = \mathbf{R}$, and no inequality constraints.

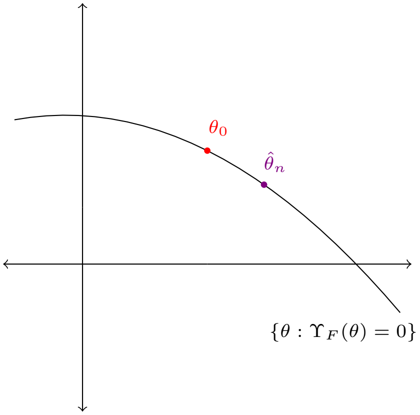
... but still provide a good approximation in a neighborhood of zero!

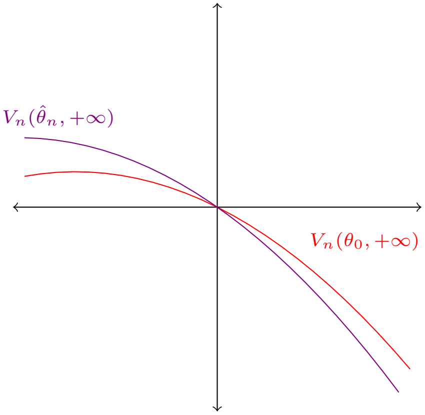


Local Space: Equalities

Example: Suppose $\mathbf{B} = \mathbf{R}^2$ and $\mathbf{F} = \mathbf{R}$, and no inequality constraints.

Moreover, as $\hat{\theta}_n$ approaches θ_0 ...

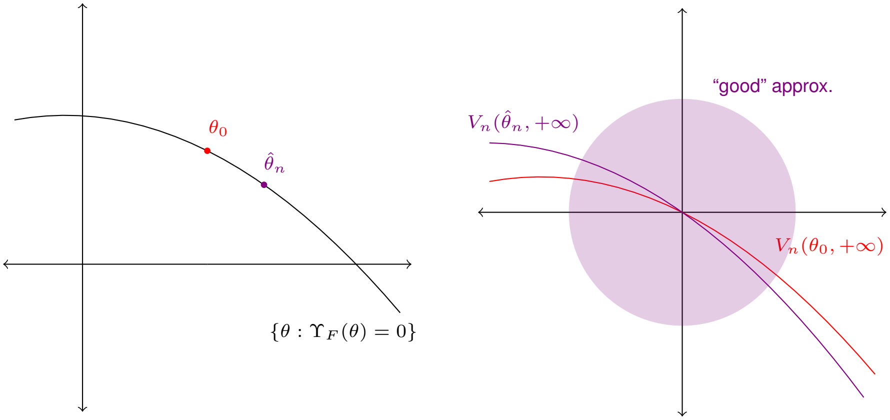




Local Space: Equalities

Example: Suppose $\mathbf{B} = \mathbf{R}^2$ and $\mathbf{F} = \mathbf{R}$, and no inequality constraints.

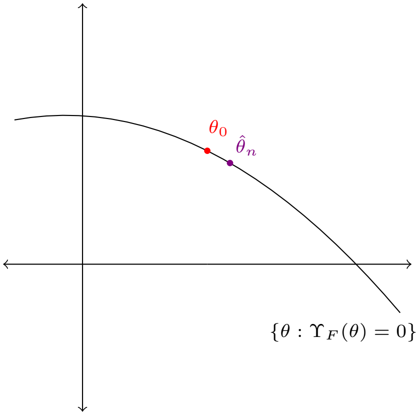
... the "reliable" neighborhood becomes larger.

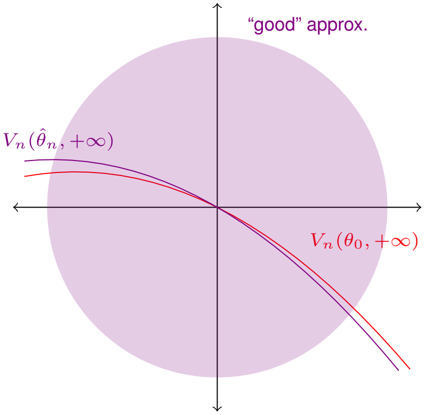


Local Space: Equalities

Example: Suppose $\mathbf{B} = \mathbf{R}^2$ and $\mathbf{F} = \mathbf{R}$, and no inequality constraints.

