24/01/23

Limit Theory of OE - Consistency (Theory and Examples)

Remember that Z_n denotes the sample, and we have at our disposal a (possibly semi-) parametric model for θ_0, and a measurable criterion $C_n : \mathbb{R}^{k \times n} \times \Theta \to \mathbb{R}$ so that

$$(*) \qquad \hat{\theta}_n \in \underset{\theta \in \Theta}{\arg\min}\, C_n(\theta)$$

[Remark: A generalization of (*) can accommodate the case that the optimization is approximate, or $\underset{\theta\in\Theta}{\arg\min}\, C_n(\theta) = \emptyset$.

If u_n is a random variable with $\mathbb{P}(u_n \geq 0) = 1$, then an approximate minimizer can be defined by:

$$(**) \qquad C_n(\hat{\theta}_n) \leq \inf_{\theta \in \Theta} C_n(\theta) + u_n$$

(**) is suitable for analysis of estimators defined by numerical optimization] [See also endnote 2 in the previous set of notes]

Question: Can we obtain mild conditions under which $\hat{\theta}_n$ defined by (*) or (**) is weakly consistent?

In what follows C^* denotes a function $\Theta \to \mathbb{R}$ that does not depend on Z_n.

— In what follows every convergence occurs as $n \to +\infty$.

Definition. We say that C_n converges in C^* locally uniformly in Θ, and in probability iff $\forall \theta \in \Theta$, $\forall \theta_n \to \theta$, $\forall \varepsilon > 0$,

$$\lim_{n\to+\infty} \mathbb{P}\left(|C_n(\theta_n) - C^*(\theta)| > \varepsilon \right) = 0. \quad (\text{lu})$$

in the previous version there was θ_n instead of θ - typo!

Definition. We say that c_n converges to c^* pointwisely in Θ, and in probability iff $\forall \theta \in \Theta$, $\forall \varepsilon > 0$

$$\lim_{n \to +\infty} \mathbb{P}(|c_n(\theta) - c^*(\theta)| > \varepsilon) = 0 \quad (p).$$

Remark. Notice that $(lu) \Rightarrow (p)$ (simply consider only the constant sequences), but $(p) \not\Rightarrow (lu)$.

The following result gives a sufficient condition (lc), such that $(lc) + (p) \Rightarrow (lu)$.

Lemma. Suppose that (i) c_n converges to c^* pointwisely in Θ and in probability, and (ii) $\forall \theta, \theta^* \in \Theta$, $|c_n(\theta) - c_n(\theta^*)| \leq k_n \|\theta - \theta^*\|$, $\exists M > 0 : \lim_{n \to +\infty} \mathbb{P}(k_n > M) = 0$. Then c_n converges to c^* locally uniformly in Θ and in probability.
(*)

Proof. Let $\theta \in \Theta$, $\Theta \ni \theta_n \to \theta$, $\varepsilon > 0$. Notice that

$$(a) \quad |c_n(\theta_n) - c^*(\theta)| \leq |c_n(\theta_n) - c_n(\theta)| + |c_n(\theta) - c^*(\theta)|$$

due to the triangle inequality ($|a+b| \leq |a| + |b|$).

Then, due to (ii)

$$|c_n(\theta_n) - c_n(\theta)| \leq k_n \|\theta_n - \theta\| \quad (b)$$

Hence (a),(b) $\Rightarrow$

$$|c_n(\theta_n) - c^*(\theta)| \leq k_n \|\theta_n - \theta\| + |c_n(\theta) - c^*(\theta)| \quad (c)$$

Notice that (*) is weaker than the condition, $\exists M > 0$: $\mathbb{E}(k_n) \leq M$ of the previous version. E.g. (*) holds when k_n converges in probability

Due to (6) we have that:

$$\mathbb{P}\left(|C_n(\theta_n) - C^*(\theta)| > \varepsilon \right) \overset{(1)}{\leq}$$

$$\mathbb{P}\left(K_n \|\theta_n - \theta\| + |C_n(\theta) - C^*(\theta)| > \varepsilon \right)$$

[$\overset{(1)}{\leq}$ is obtained from the elementary fact that $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ – why?]

$$\overset{(2)}{\leq} \mathbb{P}\left(K_n \|\theta_n - \theta\| > \varepsilon/2 \right) + \mathbb{P}\left(|C_n(\theta) - C^*(\theta)| > \varepsilon/2 \right)$$

[$\overset{(2)}{\leq}$ follows from that $\mathbb{P}(\alpha + \beta > \varepsilon) = \underbrace{\mathbb{P}(\alpha + \beta > \varepsilon, \alpha > \varepsilon/2)}_{\leq \mathbb{P}(\alpha > \varepsilon/2)}$

$+ \mathbb{P}(\alpha + \beta > \varepsilon, \alpha \leq \varepsilon/2)$

$= \mathbb{P}(\beta > \varepsilon - \alpha, -\alpha \geq -\frac{\varepsilon}{2}) \leq \mathbb{P}(\beta > \varepsilon/2, -\alpha \geq -\varepsilon/2)$

