Posterior consistency of nonparametric instrumental variable regression

Yuan Liao

Northwestern University

Presented at CEMMAP

June 12, 2010

Outline

Conditional Moment Restricted Model

$$E[\rho(X, g_0)|W] = 0$$

Single Index Model $E(Y|W) = E[h_0(W^T\theta_0)|W]$

Partially Linear Model

$$E(Y|W_1, W_2) = E[h_0(W_1) + W_2^T\theta_0|W_1, W_2]$$

Nonparametric IV $E(Y - g_0(X)|W) = 0$

Identification

- Consider $E[g_0(X)|W] = E(Y|W)$,
  g_0 is identified iff $X|W$ is complete. Newey and Powell (2003)
- The completeness condition is easy to fail:
  Severini and Tripathi (2006):
  $X = W + U$: W and U are iid Uniform $[-\frac{1}{2}, \frac{1}{2}]$.
  $E[g(X)|W] = 0$ iff
  $g(x) = g(1 + x)\ \forall x \in [-1, 0]$ and $\int_{-1}^{0} g(x)dx = 0$.

- Sometimes instead of g_0, we are interested in linear functional $h(g_0)$.
- Severini and Tripathi (2006): If $\exists m(W)$ s.t. $v(X) = E[m(W)|X]$, then

$$h(g_0) = E[v(X)g_0(X)] \text{ is identified.}$$

- We do not assume g_0 is identified.

$$\Theta_I = \{g \in \Theta : E(\rho(X, g)|W) = 0\}$$

- Santos (2007, 2008a,b), Kovchegov and Yildiz (2010)

Ill-Posed Inverse Problem (partial identification)

- Define $T(g) = E[g(X)|W]$, and $\mu(w) = E(Y|W = w)$.

$$Tg_0 = \mu$$

- For any $g_1 \in L^2(X)$, define $[g_1] = \{g : Tg = Tg_1\}$. Define A on quotient space $L^2(X)/\mathcal{N}(T)$:

$$A[g] = Tg$$

But $A^{-1} : L^2(W) \to L^2(X)/\mathcal{N}(T)$ is not continuous.

Two Regularization Approaches

Approach I Assume $g_0 \in \Theta$, where Θ is compact

$\Theta = \{g : \|g\| \leq B\}$ for some known $\|.\|$ and $B > 0$.

- Result: If $T : \Theta \to L^2$, Θ is compact and T is compact, then T^{-1} is continuous.
- Newey and Powell (2003), Ai and Chen (2003), etc.

Approach II

- If $Q_n \Rightarrow Q \geq 0$ uniformly,

$$\hat{g} = \arg\min Q_n(g) + a_n ||g||^2, \quad a_n \to 0$$

  under regularity conditions, $\hat{g} \to g_0$ if $\arg\min Q = \{g_0\}$.

- Hall and Horowitz (2005), Chen and Pouzo (2009), etc.

Contributions

Bayesian
- Obtain the posterior distribution of conditional moment restricted model
- Construct the posterior in a way that is robust to the distributional assumption

Consistency Show posterior consistency with partial identification

Traditional Nonparametric Bayesian Regression

$$y = g(X) + \epsilon, E(\epsilon|X) = 0$$

1. $\epsilon \sim N(0, \sigma^2)$
2. approximate $g \approx \sum_{j=1}^{q} \beta_j \phi_j(x)$
3. priors: $\beta_j|\sigma, q \sim N(0, v\sigma^2)$, $\sigma \sim$ Inverse Gamma;

   q: either $\rightarrow \infty$ or $Uniform\{1, 2, ..., n\}$

Coram and Lalley (2006): posterior$= \int P_g d\pi(g)$.

Proposed Bayesian Approach

- Assume W is supported on $[0, 1]$.
- $E[\rho(X, g_0)|W] = 0$ implies

$$E\underbrace{\left(\rho(X, g_0)I(W \in [\frac{i-1}{k}, \frac{i}{k}]\right)}_{m_i(D, g_0)} = 0, \quad i = 1, ..., k$$

$m = (m_1, ..., m_k)^T$.

