Optimal Sup-norm Rate, Adaptive Estimation, and Inference on NPIV

Xiaohong Chen (Yale) and Tim Christensen (NYU)

Cemmap Celebration Conference | Andrew's Birthday Conference
November 14-16, 2014

Introduction (1)

- we consider nonparametric instrumental variables (NPIV) regression:

$$\begin{aligned} Y_i &= h_0(X_i) + u_i \\ E[u_i|X_i] &\neq 0 \\ E[u_i|W_i] &= 0 \end{aligned}$$

  - endogeneity is an important issue in economics
  - nonparametric in h_0 avoids functional form misspecification

- h_0 is identified via the conditional moment restriction

$$E[Y_i|W_i] = E[h_0(X_i)|W_i]$$

  - this "smoothes out" features of h_0, making h_0 difficult to recover
  - NIPV is an ill-posed inverse problem with unknown operator

Introduction (2)

- there is a large and growing literature on NPIV:
  1. identification/consistency: Newey & Powell (03); Carrasco, Florens & Renault (07); Andrews (11);...
  2. convergence rates in L^2 norm: Hall & Horowitz (05); Blundell, Chen & Kristensen (BCK, 07); Chen & Reiß (11); Darolles, Fan, Florens & Renault (11);...
  3. almost rate-adaptive estimation in L^2: Horowitz (14)
  4. almost rate-adaptive estimation of linear functionals: Breunig & Johannes (13)
  5. inference on linear functionals of h_0: Ai & Chen (AC, 03, 07); Carrasco, Florens & Renault (07) Horowitz & Lee (13)
  6. inference on nonlinear functionals of h_0: Chen & Pouzo (14)
  7. testing: Horowitz (12); Canay, Santos & Shaikh (13); Breunig (13);...
  8. partial identification: Santos (12); Freyberger & Horowitz (13);...
- all the existing published results on NPIV are based on L^2 norm.

Contributions of this paper

1. we derive the upper bound on sup-norm convergence rates for general sieve NPIV estimators.

2. we derive minimax lower bounds in sup-norm loss over Holder class of functions for NPIR (nonparametric indirect regression) and NPIV.

3. we show that spline and wavelet sieve NPIV estimators attain the sup-norm minimax lower bounds, and hence attain the optimal sup-norm convergence rates.

4. we introduce a data-driven procedure for choosing the dimension of the sieve NPIV that is sup-norm rate-adaptive

5. we provide inference theory for plug-in sieve NPIV estimators of nonlinear functionals of h_0 under mild conditions.

- An application: inference on exact consumer surplus in nonparametric demand estimation when both price and income are endogenous.

Parametric vs nonparametric IV

- parametric IV model

$$\begin{aligned} Y_i &= X_i'\beta_0 + u_i \\ E[u_i X_i] &\neq 0 \\ E[u_i W_i] &= 0 \end{aligned}$$

- identified if $\mathrm{rank}(E[X_i W_i']) = \dim(\beta_0)$

- nonparametric IV model

$$\begin{aligned} Y_i &= h_0(X_i) + u_i \\ E[u_i | X_i] &\neq 0 \\ E[u_i | W_i] &= 0 \end{aligned}$$

- identified if $h \mapsto E[h(X_i)|W_i = \cdot]$ is injective

Parametric vs sieve nonparametric IV

- A parametric IV model can be estimated via 2SLS:

$$\widehat{\beta} = [X'W(W'W)^{-1}W'X]^{-1}[X'W(W'W)^{-1}W'Y]$$

- NP (03), AC (03), BCK (07): A nonparametric IV model can be estimated via sieve NPIV, i.e., 2SLS on basis functions

$$\begin{aligned} \widehat{h}(x) &= \psi^J(x)'\widehat{c} \\ \widehat{c} &= [\Psi'B(B'B)^{-1}B'\Psi]^{-1}\Psi'B(B'B)^{-1}B'Y \end{aligned}$$

$$\begin{aligned} \psi^{J}(x) &= (\psi_{J1}(x), \ldots, \psi_{JJ}(x))', \ \Psi = (\psi^J(X_1), \ldots, \psi^J(X_n))' \\ b^{K}(w) &= (b_{K1}(w), \ldots, b_{KK}(w))', \ B = (b^K(W_1), \ldots, b^K(W_n))' \end{aligned}$$

- $K \geq J$, with $J =$ sieve number of endogenous regressors (the key smoothing parameter), $K =$ sieve number of instruments.
- Horowitz (11): modified sieve NPIV: $K = J$ and $b^K = \psi^J$=orthognormal series of $L^2([0,1]^d)$.

Outline

1. Optimal sup-norm rates

2. Sup-norm rate-adaptive estimation

3. MC study I: Adaptive estimation procedure

4. Application: Asymptotic normality of plug-in NPIV of nonlinear functionals

5. MC study II: Bootstrap uniform confidence sets

Outline

Preliminaries: measuring ill-posedness

- let $\Pi_K : L^2(W) \to B_K$ denote the orthogonal projection onto the sieve space $B_K = \vee\{b_{K1}, \ldots, b_{KK}\}$
- weak norm $\|h\|_{w,2} = \|\Pi_K Th\|_{L^2(W)}$ where $Th(W_i) = E[h(X_i)|W_i]$
- BCK (07): sieve measure of ill-posedness

$$s_{JK}^{-1} = \sup_{h \in \Psi_J : \|h\|_{w,2} \neq 0} \frac{\|h\|_{L^2(X)}}{\|h\|_{w,2}} = \frac{1}{s_{\min}(G_\psi^{-1/2} S' G_b^{-1/2})}$$

where $G_b = G_{b,K} = E[b^K(W_i)b^K(W_i)']$,
$G_\psi = G_{\psi,J} = E[\psi^J(X_i)\psi^J(X_i)']$,
$S' = S'_{JK} = E[\psi^J(X_i)b^K(W_i)']$.

