Tutorial in Econometrics Part II(c): Sieve QLR Inference on Semi/Nonparametric Conditional Moment Restrictions

Xiaohong Chen (Yale) and Demian Pouzo (UC Berkeley)

NUS, IMS, May 16, 2014

May 16, 2014

Results in the Paper

- For possibly *irregular* (slower than root-n estimable) functionals, $\phi()$, of semi/nonparametric conditional moment restrictions $E[\rho(Z, \alpha_0)|X] = 0$, we establish:

1. asymp. normality of plug-in PSMD estimators of $\phi(\alpha_0)$;
2. consistency of sieve variance estimators of plug-in PSMD $\hat{\phi}$;
3. asymp. chi-square of an optimally weighted SQLR statistic;
4. tight limiting dist. of possibly non-optimally weighted SQLR;
5. consistency of the nonparametric and the weighted bootstrap (possibly non-optimally weighted) SQLR, sieve Wald, and sieve score statistics.

Example I: Partially linear quantile IV

- $Y_3 = \beta_0' Y_2 + h_0(Y_1) + U$ where $E[1\{U \leq 0\}|X] = \gamma$.
- parameters $\alpha = (\beta, h) \in \mathcal{A} \equiv B \times \mathcal{H} \subseteq \mathcal{R}^{d_\beta} \times L^2_{f_{Y_1}}$.

Example I: Partially linear quantile IV

- $Y_3 = \beta_0' Y_2 + h_0(Y_1) + U$ where $E[1\{U \leq 0\}|X] = \gamma$.
- parameters $\alpha = (\beta, h) \in \mathcal{A} \equiv B \times \mathcal{H} \subseteq \mathcal{R}^{d_\beta} \times L^2_{f_{Y_1}}$.
- In terms of conditional moment restrictions,

$$\rho(Z, \alpha) \equiv 1\{Y_3 \leq \beta' Y_2 + h(Y_1)\} - \gamma,$$

$$m(X, \alpha_0) \equiv E[\rho(Z, \alpha_0)|X] = 0, \ a.s. - X.$$

Example I: Partially linear quantile IV

- $Y_3 = \beta_0' Y_2 + h_0(Y_1) + U$ where $E[1\{U \leq 0\}|X] = \gamma$.
- parameters $\alpha = (\beta, h) \in \mathcal{A} \equiv B \times \mathcal{H} \subseteq \mathcal{R}^{d_\beta} \times L^2_{f_{Y_1}}$.
- In terms of conditional moment restrictions,

$$\rho(Z, \alpha) \equiv 1\{Y_3 \leq \beta' Y_2 + h(Y_1)\} - \gamma,$$

$$m(X, \alpha_0) \equiv E[\rho(Z, \alpha_0)|X] = 0, \ a.s. - X.$$

- functionals of interest: $\phi(\alpha_0) = \lambda' \beta_0$, $E[\nabla h_0(Y_1)]$ or $\nabla h_0(y_1^*)$.
- Y are endogenous variables and X are the IV or conditioning instruments.

Example II: Additive NPIV

- parameters $h = (h_1, h_2) \in \mathcal{H}_1 \times \mathcal{H}_2 \subseteq L^2_{f_{Y_1}} \times L^2_{f_{Y_2}}$.

$$\rho(Z,\alpha) \equiv Y_3 - h_1(Y_1) - h_2(Y_2)$$
$$m(X,\alpha_0) \equiv E[\rho(Z,\alpha_0)|X] = 0, \; a.s. - X.$$

- functional of interest is

$$\phi(h_0) = b_1 \nabla h_{10}(y_1^*) + b_2 \nabla h_{20}(y_2^*)$$

where $b = (b_1, b_2)$ are known weights and (y_1^*, y_2^*) is a given point in the support of $Y_1 \times Y_2$.

  - E.g. Y_3 log demand and Y_i are log-prices of firm i. $\nabla h_i(y_i^*)$ is the price elasticity.

Example III: Asset Pricing

- parameters $\alpha = (\beta, h) \in \mathcal{A} \equiv B \times \mathcal{H}$.
- $m(X, \alpha_0) \equiv E[\rho(Z, \alpha_0)|X] = 0, \ a.s. - X.$, where

$$\rho(Z, \alpha) \equiv SDF(\beta, h) \times PAY(\beta, h) - PRICE,$$

  - $SDF(\beta, h)$: Stochastic Discount factor.
  - $PAY(\beta, h)$: payoff of the bond.
- External habit: $SDF(\alpha) = \left(c' \times \frac{1-h(c')}{1-h(c)}\right)^{-\beta}$ and $PAY(\alpha) = 1$, where (c, c') are today and tomorrow's consumption growth rate resp.
- Is the habit function, h, linear?
- functional of interest: $\phi(\alpha_0) = ||\nabla^2 h_0||_{L^2}$.

Example III: Asset Pricing

- parameters $\alpha = (\beta, h) \in \mathcal{A} \equiv B \times \mathcal{H}_2 \subseteq \mathcal{R} \times \mathbf{H}$.
- $m(X, \alpha_0) \equiv E[\rho(Z, \alpha_0)|X] = 0, \ a.s. - X.$, where

$$\rho(Z, \alpha) \equiv SDF(\beta, h) \times PAY(\beta, h) - PRICE,$$

  - $SDF(\beta, h)$: Stochastic Discount factor.
  - $PAY(\beta, h)$: payoff of the bond.
- Defaultable assets: $SDF(\alpha) = (c')^{-\beta}$ and $PAY(\alpha) = 1\{Y_i' \geq h(L_i')\}$ with Y' tomorrow's income.
- Will more liabilities (L') lead to higher likelihood of default for a given level of liabilities l^*?
- functional of interest: $\phi(\alpha_0) = \nabla h_0(l^*)$.

