.

INFERENCE BASED ON CONDITIONAL MOMENT INEQUALITIES

Donald W. K. Andrews
Cowles Foundation
Yale University

Xiaoxia Shi
Department of Economics
University of Wisconsin

Introduction

- model:
  - true value θ_0 ($\in \Theta \subset R^d$) satisfies conditional moment inequalities &/or equalities:

$$\begin{aligned} E_{F_0}(m_j(W_i,\theta_0)\,|X_i) &\geq 0 \text{ a.s. } [F_{X,0}] \text{ for } j = 1,...,p \\ E_{F_0}(m_j(W_i,\theta_0)\,|X_i) &= 0 \text{ a.s. } [F_{X,0}] \text{ for } j = p+1,...,p+v \end{aligned}$$

  - $\{W_i : i \geq 1\}$ are i.i.d. w/ dist'n F_0

- key feature: true value θ_0 is not (necessarily) identified

- we are interested in confidence sets for θ_0

- for simplicity, all formulae below take $v = 0$, i.e., no mom equalities

- some examples in econometrics:
  - game theory models w/ multiple equilibria: using necessary conditions for Nash equilibria,
    e.g., see Ciliberto & Tamer (2003), Andrews, Berry, & Jia (2004), Pakes, Porter, Ho, & Ishii (2004), & Bajari, Benkard, & Levin (2008)
  - sufficient conditions for Nash equilibria, e.g., see Ciliberto & Tamer (2003), Beresteanu & Molinari (2008)
  - data censoring, e.g., when continuous variable only observed to lie in interval, see Manski & Tamer (2002)

- missing data, see Imbens & Manski (2004)
- sign restrictions, see Moon & Schorfheide (2006)
- other industrial organization examples, see Pakes, Porter, Ishii, & Ho (2004) & Eisenberg (2008)

Outline

- approach: transform cond'l moment inequalities/equalities into ∞ number of uncond'l ones
  - do so w/ no loss of identification power

- construct CS's by inverting Cramér-von Mises-type or Kol-Smirn-type tests

- crit vals obtained by generalized moment selection (GMS), (estimated) least favorable dist'n, or subsampling
  - GMS crit vals are preferred

- show CS's have correct uniform asy cov prob's
  - new methods are required b/c ∞-dimensional nuisance parameter affects asy dist'ns

- show tests are consistent against all fixed alternatives

- show tests have power against some $n^{-1/2}$-local alternatives
  - but not all such alternatives

- for computational purposes, extend results to allow truncated sums & simulated integrals for CvM tests

- extend results to allow preliminary est'n of add'l parameters τ that are identified when true θ is known
  - often arises w/ game theory models
  - arises in one model considered in simulations

- simulation results for 3 models:
  - quantile model w/ selection—quantile monotone IV
  - interval outcome reg'n model
  - entry game model w/ multiple equilibria

- simulation results yield suitable values for tuning parameters

- also, provide comparisons of different forms of test stats and crit vals

- summary of numerical work:
  - size properties are excellent in many cases, pretty good in all cases
  - CvM better than KS in terms of lower false cov probs
  - in some models, choice between Sum, QLR, & Max S function doesn't matter
    - when it matters, Max performs best
  - GMS crit vals better than least fav cv or subsampling in terms of false cov probs
  - asy GMS & bootstrap GMS crit vals perform fairly similarly
  - sentivitivity to tuning parameters is relatively low in most cases
    - quite low in many cases

Extensions

- separate paper 1: extend results to allow for ∞ number, or finite but large number, of cond'l or uncond'l moment inequalities/equalities
  - allows one to cover complicated game theory models, tests of stochastic dominance, cond'l stoch dominance, cond'l treatment effects, ...
- separate paper 2: extend results to allow for **nonparametric** parameters of interest
  - allows for nonpar reg'n or nonpar quantile reg'n w/ sel'n
- in both papers,
  - obtain uniform asy size results
  - no loss of information—power agst all fixed alternatives
  - local power results

Related Literature

- most work in literature is on uncond'l mom ineq's
  - e.g., Imbens & Manski (2004), Moon & Schorfheide (2006), Chernozhukov, Hong, & Tamer (2007), Andrews & Jia (2008), Beresteanu & Molinari (2008), Romano & Sheik (2008, 2010), Rosen (2008), Andrews & Guggenberger (2009), Andrews & Soares (2010), Bugni (2010), Canay (2010), Stoye (2010)

- cond'l mom ineq's can reduce size of identified set

- so, first-order difference between uncond'l & cond'l mom ineq's
  - different from situation w/ mom equalities

- re cond'l mom ineq's: Chernozhukov, Lee, & Rosen (2008), Fan (2008), Kim (2008)

- many moment condn's: Menzel (2008)

Confidence Sets

- we are interested in confidence sets (CS's) for true value θ_0
  - as opposed to CS for identified set

- we consider CS obtained by inverting test

- notation: nominal level $1 - \alpha$ CS for θ is

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_{1-\alpha}(\theta)\}$$

  where $c_{1-\alpha}(\theta)$ is a data-dependent critical value

Estimation

- re estimation: use CS with $\alpha = .5$
  - yields half-median unbiased est'r asy'ly in unif'm sense
  - i.e., prob of including a bdy point of the identified set is $\geq .5$ asy'ly
  - solves inward-bias problem
  - circumvents need for somewhat arbitrary sequence $\varepsilon_n \downarrow 0$
  - related to method in Chernozhukov, Lee, & Rozen (2008)

Parameter Space

- uniform results require precise parameter space $\mathcal{F}$ for (θ_0, F_0)

