ASYMPTOTIC NORMALITY OF A NONPARAMETRIC INSTRUMENTAL VARIABLES ESTIMATOR

by

Joel L. Horowitz
Department of Economics
Northwestern University
Evanston, IL
USA

INTRODUCTION

- Paper is about nonparametric estimation of the function g in model

  $$Y = g(X) + U$$

  - Assume that $\boldsymbol{E}(U \mid X) \neq 0$.
  - Instrument W is available and satisfies $\boldsymbol{E}(U \mid W) = 0$.
  - X and W may be vectors.
  - Some components of X may be exogenous, in which case they are included in W.

- Talk deals only with case of scalar X and W

- If g is known up to finite-dimensional parameter θ, then θ can be estimated by GMM.

  - GMM estimator has $n^{-1/2}$ rate of convergence and is asymptotically normal.
  - Situation more complicated when g nonparametric.

BACKGROUND

- Newey, Powell, and Vella (1999) considered triangular-array version of model.
  - In triangular array,

    $$X = \boldsymbol{E}(X \mid W) + V$$

  - V and U related by $\boldsymbol{E}(U \mid X, V) = \boldsymbol{E}(U \mid V)$.
  - This generates additive nonparametric mean-regression model

    $$\boldsymbol{E}(Y \mid X, V) = g(X) + h(V).$$

- Newey and Powell (2002) developed series estimator of g for $Y = g(X) + U$ without triangular-array restrictions.
- Darolles, Florens, and Renault (2002) developed kernel estimator of g for special case in which there are no exogenous components of X.
- Hall and Horowitz (2004) developed kernel estimator that applies with or without exogenous components of X.
  - Estimator converges at fastest possible rate under assumptions of Hall and Horowitz.

CONTRIBUTION OF TALK

- Nonparametric IV estimation is difficult because of ill-posed inverse problem.

- Results to date give conditions for consistency and, in some cases, rate of convergence of estimator of unknown function g in $Y = g(X) + U$

    - No results so far on asymptotic distribution of estimator of g

    - Darolles, et al. (2002)give conditions under which certain integrals of g are asymptotically normal.

- So far, no NPIV estimator has known pointwise asymptotic distribution.

- This talk gives conditions under which Studentized estimator of Hall and Horowitz is asymptotically standard normal.

- This enables estimator to be used for inference in large samples.

ORGANIZATION

- Identification (population version of estimation problem)
- Ill-posed inverse problem
- Review kernel estimator of Hall and Horowitz (2004)
- Give conditions for asymptotic normality.
- No Monte Carlo results yet

IDENTIFICATION

- Assume that X and W are scalars.
    - Support of (X,W) is $[0,1]^2$.
    - This can always be achieved by (if necessary) carrying out monotone transformations of X and W.
- Let f_{XW} denote joint density of (X,W) and f_W denote marginal density of W
- Define

$$t(x,z) = \int_0^1 f_{XW}(x,w) f_{XW}(z,w) dw$$

- Define T as the integral operator on $L_2[0,1]$ such that

$$(T\psi)(z) = \int_0^1 t(x,z)\psi(x)dx .$$

IDENTIFICATION (cont.)

- $\boldsymbol{E}(Y \mid W = w) = \boldsymbol{E}[g(X) \mid W = w] = \int g(x) \frac{f_{XW}(x,w)}{f_W(w)} dx$.

- Therefore,

$$\boldsymbol{E}_W[\boldsymbol{E}(Y \mid W) f_{XW}(z,W)] = \int \boldsymbol{E}(Y \mid W = w) f_W(w) f_{XW}(z,w) dw$$

$$= \int g(x) f_{XW}(x,w) f_{xw}(z,w) dx dw$$

$$= (Tg)(z).$$

- Assume that T is invertible (its eigenvalues are all strictly positive).

- Then

$$g(z) = \boldsymbol{E}_W[\boldsymbol{E}(Y \mid W)(T^{-1} f_{XW})(z,W)]$$

  - This relation identifies g

- The expectations, f_{XW}, and T can be estimated nonparametrically using standard methods (e.g., kernels).