- the NPIV model is said to be
  - mildly ill-posed if $s_{JK}^{-1} = O(J^{\varsigma/d})$ for some $\varsigma > 0$
  - severely ill-posed if $s_{JK}^{-1} = O(\exp(\frac{1}{2}J^{\varsigma/d}))$ for some $\varsigma > 0$

Preliminaries: roughness properties of the sieve

- following Newey (97), we define $\zeta(J) = \zeta_b(K) \vee \zeta_\psi(J)$,

$$\begin{aligned}\zeta_b(K) &:= \sup_w \|G_b^{-1/2} b^K(w)\|_{\ell^2} \\ \zeta_\psi(J) &:= \sup_x \|G_\psi^{-1/2} \psi^J(x)\|_{\ell^2}\end{aligned}$$

- we also introduce

$$\xi_\psi(J) := \sup_x \|\psi^J(x)\|_{\ell^1}$$

which is better suited to studying sup-norm rates

- sup-norm variance term depends on $\xi_\psi(J)$, $e_J = \lambda_{\min}(G_{\psi,J})$, s_{JK}^{-1}

Assumptions imposed for sup-norm rate

1. (i) $\{(X_i, Y_i, W_i)\}_{i=1}^n$ is an i.i.d. sample; (ii) X has compact support $\mathcal{X} \subset \mathbb{R}^d$ with nonempty interior; W has support $\mathcal{W} \subset \mathbb{R}^{d_w}$; (iii) $\sup_x |h_0(x)| < \infty$; (iv) $h \mapsto E[h(X)|W = \cdot]$ is injective on $L^\infty(X)$

2. (i) $\sup_w E[u_i^2|W_i = w] \leq \overline{\sigma}^2$; (ii) $E[|u_i|^{(2+\delta)}] < \infty$ for some $\delta > 0$.

3. (i) $\lambda_{\min}(G_{b,K}) > 0$; $e_J = \lambda_{\min}(G_{\psi,J}) > 0$; $J \leq K$;
   (ii) $s_{JK}^{-1}\zeta(J)\sqrt{(J \log J)/n} = o(1)$;
   (iii) $\zeta_b(K)^{(2+\delta)/\delta}\sqrt{(\log J)/n} = o(1)$

4. there exists $\pi_J h_0 \in \Psi_J$ such that: (i) $\|h_0 - \pi_J h_0\|_\infty \leq C^* J^{-p/d}$;
   (ii) $s_{JK}^{-1}\|h_0 - \pi_J h_0\|_{w,2} \leq C_2^* \|h_0 - \pi_J h_0\|_{L^2(X)}$;
   (iii) $\|Q_J(h_0 - \pi_J h_0)\|_\infty \leq C_\infty^* \|h_0 - \pi_J h_0\|_\infty$
   with $Q_J : L^2(X) \to \Psi_J$ the oblique projection
   $Q_J h(x) = \psi^J(x)[S' G_b^{-1} S]^{-1} S' G_b^{-1} E[b^K(W_i)h(X_i)]$.

Upper bound (1)

Theorem (Upper bound for NPIV)

Let Assumptions 1–4 hold. Then:

$$\|\widehat{h} - h_0\|_\infty = O_p\left(J^{-p/d} + s_{JK}^{-1}\xi_\psi(J)\sqrt{(\log J)/(ne_J)}\right) .$$

- For Cohen-Daubechies-Vial (CDV) wavelets and B-splines, we show that $[\xi_\psi(J)]^2/e_J = O(J)$, hence

$$\|\widehat{h} - h_0\|_\infty = O_p\left(J^{-p/d} + s_{JK}^{-1}\sqrt{(J\log J)/n}\right) .$$

Upper bound (2)

Corollary

Let $\mathcal{X} = [0,1]^d$, $0 < \inf_x f(x), \sup_x f(x) < \infty$, *and let* Ψ_J *be spanned by a CDV wavelet basis or B-spline basis of sufficient regularity. Then:*

Mildly ill-posed case: *Choosing* $J \asymp K \asymp (n/\log n)^{d/(2(p+\varsigma)+d)}$ *yields:*

$$\|h_0 - \widehat{h}\|_\infty = O_p\Big((n/\log n)^{-p/(2(p+\varsigma)+d)}\Big).$$

Severely ill-posed case: *Choosing* $J = c_0'(\log n)^{d/\varsigma}$ *for any* $c_0' \in (0,1)$ *and* $K = c_0 J$ *for some finite* $c_0 \geq 1$ *yields:*

$$\|h_0 - \widehat{h}\|_\infty = O_p\Big((\log n)^{-p/\varsigma}\Big).$$

Optimality (1)