- $\mathcal{F}$ is collection of (θ, F) such that

$$\begin{aligned}
&\text{(i) } \theta \in \Theta \subset R^d \\
&\text{(ii) } \{W_i : i \geq 1\} \text{ are i.i.d. under } F \\
&\text{(iii) } E_F(m_j(W_i,\theta)\,|X_i) \geq 0 \text{ a.s. } [F_X]\ \forall j \leq p \\
&\text{(iv) } 0 < Var_F(m_j(W_i,\theta)) < \infty\ \forall j \leq p \\
&\text{(v) } E_F\left|m_j(W_i,\theta)/\sigma_{F,j}(\theta)\right|^{2+\delta} \leq B\ \forall j \leq p
\end{aligned}$$

  for some $B < \infty$ & $\delta > 0$

- identified set given F_0:

$$\Theta_{F_0} = \{\theta : (\theta, F_0) \in \mathcal{F}\}$$

Conditional $\to$ Unconditional Moment Conditions

$$E_{F_0} m_j (W_i, \theta) g_j (X_i) \geq 0, \ \forall j \leq p \text{ for } g = (g_1, ..., g_p)' \in \mathcal{G} \tag{1}$$

- where $g = (g_1, ..., g_p)'$ are functions of X_i
  - $\mathcal{G}$ is infinite
- identified set, $\Theta_{F_0}(\mathcal{G})$, defined by these uncond'l mom ineq's:

  $\Theta_{F_0}(\mathcal{G}) = \{\theta \in \Theta : (1) \text{ holds } \& \ (\theta, F_0) \text{ satisfies (i), (ii), (iv), (v) of } \mathcal{F}\}$

- choose $\mathcal{G}$ so that $\Theta_{F_0}(\mathcal{G}) = \Theta_{F_0}$

Class of Test Statistics

- notation:

$$\overline{m}_n(\theta, g) = n^{-1} \sum_{i=1}^{n} \begin{pmatrix} m_1(W_i, \theta) g_1(X_i) \\ m_2(W_i, \theta) g_2(X_i) \\ \vdots \\ m_p(W_i, \theta) g_p(X_i) \end{pmatrix} \text{ for } g \in \mathcal{G}$$

- sample variance-covariance matrix of $n^{1/2}\overline{m}_n(\theta, g)$:

$$\widehat{\boldsymbol{\Sigma}}_n(\theta, g) = n^{-1} \sum_{i=1}^{n} \left(m(W_i, \theta, g) - \overline{m}_n(\theta, g)\right) \left(m(W_i, \theta, g) - \overline{m}_n(\theta, g)\right)'$$

- nonsingular adjustment:

$$\overline{\Sigma}_n(\theta, g) = \widehat{\Sigma}_n(\theta, g) + \varepsilon \cdot Diag(\widehat{\Sigma}_n(\theta, 1_p))$$

- in simulations, use $\varepsilon = 5/100$

- Cramér-von Mises-type (CvM) statistic:

$$T_n(\theta) = \int S(n^{1/2}\overline{m}_n(\theta, g), \overline{\mathbf{\Sigma}}_n(\theta, g))dQ(g)$$

  - S is non-negative function
  - Q is a probability measure on $\mathcal{G}$ (weight function)
  - integral is over $\mathcal{G}$

- Kolmogorov-Smirnov-type (KS) statistic:

$$T_n(\theta) = \sup_{g \in \mathcal{G}} S(n^{1/2}\overline{m}_n(\theta, g), \overline{\mathbf{\Sigma}}_n(\theta, g))$$

- Examples: **Sum** test function

$$S_1(m, \boldsymbol{\Sigma}) = \sum_{j=1}^{p} [m_j/\sigma_j]_-^2$$

QLR test function

$$S_2(m, \boldsymbol{\Sigma}) = \inf_{t \in R_+^p} (m - t)' \boldsymbol{\Sigma}^{-1} (m - t)$$

Max test function

$$S_3(m, \boldsymbol{\Sigma}) = \max\{[m_1/\sigma_1]_-^2, ..., [m_p/\sigma_p]_-^2\}$$

$\mathbf{S}$ Function

(a) $S(m, \boldsymbol{\Sigma})$ is non-increasing in m

(b) $S(m, \boldsymbol{\Sigma}) = S(Dm, D\boldsymbol{\Sigma}D)$ $\forall$pd diagonal $D \in R^{p \times p}$

(c) $S(m, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_1) \leq S(m, \boldsymbol{\Sigma})$ $\forall p \times p$ psd matrices $\boldsymbol{\Sigma}_1$

(d) $S(am, \boldsymbol{\Sigma}) = a^{\chi} S(m, \boldsymbol{\Sigma})$ for some $\chi > 0$, $\forall$ scalars $a > 0$, $\forall m$, & $\forall \boldsymbol{\Sigma}$

(e) $S(m, \Omega) \geq 0$

(f) $S(m, \Omega) > 0$ iff $m_j < 0$ for some $j \leq p$

(g) $S(m, \boldsymbol{\Sigma})$ is uniformly continuous

Collection $\mathcal{G}$

- $\mathcal{G}$ must satisfy 2 assumptions
- given (θ, F), let

$$\mathcal{X}_F(\theta) = \{x \in R^{d_x} : E_F(m_j\,(W_i, \theta)\,|X_i = x) < 0 \text{ for some } j \leq p\}$$

- **Assumption CI.** For any $\theta \in \Theta$ & F for which $P_F(X_i \in \mathcal{X}_F(\theta)) > 0$ & $E_F||m(W_i, \theta)|| < \infty$, there exists $g \in \mathcal{G}$ such that

$$E_F m_j(W_i, \theta) g_j(X_i) < 0 \text{ for some } j \leq p$$

- note (θ, F) are not in $\mathcal{F}$
- simple result: Assumption CI implies $\Theta_F(\mathcal{G}) = \Theta_F$ for all F w/ $\sup_{\theta \in \Theta} E_F||m(W_i, \theta)|| < \infty$

