ECO 519. Notes on Newey-McFadden

Note Title 3/1/2006

- Most of the chapter focuses on extremum estimators of the form

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \; \hat{Q}_n(\theta)$$

- Some specific examples are analyzed throughout:

MLE: $\hat{Q}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ln f(z_i | \theta)$

NLS: $\hat{Q}_n(\theta) = -\frac{1}{n} \sum_{i=1}^{n} [y_i - h(x_i, \theta)]^2$

GMM: $\hat{Q}_n(\theta) = -[\frac{1}{n} \sum_{i=1}^{n} g(z_i, \theta)]' \hat{W} [\frac{1}{n} \sum_{i=1}^{n} g(z_i, \theta)]$

CMD: $\hat{Q}_n(\theta) = -[\hat{\pi} - h(\theta)]' \hat{W} [\hat{\pi} - h(\theta)]$

- Classical Minimum Distance (CMD) and GMM belong to a more general class of "minimum distance" estimators.

- CMD estimators are inspired by problems in which there is a relation between a vector of "reduced-form" parameters π, and a vector of "structural" parameters of interest θ, given by $\pi = h(\theta)$. The idea is to plug-in a first-stage vector of consistent estimators $\hat{\pi}$.

Consistency

- There are various consistency theorems, whose assumptions vary depending for example on whether or not Θ is compact, or the objective function is concave, or if it has some behavior equivalent to concavity (as in Theorem 1 in Huber (1967)).

- Consistency is established without relying on "first-order conditions", but instead sticking to the definition of $\hat{\theta}$ as maximizer of $\hat{Q}_n(\theta)$.

- Theorem 2.1 is perhaps the most restrictive case.

Identification

- Letting $Q_0(\theta)$ be the probability limit of $\hat{Q}_n(\theta)$, identification will hinge on the assumption that $Q_0(\cdot)$ has a unique global maximum (in Θ) at $\theta = \theta_0$.

Theorem 2.1

Assumptions:

(i) $Q_0(\theta)$ uniquely maximized at θ_0

(ii) Θ is compact.

(iii) $Q_0(\theta)$ is continuous

(iv) $\hat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta)$ uniformly in Θ.

- Then $\hat{\theta} \xrightarrow{p} \theta_0$.

Notes on Thm 2.1: $\hat{Q}_n(\cdot)$ is not required to be continuous; we can replace (iii) with the assumption that $Q_0(\cdot)$ is semi-continuous (see below); we can also relax the assumption that $\hat{\theta}$ actually maximizes $\hat{Q}_n(\cdot)$ for every sample size. We can assume instead that:

$$\hat{Q}_n(\hat{\theta}) \geq \sup_{\theta \in \Theta} \hat{Q}_n(\theta) + o_p(1)$$

- The condition $\sup_{\Theta} \| \hat{Q}_n(\theta) - Q_0(\theta) \| = o_p(1)$

can be relaxed to:

a) $\hat{Q}_n(\theta_0) \xrightarrow{p} Q_0(\theta_0)$

b) $\forall \varepsilon > 0,\ \forall \theta \in \Theta,\ \theta \neq \theta_0$:

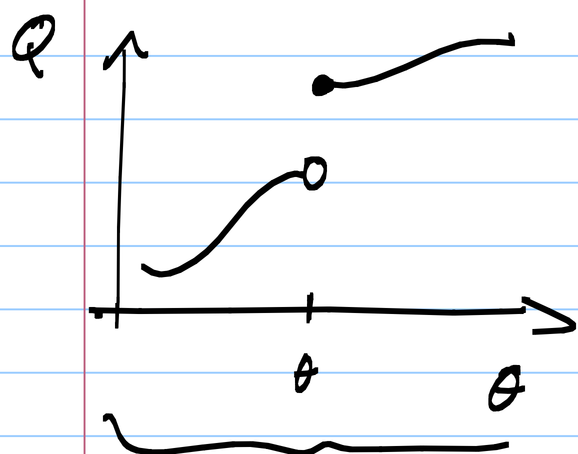
$$\Pr(\hat{Q}_n(\theta) < Q_0(\theta) + \varepsilon) \longrightarrow 1$$