$\leq \mathbb{P}(\beta > \varepsilon/2)$]

$$\overset{(3)}{\leq} \mathbb{P}\left(M \|\theta_n - \theta_0\| > \varepsilon/2 \right) + \mathbb{P}\left(|C_n(\theta) - C^*(\theta)| > \varepsilon/2 \right)$$

[$\overset{(3)}{\leq}$ is obtained from (*) and again the elementary $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$]

$$= \mathbb{P}\left(\|\theta_n - \theta\| > \frac{\varepsilon}{2M} \right) + \mathbb{P}\left(|C_n(\theta) - C^*(\theta)| > \varepsilon/2 \right) := M_{1_n} + M_{2_n}$$

Then $M_{1_n} \to 0$ since $\theta_n \to \theta$, and $M_{2_n} \to 0$ due to (p). (explain!)

hence $\lim\limits_{n\to+\infty} \mathbb{P}\left(|G_n(\theta_n) - C^*(\theta)| > \varepsilon \right) = 0$ and the result follows since $\theta, \theta_n, \varepsilon$ are arbitrary. $\square$

Remark. Hence (lu) follows from (p) complemented by some sort of strong (joint w.r.t. n) continuity property of G_n. $\square$

Remark. It can be proven that (lu) respects optimization, i.e. if (lu) holds then

$$(\mathrm{I}) \qquad \forall \varepsilon>0 \quad \lim_{n\to+\infty} \mathbb{P}\left(\left| \inf_{\theta\in\Theta} C_n(\theta) - \inf_{\theta\in\Theta} C^*(\theta) \right| > \varepsilon \right) = 0$$

as long as $\inf\limits_{\Theta} C^*(\theta)$ is well defined. $\boxed{\equiv}$

Remark. It can be also proven that if Θ is compact, i.e. closed and bounded, and (lu) holds, then for any $\theta_n^* \in \Theta$ that may depend on Z_n,

$$(\mathrm{II}) \qquad \forall \varepsilon>0 \quad \lim_{n\to+\infty} \mathbb{P}\left(|C_n(\theta_n^*) - C^*(\theta_n^*)| > \varepsilon \right) = 0. \boxed{\equiv}$$

Then (lu) to a limit that guarantees asymptotic identification when the parameter space is compact implies weak consistency:

Theorem. Suppose that there exists some $C^*: \Theta \to \mathbb{R}$ such that a. G_n converges to C^* locally uniformly in θ and in probability, and, b. $\forall \delta>0$, $\inf\limits_{\|\theta-\theta_0\|>\delta} C^*(\theta) > C^*(\theta_0)$, and, c. Θ is compact. Then θ_n is weakly consistent.

Proof. Let $\varepsilon > 0$ and consider the event $\|\theta_n - \theta_0\| > \delta$. This and D. imply that $\exists\, \varepsilon > 0 : \; c^*(\theta_n) - c^*(\theta_0) > \varepsilon$ hence

[See the note below]

$$\mathbb{P}(\|\theta_n - \theta_0\| > \delta) \le \mathbb{P}(c^*(\theta_n) - c^*(\theta_0) > \varepsilon) =$$

$$\mathbb{P}(|c^*(\theta_n) - c^*(\theta_0)| > \varepsilon) = \mathbb{P}(|c^*(\theta_n) \pm c_n(\theta_n) - c^*(\theta_0)| > \varepsilon)$$

$$\le \mathbb{P}(|c^*(\theta_n) - c_n(\theta_n)| > \varepsilon/2) + \mathbb{P}(|c^*(\theta_0) - c_n(\theta_n)| > \varepsilon/2)$$

why?

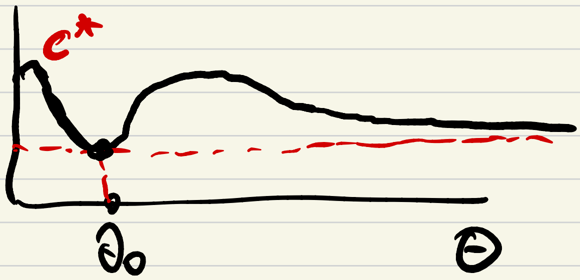
$$= \mathbb{P}(|c^*(\theta_n) - c_n(\theta_n)| > \varepsilon/2) + \mathbb{P}\left(\left|\inf_{\theta\in\Theta} c^*(\theta) - \inf_{\theta\in\Theta} c_n(\theta)\right| > \frac{\varepsilon}{2}\right)$$

$$:= Q_{1n} + Q_{2n}.$$

We have that $Q_{1n} \to 0$ due to (II) and $Q_{2n} \to 0$ due to (I). The result follows since ε is arbitrary. ∎

Remark.