- use "manageability" condition, **Assumption M**, on stoc processes $\{g(X_{n,i}) : g \in \mathcal{G},\, i \leq n, n \geq 1\}$
  - from Pollard (1990)
  - regulates complexity of $\mathcal{G}$
  - ensures that $\{n^{1/2}(\overline{m}_n(\theta, g) - E_{F_n}\overline{m}_n(\theta, g)) : g \in \mathcal{G}\}$ satisfies FCLT under drifting sequences $\{F_n : n \geq 1\}$

Collections $\mathcal{G}$ that satisfy Assumptions CI & M

- **Example 1. (Countable Hypercubes).**
  - transform regr's to $[0,1]^{d_x}$
  - let

$$\mathcal{G}_{c\text{-}cube} = \{g(x) : g(x) = 1(x \in C) \cdot 1_p \text{ for } C \in \mathcal{C}_{c\text{-}cube}\}$$

  - $\mathcal{C}_{c\text{-}cube}$ contains cubes in $[0,1]^{d_x}$ with side-lengths $(2r)^{-1}$ for integers $r = r_0, r_0 + 1, \ldots$
  - this class is countable & countable center points

- **Example 2 (Uncountable Boxes).**

- **Example 3 (Data-dependent Boxes).**

- **Example 4. (B-splines & Finite-Support Kernels).**

- **Example 5. (Continuous/Discrete Regressors)**

- **Result:** Assumptions CI and M hold for $\mathcal{G}_{c\text{-}cube}$, $\mathcal{G}_{box}$, $\mathcal{G}_{box,dd}$, $\mathcal{G}_{B\text{-}spline}$, and $\mathcal{G}_{c/d}$

Weight Function Q

- for test to have power against all fixed alternatives, Q cannot "ignore" any elements $g \in \mathcal{G}$

- let ρ_X be L^2 pseudo-metric on $\mathcal{G}$:

$$\rho_X(g, g^*) = (E_{F_{X,0}} ||g(X_i) - g^*(X_i)||^2)^{1/2} \text{ for } g, g^* \in \mathcal{G}$$

  where $F_{X,0}$ is distribution of X_i under F_0

- **Assumption Q.** Support of Q under pseudo-metric ρ_X contains $\mathcal{G}$.
  I.e., $\forall \delta > 0$, $Q(\mathcal{B}_{\rho_X}(g, \delta)) > 0 \ \forall g \in \mathcal{G}$.

- $\mathbf{Q}$ **for** $\mathcal{G}_{\mathbf{c\text{-}cube}}$
  - $\exists$ 1-1 mapping $\Pi_{c\text{-}cube} : \mathcal{G}_{c\text{-}cube} \rightarrow AR = \{(a, r) \in \{1, ..., 2r\}^{d_x} \times \{r_0, r_0 + 1, ...\}\}$
  - let Q_{AR} be a probability measure on AR w/ full support
  - e.g., uniform on $a \in \{1, ..., 2r\}^{d_x}$ & dist'n for r with pmf $\{w(r) : r = r_0, r_0 + 1, ...\}$
  - simulations use $w(r) = (r^2 + 100)^{-1}$
  - then, $Q = \Pi^{-1}_{c\text{-}cube} Q^*$ is probability measure on $\mathcal{G}_{c\text{-}cube}$

- for $\mathbf{Q}$ **for** $\mathcal{G}_{\mathbf{c\text{-}cube}}$, test statistic is

$$T_n(\theta) = \sum_{r=r_0}^{\infty} w(r) \sum_{a \in \{1,\ldots,2r\}^{d_x}} (2r)^{-d_x} S(n^{1/2}\overline{m}_n(\theta, g_{a,r}), \overline{\Sigma}_n(\theta, g_{a,r}))$$

where $g_{a,r}(x) = 1(x \in C_{a,r}) \cdot 1_k$ for $C_{a,r} \in \mathcal{C}_{c\text{-}cube}$

Computation

- test statistic $T_n(\theta)$ involves ∞ sum or integral wrt Q

- analogous ∞ sum or integral appear in defn's of crit vals
  - can't compute them exactly

- ∞ sums can be approx'd by truncation & integrals by simulation or quasi-Monte Carlo methods

- approx test stat for count hyper-cubes:

$$\overline{T}_n(\theta) = \sum_{r=r_0}^{s_n} w(r) \sum_{a \in \{1,...,2r\}^{d_x}} (2r)^{-d_x} S(n^{1/2}\overline{m}_n(\theta, g_{a,r}), \overline{\mathbf{\Sigma}}_n(\theta, g_{a,r}))$$

where $g_{a,r}(x) = 1(x \in C_{a,r}) \cdot 1_k$ for $C_{a,r} \in \mathcal{C}_{c\text{-}cube}$

- for simulated test stat, let $\{g_1, ..., g_{s_n}\}$ be s_n i.i.d. functions drawn from $\mathcal{G}$ according to the distribution Q

- simulated test statistic is

$$\widehat{T}_{n,s_n}(\theta) = s_n^{-1} \sum_{\ell=1}^{s_n} S(n^{1/2}\overline{m}_n(\theta, g_\ell), \overline{\mathbf{\Sigma}}_n(\theta, g_\ell))$$

- we show if $s_n \to \infty$ as $n \to \infty$ uniform asymptotic validity of tests & CS's hold
  - main issue is uniformity
  - asymptotic power results under fixed alternatives hold
  - most results under $n^{-1/2}$-local alternatives hold