Definition: Upper-Semicontinuity

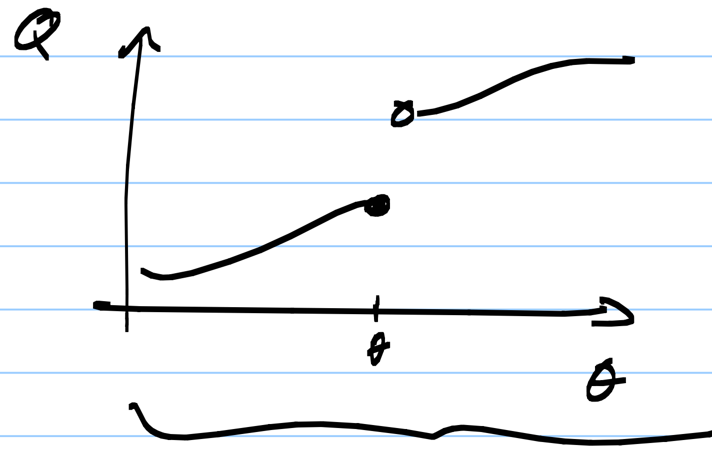
- The function $Q(\theta)$ is upper semi-continuous at θ if $\forall\ \varepsilon > 0$:

$$\sup_{\|\theta' - \theta\| < \varepsilon} Q(\theta') \longrightarrow Q(\theta) \text{ as } \varepsilon \rightarrow 0$$

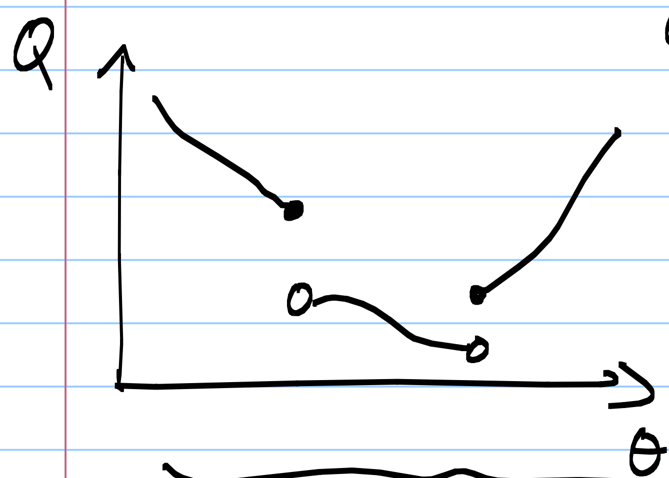
- Semi-continuous functions can have a finite number of discontinuities of a particular form. To gain some intuition, look at the following graphs in 1/2



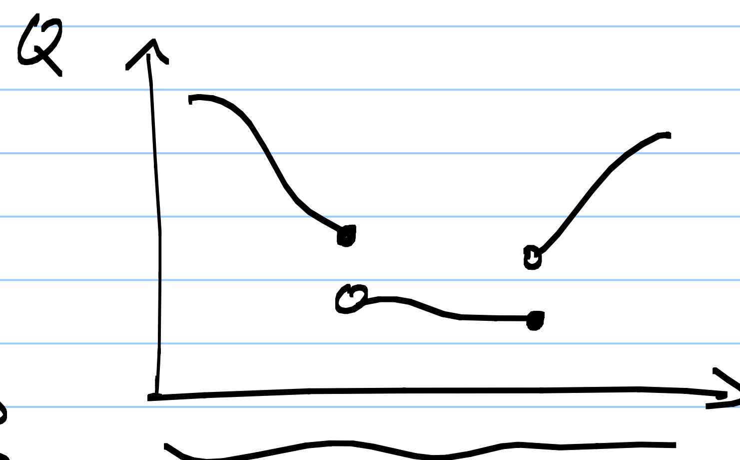
Upper semi-continuous



Not upper semi-cont.