1. The peculiar identification condition not only implies $\theta_0 = \underset{\theta\in\Theta}{\operatorname{argmin}}\, c^*(\theta)$. But also excludes pathological cases where θ_0 despite being the unique minimiser is not "distinguishable" from the other Θ, e.g.



It holds whenever Θ is compact, c^* is continuous and

(*) θ_0 is the unique minimizer, or when Θ is convex and and closed

G_n is a strictly convex function. Try to show these!

2. The compactness of Θ is not required when there is more structure. Eg. when Θ closed and convex and G_n is convex then the result holds without compactness at the cost of a slightly more involved proof. [actually in such cases pointwise convergence works]

3. (ii) can be further weakened to other forms of functional convergence tailored for the approximation of optimization problems - e.g. epi-convergence (completely out of the scope of the course)

4. As mentioned in the previous notes this limit theory does not depend on an explicit expression for $\hat{\theta}_n$ as a function of Z_n - this is most usually unavailable, but on properties of the optimization procedure.

Question. How are the above specialized in our examples?

Example: Consider the linear model $Y_n = X_n\theta + \varepsilon_n$,

$\mathbb{E}\,\varepsilon_n / \sigma(X_n) = 0_{n\times 1}$, $\mathrm{Var}(X_n / \sigma(X_n)) = I_{n\times n}$,

$\mathrm{rank}\, X_n = p$ (at least with probability 1).

- When $\Theta = \mathbb{R}^p$, we have that the OLSE has a known analytical form,

$$\hat{\theta}_n = (X_n'X_n)^{-1} X_n' Y_n \qquad [\text{using } Y_n = X_n\theta_0 + \varepsilon_n]$$
$$= (X_n'X_n)^{-1} X_n'(X_n\theta_0 + \varepsilon_n)$$

$$= \underbrace{(X_n'X_n)^{-1}(X_n'X_n)}_{= I_{p\times p}}\theta_0 + (X_n'X_n)^{-1}X_n'\varepsilon_n$$

$$= \theta_0 + (X_n'X_n)^{-1}X_n'\varepsilon_n$$ [convenient for analysis, not for evaluation of the OLSE]

(x) $$= \theta_0 + \left(\frac{X_n'X_n}{n}\right)^{-1}\frac{X_n'\varepsilon_n}{n}$$ [Convenient for establishing Asymptotic properties]

Consider the $\underbrace{\text{high level}}$ conditions

→ Not very detailed on the probabilistic properties on the random elements involved

i. $$\frac{X_n'\varepsilon_n}{n} \xrightarrow{P} 0_{p\times 1}$$

[Notice that $\frac{X_n'\varepsilon_n}{n} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_{i1}\varepsilon_i \\ \frac{1}{n}\sum_{i=1}^{n} X_{i2}\varepsilon_i \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_{ip}\varepsilon_i \end{pmatrix}$ and

due to that $\mathbb{E}(\varepsilon_n / \mathcal{G}(X_n)) = 0$, $\mathbb{E}(X_{ij}\varepsilon_i) = 0$, $\forall i = 1, \dots, n$ $\forall j = 1, \dots, p$

Thus i, would follow by any valid Law of Large Numbers

– Can you provide with an example?]

ii. There exists a deterministic $q \times p$ matrix, $M_{x'x}$, such that

$$\frac{X_n' X_n}{n} \xrightarrow{P} M_{x'x}$$

[Similarly $\frac{X_n' X_n}{n} = \left(\frac{1}{n}\sum_{i=1}^{n} x_{ij} x_{ij'}\right)_{\substack{j=1,\dots,p \\ j'=1,\dots,p}}$

hence if $\mathbb{E}(x_{ij} x_{ij'})$ exists for all $j, j' = 1, \dots, p$ and it is independent of the index i, then ii would follow as long as a law of large numbers were valid. This is not necessary though. Can you provide with an example?]

iii. $M_{x'x}$ is invertible ($\Leftrightarrow$ rank $M_{x'x} = p$)

[This is stronger than rank $X_n = p$ [$\Rightarrow$ rank $\frac{X_n' X_n}{n} = p$] (why?). It is some sort of a condition of asymptotic identification (see below). Since $\frac{1}{n} X_n' X_n$ is actually a Gram matrix, it would follow - given ii - as long as the columns of X_n remain asymptotically linearly independent]

[Actually in what follows $M_{x'x}$ need not be deterministic as long as $\mathbb{P}(\text{rank } M_{x'x} = p) = 1$.]

Notice that ii, iii, and the continuous mapping Theorem (CMT) imply that

$$\left(\frac{X_n' X_n}{n}\right)^{-1} \xrightarrow{P} M_{x'x}^{-1} \quad (**)$$

And then (**), i, CMT $\Rightarrow$

$$\left(\frac{X_n' X_n}{n}\right)^{-1} \frac{X_n' \varepsilon_n}{n} \xrightarrow{P} M_{x'x}^{-1} 0_{p\times 1}$$
$$= 0_{p\times 1}$$
(***)

And then (***), CMT $\Rightarrow$

$$\hat{\theta}_n \xrightarrow{P} \theta_0 + 0_{p\times 1} = \theta_0$$

Hence i, ii, iii, are sufficient for weak consistency, when $\Theta = \mathbb{R}^p$.