Pointwise Vs Uniform Asymptotics

- asy distn's of $T_n(\theta)$ are discont. in F
  - due to mom. ineq. slackness function
  - get different **pointwise** asy dist'n depending on

$$n^{1/2} E_F m_j(W_i,\theta) g(X_i) \begin{cases} = 0 \ \forall n & \text{if } E_F m_j(W_i,\theta) g(X_i) = 0 \\ \to \infty & \text{if } E_F m_j(W_i,\theta) g(X_i) > 0 \end{cases}$$

  - this does not reflect finite-sample situation
    - ∗ no discont'y in finite samples

- pointwise asy dist'ns do not capture finite-sample behavior

- effect of asy discont'y greater w/ cond'l mom ineq's than uncond'l mom. ineq.s
  - in several respects

- e.g., if cond'l mean function $\mu_j(x, \theta) = E_F m_j(W_i, \theta)|X_i = x)$ is cont. in x, then at bdy pts θ there are always points x for which $\mu_j(x, \theta)$ is positive, but arbitrarily close to 0

- so, there is always a uniformity issue

- second, we want to show:

$$\liminf_{n\to\infty} \inf_{(\theta,F)\in\mathcal{F}} P_F(\theta \in CS_n) = 1 - \alpha \tag{2}$$

- with finite # of uncond'l mom. ineq.s, it is sufficient to consider certain seq.s of drifting dist'ns, see Andrews & Guggenberger (2009) or Andrews, Cheng, & Guggenberger (2009)

- w/ cond'l mom ineq.s, this is not sufficient
  - b/c ∞ dim'l nuisance par affects asy dist'n

- different method is required to show (2)

Uniform Asymptotic Distribution of $\mathbf{T_n}(\theta)$

- let

$$\begin{aligned}\nu_{n,F}(\theta,g) &= n^{1/2}[\overline{m}_n(\theta,g) - E_F m(W_i,\theta,g)]\\ h_{1,n,F}(\theta,g) &= n^{1/2} E_F m(W_i,\theta,g)\\ h_{2,F}(\theta,\cdot\,\cdot) &= CovKernel \text{ of } \nu_{n,F}(\theta,\cdot) \text{ under } F\end{aligned}$$

- $h_{1,n,F}(\cdot)$ is function from $\mathcal{G}$ to R_+^p that depends on slackness of moment inequalities & n
  - let $h_{n,F}(\theta,\cdot) = (h_{1,n,F}(\cdot), h_{2,F}(\theta,\cdot\,\cdot))'$

- write test stat as

$$\begin{aligned} T_n(\theta) &= \int S\left(n^{1/2}\overline{m}_n(\theta, g), \overline{\Sigma}_n(\theta, g)\right) dQ(g) \\ &= \int S(\nu_{n,F}(\theta, g) + h_{1,n,F}(\theta, g), h_{2,F}(\theta, g, g) + \varepsilon I_p + o_p(1)) dQ(g) \end{aligned}$$

- let $\{\nu_{h_2}(g) : g \in \mathcal{G}\}$ be mean zero R^p-valued Gaussian process with cov kernel $h_2(\cdot,\cdot)$ on $\mathcal{G} \times \mathcal{G}$
  - let $\mathcal{H}_2$ be parameter space for $h_2(\cdot,\cdot)$ in model given $\mathcal{F}$

- let $h_1(\cdot)$ be any function from $\mathcal{G}$ to $R^p_{+,\infty}$

- for $h = (h_1, h_2)$, let

$$T(h) = \int S(\nu_{h_2}(g) + h_1(g), h_2(g,g) + \varepsilon I_k) dQ(g)$$

- **Result:** Under Assumptions M & S, $\forall$ compact subsets $\mathcal{H}_{2,cpt}$ of $\mathcal{H}_2$, $\forall$ constants $x_{h_{n,F}(\theta)} \in R$ that may depend on $h_{n,F}(\theta)$, & $\forall \delta > 0$,

$$\limsup_{n\to\infty} \sup_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} \Big[P_F(T_n(\theta) > x_{h_{n,F}(\theta)}) - P(T(h_{n,F}(\theta)) + \delta > x_{h_{n,F}(\theta)})\Big] \leq 0$$

$$\liminf_{n\to\infty} \inf_{\substack{(\theta,F)\in\mathcal{F}: \\ h_{2,F}(\theta)\in\mathcal{H}_{2,cpt}}} \Big[P_F(T_n(\theta) > x_{h_{n,F}(\theta)}) - P(T(h_{n,F}(\theta)) - \delta > x_{h_{n,F}(\theta)})\Big] \geq 0$$

Critical Values

- denote $1 - \alpha$ quantile of $T(h_{n,F}(\theta))$ by

$$c_0(h_{1,n,F}(\theta), h_{2,F}(\theta), 1 - \alpha)$$

- $h_{1,n,F}(\theta)$ & $h_{2,F}(\theta)$ are not known
- replace $h_{2,F}(\theta)$ by uniformly consistent estimator $\widehat{h}_{2,n}(\theta)$ $(= \widehat{h}_{2,n}(\theta, \cdot, \cdot))$
- $h_{1,n,F}(\theta)$ $(= h_{1,n,F}(\theta, \cdot))$ cannot be consistently estimated
- can replace $h_{1,n,F}(\theta)$ by zero function, $0_{\mathcal{G}}$, on $\mathcal{G}$
  - least-favorable choice
  - or worse than least favorable
- poor power properties

- subsampling crit vals
  - usual definition
  - for uncond'l mom ineq's, Bugni (2010) & Andrews & Soares (2010) show that subsampling is dominated by generalized moment selection (GMS) crit values re asy size & power

- focus on generalized moment selection (GMS) crit vals

GMS Crit Values

- replace $h_{1,n,F}(\theta)$ by data-dependent function $\varphi_n(\theta)$ $(= \varphi_n(\theta, \cdot))$ on $\mathcal{G}$
  - $\varphi_n(\theta, g)$ is constructed to be $\leq h_{1,n}(\theta, g)$ $\forall g \in \mathcal{G}$ w/ prob $\to 1$

