A Kernel Weighted Smoothed Maximum Score Estimator for the Endogenous Binary Choice Model

Jerome Krief

Department of Economics - LSU

11 June 2010

The Model

$$\begin{aligned} u &= \dot{x}'\beta + \varepsilon \\ A &= w'\pi + v \\ y &= \mathbb{1}[u \geq 0] \end{aligned}$$

Observable: response y, $\dot{x}' \equiv (z', A)$ a $1 \times K$ and w a $d \times 1$ vector

Not observable: (ε, v)

- $s \equiv (z', w')$ exogenous "instruments" and A **endogenous** due to:

$$\operatorname{cov}(\varepsilon, v) \neq 0$$

- π unknown parameter and β a **key parameter of interest**

Motivation

• Can we construct a **consistent** estimator under a **weak median restriction** which also allows for **testing**?

Contribution:

1. Solve this

2. Offers a robust estimation theory for a semi linear random utility model

Estimation strategy

- Impose key CF median restriction:

$$med(\varepsilon|z,w,v) = med(\varepsilon|v) a.s.$$

$\Rightarrow Med(u|\dot{x},v) = \dot{x}'\beta + \phi(v)$ $a.s.$ where $\phi(v) = med(\varepsilon|v)$

- Fix $v \Rightarrow Med(y|\dot{x},v) = \dot{x}'\beta + \bar{\phi}$ $a.s.$

Identification

- w contains one component not measurable in z

- There exists a partition of $\dot{x}' = (x_1, \tilde{x}')$ where dim x_1 =1 and such that the distribution function of $x_1|\tilde{x}$ has **everywhere positive density** with respect to the Lebesgue measure almost surely.

- The slope coefficient of x_1, noted β_1, is strictly positive (WLG).

- The distribution function of $v|\dot{x}$ is **absolutely continuous** with respect to the Lebesgue measure and its density evaluated at $\bar{v}$ exists almost surely.

Define $\theta_0' \equiv \frac{1}{\beta_1}(\bar{\phi}, \tilde{\beta}')$ where $\tilde{\beta}$ slope coefficient of $\tilde{x}$ and $I \equiv X_1 + X'\theta_0$.

$$\Rightarrow \theta_0 \equiv Argmax_{\theta \in \mathbb{R}^K} E[p(x, I, \bar{v})\mathbb{1}[I + x'\delta \geq 0]f(\bar{v}|x, I)],$$

$p(x, I, \bar{v}) \equiv 2P[y = 1|x, I, \bar{v}] - 1$, $\delta \equiv \theta - \theta_0$.

Description of the KWSMSE

- Let $\{y_i, \dot{x_i}\}_{i=1}^{n}$ be a iid sequence of observations
- Let $\hat{\pi}$ be some given estimator such that $\sqrt{n}(\hat{\pi} - \pi) = O_p(1)$ and write $\hat{v}_i \equiv A_i - \hat{\pi}' w_i$ for $i = 1...n$. For any $\theta \in \mathbb{R}^K$ define:

$$\widetilde{S_n}(\theta) = \frac{1}{nh_q} \sum_{i=1}^{n} (2y_i - 1) D(\frac{x_{1,i} + x_i' \theta}{h}) k(\frac{\hat{v}_i - \overline{v}}{h_q}),$$

- Let θ_0 be an interior point of $\Theta \subset \mathbb{R}^K$ compact and define:

$$\widetilde{\theta}_n \equiv Argmax_{\Theta} \widetilde{S_n}(\theta)$$

KWSMSE Consistency

- $D(.)$ is some chosen bounded function from the real line into itself meeting:

$$lim_{t \to -\infty} D(t) = 0, \; lim_{t \to \infty} D(t) = 1,$$

and

$$D' = K \text{ everywhere with } ||K||_{sup} < \infty.$$

- $k(.)$ is a given kernel satisfying notably:

$$\int k(t)dt = 1, \; \int t^u k(t)dt = 0 \text{ for } u = 1, ..., m-1 \text{ and}$$
$$\int |t^u k(t)|dt < \infty \text{ for } u = 0, m \text{ for some } m \geq 2, \; \int |k(t)|^2 dt < \infty,$$

$$k \text{ is differentiable everywhere with } ||k^{(1)}||_{sup} < \infty,$$

KWSMSE Asymptotic properties-key conditions

Introduce $F[.|x, I, v]$ the cumulative distribution function of $\varepsilon|x, I, v$ and $f(.|x, I)$ the density of $v|x, I$.