Upper semi-continuous



not upper semi-cont

Proof of Theorem 2.1 -- The complete proof is in the chapter, the argument always relies on the fact that:

$$\hat{Q}_n(\hat{\theta}) \geq \hat{Q}_n(\theta_0) \quad \text{and} \quad Q_0(\theta_0) \geq Q_0(\hat{\theta})$$

$$\text{and} \quad \hat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta) \text{ uniformly.}$$

which enables us to show that eventually, $\hat{\theta}$ must belong to any open set containing θ_0 with probability approaching 1.

Strong Consistency -- Results for strong consistency (i.e., $\hat{\theta} \xrightarrow{as} \theta_0$) follow by replacing all the "in-probability" assumptions with their "almost-surely" equivalents.

<u>Note on Measurability</u>: – There are non-trivial technicalities that determine whether or not $\max_{\theta \in \Theta} \hat{Q}_n(\theta)$ is in fact a measurable function. Additional assumptions that guarantee measurability may be introduced (e.g. Dobb-separability in Huber (1967)), or we can re-define convergence in probability using <u>outer-measure</u> (we'll discuss this concept in the Empirical Process section, see for example Andrews's chapter about Stochastic Equicontinuity in the Handbook).

– Consistency without Compactness of Θ

– We can drop the assumption that Θ is compact if we introduce conditions that bound the behavior of $\hat{Q}_n(\theta)$ as $\theta \longrightarrow \pm\infty$. Concavity of $\hat{Q}_n(\cdot)$ is an option

Theorem 2.7

- Assumptions:
  - (i) $Q_0(\theta)$ uniquely maximized at θ_0
  - (ii) $\theta_0 \in$ interior Θ, where Θ is convex and $\hat{Q}_n(\cdot)$ is concave
  - (iii) $\hat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta) \quad \forall \ \theta \in \Theta$
- Then $\hat{\theta}_n$ exists w.p approaching one and $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Theorem 1 Huber (1967)

- Let $\hat{Q}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \rho(x_i, \theta)$, and suppose that:

$$\hat{Q}_n(\hat{\theta}) - \sup_{\theta} \hat{Q}_n(\theta) = o_p(1)$$

- Suppose also that Θ is a locally compact space (every neighborhood around any point $\theta \in \Theta$ contains a compact neighborhood, trivially satisfied by Euclidean spaces). This means that Θ does not need to be compact (or convex)...

- $\ell(x,\theta)$ is measurable and Doob-separable (to ensure $\sup_\theta (\ell(x,\theta))$ is measurable).
- $\ell(x,\cdot)$ is upper semicontinuous
- $E[|\ell(x,\theta)|] < \infty \quad \forall\ \theta \in \Theta$
- There is a θ_0 such that $E[\ell(x,\theta)]$ is uniquely maximized at θ_0
- There exists a continuous function $b(\theta) > 0$ such that:

replace the assumption of "concavity" of $\tilde{Q}_n(\theta)$

$$\begin{cases} (i)\ \sup\limits_{\theta \in \Theta} \dfrac{\ell(x,\theta)}{b(\theta)} \le u(x) \\ (ii)\ \lim\limits_{\|\theta\| \to \infty} \sup b(\theta) < E[\ell(x,\theta_0)] \\ (iii)\ E\left[\lim\limits_{\theta \to \infty} \sup \dfrac{\ell(x,\theta)}{b(\theta)}\right] \le 1 \end{cases}$$

- If these assumptions are satisfied, then $\hat{\theta} \xrightarrow{p} \theta_0$.

- The assumptions in both Theorem 2.7 in Newey-McFadden, and Theorem 1 in Huber enable us to show that there exists a compact set $A \subset \Theta$ such that

$$\Pr[\hat{\theta} \in A] \longrightarrow 1$$

- Having established this result, using the same reasoning as in the proof of Theorem 2.1, it's easy to show that with probability approaching one, $\hat{\theta}$ must be inside any arbitrary open set containing θ_0. This leads to $\hat{\theta} \xrightarrow{p} \theta_0$.