- Chen and Reiß (11) showed that the $L^2(X)$ rates
  - $\|\widehat{h} - h_0\|_{L^2(X)} = O_p(n^{-p/(2(p+\varsigma)+d)})$ in the mildly ill-posed case
  - $\|\widehat{h} - h_0\|_{L^2(X)} = O_p((\log n)^{-p/\varsigma})$ in the severely ill-posed case

  are optimal in a L^2- minimax sense

- sup norm $\geq L^2$ norm
- therefore our sup-norm rate are optimal in the severely ill-posed case

- what about the mildly ill-posed case?
- now derive the minimax lower bound in sup-norm loss, i.e. the rate r_n over a parameter space $\mathcal{H}$ s.t.

$$\liminf_{n\to\infty} \inf_{\widetilde{h}_n} \sup_{h\in\mathcal{H}} \mathbb{P}_h\left(\|h - \widetilde{h}_n\|_\infty \geq c r_n\right) \geq c' > 0,$$

  for constants c, c'.

Optimality (2)

- trick: rewrite the NPIV model in terms of a nonparametric indirect regression (NPIR) model:

$$\begin{aligned} Y_i &= E[h_0(X_i)|W_i] + \varepsilon_i \\ E[\varepsilon_i|W_i] &= 0 \\ \varepsilon_i &\sim N(0, \sigma_0(W_i)^2) \end{aligned}$$

  where $E[\,\cdot\,|W_i]$ is known and $\sigma_0(\cdot)^2 \geq \underline{\sigma}_0^2 > 0$
- NPIV:

$$Y_i = h_0(X_i) + \underbrace{E[h_0(X_i)|W_i] - h_0(X_i) + \varepsilon_i}_{=:u_i}$$

  where by construction $E[u_i|W_i] = 0$

- NPIR is more informative than NPIV
- implication: lower bound for NPIV $\geq$ lower bound for NPIR

Lower bound for NPIR

Assumption (S)

(i) $h_0 \in B^p_{\infty,\infty}([0,1]^d)$, *(ii) there is a* $\varsigma > 0$ *such that*

$$\|Th\|_{L^2(X)} \lesssim \|h\|_{B^{-\varsigma}_{2,2}}$$

for all $h \in B(p,L) := \{h \in B^p_{\infty,\infty}([0,1]^d) : \|h\|_{B^p_{\infty,\infty}} \leq L\}$.

Theorem (Lower bound for NPIR)

Let Assumption S hold for the NPIR model with a random sample $\{(Y_i, W_i)\}_{i=1}^n$. *Then:*

$$\liminf_{n\to\infty} \inf_{\widetilde{h}_n} \sup_{h\in B(p,L)} \mathbb{P}_h \left(\|h - \widetilde{h}_n\|_\infty \geq c(n/\log n)^{-p/(2(p+\varsigma)+d)} \right) \geq c' > 0,$$

where $\inf_{\widetilde{h}_n}$ *denotes the infimum over all estimators based on the sample of size* n, *and the constants* c, c' *depend only on* $p, L, d, \varsigma, \underline{\sigma}_0$.

Lower bound for NPIV

Corollary (Lower bound for NPIV)

Let Assumption S hold for the NPIV model with a random sample $\{(X_i, Y_i, W_i)\}_{i=1}^n$ *and* $\inf_w E[u^2|W = w] \geq \underline{\sigma}_0^2$. *Then:*

$$\liminf_{n\to\infty} \inf_{\widetilde{h}_n} \sup_{h \in B(p,L)} \mathbb{P}_h \left(\|h - \widetilde{h}_n\|_\infty \geq c(n/\log n)^{-p/(2(p+\varsigma)+d)} \right) \geq c' > 0,$$

where $\inf_{\widetilde{h}_n}$ *denotes the infimum over all estimators based on the sample of size* n, *and the constants* c, c' *depend only on* p, L, d, ς.

Outline

Adaptive estimation for NPIV

- must choose J optimally to attain optimal rates
- optimal choice depends on the unknown p and s_{JK}^{-1}
- want a data-driven method for choosing J optimally

- existing methods focus on L^2 loss, minimizing a MSE-type criterion
  - Horowitz (14): modified sieve NPIV: $K = J$ and $b^K = \psi^J$=orthonormal series of $L^2([0,1]^d)$.
  - optimal in L^2 up to a $\log n$ factor
  - Liu & Tao (14): Mallows C_p model selection of sieve NPIV assuming homoskedastic error.
- CV/AIC/BIC/Mallows criteria aren't well suited to sup-norm rates

- we introduce a sup-norm adaptive Lepski-type procedure

Lepski-type procedure

- set $K = K(J) \asymp J$ deterministically (e.g. $K = c_0 J + a$)
- choose J by the following method. Define the sets:

$$
\begin{aligned}
\mathcal{J}_0 &= \left\{ j \in [J_{\min}, \underline{J}_{\max}] : j^{-p/d} \le C_0 V_{\sup}(j) \right\} \\
\widehat{\mathcal{J}} &= \left\{ j \in [J_{\min}, \widehat{J}_{\max}] : \|\widehat{h}_j - \widehat{h}_l\|_\infty \le \sqrt{2}\overline{\sigma}[\widehat{V}_{\sup}(j) + \widehat{V}_{\sup}(l)] \right. \\
&\qquad \left. \forall\, l \in (j, \widehat{J}_{\max}] \right\}
\end{aligned}
$$

where

$$
\begin{aligned}
V_{\sup}(j) &= s_{jK(j)}^{-1} \xi_\psi(j) \sqrt{(\log n)/(n e_j)} \\
\widehat{V}_{\sup}(j) &= \widehat{s}_{jK(j)}^{-1} \xi_\psi(j) \sqrt{(\log n)/(n \widehat{e}_j)}
\end{aligned}
$$

$\widehat{s}_{JK(J)} = s_{\min}((\Psi'\Psi)^{-1/2}(\Psi' B)(B'B)^{-1/2})$, $\widehat{e}_J = \lambda_{\min}(\Psi'\Psi/n)$.