• There exists a real constant $M < \infty$ and some open neighborhood of $\bar{v}$ such that almost surely:

$f(v|x, I)$ exists everywhere with $|f(v|x, I)| < M$, and on that neighborhood $F[-\beta_1 I + \bar{\phi}|x, I, v]$ and $f(v|x, I)$ are continuous as functions of v.

• The bandwidth sequence h_q is chosen to satisfy lim $nh_q^4 = \infty$ and lim $\frac{nh^2 h_q^2}{log(n)} = \infty$ as $n \to \infty$.

$$\Rightarrow \text{plim } \widetilde{\theta}_n = \theta_0 \text{ as } n \to \infty.$$

Asymptotic Normality-Differentiability

- suppose that as function of v,

$F[-\beta_1 I + \bar{\phi}|x, I, \bar{v}]$ and $f(v|x, I)$ are m times continuously differentiable in some open neighborhood of $\bar{v}$ for some $m \geq 7$ almost surely.

and as functions of I,

$F[-\beta_1 I + \bar{\phi}|x, I, \bar{v}]$, $f(\bar{v}|x, I)$ and $f(I|x)$ are r times continuously differentiable everywhere for some $r \geq 2$ almost surely.

Asymptotic Normality-The Kernel K is smooth

- Choose K to satisfy notably,

$$\int K(t)dt = 1,\ \int t^u K(t)dt = 0 \text{ for } u = 1, ..., r-1 \text{ and}$$
$$\int |t^u K(t)|dt < \infty \text{ for } u = 0, r,$$

K is symmetrical, twice differentiable everywhere with $||K^{(j)}||_{sup} < \infty$ for $j = 1, 2$ and $\int |K^{(1)}(t)|^2 dt < \infty$.

- $h \propto n^{-a}$ and $h_q \propto n^{-\eta a}$ for where a and η are chosen according to:

$$a \in (Max\{\tfrac{1}{1+\eta+2\eta m}; \tfrac{1}{1+\eta+2r}\}, \tfrac{1}{4+4\eta}) \text{ and } \eta \in (\tfrac{3}{2m-3}, \tfrac{1}{3}).$$

$$\Rightarrow \sqrt{nhh_q}(\widetilde{\theta}_n - \theta_0) \to_d \mathcal{N}(0, \Omega),\ \Omega = H_0^{-1}\Sigma_0 H_0^{-1}$$

Inferential feasibility

Ω can be estimated consistently from data according to:

Let $\widetilde{H_n} = \frac{1}{nh^2 h_q} \sum_{i=1}^{n} (2y_i - 1) x_i x_i' K^{(1)}(\frac{x_{1,i} + x_i' \widetilde{\theta}_n}{h}) k(\frac{\hat{v}_i - \overline{v}}{h_q})$,

and

$\widetilde{\Sigma_n} = \frac{1}{nh^{\gamma_1} h_q^{\gamma_2}} \sum_{i=1}^{n} x_i x_i' |K(\frac{x_{1,i} + x_i' \widetilde{\theta}_n}{h^{\gamma_1}})|^2 |k(\frac{\hat{v}_i - \overline{v}}{h_q^{\gamma_2}})|^2$,

for some chosen constants $\gamma_1 \in (0, 3/4]$ and $\gamma_2 \in (0, 1]$. Then under the assumptions yielding asymptotic normality:

$$\widetilde{H_n} \longrightarrow_p H_0.$$

and

$$\widetilde{\Sigma_n} \longrightarrow_p \Sigma_0.$$

Remarks about the KWSMSE

- $\widetilde{\theta}_n - \theta_0 = O_p(n^{-3/8})$ at least and parametric rate i.e. $O_p(n^{-1/2})$ whenever $Min\{m, r\}$ approaches infinity.