Main Features of these Examples

- General semi/non-parametric conditional moment restrictions. The residual function, ρ, can be linear (e.g., NPIV); non-linear pointwise smooth (e.g., asset pricing); pointwise non-smooth wrt parameters (e.g., NPQIV, defaultable assets);...
- The functional of interest, ϕ, can be linear (e.g., Euclidean parameter, evaluation functional, weighted integration functional,...); non-linear (e.g., L^2 norm,...).
- The $\phi(\alpha_0)$ can be regular (i.e., root-n estimable) or irregular (i.e., slower than root-n estimable). Sometimes difficult to check which is the case

Roadmap

- The Model, Functional of interest, Estimator.
- Asymptotic Normality of plug-in PSMD estimator
- Asymp. chi-square of optimally weighted SQLR.
- Consistency of sieve variance estimator
- Consistency of bootstrap generally weighted SQLR.
- CS based on asymp. critical values
- CS based on bootstrap critical values
- Monte Carlo and Empirical Illustration of NPQIV
- Related literature, Extensions.

The Model

- The Semi/nonparametric conditional moment model is:

$$m(X, \alpha_0) \equiv E[\rho(Z, \alpha_0)|X] = 0, \ a.s. - X.$$

True parameter values: $\alpha_0 = (\beta_0, h_0) \in \mathcal{A} \equiv B \times \mathcal{H}$,

- $B \subseteq \mathcal{R}^{d_\beta}$ and $\mathcal{H} \subseteq \mathbf{H}$.
- $||\alpha||_s = ||\beta||_e + ||h||_{h,s}$ where $(\mathbf{H}, ||\cdot||_{h,s})$ is a Banach Space.

And

- $\rho : \mathcal{Z} \times \mathcal{A} \to \mathcal{R}^{d_\rho}$ is the generalized residual function.
- $F_{Z|X}$ is unknown (nuisance parameter).

The Model

- The Semi/nonparametric conditional moment model is:

$$m(X, \alpha_0) \equiv E[\rho(Z, \alpha_0)|X] = 0, \ a.s. - X.$$

  True parameter values: $\alpha_0 = (\beta_0, h_0) \in \mathcal{A} \equiv B \times \mathcal{H}$,
  - $B \subseteq \mathcal{R}^{d_\beta}$ and $\mathcal{H} \subseteq \mathbf{H}$.
  - $||\alpha||_s = ||\beta||_e + ||h||_{h,s}$ where $(\mathbf{H}, ||\cdot||_{h,s})$ is a Banach Space.

  And
  - $\rho : \mathcal{Z} \times \mathcal{A} \to \mathcal{R}^{d_\rho}$ is the generalized residual function.
  - $F_{Z|X}$ is unknown (nuisance parameter).
- Let $\phi : \mathcal{A} \to \mathcal{R}$. The functional of interest is:

$$\phi_0 \equiv \phi(\alpha_0) \in \mathcal{R}.$$

Difficulty in Estimation and Inference of the Model

- Let $\Sigma(X)$ be a positive definite weighting matrix; let $Q(\alpha) \equiv E[m(X,\alpha)'\Sigma^{-1}(X)m(X,\alpha)]$.

Difficulty in Estimation and Inference of the Model

- Let $\Sigma(X)$ be a positive definite weighting matrix; let $Q(\alpha) \equiv E[m(X,\alpha)'\Sigma^{-1}(X)m(X,\alpha)]$.
- Estimation of α_0 could be (globally) ill-posed, i.e.,
  - $||\alpha - \alpha_0||_s$ is not continuous w.r.t. the criterion $Q(\alpha)$.
- That is, $\exists(\alpha_k) \subseteq \mathcal{A}$ such that $Q(\alpha_k) \to 0$ *but* $\liminf_k ||\alpha_k - \alpha_0||_s > 0$. "Identifiable uniqueness" fails!

Difficulty in Estimation and Inference of the Model

- Let $\Sigma(X)$ be a positive definite weighting matrix; let $Q(\alpha) \equiv E[m(X,\alpha)'\Sigma^{-1}(X)m(X,\alpha)]$.
- Estimation of α_0 could be (globally) ill-posed, i.e.,
  - $||\alpha - \alpha_0||_s$ is not continuous w.r.t. the criterion $Q(\alpha)$.
- That is, $\exists(\alpha_k) \subseteq \mathcal{A}$ such that $Q(\alpha_k) \to 0$ *but* $\liminf_k ||\alpha_k - \alpha_0||_s > 0$. "Identifiable uniqueness" fails!
- NPIV: $E[Y_2 - h_0(Y_1)|X] = 0$.
  $||h - h_0||_s^2 = E[(h(Y_1) - h_0(Y_1))^2]$.
  $Q(h) = E[(E[h(Y_1) - h_0(Y_1)|X])^2]$.