- GMS crit val is

$$c_0(\varphi_n(\theta), \widehat{h}_{2,n}(\theta), 1 - \alpha + \eta) + \eta$$

  for infinitessimal uniformity factor $\eta > 0$

- bootstrap version: replace estimated Gaussian process $\nu_{\widehat{h}_{2,n}(\theta)}(\cdot)$ by bootstrap emp'l process $\nu_n^*(\cdot)$ & replace estimated variance process $\widehat{h}_{2,n}(\theta)$ by bootstrap version
  - no higher-order improvements—test stat not asy'ly pivotal

- definition of $\varphi_n(\theta, g)$:
  - measure of slackness of mom. ineq.:

$$\xi_n(\theta, g) = \kappa_n^{-1} \widehat{D}_n^{-1/2}(\theta) n^{1/2} \overline{m}_n(\theta, g)$$

$$\varphi_{n,j}(\theta, g) = \begin{cases} B_n & \text{if } \xi_{n,j}(\theta, g) > 1 \\ 0 & \text{if } \xi_{n,j}(\theta, g) \leq 1 \end{cases}$$

  - not "pure" moment selection b/c $B_n < \infty$

- Assumption GMS. $\kappa_n - \zeta B_n \to \infty$ as $n \to \infty$ for some $\zeta > 1$

- in simulations, use $\kappa_n = (0.3 \ln(n))^{1/2}$ & $B_n = (0.4 \ln(n) / \ln \ln(n))^{1/2}$

Uniform Asymptotic Coverage Probability Results

- **Main Result:** Under Assumptions M, S, & GMS, $\forall$ compact subset $\mathcal{H}_{2,cpt}$ of $\mathcal{H}_2$, GMS confidence sets CS_n satisfy

(a) $$\liminf_{n \to \infty} \inf_{\substack{(\theta,F) \in \mathcal{F}: \\ h_{2,F}(\theta) \in \mathcal{H}_{2,cpt}}} P_F(\theta \in CS_n) \geq 1 - \alpha$$

(b) if Assumption GMS2 also holds,

$$\lim_{\eta \to 0} \liminf_{n \to \infty} \inf_{\substack{(\theta,F) \in \mathcal{F}: \\ h_{2,F}(\theta) \in \mathcal{H}_{2,cpt}}} P_F(\theta \in CS_n) = 1 - \alpha$$

Asymptotic Power Against Fixed Alternatives

- show that power of GMS tests against "all" fixed alternatives $\longrightarrow 1$ as $n \to \infty$

- this implies that given fixed true F_0 & any θ_* *not* in identified set Θ_{F_0}, GMS CS's do not include θ_* with prob $\to 1$

- here is where Assumptions CI (re $\mathcal{G}$) & Q (re weight measure) are used

Asymptotic Local Power

- show GMS tests have power against some, but not all, $n^{-1/2}$-local alternatives
  - depends on seq. $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ from which perturbations are taken
  - where θ_n is true par value

Assumption LA1.

(a) $\theta_{n,*} = \theta_n + \lambda n^{-1/2}(1 + o(1))$ for some $\lambda \in R^{d_\theta}$, $\theta_{n,*} \to \theta_0$, & $F_n \to F_0$ for some $(\theta_0, F_0) \in \mathcal{F}$

(b) $n^{1/2} E_{F_n} m_j(W_i, \theta_n, g)/\sigma_{F_n,j}(\theta_n) \to h_{1,j}(g)$ for some $h_{1,j}(g) \in R_{+,\infty}$ $\forall j \leq p$ & $g \in \mathcal{G}$

Assumption LA2. The $p \times d$ matrix $\Pi_F(\theta, g) = (\partial/\partial\theta')[D_F^{-1/2}(\theta) E_F m(W_i, \theta, g)]$ exists & is cont. in nghd of (θ_0, F_0) $\forall g \in \mathcal{G}$

- for KS test:

Assumption LA3.

For some $g \in \mathcal{G}$, $h_{1,j}(g) < \infty$ & $\Pi_{0,j}(g)'\lambda < 0$ for some $j \leq p$

- for CvM test:

Assumption LA3$'$.

$Q(\{g \in \mathcal{G} : h_{1,j}(g) < \infty$ & $\Pi_{0,j}(g)'\lambda < 0$ for some $j \leq p) > 0$

Result: Suppose $\lambda = \beta\lambda_0$ for $\beta \in R$ & $\lambda_0 \in R^d$ fixed, then $\lim_{\beta\to\infty} \lim_{n\to\infty} Power_{n,\beta}(GMS \text{ test}) = 1$

Simulation Results

- 3 models: quantile sel'n, interval outcome reg'n, entry game
- **quantile sel'n model:**
- conditional τ-quantile of a treatment response given value of covariate X_i
- use *quantile* monotone instrumental variable (QMIV) condition
  - variant of Manski and Pepper's (2000) Monotone Instrumental Variable (MIV) condition
  - bounds on quantiles: Manski (1994), Lee & Melenberg (1998), & Blundell, Gosling, Ichimura, & Meghir (2007)
- model set-up is quite similar to that in Manski and Pepper (2000)

- obs are i.i.d. for $i = 1, ..., n$
- $y_i(t)$ is individual i's "conjectured" response given treatment $t \in \mathcal{T}$
- T_i is realization of treatment for individual i
- observed outcome variable is $Y_i = y_i(T_i)$
- X_i is a covariate
- $\theta =$ cond'l τ-quantile of $y_i(t_0)$ given $X_i = x_0$ for some $t_0 \in \mathcal{T}$ & x_0
  - denoted $\theta = Q_{y_i(t_0)|X_i}(\tau|x_0)$