- $J_0 = \min_{j \in \mathcal{J}_0} j$ is optimal but infeasible
- $\widehat{J} = \min_{j \in \widehat{\mathcal{J}}} j$ is our data-driven estimator of J
- $\widehat{h}_{\widehat{J}}$ denotes the sieve NPIV estimator with $J = \widehat{J}$, $K = K(\widehat{J})$

Heuristic argument

- Suppose $\mathcal{J}_0 \subseteq \widehat{\mathcal{J}}$. Then:

$$\widehat{J} := \min \widehat{\mathcal{J}} \leq J_0 := \min \mathcal{J}_0$$

and:

$$\begin{aligned}
\|\widehat{h}_{\widehat{J}} - h_0\|_\infty &\leq \|\widehat{h}_{J_0} - h_0\|_\infty + \|\widehat{h}_{\widehat{J}} - \widehat{h}_{J_0}\|_\infty \\
&\leq \|\widehat{h}_{J_0} - h_0\|_\infty \\
&\quad + 2\sqrt{2}\overline{\sigma}\widehat{s}^{-1}_{J_0 K(J_0)} \xi_\psi(J_0) \sqrt{(\log n)/(n\widehat{e}_{J_0})} \\
&\lesssim \|\widehat{h}_{J_0} - h_0\|_\infty \\
&\quad + s^{-1}_{J_0 K(J_0)} \xi_\psi(J_0) \sqrt{(\log n)/(n e_{J_0})} \quad \text{wpa1}
\end{aligned}$$

$$\Rightarrow \quad \|\widehat{h}_{\widehat{J}} - h_0\|_\infty = O_p(J_0^{-p/d} + s^{-1}_{J_0 K(J_0)} \xi_\psi(J_0) \sqrt{(\log n)/(n e_{J_0})})$$

- implication: $\widehat{J}$ is rate adaptive to the oracle J_0

Choosing $J_{\max}$

- still need to choose $J_{\max}$
- data-driven estimator of $J_{\max}$:

$$\widehat{J}_{\max} = \min\{J > J_{\min} : \widehat{s}_{JK(J)}^{-1}\zeta(J)\sqrt{(JL(J)\log n)/n} \geq 1\}$$

where $L(J) = a\log(\log(J))$ for some constant $a > 0$.

Oracle property

- now consider the special case with a CDV wavelet or B-spline sieve, rectangular support, and well-behaved density

Theorem (Adaptivity)

Let Assumptions 1–4 hold and $s^{-1}_{\bar{J}_{\max}K(\bar{J}_{\max})}\sqrt{(\bar{J}^2_{\max}\log n)/n} = o(1)$.
Then: $\underline{J}_{\max} \leq \widehat{J}_{\max} \leq \bar{J}_{\max}$ *wpa1; and*

$$\mathcal{J}_0 \subseteq \widehat{\mathcal{J}} \quad \textit{wpa1}$$

and so:

$$\|\widehat{h}_{\widehat{\mathcal{J}}} - h_0\|_\infty = O_p(J_0^{-p/d} + s^{-1}_{J_0K(J_0)}\sqrt{(J_0\log n)/n}) \, .$$

- implication: sup-norm rate adaptive in the mildly and severely ill-posed cases; no loss of $\log n$ factor.
- automatically implies $L^2(X)$-norm rate adaptive in the severely ill-posed case, and almost adaptive in the mildly ill-posed case (up to $\log n$ factor).