Difficulty in Estimation and Inference of the Model

- Let $\Sigma(X)$ be a positive definite weighting matrix; let $Q(\alpha) \equiv E[m(X,\alpha)'\Sigma^{-1}(X)m(X,\alpha)]$.
- Estimation of α_0 could be (globally) ill-posed, i.e.,
  - $||\alpha - \alpha_0||_s$ is not continuous w.r.t. the criterion $Q(\alpha)$.
- That is, $\exists(\alpha_k) \subseteq \mathcal{A}$ such that $Q(\alpha_k) \to 0$ *but* $\liminf_k ||\alpha_k - \alpha_0||_s > 0$. "Identifiable uniqueness" fails!
- NPIV: $E[Y_2 - h_0(Y_1)|X] = 0$.
  $||h - h_0||_s^2 = E[(h(Y_1) - h_0(Y_1))^2]$.
  $Q(h) = E[(E[h(Y_1) - h_0(Y_1)|X])^2]$.
- One has to "regularize" the problem via methods of sieves and/or penalization.

The PSMD Estimator

- PSMD (Chen-Pouzo '09, 12): $\hat{\alpha}_n$ solves:

$$\inf_{\mathcal{A}_n} n^{-1} \sum_{i=1}^{n} \{\hat{m}(X_i, \alpha)' [\hat{\Sigma}(X_i)]^{-1} \hat{m}(X_i, \alpha)\} + \lambda_n Pen(h)$$

- $\widehat{m}(X, \alpha)$ is any consistent nonparametric estimator of $m()$.
- $\hat{\Sigma}$ is any consistent nonparametric estimators of $\Sigma(X)$.
- $\mathcal{A}_n \equiv B \times \mathcal{H}_n$, $\mathcal{H}_n$ is a finite dimensional linear sieve.
- $Pen(\cdot) > 0$ is a penalty; $\lambda_n \to 0$ fast.

The PSMD Estimator

- PSMD (Chen-Pouzo '09, 12): $\hat{\alpha}_n$ solves:

$$\inf_{\mathcal{A}_n} n^{-1} \sum_{i=1}^{n} \{\hat{m}(X_i, \alpha)' [\hat{\Sigma}(X_i)]^{-1} \hat{m}(X_i, \alpha)\} + \lambda_n Pen(h)$$

- $\widehat{m}(X, \alpha)$ is any consistent nonparametric estimator of $m()$.
- $\hat{\Sigma}$ is any consistent nonparametric estimators of $\Sigma(X)$.
- $\mathcal{A}_n \equiv B \times \mathcal{H}_n$, $\mathcal{H}_n$ is a finite dimensional linear sieve.
- $Pen(\cdot) > 0$ is a penalty; $\lambda_n \to 0$ fast.
- Plug-in PSMD estimator of ϕ_0: $\hat{\phi}_n = \phi(\hat{\alpha}_n)$.

Examples of Sieves

- Finite-dimensional linear sieves $\mathcal{H}_n$ is of the form $\{h(.) = \sum_{k=1}^{k_n} a_k q_k(.) : a_k \in \mathcal{R}\}$, with $q_k(.)$ a known basis, e.g.

  1. Polynomials: $q_k(Y) = Y^k$
  2. Sine (or Cosine): $q_k(Y) = Sin(k\pi Y)$ $(or\ Cos(k\pi Y))$
  3. B-Splines: $q_k(X) = 2^{k_{1n}/2} B_r(2^{k_{1n}} Y - k)$
  4. Polynomial splines, wavelets, finite elements, Hermite poly., Laguerre poly., radial basis,...

Convergence Rates in "Weak" and "Strong" Norms

- For nonlinear illposed inverse problems, Chen-Pouzo (12) establish convergence rates in "weak" and "strong' norms.
- "Strong Norm": $||\cdot||_s = |\cdot|_e + ||\cdot||_{h,s}$
  - $||\cdot||_{h,s}$ is a Banach norm, e.g., L^2.
  - Estimation of α_0 may be ill-posed under this strong norm.

Convergence Rates in "Weak" and "Strong" Norms

- For nonlinear illposed inverse problems, Chen-Pouzo (12) establish convergence rates in "weak" and "strong' norms.
- "Strong Norm": $||\cdot||_s = |\cdot|_e + ||\cdot||_{h,s}$
  - $||\cdot||_{h,s}$ is a Banach norm, e.g., L^2.
  - Estimation of α_0 may be ill-posed under this strong norm.
- "Weak Norm": $||\cdot||$ where

$$||\alpha - \alpha_0||^2 = E\left[\frac{dm(X,\alpha_0)}{d\alpha}[\alpha - \alpha_0]'\Sigma(X)^{-1}\frac{dm(X,\alpha_0)}{d\alpha}[\alpha - \alpha_0]\right]$$

with

$$\frac{dm(X,\alpha_0)}{d\alpha}[\alpha - \alpha_0] = \frac{dm(X,\alpha_0)}{d\beta}(\beta - \beta_0) + \frac{dm(X,\alpha_0)}{dh}[h - h_0]$$

$$\text{and} \quad \frac{dm(X,\alpha_0)}{dh}[h - h_0] \equiv \lim_{t\to 0}\frac{E[\rho(Z,(\beta_0, h_0 + t(h - h_0)))]}{t}$$