- examples: (i) $y_i(t)$ is conjectured wages of individual i for t years of schooling
  - T_i is realized years of schooling
  - X_i is measured ability or wealth

- (ii) $y_i(t)$ is conjectured wages when individual i is employed, say $t = 1$
  - X_i is measured ability or wealth
  - selection occurs due to elastic labor supply

- (iii) $y_i(t)$ is some health response of individual i given treatment t
  - T_i is the realized treatment—non-randomized or randomized but subj to imperfect compliance
  - X_i is some characteristic of individual i, such as weight, blood pressure

- quantile monotone IV assumption is:

Assumption QMIV. If $x_1 \leq x_2$,

$$Q_{y_i(t)|X_i}(\tau|x_1) \leq Q_{y_i(t)|X_i}(\tau|x_2)$$

- for Monte Carlo simulations, DGP:

$$\begin{aligned} y_i(1) &= \mu(X_i) + \sigma(X_i)\, u_i, \text{ where } \partial\mu(x)/\partial x \geq 0 \text{ and } \sigma(x) \geq 0 \\ T_i &= 1\{\varphi(X_i) + \varepsilon_i \geq 0\}, \text{ where } \partial\varphi(x)/\partial x \geq 0 \\ X_i &\sim Unif[0,2], \ (\varepsilon_i, u_i) \sim N(0, I_2), \ X_i \perp (\varepsilon_i, u_i) \\ Y_i &= y_i(T_i), \ \& \ t = 1 \end{aligned}$$

- consider the median, $\tau = 0.5$, & $x_0 = 1.5$

- conditional moment inequalities:

$$\theta \geq \underline{\theta}(x) = \mu(x) + \sigma(x)\,\Phi^{-1}\left(1 - [2\Phi(\varphi(x))]^{-1}\right), \ \forall x \leq 1.5$$
$$\theta \leq \bar{\theta}(x) = \mu(x) + \sigma(x)\,\Phi^{-1}\left([2\Phi(\varphi(x))]^{-1}\right), \ \forall x \geq 1.5$$

- identified set for quantile selection model:

$$\left[\sup_{x \leq x_0} \underline{\theta}(x), \quad \inf_{x \geq x_0} \bar{\theta}(x)\right]$$

- shape of lower & upper bound functions depends on the shape of φ, μ, and σ functions

- consider 2 specifications: flat bd functions & kinky bd function

- **0.1** sec for 2 tests using 5000 crit val reps
  - CvM/Max/GMS/Asy & CvM/Max/PA/Asy

Table I. Quantile Selection Model: Basecase Comparisons

(a) Cov Probs

DGP	Statistic: / Crit Val	CvM/ Sum	CvM/ QLR	CvM/ Max	KS/ Sum	KS/ QLR	KS/ Max
Flat Bd	PA/Asy	.979	.979	.976	.972	.972	.970
	GMS/Asy	.953	.953	.951	.963	.963	.960
Kinky Bd	PA/Asy	.999	.999	.999	.994	.994	.994
	GMS/Asy	.983	.983	.983	.985	.985	.984

(b) False Cov Probs

DGP	Crit Val	CvM/ Sum	CvM/ QLR	CvM/ Max	KS/ Sum	KS/ QLR	KS/ Max
Flat Bd	PA/Asy	.51	.50	.48	.68	.67	.66
	GMS/Asy	.37	.37	.37	.60	.60	.59
Kinky Bd	PA/Asy	.65	.65	.62	.68	.68	.67
	GMS/Asy	.35	.35	.34	.53	.53	.52

Table II. Quantile Selection Model w/ Flat Bound: Variations on Basecase

		(a) Cov Prob's	(b) FCP's (CP cor)
	Statistic:	CvM/Max	CvM/Max
Case	Crit Val:	GMS/Asy	GMS/Asy
Basecase ($n = 250, r_1 = 7$)		.951	.37
$n = 100$		.957	.40
$n = 500$		.954	.36
$n = 1000$		.948	.34
$r_1 = 5$		.949	.36
$r_1 = 9$		.951	.37
$r_1 = 11$		.951	.37
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.948	.38
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.967	.38
$\varepsilon = 1/100$		.949	.37
$\alpha = .5$		.518	.03
$\alpha = .5$ & $n = 500$		.513	.03

Interval Outcome Regression Model

- Manski & Tamer (2002)
- $Y_i^* = \theta_1 + X_i\theta_2 + U_i$, where $E(U_i|X_i) = 0$ a.s.
- observe Y_{Li} & Y_{Ui}, where $Y_{Li} \leq Y_i^* \leq Y_{Ui}$
- inequalities:

$$\begin{aligned} E(\theta_1 + X_i\theta_2 - Y_{Li}|X_i) &\geq 0 \text{ a.s.} \\ E(Y_{Ui} - \theta_1 - X_i\theta_2|X_i) &\geq 0 \text{ a.s.} \end{aligned}$$

- basecase: $n = 250$, $r_1 = 7$, $\varepsilon = 5/100$
- $U_i \sim N(0,1)$, $X_i \sim U[0,1]$

- **0.1** sec for 2 tests using 5000 crit val reps
  - CvM/Max/GMS/Asy & CvM/Max/PA/Asy

Table IV. Interval Outcome Regression Model: Basecase

	(a) Coverage Probs					
	Stat: CvM	CvM	CvM	KS	KS	KS
Crit. Value:	Sum	QLR	Max	Sum	QLR	Max
PA/Asy	.990	.993	.990	.989	.990	.989
GMS/Asy	.950	.950	.950	.963	.963	.963
	(b) False Coverage Probs					
PA/Asy	.62	.66	.61	.78	.80	.78
GMS/Asy	.37	.37	.37	.61	.61	.61