Outline

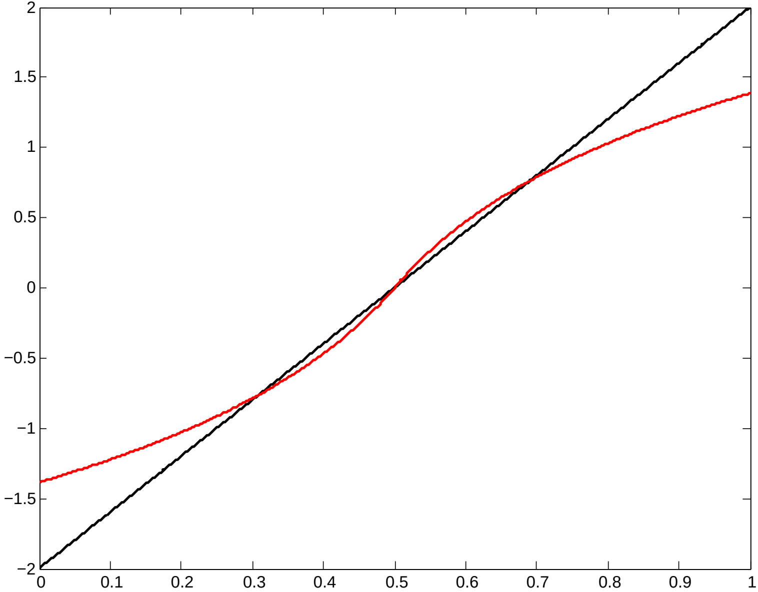
MC design

- Newey and Powell (03) design, but with compact support: generate

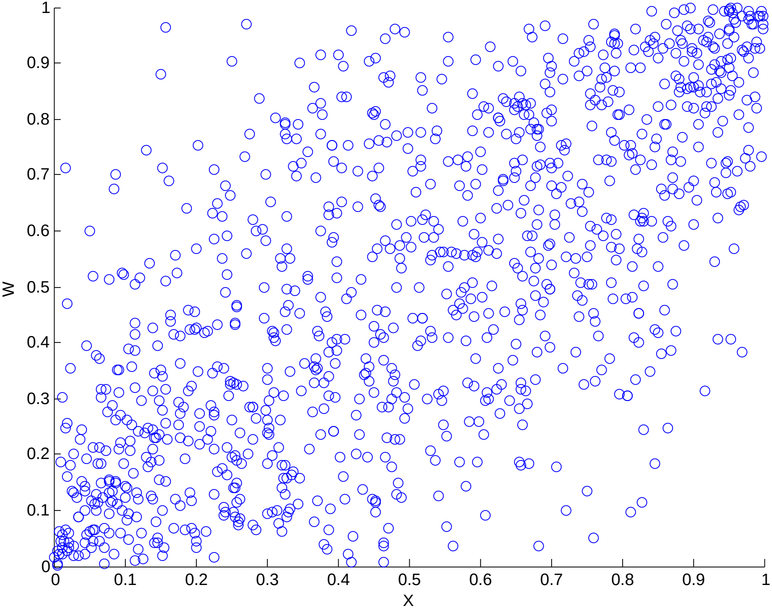
$$\begin{pmatrix} U_i \\ V_i^* \\ W_i^* \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

  and set $X_i = \Phi((W_i^* + V_i^*)/\sqrt{2})$ and $W_i = \Phi(W_i^*)$
- linear design: $h_0(x) = 4x - 2$
- nonlinear design: $h_0(x) = \log(|6x - 3| + 1)\mathrm{sgn}(x - \frac{1}{2})$
- generate 1000 samples of length 1000
- implement with cubic/quartic B-splines (with nested knots) and Legendre polynomials
- use $\overline{\sigma} = 1$ (true $\overline{\sigma}$) and $\overline{\sigma} = .1$
- take $L(J) = \frac{1}{10} \log \log J$ in definition of $\widehat{J}$
- compare sup-norm and L^2-norm error of Lepski procedure against infeasible choice of J which minimizes sup-norm error in each sample

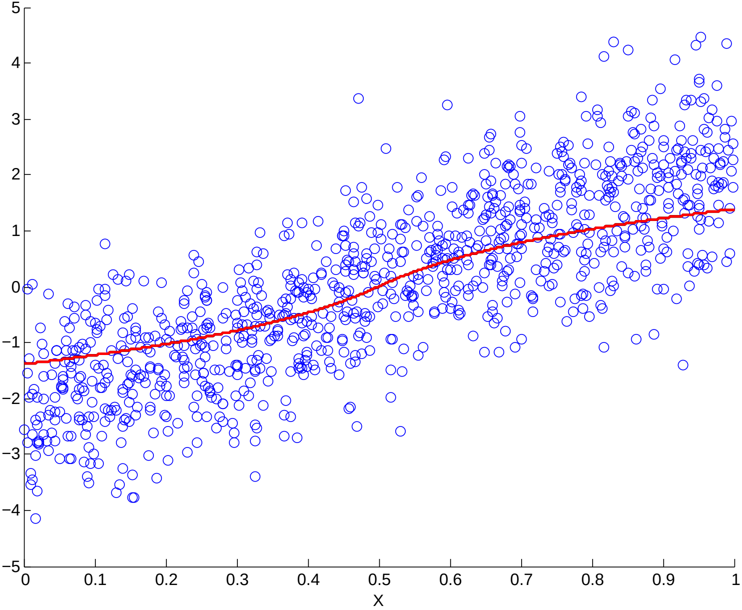
MC design: linear h_0 (black), nonlinear h_0 (red)



MC design: scatter plot of (X_i, W_i)