Convergence Rates in "Weak" and "Strong" Norms

- For nonlinear illposed inverse problems, Chen-Pouzo (12) establish convergence rates in "weak" and "strong' norms.
- "Strong Norm": $||\cdot||_s = |\cdot|_e + ||\cdot||_{h,s}$
  - $||\cdot||_{h,s}$ is a Banach norm, e.g., L^2.
  - Estimation of α_0 may be ill-posed under this strong norm.
- "Weak Norm": $||\cdot||$ where

$$||\alpha - \alpha_0||^2 = E\left[\frac{dm(X,\alpha_0)}{d\alpha}[\alpha-\alpha_0]'\Sigma(X)^{-1}\frac{dm(X,\alpha_0)}{d\alpha}[\alpha-\alpha_0]\right]$$

with

$$\frac{dm(X,\alpha_0)}{d\alpha}[\alpha-\alpha_0] = \frac{dm(X,\alpha_0)}{d\beta}(\beta-\beta_0) + \frac{dm(X,\alpha_0)}{dh}[h-h_0]$$

$$and\ \ \frac{dm(X,\alpha_0)}{dh}[h-h_0] \equiv \lim_{t\to 0}\frac{E[\rho(Z,(\beta_0, h_0 + t(h-h_0)))]}{t}$$

- Key: $E[m(X,\alpha)'\Sigma(X)^{-1}m(X,\alpha)] \asymp ||\alpha-\alpha_0||^2$ locally.

Local Neighborhood

- Our proof of asymp. normality of $\widehat{\phi}_n$ relies on local linear approximations.
- Given the consistency and convergence rate of PSMD estimator $\hat{\alpha}_n$ (Chen-Pouzo, 12), we can restrict to a shrinking $||\cdot||_s-$neighborhood around α_0. Let

$$\mathcal{A}_{os} \equiv \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_s = o(1) \textit{ and } ||\alpha - \alpha_0|| = o(1)\}$$

$$\mathcal{N}_{os} \equiv \{\alpha \in \mathcal{A} : ||\alpha - \alpha_0||_s = O(\delta_{s,n}) \textit{ and } ||\alpha - \alpha_0|| = O(\delta_n)\}$$

$$\mathcal{N}_{osn} \equiv \mathcal{A}_n \cap \mathcal{N}_{os}$$

where δ_n and $\delta_{s,n}$ are the convergence rates under "weak" and "strong" norm, respectively.

Local Characterization of $\phi()$

- Let $\mathbf{V} = clsp_{||.||}\{\mathcal{A}_{os} - \{\alpha_0\}\}$.
- $\frac{d\phi(\alpha_0)}{d\alpha}[v]$ is a linear functional $\mathbf{V} \to L^2(f_X)$.

Local Characterization of $\phi()$

- Let $\mathbf{V} = clsp_{||\cdot||}\{\mathcal{A}_{os} - \{\alpha_0\}\}$.
- $\frac{d\phi(\alpha_0)}{d\alpha}[v]$ is a linear functional $\mathbf{V} \to L^2(f_X)$.
- Under mild conditions,

$$\sqrt{Asym.\ Var.(n^{1/2}[\phi(\hat{\alpha}_n) - \phi(\alpha_0)])} \asymp ||v^*|| \equiv \sup_{v \in \mathbf{V}} \frac{|\frac{d\phi(\alpha_0)}{d\alpha}[v]|}{||v||}.$$

Local Characterization of $\phi()$

- Let $\mathbf{V} = clsp_{||\cdot||}\{\mathcal{A}_{os} - \{\alpha_0\}\}$.
- $\frac{d\phi(\alpha_0)}{d\alpha}[v]$ is a linear functional $\mathbf{V} \to L^2(f_X)$.
- Under mild conditions,

$$\sqrt{Asym.\ Var.(n^{1/2}[\phi(\hat{\alpha}_n) - \phi(\alpha_0)])} \asymp ||v^*|| \equiv \sup_{v \in \mathbf{V}} \frac{|\frac{d\phi(\alpha_0)}{d\alpha}[v]|}{||v||}.$$

- If $||v^*|| < \infty$, $\phi_0 \equiv \phi(\alpha_0)$ could be root-n estimable, *regular*.
- If $||v^*|| = \infty$ then $\phi_0 \equiv \phi(\alpha_0)$ could at best estimated at a slower than root-n rate, *irregular*.

Local Characterization of $\phi()$

- Our solution: regardless of whether $||v^*||$ is finite or infinite, it can always be approximated by a sequence of finite $||v_n^*||$ that are computed in sieve spaces.
- Since $\mathbf{V}_n \equiv \Pi_n \mathbf{V} = clsp\{\mathcal{A}_n\} - \{\alpha_{0,n}\}$ is of finite dimension, $\frac{d\phi(\alpha_0)}{d\alpha}[v]$ is linear and hence is *always bounded* in $\mathbf{V}_n$, i.e.,

$$||v_n^*|| = \sup_{v \in \mathbf{V}_n} \frac{|\frac{d\phi(\alpha_0)}{d\alpha}[v]|}{||v||} < \infty.$$

- $||v_n^*||$ can be computed in closed form via population least squares regression.

NPQIV Example

- NPQIV: $E[1(Y_2 \leq h_0(Y_1))|X] = \gamma$. Let $\phi_0 = \int \nabla h_0(y)w(y)dy = E_w[\nabla h_0]$ where w is a weight function.
- Let $h \mapsto T[h](X) \equiv E[f_{U|Y_1,X}(0; Y_1, X)h(Y_1)|X]$, where $T : \mathcal{H} \subseteq \mathbf{H} \to L^2(\mathcal{X}, f_x)$.
- Assume T is compact, then
  - $\exists \{\mu_k, \psi_k(Y), b_k(X)\}_k$ such that $T[\psi_k](X) = \mu_k b_k(X)$.

NPQIV Example

- It can be shown that $||v^*||^2$, is proportional to,

$$\sum_{k=1}^{\infty} \mu_k^{-2} (E_w[\nabla \psi_k])^2.$$

If it is ∞, then ϕ_0 is not root-n estimable.