... the “reliable” neighborhood becomes larger.





Local Space: Equalities

$$F_n(\hat{\theta}_n) \equiv \{\frac{h}{\sqrt{n}} : \Upsilon_F(\hat{\theta}_n + \frac{h}{\sqrt{n}}) = 0 \text{ and } \|\frac{h}{\sqrt{n}}\|_{\mathbf{B}} \leq \ell_n\}$$

where we choose $\ell_n \downarrow 0$ sufficiently fast to justify the outlined argument.

Comments

- Bandwidth ℓ_n has multiple roles (derivative, nonlinearity in Υ_G, and Υ_F)

  $\Rightarrow \ell_n$ is necessary even if Υ_F is linear.
- Can study nonlinear functionals despite slow convergence of $\hat{\theta}_n$.
- Stronger requirements to handle Υ_G and Υ_F when latter is nonlinear.

Bootstrap Approximation

Goal: Build bootstrap estimate of the distribution for approximation to $I_n(R)$

$$U_n(R) \equiv \inf_{\frac{h}{\sqrt{n}} \in V_n(\theta_0, \ell_n)} \|\mathbb{W}_n + E[\nabla_\theta \rho(X_i, \theta_0)[h] q^{k_n}(Z_i)]\|$$

Strategy

- Replace $\mathbb{W}_n$ by its bootstrap analogue $\hat{\mathbb{W}}_n$.
- Replace the derivative by its estimator.
- Replace local parameter space by outlined construction.

$$\hat{U}_n(R) \equiv \inf_{\frac{h}{\sqrt{n}}} \|\hat{\mathbb{W}}_n + \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta \rho(X_i, \hat{\theta}_n)[h] q^{k_n}(Z_i)\|$$

$$\text{s.t. (i) } \frac{h}{\sqrt{n}} \in G_n(\hat{\theta}_n), \;\; \text{(ii) } \frac{h}{\sqrt{n}} \in F_n(\hat{\theta}_n), \;\; \text{(iii) } \frac{h}{\sqrt{n}} \in \overline{\text{span}}\{\Theta_n \cap R\}$$

Running Example

$$Y_i = \theta_0(W_i) + \epsilon_i \qquad\qquad E[\epsilon_i | Z_i] = 0$$

Goal: Test null hypothesis $\theta_0(w_0) = c_0$ while imposing that θ_0 be monotone.

General Bootstrap

$$\hat{U}_n(R) \equiv \inf_{\frac{h}{\sqrt{n}}} \|\hat{\mathbb{W}}_n + \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta \rho(X_i, \hat{\theta}_n)[h] q^{k_n}(Z_i)\|$$

s.t. (i) $\frac{h}{\sqrt{n}} \in G_n(\hat{\theta}_n)$, (ii) $\frac{h}{\sqrt{n}} \in F_n(\hat{\theta}_n)$, (iii) $\frac{h}{\sqrt{n}} \in \overline{\text{span}}\{\Theta_n \cap R\}$

In This Example

$$\hat{U}_n(R) \equiv \inf_{\beta \in \mathbf{R}^{j_n}} \|\hat{\mathbb{W}}_n - \frac{1}{n}\sum_{i=1}^{n} q^{k_n}(Z_i) p^{j_n}(W_i)'\beta\|$$

s.t. (i) $\nabla p^{j_n\prime}\beta \leq 0 \vee (-r_n \mathbf{1_G} - \hat{\theta}_n)$, (ii) $p^{j_n}(w_0)'\beta = 0$ and $\|\beta\| \leq \sqrt{n}\ell_n$

Bootstrap Validity

$$\hat{c}_{n,1-\alpha} \equiv \inf\{u : P(\hat{U}_n(R) \leq u | \{X_i, Z_i\}_{i=1}^n) \geq 1-\alpha\}$$

Theorem: Under Assumptions L, M, R, and regularity conditions we obtain

$$\limsup_{n\to\infty} P(I_n(R) > \hat{c}_{n,1-\alpha}) \leq \alpha$$

Comments

- Theorem is actually uniform in underlying distribution P of the data.
- Bandwidth ℓ_n unnecessary if rate of convergence sufficiently fast.
- Bandwidth r_n unfortunately necessary for inequality constraints.
- Consistency and characterization of local power in paper.