Does this also hold when $\Theta \subsetneq \mathbb{R}^p$?
(Remember that we globally assume correct specification hence always $\theta_0 \in \Theta$).
When $\Theta \subsetneq \mathbb{R}^p$ then we may not have an analytical form of the estimator to work with; we could try to rely to work with results like the previous theorem:

1. We have to identify c^*. Remember that

$$L_n(\theta) = \frac{1}{n}(Y_n - X_n\theta)'(Y_n - X_n\theta)$$ 1

$$= \dots = (\theta-\theta_0)' \frac{X_n' X_n}{n}(\theta-\theta_0) - 2\frac{X_n'\varepsilon_n}{n}(\theta-\theta_0) + 1$$

Given i, ii we may be tempted to assume that
$c^*(\theta) = (\theta-\theta_0) M_{x'x}(\theta-\theta_0) + 1$.

2. Due to the CLLT, we certainly have pointwise convergence in probability for arbitrary Θ.

Notice that $\forall \theta \in \Theta$, $|C_n(\theta) - C^*(\theta)|$

$$= \left| (\theta-\theta_0)'\left(\frac{X_n'X_n}{n} - U_{xx'}\right)(\theta_n-\theta_0) - 2\frac{X_n'\varepsilon_n}{n}(\theta-\theta_0) \right|$$

$$\le \left|(\theta-\theta_0)'\left(\frac{X_n'X_n}{n} - U_{x'x}\right)(\theta_n-\theta_0)\right| + 2\left|\frac{X_n'\varepsilon_n}{n}(\theta-\theta_0)\right|$$

$= A_n + B_n$. $\;$ ii+CLLT $\Rightarrow A_n \xrightarrow{P} 0$, ii+CLLT $\Rightarrow B_n \xrightarrow{P} 0$.

Hence $\forall \theta \in \Theta$, $C_n(\theta) \xrightarrow{P} C^*(\theta)$.

3. When Θ is compact, we can show that $\forall \theta, \theta^* \in \Theta$

$$|C_n(\theta) - C_n(\theta^*)| = \left|(\theta-\theta_0)'\frac{X_n'X_n}{n}(\theta-\theta_0) - (\theta^*-\theta_0)'\frac{X_n'X_n}{n}(\theta^*-\theta_0)\right.$$

$$\left. -2\frac{X_n'\varepsilon_n}{n}(\theta-\theta^*)\right| \le \left|(\theta-\theta^*)'\frac{X_n'X_n}{n}(\theta-\theta_0)\right|$$

$$+ \left|(\theta-\theta^*)'\frac{X_n'X_n}{n}(\theta^*-\theta_0)\right| + 2\left|\frac{X_n'\varepsilon_n}{n}(\theta-\theta^*)\right|$$

$$= A_1 + A_2 + A_3$$

Due to the compactness of Θ, we can show that there exists some $C>0$: (employing some inequalities involving norms) $A_1 \le C \left\| \frac{X_n'X_n}{n} \right\| \|\theta-\theta^*\|$

this is a Matrix Norm called Frobenius Norm; submultiplicative and $A_2 \leq C \left\| \frac{X_n' X_n}{n} \right\| \, \| \theta^* - \theta \|$

and $A_3 \leq 2 \left\| \frac{X_n' \varepsilon_n}{n} \right\| \, \| \theta^* - \theta \|$.

Hence we obtain the existence of a GO such that

$$|C_n(\theta) - C_n(\theta^*)| \leq 2\left(C \left\| \frac{X_n' X_n}{n} \right\| + \left\| \frac{X_n' \varepsilon_n}{n} \right\| \right) \|\theta - \theta^*\|, \ \forall \theta, \theta^* \in \Theta$$

Furthermore i, ii, CLT $\Rightarrow 2C \left\| \frac{X_n' X_n}{n} \right\| + \left\| \frac{X_n' \varepsilon_n}{n} \right\|)$

$$\xrightarrow{P} 2C \| M_{x'x} \|.$$

Thereby, when Θ is compact, C_n satisfies $(*)$

with $K_n = 2\left(C \left\| \frac{X_n' X_n}{n} \right\| + \left\| \frac{X_n' \varepsilon_n}{n} \right\| \right)$.

Hence, when Θ is compact, α, c and the above established Lemma imply that:

C_n converges to $(\theta - \theta_0)' M_{x'x} (\theta - \theta_0) + L$ locally uniformly and in probability.

Hence we have established conditions α and c of our Theorem. What about condition b;

4. Notice that $(\theta - \theta_0)' M_{x'x} (\theta - \theta_0)$ is continuous in θ, as a quadratic function. Due to iii, when $\Theta = \mathbb{R}^p$, $\underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \, (\theta - \theta_0)' M_{x'x} (\theta - \theta_0) = \theta_0$.