- Estimation consists of replacing these population quantities with sample analogs

ILL-POSED INVERSE PROBLEM

- Relation

  $$(Tg)(z) = \boldsymbol{E}_W[\boldsymbol{E}(Y \mid W) f_{XW}(z,W)]$$

  is Fredholm integral equation of first kind.

- Generates "ill-posed inverse problem" if, as is usually case, 0 is a limit point of eigenvalues of T.
  - Problem is that T^{-1} is unbounded and discontinuous.
  - Therefore, $T^{-1}(\psi_1 - \psi_2)$ is not necessarily close to 0 even if $\psi_1 - \psi_2$ is close to 0.
  - Implication: Replacing $\boldsymbol{E}_W[\boldsymbol{E}(Y \mid W) f_{XW}(z,W)]$ with consistent estimator does not necessarily give consistent estimator of $\boldsymbol{E}_W[\boldsymbol{E}(Y \mid W)(T^{-1} f_{XW})(z,W)]$.

- Solution: Replace T with $T + a_n I$, where
  - I is identity operator
  - $\{a_n\}$ is sequence of strictly positive numbers that converges to 0 as $n \to \infty$.

- Identifying relation becomes

  $$g_n(z) = \boldsymbol{E}_W[\boldsymbol{E}(Y \mid W)(T + a_n I)^{-1} f_{XW}(z,W)].$$

  - g_n is estimated population quantity
  - $g_n \to g$ as $n \to \infty$

HALL-HOROWITZ KERNEL ESTIMATOR

- Estimation consists of replacing unknown population quantities with sample analogs

- Data are independent random sample $\{Y_i, X_i, W_i : i = 1,...,n\}$

- Let $\hat{f}_{XW}$ and $\hat{T}$ be estimators of f_{XW} and T.
  - Define $\hat{T}^+ = (\hat{T} + a_n I)^{-1}$

- Estimate $g(z)$ by

$$\hat{g}(z) = n^{-1} \sum_{i=1}^{n} Y_i (\hat{T}^+ \hat{f}_{XW})(z, W_i)$$

- We need two estimators of f_{XW}, a "regular" estimator and a leave-one-out estimator.

- Let K be kernel function
  - Define $K_h(v) = K(v/h)$, where h is bandwidth parameter.
  - In general, K is "boundary kernel" to deal with possibility that f_{XW} does not go smoothly to 0 on boundaries of its support
  - Here, for simplicity, assume that K is symmetrical probability density function on [-1,1].

ESTIMATION (cont.)

- Kernel estimators

$$\hat{f}_{XW}(x,w) = \frac{1}{nh^2}\sum_{i=1}^{n} K_h(x - X_i)K_h(w - W_i)$$

$$\hat{f}_{XW}^{(-i)}(x,w) = \frac{1}{nh^2}\sum_{\substack{j=1 \\ j \neq i}}^{n} K_h(x - X_j)K_h(w - W_j)$$

- Estimator of T:
  - Define

    $$\hat{t}(x,z) = \int \hat{f}_{XW}(x,w)\hat{f}_{XW}(z,w)dw$$

  - Estimator of T is operator $\hat{T}$ on $L_2[0,1]$ that is defined by

    $$(\hat{T}\psi)(z) = \int \hat{t}(x,z)\psi(x)dx$$

  - Set $\hat{T}^+ = (\hat{T} + a_n I)^{-1}$.
- Estimator of g is

$$\hat{g}(z) = n^{-1}\sum_{i=1}^{n} Y_i(\hat{T}^+ \hat{f}_{XW}^{(-i)})(z, W_i)$$

ASSUMPTIONS (1)