*<u>Note</u>: If θ_0 were not unique, our consistency results would state that with probability approaching one, $\hat{\theta}$ must be inside any open set that contains the set of θ's that maximize $Q_0(\theta)$.

***Note:** Theorem 2.7 does not explicitly impose uniform convergence in probability of the objective function as an additional condition because it is guaranteed by concavity of the limiting objective function: pointwise convergence of concave functions implies uniform convergence.

Conditions for Uniform Convergence

- So all consistency results require uniform convergence, either as an explicit assumption, or as the result of other assumptions (e.g., Theorem 2.7).

- Lemma 2.4 details how continuity and dominance yield uniform convergence

Lemma 2.4

- Assumptions:
  - $\{z_i\}_{i=1}^{n}$ are iid
  - Θ is compact
  - $a(z_i,\theta)$ is continuous $\forall\, \theta \in \Theta$ with probability one
  - There exists a random variable $d(z)$: $\|a(z,\theta)\| \leq d(z)$ $\forall\, \theta \in \Theta$ and $E[d(z)] < \infty$

- If these assumptions are satisfied, then:

  i) $E[a(z,\theta)]$ is continuous

  ii) $\sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} a(z_i,\theta) - E[a(z,\theta)] \right\| \xrightarrow{p} 0$

*Note: $a(\cdot,\theta)$ can be discontinuous at a set A of values of z, as long as:

$$\Pr(z \in A) = 0$$

Stochastic Equicontinuity and Uniform Convergence

- There is a more general result (property) that yields uniform convergence in probability.

(*Note: We will discuss the concept of stochastic Equicontinuity at length when we read Andrews' chapter in the Handbook.)

Stochastic Equicontinuity (def)

$\hat{Q}_n(\cdot)$ is stochastic equicontinuous if for all $\varepsilon, \eta > 0$, there exists a sequence of random variables $\hat{\Delta}_n$ and a $n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$, $\Pr(|\hat{\Delta}_n| > \varepsilon) < \eta$ and for each $\theta \in \Theta$, there exists an open set $\mathcal{N}$ containing θ such that

$$\sup_{\theta' \in \mathcal{N}} |\hat{Q}_n(\theta') - \hat{Q}_n(\theta)| \leq \hat{\Delta}_n \quad \forall n \geq n_0$$

- This property is a stochastic generalization of the concept of <u>equicontinuity</u>. Equicontinuous functions are continuous functions that share an equal bound on their variation over any given neighborhood: A sequence of functions $f_n : X \times \mathbb{N} \longrightarrow \mathbb{R}$ is equicontinuous if:

$\forall \varepsilon > 0$ and $\forall x \in X$, $\exists$ open set $\mathcal{N}$, and $n_0 \in \mathbb{N}$ such that:

$$\forall n \geq n_0, \quad x' \in \mathcal{N} \implies |f_n(x) - f_n(x')| < \varepsilon$$

- Stochastic equicontinuity generalizes this concept to a "random epsilon".

- Lemma 2.8 links stochastic equicontinuity and uniform convergence in probability

Lemma 2.8

- If Θ is compact and $Q_0(\theta)$ is continuous, then

$$\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$$

if and only if $\hat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta)$ for all $\theta \in \Theta$ [pointwise converg], <u>and</u> $\hat{Q}_n(\theta)$ is stochastically equicontinuous.

- "In-probability Lipschitz condition" is sufficient for stochastic equicontinuity:

Lemma 2.9

- If Θ is compact, $Q_0(\theta)$ is continuous, $\hat{Q}_n(\theta) \longrightarrow Q_0(\theta)$ for each $\theta \in \Theta$ and $\exists$ $\alpha > 0$ and $\hat{B}_n = O_p(1)$ such that $\forall$ $\theta', \theta \in \Theta$, $|\hat{Q}_n(\theta') - \hat{Q}_n(\theta)| \leq \hat{B}_n \|\theta' - \theta\|^{\alpha}$, then:

$$\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0.$$

* <u>Note</u>: Lemma 3 in Huber (1967) is all about showing that if some conditions are satisfied, a stochastic equicontinuity condition follows.