- $Em(D, g_0) = 0$

Question: How to derive the posterior?

- $p(g|Data) \propto p(g) \times$likelihood.

  Construct the likelihood first.

- In practice,

  What is known: moment conditions. $Em(D, g_0) = 0$

  What is not known: the true likelihood function

- Derive from $Em(D, g_0) = 0$.

- Define $\bar{m}(g) = \frac{1}{n}\sum_{i=1}^{n} m(D_i, g)$.
- CLT $\Rightarrow \sqrt{n}\bar{m}(g_0) \to^d N(0, V)$

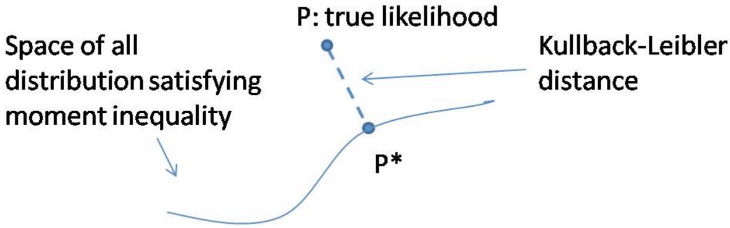
$$L(g) \propto \exp\left(-\frac{n}{2}\bar{m}(g)^T V^{-1}\bar{m}(g)\right)$$

- QUESTION: Does it have a likelihood interpretation?

  Likelihood$=p(Data|\theta)$.

  Answer: It is **Best Approximation to True Likelihood** subject to moment restrictions. Kim(2002) “limited information likelihood”

Limited Information Likelihood



$$L(g) \rightarrow^{p} P^{*} \quad Kim(2002)$$

$L(g)$ is called the “limited information likelihood”.

- $V = Var(m(D, g_0))$, which depends on the unknown distribution of D and g_0, Kim(2002) suggested replace V by a constant variance covariance matrix (may depend on n).
- It can be shown that $Var(m_i(D, g_0)) = O(k^{-1})$, we set $V = k^{-1}I$. Such a choice of V will not affect the consistency result.

- Suppose $g_0(x) \approx \sum_{i=1}^{q} b_i \phi_i(x)$, as $q \to \infty$
- Put priors on $\{b_1, ..., b_q\}$, fixing q. Let $g_q = \sum_{i=1}^{q} b_i \phi_i$
- $p(g_q|D) \propto p(b) \exp[-n\bar{m}(g_q)^T V \bar{m}(g_q)]$
- Outline of the proof of consistency:

$\bar{G}(g) = \bar{m}(g_q)^T V \bar{m}(g_q)$, $G(g) = \int \frac{E(\rho(Z,g)|w)^2}{E(\rho(Z,g_0)|w)} dF_W(w)$

$$\sup_{g \in \Theta_q} |\bar{G}(g_q) - G(g_q)| = o_p(1)$$

Jiang and Tanner (2008): for any $\delta > 0$,

$$E(P(|G(g_q) - \inf_{\Theta_q} G(g_q)| > \delta | Data) \leq P(\sup_{\Theta_q} |\bar{G}(g) - \inf_{\Theta_q} G(g)| > \frac{\delta}{5})$$

$$+ \frac{\exp(-\frac{2}{5} n\delta)}{Prior(G(g_q) - \inf_{\Theta_q} G(g) < \delta)}$$

Assumptions

1. Θ is compact under some norm $\|.\|_H$.
2. (i) $P(W \in [\frac{j}{k}, \frac{j+1}{k}]) = O(k^{-1})$, $k = o(n^{2/5})$.

   (ii) $q = o(n)$.
3. prior $p(g_q)$ has support around Θ_I.
4. $EY^2 < \infty$, $E\|m(D,g)\|^4 < B$ uniformly on Θ.
5. $\max_{j \le k} \sqrt{n}|\bar{m}_j(g) - Em_j(g, X)|$ is stochastic equicontinuous on Θ.