- Asymptotically centered not optimal from an asymptotic MSE perspective Vs regularity conditions and Bootstrapping (Horowitz 2002)

- Maximization of the objective via annealing procedure (see Horowitz 1992 and Szu et al. 1987).

Accelerating Convergence: a SASMSE

- Combine n locals KWSMSE at $v = 1/n, 2/n, ..., 1$ and recover the CF via Sieves
- Use asymptotic representation for the feasible SMSE
- Under certain conditions a faster rate of convergence in probability is achieved

Constructing the SASMSE

Define $e'_K = [O, I_{K-1}]$ the $K-1 \times K$ matrix where the first column is the zero vector while I_{K-1} represents the $K-1 \times K-1$ identity matrix and e'_1 the $1 \times K$ vector whose first entry is 1 and zero elsewhere.

$$\tilde{\theta}(v) \equiv Argmax_{\Theta} \frac{1}{nh_q} \sum_{i=1}^{n} \alpha_i D(\frac{X_{1,i} + X_i'\theta}{h}) k(\frac{\hat{V}_i - v}{h_q}),$$

and

$$\tilde{\beta}(v) \equiv e'_K \tilde{\theta}(v) \text{ while } \tilde{\phi}(v) \equiv e'_1 \tilde{\theta}(v),$$

Choose a Linear Sieves Basis

- Let $\{f_j\}_{j\geq 1}$ a given basis of functions on $[0,1]$ such that $E_\rho = \{f : [0,1] \to \mathbb{R}, f = \sum_{j=1}^{\rho} b_j f_j, b_j \in \mathbb{R}\}$ is dense for a smooth function $\phi(.)$ in the sense that for any $\zeta > 0$,

$$inf_{f\in E_\rho} \, ||f - \phi||_{sup[0,1]} < \zeta \text{ if } \rho \text{ large enough}$$

- examples: Power series, Splines, Trigonometric series, Wavelets (see Chen 2007).

Stage 1: artificial LS

Define $p_n(.)' = (f_1(.), ..., f_{\rho(n)}(.))$ where $\rho(n)$ is some chosen deterministic sequence of natural numbers satisfying $\rho(n) \to \infty$ as $n \to \infty$ but $\rho(n) < n$.

write Λ_n the $n \times \rho(n)$ matrix whose i^{th} row is $p_n(i/n)'$ and $\tilde{\phi}_n$ the $n \times 1$ vector whose i^{th} entry is $\tilde{\phi}(i/n)$.

- Retrieve the following:

$$b_n \equiv Argmin_{b \in \mathbb{R}^{\rho(n)}} ||\tilde{\phi}_n - \Lambda_n b|| \equiv (\Lambda_n' \Lambda_n)^{-1} \Lambda_n' \tilde{\phi}_n.$$

- Essence of the SASMS estimator since for $\rho(n)$ well chosen and under some regularity conditions then,

$$\text{plim } sup_{v \in [0,1]} |b_n' p_n(v) - \tilde{\phi}_0(v)| = 0, \ \tilde{\phi}_0(.) = \frac{1}{\beta_1} \phi(.)$$

Stage 2: "plug in"

- So estimate $\tilde{\phi}_0(v_i)$ with $b_n' p_n(\hat{v}_i)$

- $\Psi(.)$ a kernel from the real line into itself whose derivative exists everywhere and define for an arbitrary β:

$$G_n[\beta] \equiv \frac{1}{nh_*} \sum_{i=1}^{n} \tau(\hat{v}_i)(2y_i - 1)\tilde{x}_i \Psi(\frac{x_{1,i} + \tilde{x}_i{}' \beta + b_n' p_n(\hat{v}_i)}{h_*}),$$

and

$$H_n[\beta] \equiv \frac{1}{nh_*^2} \sum_{i=1}^{n} \tau(\hat{v}_i)(2y_i - 1)\tilde{x}_i \tilde{x}_i{}' \Psi^{(1)}(\frac{x_{1,i} + \tilde{x}_i{}' \beta + b_n' p_n(\hat{v}_i)}{h_*}),$$

where $\tau(.) = \mathbb{1}[0 \leq . \leq 1]$ and h_* is a deterministic strictly positive sequence of real numbers meeting lim $h_* = 0$ as $n \to \infty$.