Simulation Design

$$\begin{pmatrix} X_i^* \\ Z_i^* \\ \epsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & 0.5 & 0.3 \\ 0.5 & 1 & 0 \\ 0.3 & 0 & 1 \end{bmatrix} \right)$$

and set $X_i = \Phi(X_i^*)$, instrument $Z_i = \Phi(Z_i^*)$, with Y_i generated according to

$$Y_i = \sigma\{1 - 2\Phi\left(\frac{X_i - 0.5}{\sigma}\right)\} + \epsilon_i$$

Comments

- Function θ_0 can be constant ($\sigma \approx 0$) or strictly monotonic $(\sigma \approx 1)$.
- For sieve (p^{j_n}) and moments (q^{k_n}) use b-splines of order 3.
- 500 observations, 200 bootstrap samples, 5000 replications.

The Hypothesis

$$H_0 : \theta_0 \in R \qquad\qquad H_1 : \theta_0 \notin R$$

Goal: Test whether $\theta_0(0.5) = 0$ (true) employing two different approaches ...

Without Exploiting Monotonicity

- Set $\Upsilon_F(\theta) = \theta(0.5)$, $\mathbf{F} = \mathbf{R}$, and no restriction Υ_G to get the set

$$R = \{\theta : \theta(0.5) = 0\}$$

Imposing Monotonicity

- Set $\Upsilon_F(\theta) = \theta(0.5)$, $\mathbf{F} = \mathbf{R}$, $\Upsilon_G(\theta) = \theta'$, $\mathbf{G} = C([0, 1])$ to get the set

$$R = \{\theta : \theta(0.5) = 0 \text{ and } \theta'(w) \leq 0 \text{ for all } w \in [0, 1]\}$$

Implementation

Step 1

- Compute full sample statistic (using 1^{st}-stage GMM weighting matrix)

$$I_n(R) = \inf_{\beta \in \mathbf{R}^{j_n}} \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - p^{j_n}(W_i)'\beta) q^{k_n}(Z_i) \|$$

$$\text{s.t. (i) } p^{j_n}(w_0)'\beta = 0, \text{ (ii) } \nabla p^{j_n}(w)'\beta \leq 0 \text{ for all } w \in [0,1]$$

Step 2

- Compute 200 weighted bootstrap $\hat{\mathbb{W}}_n$ and solve the optimizations

$$\hat{U}_n(R) \equiv \inf_{\beta \in \mathbf{R}^{j_n}} \| \hat{\mathbb{W}}_n - \frac{1}{n} \sum_{i=1}^{n} q^{k_n}(Z_i) p^{j_n}(W_i)'\beta \|$$

$$\text{s.t. (i) } \nabla p^{j_n \prime}\beta \leq 0 \vee (-r_n \mathbf{1}_{\mathbf{G}} - \hat{\theta}_n), \text{ (ii) } p^{j_n}(w_0)'\beta = 0, \text{ (iii) } \| \frac{\beta}{\sqrt{n}} \|_\infty \leq \ell_n$$

Implementation

Step 3

- Given 200 bootstrap statistics $\{\hat{U}_{n,b}(R)\}_{b=1}^{200}$ reject null hypothesis if

$$\frac{1}{200}\sum_{b=1}^{200} 1\{\hat{U}_{n,b}(R) > I_n(R)\} < \alpha$$

Examine Sensitivity

- To choice of k_n (number of moments).
- To choice of j_n (dimension of sieve).
- To choice of r_n (imposing monotonicity).
  Data Driven: bootstrap quantile $\|\hat{\theta}_n - \theta_0\|_{1,\infty} \Rightarrow$ small $\approx$ aggressive.
- To choice of ℓ_n (local parameter space).
  Data Driven: bootstrap quantile der. est. error $\Rightarrow$ small $\approx$ aggressive.