Asymptotic Normality with Smooth Objective functions

- Establishing asymptotic normality for extremum estimators of smooth, differentiable objective functions is straightforward via Taylor approximations:

<u>Theorem 3.1</u>

- Suppose $\hat{\theta} \xrightarrow{p} \theta_0$, $\theta_0 \in$ interior Θ, $\hat{Q}_n(\theta_0)$ is C^2 in a neighborhood $\mathcal{N}$ of θ_0; $\sqrt{n}\, \nabla_\theta \hat{Q}_n(\theta_0) \xrightarrow{d} N(0, \Sigma)$; there exists a function $H(\theta)$ that is continuous at θ_0 such that

$$\sup_{\theta \in \mathcal{N}} \| \nabla_{\theta\theta'} \hat{Q}_n(\theta) - H(\theta) \| \xrightarrow{p} 0$$

and $H(\theta_0)$ is nonsingular. Then:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, H^{-1} \Sigma H^{-1}).$$

- Since $\theta_0 \in$ interior Θ, and θ_0 maximizes $Q_0(\theta)$ in Θ, then it must solve $\nabla_\theta Q_0(\theta_0) = 0$. With probability approaching one, $\hat{\theta}$ also solves FOC:

$$0 = \nabla_\theta \hat{Q}_n(\hat{\theta}) = \nabla_\theta \hat{Q}_n(\theta_0) + \nabla_{\theta\theta'} \hat{Q}_n(\tilde{\theta})(\hat{\theta} - \theta_0)$$

with $\tilde{\theta}$ between $\hat{\theta}$ and θ_0. With probability approaching one:

$$\sqrt{n}(\hat{\theta} - \theta_0) = -(\nabla_{\theta\theta'} \hat{Q}_n(\tilde{\theta}))^{-1} \sqrt{n}\, \nabla_\theta \hat{Q}_n(\theta_0)$$

$$\xrightarrow{d} N(0, H^{-1} \Sigma H^{-1})$$

One-Step Estimators

only need $\sqrt{n}$-consistency

- If we have a $\sqrt{n}$-consistent, asymptotically normal estimator $\bar{\theta}$, we can use it to obtain a new estimator that has the same limiting distribution as an extremum estimator by linearizing the objective function around $\bar{\theta}$: Suppose $\bar{\theta}$ satisfies $\sqrt{n}(\bar{\theta}-\theta_0) = O_p(1)$ and let $\bar{H} = \nabla_{\theta\theta'}\hat{Q}_n(\bar{\theta})$. Let

$$\tilde{\theta} = \bar{\theta} - \bar{H}^{-1}\nabla_\theta \hat{Q}_n(\bar{\theta})$$

$$\nabla_\theta \hat{Q}_n(\bar{\theta}) = \nabla_\theta \hat{Q}_n(\theta_0) + \nabla_{\theta\theta'}\hat{Q}_n(\theta_0)(\bar{\theta}-\theta_0) + o_p(n^{-1})$$

$$\Rightarrow \sqrt{n}(\tilde{\theta}-\theta_0) = \sqrt{n}(\bar{\theta}-\theta_0) - \bar{H}^{-1}\left[\nabla_\theta \hat{Q}_n(\theta_0) + \nabla_{\theta\theta'}\hat{Q}_n(\theta_0)(\bar{\theta}-\theta_0) + o_p(n^{-1})\right]$$

$$= \underbrace{[I - \bar{H}^{-1}\nabla_{\theta\theta'}\hat{Q}_n(\theta_0)]}_{\xrightarrow{p} 0}\underbrace{\sqrt{n}(\bar{\theta}-\theta_0)}_{O_p(1)} - \bar{H}^{-1}\sqrt{n}\nabla_\theta \hat{Q}_n(\theta_0) + o_p(n^{-1/2})$$

$$\boxed{\xrightarrow{d} N(0, H^{-1}\Sigma H^{-1}) \longrightarrow \text{same as in Theorem 3.1}}$$

— Next time, we will conclude with some efficiency results and two-step estimators.