- The SASMS estimator:

$$\bar{\beta} \equiv \tilde{\beta}(v) - H_n[\tilde{\beta}(v)]^{-1} G_n[\tilde{\beta}(v)],\ v \in [0, 1]$$

SASMSE key conditions-Differentiability

- $\phi(.)$ is p times **continuously differentiable** on $[0,1]$ for some $p \geq 5$ and there exists some finite constant C and some $\gamma \in (0,1]$ such that $|\phi^{(p)}(v_1) - \phi^{(p)}(v_2)| \leq C|v_1 - v_2|^\gamma$ for all $(v_1, v_2) \in [0,1] \times [0,1]$.

Write $\lambda_i \equiv \frac{1}{\beta_1} Med(U|\dot{x}_i, v_i)$. Define $F[.|\tilde{x}, \lambda, v]$ the cdf of $\varepsilon|\tilde{x}, \lambda, v$ $f(.|\tilde{x}, v)$ the Lebesgue density of $\lambda|\tilde{x}, v$.

- $F[-\beta_1\lambda + \phi(v)|\tilde{x}, \lambda, v]$ and $f(\lambda|\tilde{x}, v)$ are, as functions of λ, s times **continuously differentiable in some open neighborhood of the origin**for some $s \geq 4$ almost surely.

SASMSE key conditions-Kernels and Bandwidths

• Ψ is a kernel of order s meeting the same differentiability and integrability conditions as kernel K used for computing the KWSMSE.

• The bandwidth h_* satisfies $nh_*^8/log(n) \rightarrow \infty$ as $n \rightarrow \infty$ and $inf_{g \in E_{\rho(n)}} ||g - \phi||_{sup[0,1]} ||p_n||_{sup[0,1]} = o(h_*^3)$.

$$\Rightarrow \text{plim } \bar{\beta} = \tilde{\beta}_0, \tilde{\beta}_0 \equiv \frac{\tilde{\beta}}{\beta_1}$$

SASMSE key conditions-CAN

- $h_*/hh_q \to \infty$ as $n \to \infty$ and $nh_*^{2s+1} \to 0$ as $n \to \infty$.

$$\Rightarrow \sqrt{nh_*}(\bar{\beta} - \tilde{\beta}_0) \to_d \mathcal{N}(0, \aleph),\ \aleph \equiv Q^{-1}\Xi Q^{-1}.$$

- $\aleph$ can be estimated consistently from data via:

$$\hat{\Xi} \equiv \frac{1}{nh_*} \sum_{i=1}^{n} \tau(\hat{v}_i)\tilde{x}_i\tilde{x}_i{}'|\Psi(\frac{x_{1,i}+\tilde{x}_i{}'\tilde{\beta}(v)+b_n'p_n(\hat{v}_i)}{h_*})|^2,$$

$$H_n[\tilde{\beta}(v)] \longrightarrow_p Q \text{ and } \hat{\Xi} \longrightarrow_p \Xi.$$

Monte Carlo experiment

• These estimators used to estimate the parameter $\beta = 1$ when the data generating process obeys:

$$Y = 1 \text{ } if \text{ } Z + \beta A + \varepsilon \geq 0 \text{ and } Y = 0 \text{ otherwise,}$$

$$A = \pi W + V,$$

$$\varepsilon = \phi(V) + e,$$

• (Z, W) standard bivariate Normal couple of correlation coefficient ϱ , $V \sim \mathcal{N}(0, 1)$, and π set equal to 1.

3 designs

$\bullet$Design ST: $\varrho = 0.5$; $\phi(V) = exp(-V^2)$; $e = (1 + Z^2 + Z^4)T$ where T is Student with 3 degrees of freedom.