When $\Theta \subsetneq \mathbb{R}^p$ but $\theta_0 \in \Theta$ then the previous implies that $\underset{\theta \in \Theta}{\operatorname{argmin}} (\theta - \theta_0)' M_{x'x} (\theta - \theta_0) = \theta_0$.

Hence when Θ is compact and iii holds, due to (*), b holds. Thereby we have established:

Lemma [OLS - Compactness] Under i, ii, iii and if Θ is compact then the OLSE is weakly consistent.

What about other cases?

Again by a previous remark (see Remark 2 above), due to 1, and 4 we can prove that

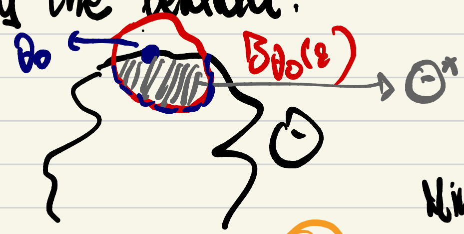
Lemma [OLS - Convexity] Under i, ii, iii and if Θ is closed and convex then the OLSE is weakly consistent.

What about the remaining cases where Θ is such that $\theta_n \in \underset{\Theta}{\operatorname{argmin}}\, G_n(\theta)$? [and thus it is not such, so that the generalized definition is needed]

When Θ is closed the OLS-compactness lemma along with a geometric argument can establish consistency:

Let $B_{\theta_0}(\varepsilon)$ for some small enough $\varepsilon>0$, be the closed ball centered at θ_0 with radius ε, and consider $\Theta^* := \Theta \cap B_{\theta_0}(\varepsilon)$. Θ^* can be shown compact. Consider the oracle OLS $\hat{\theta}_n^* \in \arg\min_{\Theta^*} G_n(\theta)$. $\hat{\theta}_n^*$ is weakly consistent by the lemma.

infeasible but useful for the derivation of properties



If $\hat{\theta}_n$ lies in Θ^* then $\hat{\theta}_n^* = \hat{\theta}_n$.
If $\hat{\theta}_n \notin \Theta^*$ then since G_n is strictly convex $\hat{\theta}_n^*$ lies on the boundary of Θ^* at a minimal distance from $\hat{\theta}_n$. But $\hat{\theta}_n^* \xrightarrow{P} \theta_0$ hence with probability converging to 1 $\hat{\theta}_n^* = \hat{\theta}_n$ and weak consistency holds.

(2)

Whenever Θ is such that $\hat{\theta}_n \in \arg\min_{\Theta} G_n$, i, ii, iii suffice for weak consistency.

Note: We return for a while to the asymptotic identification condition that appears in the theorem of consistency. Remember that it says, that for any $\delta > 0$ the restricted optimization of c^* outside the closed ball centered at θ_0 with radius δ, must not result into the approximation of $\min_{\theta\in\Theta} c^*(\theta)$ (global optimization)

(in relation to the proof of the Theorem of consistency)

This directly implies that:

a. θ_0 is the unique minimizer of c^*, since if $\theta_1 \neq \theta_0$ was also a minimizer, then

$$\inf_{\theta\in\Theta} c^*(\theta) = \underbrace{\inf_{\theta: \|\theta-\theta_0\|>\delta} c^*(\theta)}_{}$$

for any $\delta < \|\theta_1 - \theta_0\|$

↳ θ_1 would lie in there

b. Since for any $\delta>0$ $\quad C^*(\theta_0)=\inf_{\theta\in\Theta} C^*(\theta) < \inf_{\theta:\|\theta-\theta_0\|>\delta} C^*(\theta)$

then there exists $\varepsilon>0$:

$$C^*(\theta_0) < \inf_{\theta:\|\theta-\theta_0\|>\delta} C^*(\theta) - \varepsilon \qquad (**)$$

[simply choose ε as anything less than $\inf_{\theta:\|\theta-\theta_0\|>\delta} C^*(\theta) - C^*(\theta_0)$]

c. Then $(**)$ implies that for any $\delta>0$, $\exists\, \varepsilon>0$ such that for any θ: $\|\theta-\theta_0\|>\delta$

$$C^*(\theta_0) < C^*(\theta) - \varepsilon \iff$$

$$C^*(\theta) - C^*(\theta_0) > \varepsilon \xRightarrow{C^*(\theta)-C^*(\theta_0)>0}$$

$$|C^*(\theta) - C^*(\theta_0)| > \varepsilon \qquad (***)$$

d. Thus in the proof of our theorem, and since we were occuppied with the event $\|\theta_n-\theta_0\|>\delta$, $(**)$ implies the existence of $\varepsilon>0$ independent of θ_n such that

$$\|\theta_n-\theta_0\|>\delta \Rightarrow |C^*(\theta_n)-C^*(\theta_0)|>\varepsilon$$

and thus due to the monotonicity of $\mathbb{P}$

$$\mathbb{P}(\|\theta_n-\theta_0\|>\delta) \le \mathbb{P}(|C^*(\theta_n)-C^*(\theta_0)|>\varepsilon)$$ as claimed in the proof.