Table VI. Interval Outcome Regression Model: Variations on the Basecase

Case	Statistic: Crit Val:	(a) Coverage Probabilities CvM/Max GMS/Asy	KS/Max GMS/Asy
Basecase ($n = 250, r_1 = 7,$	$\varepsilon = 5/100$)	.950	.963
$n = 100$		.949	.970
$n = 500$		.950	.956
$n = 1000$		.954	.955
$r_1 = 5$ (30 cubes)		.949	.961
$r_1 = 9$ (90 cubes)		.951	.965
$r_1 = 11$ (132 cubes)		.950	.968
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.944	.961
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.958	.973
$\varepsilon = 1/100$		.946	.966
$(\theta_1, \theta_2) = (1.0, 0.5)$		.999	.996
$(\theta_1, \theta_2) = (1.5, 0.0)$		1.000	.996
$\alpha = .5$		.472	.481
$\alpha = .5$ & $n = 500$		.478	.500

Case	Statistic: Crit Val:	(b) False Cov Probs (CPcor) CvM/Max GMS/Asy	KS/Max GMS/Asy
Basecase ($n = 250, r_1 = 7,$	$\varepsilon = 5/100$)	.37	.61
$n = 100$		.39	.66
$n = 500$		.37	.60
$n = 1000$		.37	.60
$r_1 = 5$ (30 cubes)		.37	.59
$r_1 = 9$ (90 cubes)		.37	.63
$r_1 = 11$ (132 cubes)		.38	.64
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$		.40	.62
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$		.39	.65
$\varepsilon = 1/100$		.39	.69
$(\theta_1, \theta_2) = (1.0, 0.5)$		.91	.92
$(\theta_1, \theta_2) = (1.5, 0.0)$		.99	.97
$\alpha = .5$		.03	.08
$\alpha = .5$ & $n = 500$		.03	.07

Entry Game Model w/ Multiple Equilibria

- complete information simultaneous game (entry model)
- two players & n i.i.d. plays of the game
- consider Nash equilibria in pure strategies
- due to possibility of multiple equilibria, model is incomplete
- 2 cond'l moment ineq's & 2 conditional moment equal's arise
- Andrews, Berry, & Jia (2004), Beresteanu, Molchanov, & Molinari (2009), Galichon & Henry (2009b), Ciliberto & Tamer (2009)

- player b's utility/profits are

$$X'_{i,b}\tau_b + U_{i,b} \text{ if other player does not enter}$$
$$X'_{i,b}\tau_b - \theta_b + U_{i,b} \text{ if other player enters}$$

- $\theta_1 \in R$ indexes competitive effect on player 1 of entry by player 2

- θ_2 likewise

- $U_{i,b} \sim N(0,1)$ is known to both players

  - unobserved by econometrician

- econometrician observes $X_{i,1} \in R^4$, $X_{i,2} \in R^4$, $Y_{i,1}$, & $Y_{i,2}$
  - $Y_{i,b} = 1$ if player b enters & 0 otherwise for $b = 1, 2$

- unknown parameters: $\theta = (\theta_1, \theta_2)' \in [0, \infty)^2$, & $\tau = (\tau_1', \tau_2')' \in R^8$

- $X_{i,b} = (1, X_{i,b,2}, X_{i,b,3}, X_i^*)' \in R^4$
  - $X_{i,b,2} \sim$Bern(p), $X_{i,b,3} \sim N(0,1)$, $X_i^* \sim N(0,1)$

- equil'm selection rule (ESR) employed is maximum profit ESR
  - unknown to econometrician
  - i.e., if Y_i could be either $(1,0)$ or $(0,1)$ in equil'm, then $Y_i = (1,0)$ if player 1's monopoly profit exceeds that of player 2 & $Y_i = (0,1)$ otherwise

- provide some results for "player 1 first" ESR

- moment ineq. functions:

$$
\begin{aligned}
m_1(W_i,\theta,\tau) &= P(X_{i,1}'\tau_1 + U_{i,1} \geq 0,\ X_{i,2}'\tau_2 - \theta_2 + U_{i,2} \leq 0 | X_i) \\
&\quad -1(Y_i = (1,0)) \\
m_2(W_i,\theta,\tau) &= P(X_{i,1}'\tau_1 - \theta_1 + U_{i,1} \leq 0,\ X_{i,2}'\tau_2 + U_{i,2} \geq 0 | X_i) \\
&\quad -1(Y_i = (0,1))
\end{aligned}
$$

- 2 mom equalities ...

- model is identified by $(0,0)$ & $(1,1)$ outcomes

- moment ineq's provide additional info

- we estimate τ_1 & τ_2 given θ

- use mom ineq's with $\widehat{\tau}_{1n}(\theta)$ & $\widehat{\tau}_{2n}(\theta)$ plugged in

- use hypercubes in R^2 for $X_{i,1}' \widehat{\tau}_{1n}(\theta)$ & $X_{i,2}' \widehat{\tau}_{2n}(\theta)$
- transform variables to $[0,1]^2$
  - transform so sample covariance $=0$, then apply std normal cdf
- side-edge lenghts $(2r)^{-1}$ for $r=1,...,r_1$
- basecase: $r_1=3$ yields 56 cubes
  - also, $r_1=2 \Rightarrow 20$ cubes & $r_1=3 \Rightarrow 120$ cubes
- $n=500$; also, 250 & 1000
- **0.24** sec for 2 tests using 5000 crit val reps
  - CvM/Max/GMS/Asy & CvM/Max/PA/Asy