- Because of ill-posed-inverse issue, rate of convergence of $\hat{g}$ depends on rate at which eigenvalues of T converge to 0.
    - Rate of $\hat{g}$ is slower if eigenvalues converge rapidly.
    - Regularity conditions reflect importance of rate of convergence of eigenvalues.
- Let $\{\lambda_j, \phi_j : j = 1,2,...\}$ denote eigenvalues and orthonormalized eigenvectors of T ordered so that $\lambda_1 \geq \lambda_2 \geq ... > 0$.
    - Assume that $\{\phi_j\}$ forms orthonormal basis for [0,1].
- Define sequences of Fourier coefficients $\{d_{jk}\}$ and $\{b_j\}$ by

$$f_{XW}(x,w) = \sum_{j,k=1}^{\infty} d_{jk}\phi_j(x)\phi_k(w),$$

$$g(x) = \sum_{j=1}^{\infty} b_j\phi_j(x),$$

$$d_{jk} = \int f_{XW}(x,w)\phi_j(x)\phi_k(w)dxdw,$$

$$b_j = \int g(x)\phi_j(x)dx.$$

- Then

$$t(x,z) = \sum_{j=1}^{\infty} \lambda_j\phi_j(x)\phi_j(z)$$

ASSUMPTIONS

- The data $\{Y_i, X_i, W_i : i = 1,...,n\}$ are iid. The support of (X, W) is $[0,1]^2$, and $\boldsymbol{E}[Y - g(X) \mid W = w] = 0$.

- The density f_{XW} is r times differentiable on $[0,1]^2$. There is a finite constant C such that $\left| f_{XW}(x,w) \right| \le C$, $\boldsymbol{E}(Y^2 \mid W = w) \le C$, and $\boldsymbol{E}(Y^2 \mid X = x, W = w) \le C$.

- There are constants α and β with $\alpha > 1$ and $\alpha \le \beta / 3 + 1/2$ such that $|b_j| \le C j^{-\beta}$, $j^{-\alpha} \le C\lambda_j$, and $\sum_{k=1}^{\infty} |d_{jk}| \le C j^{-\alpha/2}$ for all $j \ge 1$.

- The parameters a_n and h satisfy $a_n \propto n^{-\alpha/(2\beta+\alpha)}$ and $h \propto n^{-\gamma}$ for all sufficiently large n, where

$$\frac{1}{2r}\frac{2\alpha + 2\beta - 1}{2\beta + \alpha} \le \gamma \le \min\left\{ \frac{1}{2}\frac{2\beta - \alpha}{2\beta + \alpha}, \frac{4\beta - \alpha + 1}{5(2\beta + \alpha)} \right\}.$$

- The kernel function K is bounded, supported on [-1,1], symmetrical about 0, and satisfies

$$\int_{-1}^{1} v^j K(v)\,dv = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \le j \le r - 1 \end{cases}$$

- Comment: Third assumption imposes smoothness conditions in terms of the Fourier expansions of g and f_{XW} in addition to controlling the rate of convergence of the eigenvalues $\{\lambda_j\}$.

RATE OF CONVERGENCE OF ESTIMATOR

- Let $\mathcal{G}$ denote the set of distributions that satisfies the regularity conditions for fixed values of the constants C, α, and β.

- Theorem 1: Under the regularity conditions

$$\sup_{G \in \mathcal{G}} \int_0^1 \boldsymbol{E}_G [\hat{g}(x) - g(x)]^2 \, dx = O[n^{-(2\beta-1)/(2\beta+\alpha)}]$$

- Comments:
  - Rate of convergence increases as β increases (g becomes smoother).
  - Rate of convergence decreases as α increases (faster converging eigenvalues).

- Rate of convergence is fastest possible under the assumptions that are made.
  - Let $\bar{g}$ be any estimator of g.
  - For each $z \in [0,1]^q$,

$$\liminf_{n \to \infty} n^{(2\beta-1)/(2\beta+\alpha)} \inf_{\bar{g}} \sup_{G \in \mathcal{G}} \int \boldsymbol{E}_G [\bar{g}(x) - g(x)]^2 \, dz > 0.$$

- Estimator remains mean-square consistent even if restrictions on eigenvalues and Fourier coefficients do not hold.