NPQIV Example

- It can be shown that $||v^*||^2$, is proportional to,

$$\sum_{k=1}^{\infty} \mu_k^{-2} (E_w[\nabla \psi_k])^2.$$

  If it is ∞, then ϕ_0 is not root-n estimable.

- If $(E_w[\nabla \psi_k])^2 \leq Const. < \infty$, then:
  - The faster $\mu_j \to 0$ (i.e., the "higher the ill-posedness"), the harder is to get root-n.
  - Even if $\mu_j \geq const > 0$ (e.g. no endogeneity) it may not be root-n estimable (depends on summability of $(E_w[\nabla \psi_k])^2$, e.g., on the weights w).

NPQIV Example

- It can be shown that $||v^*||^2$, is proportional to,

$$\sum_{k=1}^{\infty} \mu_k^{-2} (E_w[\nabla \psi_k])^2.$$

  If it is ∞, then ϕ_0 is not root-n estimable.
- If $(E_w[\nabla \psi_k])^2 \leq Const. < \infty$, then:
  - The faster $\mu_j \to 0$ (i.e., the "higher the ill-posedness"), the harder is to get root-n.
  - Even if $\mu_j \geq const > 0$ (e.g. no endogeneity) it may not be root-n estimable (depends on summability of $(E_w[\nabla \psi_k])^2$, e.g., on the weights w).
- Hard to know ex-ante if the functional is $\sqrt{n}$-estimable. It is desirable to perform inference without knowing this.

NPQIV Example

- Even if $||v^*|| = \infty$, we still have $||v_n^*|| < \infty$, where

$$||v_n^*||^2 \asymp \sum_{k=1}^{k(n)} \mu_k^{-2} (E_w[\nabla \psi_k])^2;$$

NPQIV Example

- Even if $||v^*|| = \infty$, we still have $||v_n^*|| < \infty$, where

$$||v_n^*||^2 \asymp \sum_{k=1}^{k(n)} \mu_k^{-2} (E_w[\nabla \psi_k])^2;$$

- $||v_n^*||$ is non-decreasing in $k(n)$.

Asymp. Normality of Plug-in PSMD Estimator

- Let $||v_n^*||_{sd}^2 = Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n S_{n,i}\right) \asymp ||v_n^*||^2$. Assume:

$$\left|\phi(\hat{\alpha}_n) - \phi(\alpha_0) - \frac{\partial\phi(\alpha_0)}{\partial\alpha}[\hat{\alpha}_n - \alpha_0]\right| = o_p\left(n^{-\frac{1}{2}}||v_n^*||\right), \quad (1)$$

$$\left|\frac{\partial\phi(\alpha_0)}{\partial\alpha}[\alpha_{0,n} - \alpha_0]\right| = o\left(n^{-\frac{1}{2}}||v_n^*||\right), \quad (2)$$

- **Theorem 1.** Under (1), (2) and some mild conditions,

$$\sqrt{n}\frac{\phi(\hat{\alpha}_n) - \phi(\alpha_0)}{||v_n^*||_{sd}} = -\sqrt{n}\mathbb{Z}_n + o_p(1) \Rightarrow N(0,1),$$

where $\mathbb{Z}_n = \frac{1}{n}\sum_{i=1}^n \frac{S_{n,i}}{||v_n^*||_{sd}}$,

$$S_{n,i} = \left(\frac{dm(X_i,\alpha_0)}{d\alpha}[v_n^*]\right)' \Sigma(X_i)^{-1}\rho(Z_i,\alpha_0)$$

Asymp. Normality of Plug-in PSMD Estimator

- For iid data,
  $$||v_n^*||_{sd}^2 = Var\left(\left(\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*]\right)' \Sigma(X)^{-1}\rho(Z,\alpha_0)\right) \asymp ||v_n^*||^2$$

(1) Root-n Case: $||v^*|| < \infty$ then:

- $||v_n^*|| \to ||v^*||$; root-n asymptotic normality.

(2) Slower than root-n Case: $||v^*|| = \infty$ then:

- $||v_n^*|| \to \infty$; slower than root-n asymptotic normality.

NPQIV Example

- By Normality Theorem, we obtain:

$$\sqrt{n}\frac{\phi(\widehat{h}_n) - \phi(h_0)}{[\gamma(1-\gamma)]^{1/2}\sigma_n} \Rightarrow N(0,1),$$

$$\sigma_n^2 = \frac{\partial\phi(h_0)}{\partial h}[q^{k(n)}(\cdot)]'\Omega^{-1}\frac{\partial\phi(h_0)}{\partial h}[q^{k(n)}(\cdot)],$$

$$\Omega = E\left(E[f_{U|Y_1,X}(0)q^{k(n)}(Y_1)|X]E[f_{U|Y_1,X}(0)q^{k(n)}(Y_1)|X]'\right)$$

- For the evaluation functional $\phi(h) = h(\overline{y}_1)$ we have $\frac{\partial\phi(h_0)}{\partial h}[q^{k(n)}(\cdot)] = q^{k(n)}(\overline{y}_1)$.
- For the weighted integration functional $\phi(h) = \int a(y_1)h(y_1)dy_1$ we have $\frac{\partial\phi(h_0)}{\partial h}[q^{k(n)}(\cdot)] = \int a(y_1)q^{k(n)}(y_1)dy_1$.