(a) Coverage Probs						
Case Stat:	CvM	CvM	CvM	KS	KS	KS
	Sum	QLR	Max	Sum	QLR	Max
$(\theta_1, \theta_2) = (0, 0)$	.979	.972	.980	.977	.975	.985
$(\theta_1, \theta_2) = (1, 0)$	.961	.980	.965	.959	.983	.972
$(\theta_1, \theta_2) = (1, 1)$	.961	.985	.961	.955	.985	.962
$(\theta_1, \theta_2) = (2, 0)$	.935	.982	.935	.944	.984	.952
$(\theta_1, \theta_2) = (2, 1)$	.943	.974	.940	.953	.987	.947
$(\theta_1, \theta_2) = (3, 0)$	.921	.975	.915	.938	.935	.984
$(\theta_1, \theta_2) = (2, 2)$	.928	.942	.913	.943	.972	.922
(b) False Coverage Probs (cov prob corrected)						
$(\theta_1, \theta_2) = (0, 0)$	.76	.99	.59	.91	.99	.83
$(\theta_1, \theta_2) = (1, 0)$	.60	.99	.42	.83	.66	.99
$(\theta_1, \theta_2) = (1, 1)$	.62	.96	.41	.82	.99	.58
$(\theta_1, \theta_2) = (2, 0)$	.51	.83	.35	.66	.96	.47
$(\theta_1, \theta_2) = (2, 1)$	.57	.57	.38	.69	.82	.44
$(\theta_1, \theta_2) = (3, 0)$	.49	.41	.36	.61	.43	.64
$(\theta_1, \theta_2) = (2, 2)$	.59	.34	.39	.65	.42	.49

Case	(a) Coverage Probs CvM/Max GMS/Asy	(a) Coverage Probs KS/Max GMS/Asy	(b) False Cov Probs CvM/Max GMS/Asy	(b) False Cov Probs KS/Max GMS/Asy
Basecase ($n = 500$, $r_1 = 3$, $\varepsilon = 5/100$)	**.961**	**.962**	**.41**	**.58**
$n = 250$	.948	.963	.39	.56
$n = 1000$	.979	.968	.52	.65
$r_1 = 2$ (20 cubes)	.962	.956	.41	.55
$r_1 = 4$ (120 cubes)	.962	.964	.42	.59
$(\kappa_n, B_n) = 1/2(\kappa_{n,bc}, B_{n,bc})$	.954	.959	.39	.57
$(\kappa_n, B_n) = 2(\kappa_{n,bc}, B_{n,bc})$	.967	.962	.42	.58
$\varepsilon = 1/100$	.926	.873	.32	.66
Reg'r Variances = 2	.964	.968	.54	.71
Reg'r Variances = 1/2	.963	.966	.29	.43
Player 1 First Eq Sel Rule	.955	.957	.39	.57
$\alpha = .5$	.610	.620	.05	.13
$\alpha = .5$ & $n = 1000$	.695	.650	.06	.16

Many Moment Inequalities

- allow for infinite # mom ineq's by indexing $m_j(W_i, \theta)$ by $t \in \mathcal{T}$
  - cond'l or uncond'l
- model is: $\forall t \in \mathcal{T}$,

$$E_F(m_j(W_i, \theta, t) | X_i) \geq 0 \text{ a.s. } [F_X] \; \forall j \leq p_t$$

- can use same functions $g \in \mathcal{G}$ & measure Q as above

- specifiy weight function $Q_{\mathcal{T}}$ on $\mathcal{T}$

- test stat is

$$T_n(\theta) = \int \int S(n^{1/2}\overline{m}_n(\theta, t, g), \overline{\mathbf{\Sigma}}_n(\theta, t, g))dQ(g)dQ_{\mathcal{T}}(t)$$

  where

$$\overline{m}_n(\theta, t, g) = n^{-1} \sum_{i=1}^{n} \begin{pmatrix} m_1(W_i, \theta, t)\, g_1(X_i) \\ \vdots \\ m_p(W_i, \theta, t)\, g_p(X_i) \end{pmatrix}$$

- use emp'l process result for $\nu_n(t, g) = n^{1/2}(\overline{m}_n(\theta, t, g) - E_F \overline{m}_n(\theta, t, g))$

- show analogous results as above hold for GMS tests & CS's
  - unif asy validity, etc.

- note: results for infinite # mom ineq's cover tests w/ no par θ

- example 1. test of stoch dominance
  - related work: Linton, Maasoumi, & Whang (2005), Linton, Song, & Whang (2008)
  - $Y_{1,i} \sim G_1(\cdot)$ & $Y_{2,i} \sim G_2(\cdot)$
  - $H_0 : G_1(t) - G_2(t) \geq 0$, $\forall t \in R$ $(= \mathcal{T})$
  - take $m(W_i, \theta, t) = 1(Y_{1,i} \leq t) - 1(Y_{2,i} \leq t)$
  - no θ appears; no functions $g(X_i)$ needed
  - get uniformity results for CvM & KS tests

- example 2. test of stoch dominance of cond'l dist'ns
  - related work: Lee & Whang (2008)
  - $Y_{1,i}|(X_i = x) \sim G_1(\cdot|x)$ & $Y_{2,i}|(X_i = x) \sim G_2(\cdot|x)$
  - $H_0 : G_1(t, x) - G_2(t, x) \geq 0,\ \forall t \in R\ (= \mathcal{T}),\ \forall x \in R^{d_x}$
  - take $m(W_i, \theta, t)$ as above, use functions $g \in \mathcal{G}$

Summary

- provide methods to construct CS's for parameters based on cond'l mom ineq's & equalites

- parameters need not be identified

- CS's based on CvM or KS-type statistics
  - allow for truncation of ∞ sums & simulation of integrals

- combine w/ generalized mom selection critical values

- establish uniform asy validity

- show CS shrinks to identified set: no info loss

- show tests have power against some $n^{-1/2}$-alternatives

- simulation results show CvM/Max w/ GMS crit val performs well in terms of cov probs & false cov probs