MC design: scatter plot of (X_i, Y_i) with nonlinear h_0



MC results: Lepski procedure, linear design

Table 1: Linear design, cubic ($r = 4$) and quartic ($r = 5$) B-spline bases

	r_J	r_K	Lepski ($\overline{\sigma} = 1$) L^∞ loss	Lepski ($\overline{\sigma} = 1$) L^2 loss	Lepski ($\overline{\sigma} = 0.1$) L^∞ loss	Lepski ($\overline{\sigma} = 0.1$) L^2 loss	Infeasible L^∞ loss	Infeasible L^2 loss
			Results with $K(J) = J - r_J + r_K$					
Mean	4	4	0.4262	0.1547	0.4262	0.1547	0.4141	0.1608
Med.	4	4	0.3828	0.1394	0.3828	0.1394	0.3708	0.1443
Mean	4	5	0.4179	0.1524	0.4209	0.1536	0.3937	0.1540
Med.	4	5	0.3681	0.1368	0.3692	0.1370	0.3476	0.1368
Mean	5	5	0.6633	0.2355	0.6633	0.2355	0.6243	0.2494
Med.	5	5	0.6007	0.2202	0.6007	0.2202	0.5646	0.2311
			Results with $K(J) = 2(J - r_J) + r_K + 1$					
Mean	4	4	0.4188	0.1526	0.4188	0.1526	0.3895	0.1552
Med.	4	4	0.3696	0.1375	0.3696	0.1375	0.3470	0.1371
Mean	4	5	0.3918	0.1439	0.3945	0.1449	0.3720	0.1486
Med.	4	5	0.3430	0.1291	0.3430	0.1291	0.3295	0.1311
Mean	5	5	0.6366	0.2277	0.6366	0.2277	0.5816	0.2352
Med.	5	5	0.5800	0.2089	0.5800	0.2089	0.5228	0.2111

MC results: Lepski procedure, linear design

Table 2: Linear design, Legendre polynomial bases

	Lepski ($\overline{\sigma} = 1$)		Lepski ($\overline{\sigma} = 0.1$)		Infeasible	
	L^∞ loss	L^2 loss	L^∞ loss	L^2 loss	L^∞ loss	L^2 loss
	Results with $K(J) = J$					
Mean	0.0882	0.0492	0.2943	0.1185	0.0869	0.0494
Med.	0.0777	0.0452	0.1674	0.0810	0.0764	0.0453
	Results with $K(J) = 2J$					
Mean	0.0878	0.0490	0.2745	0.1119	0.0862	0.0492
Med.	0.0779	0.0453	0.1640	0.0807	0.0766	0.0455

MC results: Lepski procedure, nonlinear design

Table 3: Nonlinear design, cubic ($r = 4$) and quartic ($r = 5$) B-spline bases

	r_J	r_K	Lepski ($\overline{\sigma} = 1$) L^∞ loss	Lepski ($\overline{\sigma} = 1$) L^2 loss	Lepski ($\overline{\sigma} = 0.1$) L^∞ loss	Lepski ($\overline{\sigma} = 0.1$) L^2 loss	Infeasible L^∞ loss	Infeasible L^2 loss
			Results with $K(J) = J - r_J + r_K$					
Mean	4	4	0.4343	0.1621	0.4343	0.1621	0.4233	0.1671
Med.	4	4	0.3855	0.1469	0.3855	0.1469	0.3748	0.1503
Mean	4	5	0.4262	0.1600	0.4271	0.1605	0.4030	0.1615
Med.	4	5	0.3738	0.1444	0.3744	0.1445	0.3514	0.1445
Mean	5	5	0.6726	0.2407	0.6726	0.2407	0.6318	0.2531
Med.	5	5	0.6069	0.2278	0.6069	0.2278	0.5646	0.2345
			Results with $K(J) = 2(J - r_J) + r_K + 1$					
Mean	4	4	0.4271	0.1601	0.4286	0.1609	0.3987	0.1623
Med.	4	4	0.3764	0.1445	0.3764	0.1445	0.3518	0.1443
Mean	4	5	0.4002	0.1518	0.4029	0.1528	0.3812	0.1563
Med.	4	5	0.3410	0.1384	0.3414	0.1384	0.3258	0.1402
Mean	5	5	0.6471	0.2330	0.6471	0.2330	0.5895	0.2390
Med.	5	5	0.5797	0.2143	0.5797	0.2143	0.5341	0.2141

MC results: Lepski procedure, nonlinear design

Table 4: Nonlinear design, Legendre polynomial bases

	Lepski ($\overline{\sigma} = 1$)		Lepski ($\overline{\sigma} = 0.1$)		Infeasible	
	L^∞ loss	L^2 loss	L^∞ loss	L^2 loss	L^∞ loss	L^2 loss
	Results with $K(J) = J$					
Mean	0.2494	0.1305	0.4283	0.1719	0.2297	0.1224
Med.	0.2367	0.1266	0.3210	0.1426	0.2218	0.1243
	Results with $K(J) = 2J$					
Mean	0.2475	0.1306	0.4063	0.1644	0.2241	0.1208
Med.	0.2346	0.1267	0.3132	0.1395	0.2178	0.1242

Outline

Pointwise and uniform inference

- Our sup-norm rates allow for low-level mild conditions for asymptotic normality of plug-in sieve NPIV estimators of possibly nonlinear functionals of h_0 in two cases

1. "pointwise" inference on $f(h_0)$
   - e.g.: exact consumer surplus of a price variation from p^0 to p^1 at income i

$$\begin{aligned} Q_i &= h_0(P_i, I_i) + u_i \\ f(h_0) &= S(p^0) \\ \text{where} \quad S_i'(p) &= -h_0(p, i - S_i(p)) \\ S_i(p^1) &= 0 \end{aligned}$$

   cf. Hausman & Newey (95), Vanhems (10), Blundell et al. (12)

2. "uniform" inference on $\{f_\tau(h_0) : \tau \in \mathcal{T}\}$ where $\mathcal{T} \subset \mathbb{R}^{d_\mathcal{T}}$
   - e.g.: uniform inference on consumer surplus/deadweight loss

Pointwise inference (1)

- we focus on slower than root-n functionals that are bounded wrt the sup norm:

5. (i) there exists a linear functional $Df(h_0)[\cdot]$ and constant C s.t.

$$|f(h) - f(h_0) - Df(h_0)[h - h_0]| \leq C\|h - h_0\|_\infty^2$$

for all $h \in N_n(h_0)$ where $\widehat{h} \in N_n$ wpa1;
(ii) $V_n^{-1/2}\|\widehat{h} - h_0\|_\infty^2 = o_p(n^{-\frac{1}{2}})$

- includes CS/DWL functionals and quadratic functional.
- sufficient for a more general condition of Chen and Pouzo (14).
- here $V_n \nearrow \infty$ is the sieve variance

$$\begin{aligned} V_n &= Df(h_0)[\psi^J]'\Sigma_n Df(h_0)[\psi^J] \\ \Sigma_n &= [S'G_b^{-1}S]^{-1}\left(S'G_b^{-1}\Omega G_b^{-1}S\right)[S'G_b^{-1}S]^{-1}, \end{aligned}$$

where $S = E[b^K(W_i)\psi^J(X_i)']$, $\Omega = E[u_i^2 b^K(W_i)b^K(W_i)']$

Pointwise inference (2)

Theorem (Pointwise asymptotic normality of sieve t-statistics)

Let Assumptions 1–5 (etc) hold. Then:

$$\frac{\sqrt{n}(f(\widehat{h}) - f(h_0))}{\widehat{V}_n^{1/2}} \to_d N(0,1).$$

- here $\widehat{V}_n \nearrow \infty$ is the sieve variance estimator

$$\begin{aligned}\widehat{V}_n &= Df(\widehat{h})[\psi^J]'\widehat{\Sigma}Df(\widehat{h})[\psi^J] \\ \widehat{\Sigma} &= [\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\Big(\widehat{S}'\widehat{G}_b^{-1}\widehat{\Omega}\widehat{G}_b^{-1}\widehat{S}\Big)[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\end{aligned}$$

  $\widehat{S} = B'\Psi/n$, $\widehat{G}_b = (B'B/n)$, $\widehat{\Omega} = n^{-1}\sum_{i=1}^n \widehat{u}_i^2 b^K(W_i)b^K(W_i)'$.
- just like 2SLS variance estimator but using basis functions. Chen and Pouzo (14), Newey (13).

Uniform inference (1)

- now impose a uniform (for $\tau \in \mathcal{T}$) version of Assumption 5

5′. (i) $Df_\tau(h_0)[\cdot]$ is a linear functional for each $\tau \in \mathcal{T}$, (ii) there exists a constant C s.t.

$$\sup_{\tau \in \mathcal{T}} |f_\tau(h) - f_\tau(h_0) - Df_\tau(h_0)[h - h_0]| \leq C\|h - h_0\|_\infty^2$$

for all $h \in N_n(h_0)$ where $\widehat{h} \in N_n$ wpa1;
(ii) $\sup_{\tau \in \mathcal{T}} V_{\tau,n}^{-1/2} \|\widehat{h} - h_0\|_\infty^2 = o_p(n^{-\frac{1}{2}})$

- here $V_{\tau,n} = Df_\tau(h_0)[\psi^J]'\Sigma_n Df_\tau(h_0)[\psi^J]$
- estimate with $\widehat{V}_{\tau,n} = Df_\tau(\widehat{h})[\psi^J]'\widehat{\Sigma} Df_\tau(\widehat{h})[\psi^J]$

Uniform inference (2)

Theorem (Uniform asymptotic normality of sieve t-statistics)

Let Assumptions 1–5' (etc) hold. Then there exists a sequence of tight Gaussian processes $\mathbb{G}_n$ on $\ell^\infty(\mathcal{T})$ with covariance function

$$E[\mathbb{G}_n(t_1)\mathbb{G}_n(t_2)] = \frac{Df_{t_1}(h_0)[\psi^J]'\Sigma_n Df_{t_2}(h_0)[\psi^J]}{V_{t_1,n}^{1/2} V_{t_2,n}^{1/2}}$$

and random variables $Z_n =_d \sup_{\tau \in \mathcal{T}} |\mathbb{G}_n(\tau)|$ such that

$$\sup_{\tau \in \mathcal{T}} \left| \frac{\sqrt{n}(f_\tau(\widehat{h}) - f_\tau(h_0))}{\widehat{V}_{\tau,n}^{1/2}} \right| = Z_n + o_p(1)$$

as $n, J, K \to \infty$.

- we follow Chernozhukov, Chetverikov, Kato (14) (also see Chernozhukov, Lee, Rosen (13)) construction rather than strong approximation

Example: uniform confidence bands

- $f_\tau(h_0) = h_0(\tau)$ with $\mathcal{T} = \mathcal{X}$, and $Df_\tau(h)[\psi^J] = \psi^J(\tau)$

- by previous theorem, there exists a sequence of tight Gaussian processes $\mathbb{G}_n$ on $\ell^\infty(\mathcal{X})$ with covariance function

$$E[\mathbb{G}_n(x_1)\mathbb{G}_n(x_2)] = \frac{\psi^J(x_1)'\Sigma_n\psi^J(x_2)}{V_{x_1,n}^{1/2}V_{x_2,n}^{1/2}}$$

  and random variables $Z_n =_d \sup_{x\in\mathcal{X}} |\mathbb{G}_n(x)|$ such that

$$\sup_{x\in\mathcal{X}} \left| \frac{\sqrt{n}(\widehat{h}(x) - h_0(x))}{\widehat{V}_{x,n}^{1/2}} \right| = Z_n + o_p(1)$$

  as $n, J, K \to \infty$.