Stochastic Boundedness of SQLR

- $\widehat{Q}_n(\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}(X_i,\alpha)$.
- $\mathcal{A}_n(\phi_0) = \{\alpha \in \mathcal{A}_n : \phi(\alpha) = \phi_0\}$.
- $\widehat{QLR}_n(\phi_0) = \inf_{\mathcal{A}_n(\phi_0)} \widehat{Q}_n(\alpha) - \widehat{Q}_n(\widehat{\alpha}_n)$.

Theorem

Let $\widehat{\alpha}_n$ be the PSMD estimator. Then, under the null hypothesis of $\phi(\alpha_0) = \phi_0$,

$$\widehat{QLR}_n(\phi_0) = \frac{||v_n^*||_{sd}^2}{||v_n^*||^2}\left(\sqrt{n}\mathbb{Z}_n\right)^2 + o_{P_{Z^\infty}}(1) = O_p(1).$$

Chi-square of Optimally Weighted SQLR

- Assume $\hat{\Sigma}(X)$ approximates $\Sigma_0(X) \equiv E[\rho(Z, \alpha_0)^2 | X]$.

Chi-square of Optimally Weighted SQLR

- Assume $\hat{\Sigma}(X)$ approximates $\Sigma_0(X) \equiv E[\rho(Z, \alpha_0)^2 | X]$.
- Then $||v_n^*||^2 = ||v_n^*||_{sd}^2$

Chi-square of Optimally Weighted SQLR

- Assume $\hat{\Sigma}(X)$ approximates $\Sigma_0(X) \equiv E[\rho(Z, \alpha_0)^2 | X]$.
- Then $||v_n^*||^2 = ||v_n^*||_{sd}^2$
- Theorem: Let $\widehat{\alpha}_n$ be the optimally weighted PSMD estimator.

  Then, under the null hypothesis of $\phi(\alpha_0) = \phi_0$,

$$\widehat{QLR}_n(\phi_0) = \left(\sqrt{n}\mathbb{Z}_n\right)^2 + o_{P_{Z^\infty}}(1) \Rightarrow \chi_1^2.$$

Chi-square Approximation: Confidence Intervals

- Let $c_{\chi^2}(\tau)$ the $(1-\tau)$ quantile of the χ^2 distribution. Then the $(1-\tau)$ SQLR based CS for the functional $\phi(\alpha)$ is

$$\left\{ \phi \in \mathcal{R} : n \left(\inf_{\mathcal{A}_n(\phi)} \widehat{Q}_n(\alpha) - \widehat{Q}_n(\widehat{\alpha}_n) \right) \leq c_{\chi^2}(\tau) \right\}$$

Chi-square Approximation: Confidence Intervals

- Let $c_{\chi^2}(\tau)$ the $(1-\tau)$ quantile of the χ^2 distribution. Then the $(1-\tau)$ SQLR based CS for the functional $\phi(\alpha)$ is

$$\left\{\phi \in \mathcal{R} : n\left(\inf_{\mathcal{A}_n(\phi)} \widehat{Q}_n(\alpha) - \widehat{Q}_n(\widehat{\alpha}_n)\right) \leq c_{\chi^2}(\tau)\right\}$$

- This allows to compute the CS, *without* computing $||v_n^*||$.
- It is valid regardless of whether the functional is root-n estimable or not.

Sieve Estimator of the Asymptotic Variance

- After specifying the finite dimensional linear sieve space $\mathbf{V}_n$, v_n^* and $||v_n^*|| = \sup_{v \in \mathbf{V}_n} \frac{|\frac{d\phi(\alpha_0)}{d\alpha}[v]|}{||v||}$ can be computed explicitly.
- $||v_n^*||_{st}$ can be consistently estimated via a standard plug-in sieve estimation procedure.
- See paper.

Bootstrap SQLR

- Let $\widehat{Q}_n^B(\alpha) \equiv n^{-1} \sum_{i=1}^n \widehat{m}^B(X_i, \alpha)' \widehat{\Sigma}(X_i)^{-1} \widehat{m}^B(X_i, \alpha)$ be bootstrap version of $\widehat{Q}_n(\alpha)$, where $\widehat{m}^B(X, \alpha)$ is computed in the same way as the original-sample $\widehat{m}(X, \alpha)$ except that we use the weighted residual $\omega_j \rho(Z_j, \alpha)$ instead of $\rho(Z_j, \alpha)$.

Bootstrap SQLR

- Let $\widehat{Q}_n^B(\alpha) \equiv n^{-1}\sum_{i=1}^n \widehat{m}^B(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\widehat{m}^B(X_i,\alpha)$ be bootstrap version of $\widehat{Q}_n(\alpha)$, where $\widehat{m}^B(X,\alpha)$ is computed in the same way as the original-sample $\widehat{m}(X,\alpha)$ except that we use the weighted residual $\omega_j\rho(Z_j,\alpha)$ instead of $\rho(Z_j,\alpha)$.
- For example, if $\widehat{m}(X,\alpha)$ is a series LS estimator of $m(X,\alpha) = E[\rho(Z,\alpha)|X]$, then

$$\widehat{m}^B(X,\alpha) = p^{J_n}(X)'(P'P)^{-}\sum_{j=1}^n p^{J_n}(X_j)\omega_j\rho(Z_j,\alpha).$$

- Weighted bootstrap: $(\omega_i)_i \sim iid$, $E[\omega_1] = 1$ and $\sigma_\omega^2 = Var(\omega_1) < \infty$. Or Nonparametric Bootstrap: $(\omega_1,...,\omega_n) \sim MN(1/n,...,1/n;n)$.