$\boxtimes$

What about our further examples?

Example [linear Model – IV case]

Remember that now $Y_n = X\theta_0 + \varepsilon_n$ but for W_n the Matrix of instruments $\mathbb{E}(W_n'\varepsilon_n) = 0_{q\times 1}$, and $\operatorname{rank} X_n = p$, $\operatorname{rank} W_n = q$, $n \geq \max(p,q)$, $q \geq p$.

Remember that for a strictly positive definite $q\times q$ Matrix V,

$$C_n(\theta; V) := \left(\tfrac{1}{n} W_n'(Y_n - X_n\theta)\right)' V \left(\tfrac{1}{n} W_n'(Y_n - X_n\theta)\right)$$

$$= \frac{1}{n^2}(Y_n - X_n\theta)' W_n V W_n'(Y_n - X_n\theta) \quad (a)$$

$$= \frac{1}{n^2}\left(X_n(\theta_0-\theta)+\varepsilon_n\right)' W_n V W_n'\left(X_n(\theta_0-\theta)+\varepsilon_n\right)$$

$$\left.\begin{aligned} = \ & (\theta_0-\theta)' \frac{X_n' W_n}{n} V \frac{W_n' X_n}{n} (\theta_0-\theta) \\ + \ & 2(\theta_0-\theta)' \frac{X_n' W_n}{n} V \frac{W_n' \varepsilon_n}{n} \\ + \ & \frac{\varepsilon_n' W_n}{n} V \frac{W_n' \varepsilon_n}{n} \end{aligned}\right\} \quad (b)$$

Analogously to the case of the OLSE, (a) is observable, and (b) latent, yet useful for properties derivation.

Given Θ, the IVE is defined by $\theta_n \in \operatorname*{argmin}_{\theta\in\Theta} C_n(\theta, V)$ and when $\Theta = \mathbb{R}^2$, given C_n is twice differentiable we have that the optimization problem is analytically solvable using foc:

foc: $\frac{\partial C_n(\theta,v)}{\partial \theta} = 0_{p\times 1}$ (a) $\Rightarrow$

[$\frac{\partial x'Ax}{\partial x} = Ax + (x'A)' = (A+A')x$

and when A is symmetric this reduces to $2A$ – in our case

$A = W_n V W_n'$, $A' = (W_n V W_n')' = (W_n')' V' W_n' \overset{V\, sym}{=} W_n V W_n'$

$= A$, $x = (Y_n - X_n\theta)$, and $\frac{\partial x}{\partial \theta} = \frac{\partial (Y_n - X_n\theta)}{\partial \theta} = -X_n'$

thereby

$\frac{\partial C_n(\theta,v)}{\partial \theta} = \frac{1}{n^2}\frac{\partial x}{\partial \theta}\frac{\partial x'Ax}{\partial x} = -\frac{2}{n^2} X_n' W_n V W_n' (Y_n - X_n\theta)$]

$-\frac{2}{n^2} X_n' W_n V W_n' (Y_n - X_n\theta) = 0_{p\times 1} \Rightarrow$

$\frac{2}{n^2}(X_n' W_n) V (W_n' X_n)\theta = \frac{2}{n^2}(X_n' W_n) V W_n' Y_n \Leftrightarrow$

$\theta_n = \left((X_n' W_n) V (W_n' X_n)\right)^{-1}(X_n' W_n) V W_n' Y_n$ 3

[Since the aforementioned rank and dimension conditions imply that $(X_n' W_n) V (W_n' X_n)$ is invertible – why?]

This must be the unique minimizer since $W_n V W_n'$ has rank q, and thereby it is strictly positive definite [why? use the Cholesky decomposition of V]

Then $$\hat\theta_n = \left((X_n'W_n)V(W_n'X_n)\right)^{-1}(X_n'W_n)VW_n'Y_n$$

$$= \left((X_n'W_n)V(W_n'X_n)\right)^{-1}(X_n'W_n)VW_n'(X_n\theta_0+\varepsilon_n)$$

$$= \underbrace{\left((X_n'W_n)V(W_n'X_n)\right)^{-1}(X_n'W_n)V(W_n'X_n)}_{I_{p\times p}}\theta_0 + \left((X_n'W_n)V(W_n'X_n)\right)^{-1}(X_n'W_n)VW_n'\varepsilon_n$$

$$= I_{p\times p}\theta_0 + \left((X_n'W_n)V(W_n'X_n)\right)^{-1}(X_n'W_n)VW_n'\varepsilon_n$$

$$= \theta_0 + \left((X_n'W_n)V(W_n'X_n)\right)^{-1}(X_n'W_n)VW_n'\varepsilon_n \quad (c)$$

$$= \theta_0 + \left(\left(\frac{X_n'W_n}{n}\right)V\left(\frac{W_n'X_n}{n}\right)\right)^{-1}\left(\frac{X_n'W_n}{n}\right)V\frac{W_n'\varepsilon_n}{n} \quad (c')$$

(c) directly shows that $\hat\theta_n$ need not be unbiased, e.g. we do not necessarily have that
$$\mathbb{E}\left(\varepsilon_n / \sigma(X_n, W_n)\right) = 0_{n\times 1}.$$

(c') is convenient for the derivation of asymptotic properties.