$\bullet$Design PR: $\varrho = 0.5$; $\phi(V) = 0.5V$; $e \sim \mathcal{N}(0, 1)$.

$\bullet$Design LG: $\varrho = 0$; $\phi(V) = cos(\pi V)$; $e \sim$ *Logistic*.

In addition use the (LIML) proposed in Rivers and Vuong (1988) and (2SLS) suggested in Lewbel (2000).

• A simulation for a sample size $n = 250, 500$ and 1000 and 1000 replications for all estimators but the SASMS estimator.

• For that later, experiments with $n = 1000$ not performed and 500 replications completed due to the long computational time.

KWSMSE-implementation

- The smoothing of the indicator:

$$D(t) = [0.5 + \tfrac{105}{64}(t - \tfrac{5}{3}t^3 + \tfrac{7}{5}t^5 - \tfrac{3}{7}t^7)]1[|t| \leq 1] + 1[t > 1].$$

- The weighting of the objective:

$$k(t) = \tfrac{1}{48}(105 - 105t^2 + 21t^4 - t^6)\tfrac{1}{\sqrt{2\pi}}exp(-\tfrac{1}{2}t^2),$$

- π estimated via least square and $\bar{v} = 0$ chosen.

- Bandwidths selected via a simple Silverman's like rule of thumb: $h = \hat{\sigma}_I n^{-3/16}$ and $h_q = \hat{\sigma}_V n^{-3\eta/16}$ where $\eta = 1/3$,

$\hat{\sigma}_V =$ std$\{\hat{V}_i\}_{i=1..n}$ and $\hat{\sigma}_I =$std$\{X_{1,i} + X_i'\tilde{\theta}\}_{i=1..n}$ with $\tilde{\theta}$ a previous KWSMS estimator.

SASMSE-implementation

- Collect the n locals KWSMSE's at $v = 1/n, ..., 1$ as described previously.

- The pseudo LS computed using the Trigonometric Cosines basis of $[0,1]$ i.e. $f_1(v) = 1$ and $f_2(v) = \sqrt{2}cos(2\Pi(j-1)v), j \geq 2$.

- $\Psi_6(t) = \frac{315}{2048}(15 - 140t^2 + 378t^4 - 396t^6 + 143t^8)1[|t| \leq 1]$,

- $h_* = \hat{\sigma}_\lambda n^{-1/10}$ where $\hat{\sigma}_\lambda$=std$\{x_{1,i} + \tilde{x}_i{}' \widetilde{\beta(1/n)} + b_n' p_n(\hat{v}_i)\}_{i=1..n}$

Results

Table: Losses

n=250	LIML	2SLS	KWSMS	SASMS
	RMSE	RMSE	RMSE	RMSE
ST	0.300	0.638	0.240	0.368
PR	0.178	0.676	0.939	0.786
LG	0.141	0.318	0.434	1.106
n=500				
ST	0.236	0.596	0.146	0.135
PR	0.118	0.630	0.355	0.347
LG	0.104	0.270	0.244	0.380
n=1000				
ST	0.184	0.560	0.098	
PR	0.082	0.584	0.255	
LG	0.070	0.236	0.168	

Results

Table: Sizes

n=250	KWSMS	SASMS
Nominal level	0.01—0.05—0.10	0.01—0.05—0.10
ST	0.11—0.20—0.27	0.03—0.07—0.09
PR	0.23—0.34—0.42	0.10—0.16—0.21
LG	0.26—0.38—0.45	0.09—0.17—0.20
n=500		
ST	0.07—0.12—0.19	0.01—0.02—0.06
PR	0.17—0.26—0.33	0.08—0.14—0.18
LG	0.24—0.36—0.42	0.06—0.10—0.13
n=1000		
ST	0.04—0.10—0.16	
PR	0.13—0.23—0.30	
LG	0.19—0.29—0.35	

Conclusion

- New tools to conduct hypothesis testing in the binary choice model when endogeneity is present without having to impose strong distributional assumptions.
- Robust estimator for a semi linear random utility model
- Optimal choice for the bandwidths?
- Estimation theory for endogeneity with discontinuous objective?