Bootstrap SQLR

- Result: under null hypothesis,

$$n \frac{\inf_{\mathcal{A}_n(\hat{\phi}_n)} \widehat{Q}_n^B(\alpha) - \widehat{Q}_n^B(\widehat{\alpha}_n^B)}{\sigma_\omega^2} \approx \left(\frac{||v_n^*||_{st}}{||v_n^*||} \sqrt{n} \frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega} \right)^2 . \quad (3)$$

where $\widehat{\alpha}_n^B$ and $\widehat{\alpha}_n$ the unconstrained PSMD w/ $\widehat{Q}_n^B$ and $\widehat{Q}_n$ resp. and $\hat{\phi}_n = \phi(\widehat{\alpha}_n)$, and

$$\mathbb{Z}_n^{\omega-1} \equiv n^{-1} \sum_{i=1}^{n} (\omega_i - 1) \left(\frac{dm(X_i, \alpha_0)}{d\alpha} [\frac{v_n^*}{||v_n^*||_{st}}] \right)' \Sigma(X_i)^{-1} \rho(Z_i, \alpha_0)$$

Bootstrap SQLR

- Result: under null hypothesis,

$$n\frac{\inf_{\mathcal{A}_n(\hat{\phi}_n)} \widehat{Q}_n^B(\alpha) - \widehat{Q}_n^B(\widehat{\alpha}_n^B)}{\sigma_\omega^2} \approx \left(\frac{||v_n^*||_{st}}{||v_n^*||} \sqrt{n} \frac{\mathbb{Z}_n^{\omega-1}}{\sigma_\omega} \right)^2 . \quad (3)$$

  where $\widehat{\alpha}_n^B$ and $\widehat{\alpha}_n$ the unconstrained PSMD w/ $\widehat{Q}_n^B$ and $\widehat{Q}_n$ resp. and $\hat{\phi}_n = \phi(\widehat{\alpha}_n)$, and

$$\mathbb{Z}_n^{\omega-1} \equiv n^{-1} \sum_{i=1}^{n} (\omega_i - 1) \left(\frac{dm(X_i, \alpha_0)}{d\alpha} [\frac{v_n^*}{||v_n^*||_{st}}] \right)' \Sigma(X_i)^{-1} \rho(Z_i, \alpha_0)$$

- If $\Sigma(X) = \Sigma_0(X)$ then it is chi-square distributed.
- If $\Sigma(X) \neq \Sigma_0(X)$ then the LHS in (3) has the same asy dist. as that of $n\{\inf_{\mathcal{A}_n(\phi_0)} \widehat{Q}_n(\alpha) - \widehat{Q}_n(\widehat{\alpha}_n)\}$.

Illustrative Example: Pointwise Asymptotic Normality of NPQIV

- $Y_1 = h_0(Y_2) + U$, $E[1\{U \leq 0\}|X] = \gamma$, $\phi(\alpha) = h(y_2^*)$.
- $\mathcal{H}_{k(n)} = \{h : h(y_2) = \sum_{k=1}^{k(n)} b_k B_k(y_2)\}$.
- Let $h \mapsto T[h](X) \equiv E[f_{U|YX}(0; Y, X)h(Y_2)|X]$, and $T^*T[\psi_j] = \mu_j^2 \psi_j$ and $B^{k(n)} = (\psi_1, \ldots, \psi_{k(n)})'$.
- $\Sigma = \Sigma_0 = \gamma(1 - \gamma)$.

Illustrative Example: Pointwise Asymptotic Normality of NPQIV

- $Y_1 = h_0(Y_2) + U$, $E[1\{U \leq 0\}|X] = \gamma$, $\phi(\alpha) = h(y_2^*)$.
- $\mathcal{H}_{k(n)} = \{h : h(y_2) = \sum_{k=1}^{k(n)} b_k B_k(y_2)\}$.
- Let $h \mapsto T[h](X) \equiv E[f_{U|YX}(0; Y, X)h(Y_2)|X]$, and $T^* T[\psi_j] = \mu_j^2 \psi_j$ and $B^{k(n)} = (\psi_1, \ldots, \psi_{k(n)})'$.
- $\Sigma = \Sigma_0 = \gamma(1 - \gamma)$.
- Then: $||v_n^*||^2 = \sum_{k=1}^{k(n)} \mu_k^{-2} |\psi_k(y_2^*)|^2 \asymp \sum_{k=1}^{k(n)} \mu_k^{-2}$.

Illustrative Example: Pointwise Asymptotic Normality of NPQIV

- $Y_1 = h_0(Y_2) + U$, $E[1\{U \leq 0\}|X] = \gamma$, $\phi(\alpha) = h(y_2^*)$.
- $\mathcal{H}_{k(n)} = \{h : h(y_2) = \sum_{k=1}^{k(n)} b_k B_k(y_2)\}$.
- Let $h \mapsto T[h](X) \equiv E[f_{U|YX}(0; Y, X)h(Y_2)|X]$, and $T^* T[\psi_j] = \mu_j^2 \psi_j$ and $B^{k(n)} = (\psi_1, \ldots, \psi_{k(n)})'$.
- $\Sigma = \Sigma_0 = \gamma(1-\gamma)$.
- Then: $||v_n^*||^2 = \sum_{k=1}^{k(n)} \mu_k^{-2} |\psi_k(y_2^*)|^2 \asymp \sum_{k=1}^{k(n)} \mu_k^{-2}$.
- Bias order:

$$\sqrt{n} \frac{\Pi_n h_0(y_2^*) - h_0(y_2^*)}{||v_n^*||} \leq \sqrt{n} \frac{||h_0 - \Pi_n h_0||_\infty}{||v_n^*||} = o(1).$$