Remark. Notice, in analogy to the OLS case, that $\frac{X_n'W_n}{n}$, and $\frac{W_n'E_n}{n}$ are arrays of empirical averages.

We employ the following high level conditions:

i'. $\frac{W_n'E_n}{n} \xrightarrow{P} 0$

ii'. $\frac{X_n'W_n}{n} \xrightarrow{P} M_{x'w}$ which is a deterministic $p \times q$ matrix

iii' $\operatorname{rank} M_{x'w} = p$

[We can also assume that V is stochastic and sample dependent - this could be justifiable in the case where V was needed to be optimally chosen. We would in this case require a condition of the form:

iv'. $V \xrightarrow{P} V^*$ deterministic

v.' $\operatorname{rank} V^* = q$

We will no pursue this for simplicity - Try it as an exercise]

Remark i', ii' could have lower level analogues that involve laws of large numbers, e.g. in the iid framework. Additionally, iii' could be facilitated by conditions that ensure asymptotic algebraic independence for the columns of X_n and W_n.

Under i', ii' and iii' and due to the CMT:

$$\left(\frac{X_n'W_n}{n}\right)V\left(\frac{W_n'X_n}{n}\right) \xrightarrow{P} M_{x'w}\,V\,M_{w'x}$$

Since rank $M_{x'w}=p$ and rank $V=q$ $\Rightarrow$ rank $M_{x'w}VM_{w'x}$

$$=p \overset{CMT}{\Rightarrow} \left(\left(\frac{X_n'W_n}{n}\right)V\left(\frac{W_n'X_n}{n}\right)\right)^{-1} \xrightarrow{P} \left(M_{x'w}VM_{w'x}\right)^{-1}$$

Analogously, $\left(\frac{X_n'W_n}{n}\right)V\frac{W_n'\varepsilon_n}{n} \xrightarrow{P} M_{x'w}V0_{q\times 1} = 0_{p\times 1}$.

Hence due to the CMT

$$\theta_0+\left(\left(\frac{X_n'W_n}{n}\right)V\left(\frac{W_n'X_n}{n}\right)\right)^{-1}\left(\frac{X_n'W_n}{n}\right)V\frac{W_n'\varepsilon_n}{n} \xrightarrow{P}$$

$$\theta_0+\left(M_{x'w}VM_{x'w}'\right)^{-1}M_{x'w}V0_{q\times 1}=\theta_0+0_{p\times 1}$$

$$=\theta_0$$

Hence:

Lemma If $\Theta=\mathbb{R}^p$ and i', ii', iii' hold then the IVE is weakly consistent.

What about when $\Theta\neq\mathbb{R}^p$? We can then use our general results and perform an analogous analysis to the previous example:

1'. Identify c^*. Using (b') and i', ii', iii' we have that for any $\theta\in\Theta$

$$*\ (\theta-\theta_0)'\left(\frac{X_n'W_n}{n}\right)V\left(\frac{W_n'X_n}{n}\right)(\theta-\theta_0)$$

$$\xrightarrow{P}(\theta-\theta_0)'M_{x'w}VM_{x'w}'(\theta-\theta_0),$$

$$*\ 2(\theta-\theta_0)'\left(\frac{X_n'W_n}{n}\right)V\frac{W_n'\varepsilon_n}{n}$$

$$\xrightarrow{P}2(\theta-\theta_0)'M_{x'w}V0_{q\times 1}=0,$$

$* \ \frac{\varepsilon_n' w_n}{n} V \frac{w_n' \varepsilon_n}{n} \to 0_{1\times q} V 0_{q\times 1} = 0$

Hence $C_n(\theta, V) \xrightarrow{P} C^*(\theta, V) =$
$= (\theta-\theta_0)' M_{x'w} V M_{x'w}' (\theta-\theta_0) \ \forall \theta \in \Theta.$

2! When Θ is compact it can be shown that

$\exists \ k_n > 0 : \forall \theta, \theta^* \in \Theta$

$$|C_n(\theta, V) - C_n(\theta^*, V)| \leq k_n \|\theta - \theta^*\|$$

for $k_n = C \| \frac{x_n' w_n}{n} V \frac{w_n' x_n}{n} \| + \| \frac{x_n' w_n}{n} V \frac{w_n' \varepsilon_n}{n} \|$

for some large enough $C^* > 0$ that depends on the diameter of Θ. then $k_n \xrightarrow{P} C^* \| M_{x'w} V M_{x'w}' \|$
hence it is bounded in probability, i.e. $\exists C > 0$:
$\lim_{n\to+\infty} \mathbb{P}(k_n > C) = 0.$

3! 1, 2 imply that when Θ is compact C_n converges to C^* locally uniformly and in probability.