ASYMPTOTIC NORMALITY

- Model: $Y = g(X) + U; \quad \boldsymbol{E}(U \mid W) = 0$

- Estimator: $\hat{g}(z) = n^{-1}\sum_{i=1}^{n} Y_i(\hat{T}^{+}\hat{f}_{XW}^{(-i)})(z, W_i)$

- Write

$$\hat{g}(z) = n^{-1}\sum_{i=1}^{n} U_i(\hat{T}^{+}\hat{f}_{XW}^{(-i)})(z, W_i) + n^{-1}\sum_{i=1}^{n}(\hat{T}^{+}\hat{f}_{XW}^{(-i)})(z, W_i)g(X_i)$$

$$\equiv S_{n1}(z) + S_{n2}(z)$$

- $S_{n2}(z)$ is asymptotic bias term caused by regularization.
  - As $n \to \infty$,

$$S_{n2}(z) \to \int t^{+}(x, z) f_{XW}(z, w) f_{XW}(x, w) g(x) dx$$

$$= (T + a_n)^{-1} Tg(z)$$

  - $S_{n2}(z) - g(z) \to -a_n (T + a_n)^{-1} g(z)$
  - If there were no regularization, $S_{n2}(z) - g(z)$ would be zero.
  - $S_{n2}(z)$ can be made negligible at almost every z (at cost of higher variance) by choosing a_n to converge to 0 at faster than optimal rate (under-regularizing).
- It suffices to consider asymptotic normality of $S_{n1}(z)$

ASYMPTOTIC NORMALITY (cont.)

- Now consider "random" term

$$S_{n1}(z) = n^{-1}\sum_{i=1}^{n} U_i(\hat{T}^{+}\hat{f}_{XW}^{(-i)})(z,W_i)$$

$$= n^{-1}\sum_{i=1}^{n} U_i(T^{+}f_{XW})(z,W_i)$$

$$+ n^{-1}\sum_{i=1}^{n} U_i(\hat{T}^{+}\hat{f}_{XW}^{(-i)} - T^{+}f_{XW})(z,W_i)$$

$$\equiv R_{n1}(z) + R_{n2}(z).$$

- Under the regularity conditions, $\|R_{n2}(z)\|^2 \big/ Var[R_{n1}(z)] = o_p(1)$, where $\| \ \|$ is $L_2[0,1]$ norm.
  - So $S_{n1}(z)/\sqrt{Var[R_{n1}(z)]} \to^d R_{n1}(z)/\sqrt{Var[R_{n1}(z)]}$ for almost every $z \in [0,1]$.
- $R_{n1}(z)$ is triangular array of mean-zero, iid random variables
  - By central limit theorem for triangular arrays, get

    $$S_{n1}(z)/\sqrt{Var[R_{n1}(z)]} \to^d N(0,1)$$

- Main result: For almost every $z \in [0,1]$,

$$[\hat{g}(z) - g(z) + a_n(T + a_n)^{-1}g(z)] \big/ Var[R_{n1}(z)] \to^d N(0,1)$$

ESTIMATION OF VARIANCE

- $Var[R_{n1}(z)]$ can be replaced by estimator by

  $$\hat{V}(z) = n^{-2} \sum_{i=1}^{n} \hat{U}_i^2 [(\hat{T}^+ \hat{f}_{XW}^{(-i)})(z, W_i)]^2$$

  where $\hat{U}_i = Y_i - \hat{g}(X_i)$

- This estimator may be imprecise due to slow rate of convergence of $\hat{g}$.

- May be useful to explore resampling methods for estimating asymptotic distribution of $\hat{g} - g$.

ADDITIONAL REGULARITY CONDITIONS FOR ASYMPTOTIC NORMALITY

- $\boldsymbol{E}(U^{\nu} \mid W = w) < \infty$ for almost all $w \in [0,1]$ and an even integer ν satisfying

  $$\nu > \max\left[2, \frac{2\beta + \alpha}{2(\beta - 1)}\right]$$

- Asymptotic bias is negligible if

  - $a_n \propto n^{-\rho\alpha/(2\beta+\alpha)}$ for some ρ satisfying $1 < \rho < 1.5$ (under-regularization).

  - Bandwith h is in a range that is narrower than the one specified in previous regularity conditions.

CONCLUSIONS

- Paper gives conditions for pointwise asymptotic normality of Hall-Horowitz nonparametric IV estimator
- This is first pointwise asymptotic distributional result for a nonparametric IV estimator.
- Topics for further research:
    - Use of resampling methods to estimate asymptotic distribution
    - Data-based choices of smoothing parameters
    - Extension to multivariate setting in which some components of X may be exogenous.