An Example: Pointwise Asy Normality of NPQIV

- ASYMPTOTIC NORMALITY:

$$\sqrt{n}\frac{\widehat{h}_n(y_2^*) - h_0(y_2^*)}{||v_n^*||\sqrt{\gamma(1-\gamma)}} \Rightarrow N(0,1)$$

An Example: Pointwise Asy Normality of NPQIV

- Asymptotic Normality:

$$\sqrt{n}\frac{\widehat{h}_n(y_2^*) - h_0(y_2^*)}{||v_n^*||\sqrt{\gamma(1-\gamma)}} \Rightarrow N(0,1)$$

- Confidence Region:

$$\left\{\beta \in \mathcal{R} : \beta = \phi(\alpha) \text{ and } n\left(\widehat{Q}_n(\alpha) - \widehat{Q}_n(\widehat{\alpha}_n)\right) \leq c_\chi(\tau)\right\}$$

where $c_\chi(\tau)$ is the τ-th quantile of the χ^2 distribution.

Monte Carlo: NPQIV model

- The model: $E[1\{Y_1 \leq h_0(Y_2)\}|X] = \gamma$.

$$Y_1 = h_0(Y_2) + \sqrt{0.0025}U$$
$$U = -\Phi^{-1}\left(\frac{E[h_0|X] - h_0(Y_2)}{25} + \gamma\right) + V$$
$$V \sim N(0,1).$$

- $(Y_2, X) \sim N(\mu, \Sigma)$ where μ and Σ are the sample estimators of BCK Engel curve dataset and $\rho = 0.75$.
- $h_0(Y_2) = \Phi\left(\frac{Y_2 - \mu_{Y_2}}{\sigma_{Y_2}}\right)$.
- Only present $\gamma = 0.5$; see CP (08a, 08b) for estimates of other quantiles.
- Sample size: $n \in \{250, 500\}$. MC repetitions: 1000.

Monte Carlo: NPQIV model

- Parameters of interest:

$$PAN : \phi_0 = \phi(h_0) = \nabla h_0(\mu_{Y_2}) = \frac{1}{\sqrt{2\pi}\sigma_{Y_2}}$$

$$WAD : \phi_0 = \phi(h_0) = E[W(Y_2)\nabla h_0(Y_2)] \approx 0.43$$

$$with\ W(Y_2) = 75(\Phi(Y_2))^3(1-\Phi(Y_2))^3$$

- $\mathcal{H}_n$: P-spline basis w/ $k(n) \in \{6,7\}$. $P(h)$: $||\nabla^2 h||^2_{L^2}$ w/ $\lambda_n \in \{0.0001, 0.0005\}$.
- p^{J_n}: P-Spline(6,8).

Monte Carlo Results

		BIAS	Std. Dev.
PAN	$n = 250$	0.066	0.236
	$n = 500$	0.057	0.133
WAD	$n = 250$	0.026	0.117
	$n = 500$	0.020	0.087

		95% (MC)	95% (χ^2)
WAD (Effi.)	$n = 250$	(0.35,0.48)	(0.36,0.45)
	$n = 500$	(0.37,0.46)	(0.38,0.45)

Closely Related Literature

- **Review on NPIV** ($E[Y_2 - h_0(Y_1)|X] = 0$): Carrasco-Florens-Renault (07), Horowitz (11).
- **Nonparametric consistency**: Newey-Powell (03); Chen-Pouzo (12).
- **Nonparametric convergence rate**: Chen-Pouzo (12); Horowitz-Lee (07, NPQIV); many on NPIV (e.g., Hall-Horowitz (05), Blundell-Chen-Kristensen (07), Darolles-Fan-Florens-Renault (11),...)
- **Root-n normality and efficiency of regular functionals**: Ai-Chen (03), Chen-Pouzo (09), Otsu (11).
- **Chi-square CS for regular functionals**: Chen-Pouzo (09).
- **Asy normality of unknown functions**: CFR (07, NPIV), Horowitz (07, NPIV), Horowitz-Lee (09, NPIV for CS); Gagliardini-Scaillet (12, NPQIV-PAN).

Conclusions and Extensions

- For functionals of general semi/non-parametric conditional moment models, we derive the asy normality of plug-in PSMD, asy chi-square of the optimally weighted SQLR, asy tight dist of general weighted SQLR.
- Due to ill-posedness, it is difficult to know if a functional is regular or irregular. Our paper provides simple valid inference procedures regardless of whether the functional is root-n estimable or not.
- We show the validity of weighted and nonparametric bootstrap SQLR, sieve Wald, sieve score statistics under virtually the same conditions as those for original-sample statistics.

Conclusions and Extensions

- For functionals of general semi/non-parametric conditional moment models, we derive the asy normality of plug-in PSMD, asy chi-square of the optimally weighted SQLR, asy tight dist of general weighted SQLR.
- Due to ill-posedness, it is difficult to know if a functional is regular or irregular. Our paper provides simple valid inference procedures regardless of whether the functional is root-n estimable or not.
- We show the validity of weighted and nonparametric bootstrap SQLR, sieve Wald, sieve score statistics under virtually the same conditions as those for original-sample statistics.
- Extensions: Chen, Pouzo and Tamer (2011) consider bootstrap SQLR statistic for partially identified semi/nonparametric conditional moment models.

Thanks!