- invert for uniform confidence band

Example: uniform inference on exact consumer surplus

- $f_\tau(h_0) = S_i(p)$ with $\mathcal{T} = [\underline{p}^0, \overline{p}^0] \times [\underline{i}, \overline{i}]$

$$
\begin{aligned}
Df_\tau(h)[\psi^J] &= -\int_{p^1}^{p} \psi^J(t, i - S_i(t)) e^{\int_p^s \partial_2 h_0(u, i - S_i(u))\,\mathrm{d}u}\,\mathrm{d}t \\
Df_\tau(\widehat{h})[\psi^J] &= -\int_{p^1}^{p} \psi^J(t, i - \widehat{S}_i(t)) e^{\int_p^s \partial_2 \widehat{h}(u, i - S_i(u))\,\mathrm{d}u}\,\mathrm{d}t
\end{aligned}
$$

- uniform asymptotic normality of $\{\widehat{S}_i(p) : (p, i) \in [\underline{p}^0, \overline{p}^0] \times [\underline{i}, \overline{i}]\}$ follows from previous theorem
- could equally consider uniform inference on deadweight loss

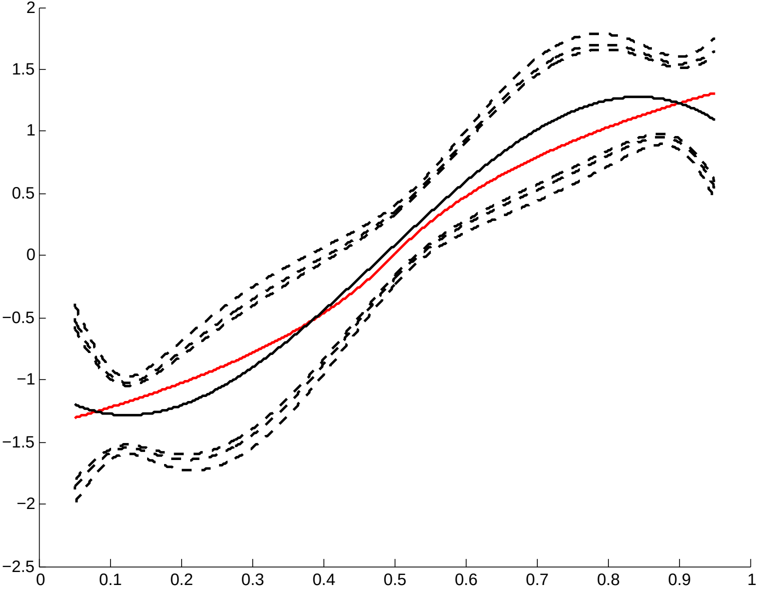
- our sup-norm rates here are critical to control bias

Outline

MC design

- same Newey and Powell (03) linear and nonlinear designs
- estimate $\widehat{J}_{\max}$ as before, use $J = \widehat{J}_{\max}$, $K = K(\widehat{J}_{\max})$ to implement sieve NPIV estimator
- estimate critical values for uniform confidence bands using the sieve score bootstrap (Chen and Pouzo, 14) with Mammen (93) two-point distribution with 1000 bootstrap replications for each sample
- computationally simpler than bootstrap sieve t stat in Chen-Pouzo (14) or the bootstrap in Horowitz-Lee (12).
- compare MC with nominal coverage probabilities

Estimated UCBs (dashed), $\widehat{h}$ (black line), h_0 (red line)



MC results: coverage probabilities

Table 5: Linear and nonlinear design, cubic ($r = 4$) and quartic ($r = 5$) B-spline bases.

	r_J	r_K	90%	95%	99%	90%	95%	99%
linear	4	4	0.933	0.966	0.996	0.944	0.971	0.994
linear	4	5	0.937	0.975	0.995	0.937	0.963	0.994
linear	5	5	0.961	0.983	0.997	0.959	0.985	0.997
nonlinear	4	4	0.884	0.945	0.987	0.912	0.956	0.989
nonlinear	4	5	0.894	0.946	0.987	0.906	0.951	0.987
nonlinear	5	5	0.956	0.978	0.995	0.951	0.979	0.996

Note: Left panel uses $K(J) = J - r_J + r_K$, right panel uses $K(J) = 2(J - r_J) + r_K + 1$.

MC results: coverage probabilities

Table 6: Linear and nonlinear design, Legendre polynomial bases

	90%	95%	99%	90%	95%	99%
linear	0.937	0.964	0.997	0.928	0.959	0.989
nonlinear	0.901	0.952	0.988	0.906	0.948	0.989

Note: Left panel uses $K(J) = J$, right panel uses $K(J) = 2J$.

Conclusions

- contributions:
  1. optimal sup-norm rates and attainability by sieve estimators
  2. Lepski procedure for adaptive estimation in sup norm
  3. pointwise and uniform inference on possibly nonlinear functionals
- first such results for NPIV (or indeed any ill-posed inverse problem with unknown operator)
- application to inference on consumer surplus in demand estimation and uniform confidence bands