Table: Level Test Imposing Monotonicity - Empirical Size

			$\sigma = 1$					
			$k_n = 6$			$k_n = 13$		
j_n	q_ℓ	q_r	10%	5%	1%	10%	5%	1%
3	5%	5%	0.077	0.037	0.008	0.092	0.043	0.008
3	5%	95%	0.053	0.026	0.005	0.075	0.033	0.008
3	95%	5%	0.077	0.037	0.008	0.092	0.043	0.008
3	95%	95%	0.053	0.026	0.005	0.075	0.033	0.008
4	5%	5%	0.055	0.026	0.006	0.073	0.033	0.008
4	5%	95%	0.055	0.026	0.006	0.073	0.033	0.008
4	95%	5%	0.055	0.026	0.006	0.073	0.033	0.008
4	95%	95%	0.055	0.026	0.006	0.073	0.033	0.008
			$\sigma = 0.01$					
			$k_n = 6$			$k_n = 13$		
j_n	q_ℓ	q_r	10%	5%	1%	10%	5%	1%
3	5%	5%	0.102	0.053	0.012	0.109	0.054	0.011
3	5%	95%	0.100	0.051	0.011	0.107	0.053	0.011
3	95%	5%	0.102	0.053	0.012	0.109	0.054	0.011
3	95%	95%	0.100	0.051	0.011	0.107	0.053	0.011
4	5%	5%	0.101	0.051	0.011	0.106	0.052	0.011
4	5%	95%	0.101	0.051	0.011	0.106	0.052	0.011
4	95%	5%	0.101	0.051	0.011	0.106	0.052	0.011
4	95%	95%	0.101	0.051	0.011	0.106	0.052	0.011

Table: Level Test Not Imposing Monotonicity - Empirical Size

		$k_n = 6$			$k_n = 13$		
σ	j_n	10%	5%	1%	10%	5%	1%
1	3	0.106	0.051	0.010	0.107	0.056	0.012
1	4	0.072	0.034	0.006	0.078	0.038	0.008
0.01	3	0.106	0.052	0.010	0.107	0.056	0.011
0.01	4	0.073	0.034	0.006	0.077	0.038	0.008

Summary

- Adequate size control across specifications.
- Imposing monotonicity test can be undersized for $\sigma = 1$.
- Test insensitive to choice of ℓ_n.

Next: Examine power performance for $j_n = 3$ and "aggressive" r_n and ℓ_n in

$$Y_i = \sigma\{1 - 2\Phi(\frac{X_i - 0.5}{\sigma})\} + \delta + \epsilon_i$$

Figure: Empirical Power - Strict Monotonicity ($\sigma = 1$)

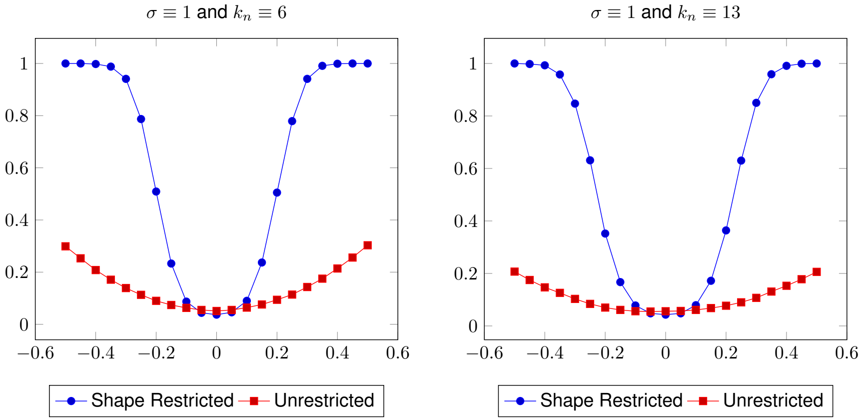
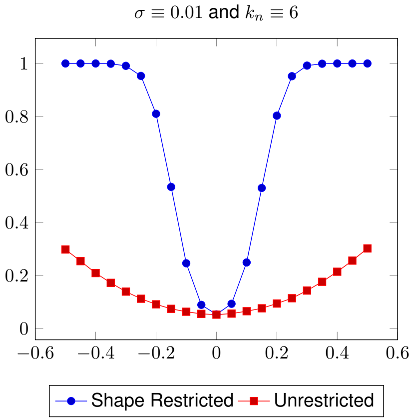


Figure: Empirical Power - Monotonicity Binds ($\sigma = 0.01$)



Conclusion

Main Contributions

- General framework for testing or imposing shape restrictions.
- Applies to general class of conditional moment restriction models.
- Uniform analysis in nonparametric/semiparametric models.
- Simulations show promising power advantages.

Open Questions

- Data driven choices of j_n and k_n – some results in estimation.
- Data driven choices of ℓ_n and r_n – accessible in special cases.
- Straightforward extension to other models (e.g. likelihood based)