Assumptions

Define $K_g(w) = E(\rho(Z,g)|w)$.

Assumption 6:

- $\{K_g(w) : g \in \Theta\}$ is equicontinuous on $w \in [0,1]$.
- $E(\rho(Z,g_0)^2|w)$ is continuous and bounded away from zero on $w \in [0,1]$
- For any $\epsilon > 0$, exists $\delta > 0$,

$$\sup_{z} \sup_{||g_1 - g_2||_H < \delta} |\rho(z,g_1) - \rho(z,g_2)| < \epsilon$$

We can obtain the sufficient conditions for Assumptions 5,6 for single index model and nonparametric IV regression.

Single Index Model (1) $f_W(w)$ is bounded away from 0

(2) For any $\epsilon > 0$, there exists $\delta > 0$,

$$\sup_{y} \sup_{|w_1 - w_2| < \delta} |f_{w|y}(w_1|y) - f_{w|y}(w_2|y)| < \epsilon$$

Nonparametric IV (1) $f_W(w)$ is bounded away from 0

(2) For any $\epsilon > 0$, there exists $\delta > 0$,

$$\sup_{x,y} \sup_{|w_1 - w_2| < \delta} |f_{w|x,y}(w_1|x, y) - f_{w|y}(w_2|x, y)| < \epsilon$$

Posterior Consistency

Define $U_\delta(\Theta_I) = \{g \in \Theta : \inf_{g^* \in \Theta_I} ||g - g^*||_H < \delta\}$

Theorem 1

Let $\Theta_I = \{g \in \Theta : E[\rho(X, g)|W]\}$. *For any* $U(\Theta_I)$,

$$P(g_q \in U(\Theta_I)|Data) \rightarrow^p 1$$

Corollary 1

If $h(g_0)$ *is identified,* $h : \Theta \rightarrow \mathbb{R}$ *is continuous, for any* $\epsilon > 0$,

$$P(|h(g_q) - h(g_0)| < \epsilon|Data) \rightarrow^p 1$$

Norm Specification

Single Index Model: $g_0 = (\psi_0, \theta_0)$

$$||\psi||_s = \sup_t |\psi(t)| + \sup_{t_1 \neq t_2} \frac{|\psi(t_1) - \psi(t_2)|}{|t_1 - t_2|}$$

$||g||_H = ||\psi||_s + \|\theta\|$

Nonparametric IV: $||g||_H = \sup_x |g(x)| + sup_{x_1 \neq x_2} \frac{|g(x_1) - g(x_2)|}{|x_1 - x_2|}$

Relaxing the Compactness in NPIV

- Specify prior $\propto e^{-na_n^2\|g\|^2}$,

  limited Information Likelihood $\propto e^{-n\hat{Q}(g)}$

$$\log posterior = -n[\hat{Q}(g) + a_n^2\|g\|^2]$$

- Require the prior variance $na_n^2 \to \infty$. Florens and Simoni (2009).
- $T(g)(w) = E(g(X)|W = w)$. To illustrate the posterior consistency, we assume g_0 is point identified.
- Let the eigenvalues of T be $\lambda_1, ...,$ ordered such that $|\lambda_1| \geq |\lambda_2| \geq ... > 0$

Assumptions

1. For some $\alpha > 0$, and and c,

$$\sup_{\|g\| \le c} |\bar{G}(g) - G(g)| = o_p(n^{-\alpha})$$

2. $a_n^2 \to 0$, and $na_n^2 \to \infty$
3. There exists $\{s_n\}_{n=1}^{\infty} \subset \mathbb{N}$, $s_n \to \infty$, such that $\sum_{j \ge s_n} g_j^2 = O(a_n^2/\lambda_{s_n}^2) = o(1)$
4. $n \succ q_n \succ \max\{n^{1-\alpha}, na_n^2/\lambda_{s_n}^2, \lambda_{q_n}^{-2}\}$.