4! $C^*(\theta, V)$ is continuous in Θ as a quadratic. Since $M_{x'w} V M_{x'w}'$ is of full rank, it is strictly positive definite. Hence $C^*(\theta, V)$ is uniquely minimized at θ_0.
Thereby when Θ is compact the asymptotic identification condition holds, due to the remark immediately after the proof of the consistency Theorem.

Hence due to 1', 2', 3', 4'

Lemma When Θ is compact, and i', ii', iii' hold then the IVE is weakly consistent.

5.' When Θ is closed convex, 1, 4 and the second part of the remark immediately after the proof of the consistency theorem imply that:

Lemma When Θ is closed and convex, and i', ii', iii' hold then the IVE is weakly consistent.

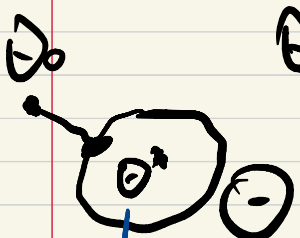
6.' Finally and for a general Θ for which

$\hat{\theta}_n \in \arg\min_{\theta \in \Theta} G_n(\theta, V)$ we can apply an analogous geometric argument similar to the one in the OLSE case to conclude that:

Lemma Under i', ii', iii' the IVE is weakly consistent.

Remark [a glipse of misspecification] Suppose that everything else holds, Θ is closed and convex but $\theta_0 \notin \Theta$. Then the previous imply that

$$\hat{\theta}_n \xrightarrow{P} \arg\min_{\Theta} C^*(\theta).$$ This is unique by iii', and the properties of Θ.



↳ pseudo true value at which $\hat{\theta}_n$ converges in probability

Using some convex analysis it is not difficult to show that $\underset{\theta\in\Theta}{\arg\min}\, C^*(\theta,v) = \underset{\theta\in\Theta}{\arg\min}\, \|\theta-\theta_0\|$.

Hence the estimator will converge to a pseudo true value defined as the unique element of Θ that is "closest" to θ_0. ∎

Example. For the GARCH(1,1) case the derivations are more complicated. It can be proven using the theorem above, that when Θ is compact, and additionally to the conditions that define the model, $\alpha_0+\beta_0<1$, and the distribution of z_0 is supported on at least three points, the Gaussian QMLE is weakly consistent. ∎

Appendix

* This colouring designates endnotes; they are indicated by numbers appearing in the main text

Endnotes

(1) Notice there $C_n(\theta)$ is not actually what appears there. Instead of the term 1, the correct term (see also its derivation in the previous set of notes) is $\varepsilon_n'\varepsilon_n/n$, i.e.

$$C_n(\theta) = (\theta-\theta_0)'\frac{X_n'X_n}{n}(\theta-\theta_0) - 2(\theta-\theta_0)'\frac{X_n'\varepsilon_n}{n} + \varepsilon_n'\varepsilon_n/n.$$

However notice that the term $\varepsilon_n'\varepsilon_n/n$ does not depend on Θ (it depends on Θ_0!), hence it does not affect the optimization of G_n w.r.t. Θ. Thereby we can equivalently ("optimization-wise") consider this "deformed"

Ⓐ $G_n(\Theta) = (\Theta-\Theta_0)' \frac{x_n' x_n}{n} (\Theta-\Theta_0) - 2(\Theta-\Theta_0)' \frac{x_n'\varepsilon_n}{n} + 1$

Notice for example that minimizing Ⓐ over $\Theta = \mathbb{R}^p$ (Ⓐ is also strictly convex - why?) results into $\Theta - \Theta_0 = \left(\frac{x_n' x_n}{n}\right)^{-1} \frac{x_n'\varepsilon_n}{n}$ (derive it!) which is the latent version of the OLSE (derive the equivalence of optimizing Ⓐ with optimizing the correct G_n for general Θ).

Notice also that:

a. the term 1 in Ⓐ is irrelevant; it can be replaced by an arbitrary $\alpha \in \mathbb{R}$, (or a random variable?), as long as c^* is analogously transformed - explain!

b. it essentially shows that the variance specifications in the linear model is irrelevant for consistency - why?

$\boxed{3}$

② There actually exists a different derivation of consistency for the OLSE that goes through the consistency - under i, ii, iii, of the unrestricted estimator $(X_n'X_n)^{-1}X_n'Y_n$. Can you find it?

③ A sufficient condition under which $\hat{\theta}_n$ does not depend on V is that $p=q$.

Please show it.