Posterior Consistency

Theorem 2

Assume g_0 is identified,

$$E[||g_q - g_0||^2 | Data] \to^p 0$$

- Prior variance $na_n^2 \to \infty$ for regularization.

  Prior mean was set to zero.

- To incorporate prior knowledge of g_0, e.g., convexity, monotonicity, use prior

$$\log p(g_q) \propto -na_n^2 ||g_q - g^*||^2$$

Bayesian Implementation

1. Fix q, k, construct likelihood $L(g_q) \propto \exp(-n\bar{G}(g_q))$, where $g_q = \sum_{j=1}^{q} b_j \phi_j(x)$
2. Put prior $(b_1, ..., b_q)$, $b_j \sim N(0, j^{-\alpha})$, for some $\alpha > 0$
3. Obtain $B = (5,000 \sim 10,000)$ draws from posterior $\propto p(b_1, ..., b_q)L(g_q)$ using MCMC algorithm. The first $1/4 \sim 1/3$ are dropped for the MCMC to "warm-up".
4. Compute the posterior mean $E(b_j|Data) \approx \frac{1}{B}\sum_i b_j^i$
5. $\hat{g}(x) = \sum_j E(b_j|Data)\phi_j(x)$

Numerical Example

$$Y = \sin(x)e^{\sqrt{|x|}} + \epsilon$$

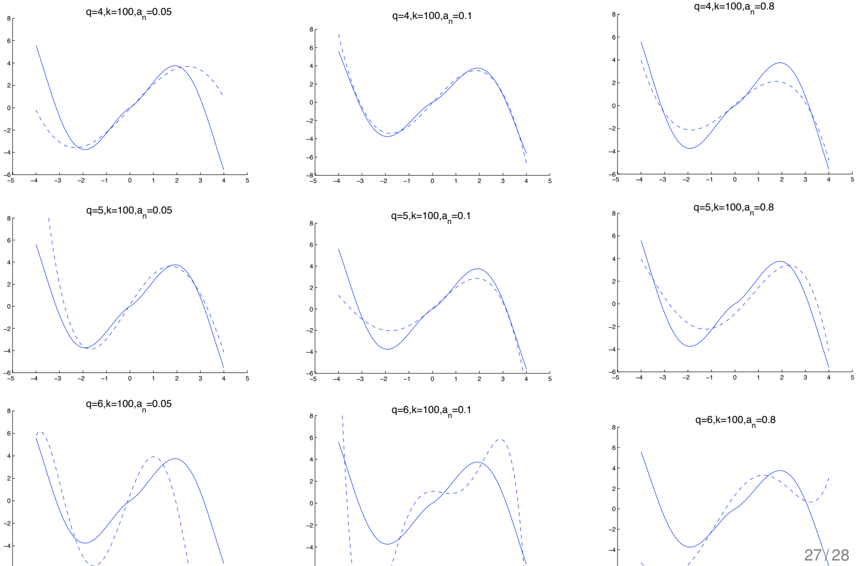
$$x = w + v$$

$w \perp \epsilon$, $Cov(v, \epsilon) = 0.6$. w is supported on $[-4, 4]$.

Hermite series approximation:

$$H_1(x) = 1, H_2(x) = x, \text{ and } H_j(x) = H_{j-1}(x) - (j-1)H_{j-2}(x)$$

Approach 2:


q=4,k=100,a_n=0.05
q=4,k=100,a_n=0.1
q=4,k=100,a_n=0.8
q=5,k=100,a_n=0.05
q=5,k=100,a_n=0.1
q=5,k=100,a_n=0.8
q=6,k=100,a_n=0.05
q=6,k=100,a_n=0.1
q=6,k=100,a_n=0.8

- We have not derived the convergence rate of $E(||g_q - g_0||^2|Data)$ yet.
- Choice of V: Suppose we are interested in functional $h(g_0)$, V can be chosen to minimize the asymptotic variance of $E(h(g_q)|Data)$.
- As this is among the first papers that consider moment condition based likelihood, to choose V from $p(h(g)|Data)$ has not been considered yet.