Semiparametrics

Introduction

Asymptotic normality

Two-step estimators

Semiparametric two-step estimators

Inference

Introduction

Example: Robinson

$$Y_i = X_i\beta + g(Z_i) + u_i$$

and

$$E(u_i|X_i, Z_i) = 0.$$

The parameter of interest is β.

Robinson estimator: idea

- Take the conditional expectation with respect to X_i:

$$E(Y_i|Z_i) = E(X_i|Z_i)\beta + g(Z_i)$$

- Subtract:

$$\begin{aligned} Y_i - E(Y_i|Z_i) &= (X_i - E(X_i|Z_i))\beta + (g(Z_i) - g(Z_i)) + u_i \\ \tilde{Y}_i &= \tilde{X}_i\beta + u_i \end{aligned}$$

- If the conditional expectations were known, use

$$\hat{\beta}^* = (\tilde{X}_i'\tilde{X}_i)^{-1}\tilde{X}_i\tilde{Y}_i$$

- In practice: estimate $\hat{E}(Y_i|Z_i)$ and $\hat{E}(X_i|Z_i)$ and call the resulting estimator $\hat{\beta}$

Example (2): Klein and Spady

[see previous slides]

Problem

Our standard proof for the consistency and asymptotic normality of extremum estimators does not cover these new estimators.

Asymptotic normality

Review: setup

An extremum estimator is defined as

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} Q_n(\theta)$$

for some real-valued function Q_n of the data $\{z_i,\ i = 1, 2, \cdots, n\}$ and a parameter θ from the space of parameters under consideration Θ.

Consistency

If the following conditions hold,

1. $Q_0(\theta_0) > Q_0(\theta)$ for any $\theta \in \Theta, \theta \neq \theta_0$
2. Q_0 is continuous on Θ
3. Θ is compact
4. $\sup_{\theta \in \Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$.

then $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Asymptotic normality

Suppose that $\hat{\theta}_n$ is consistent,

1. $\theta_0 \in \text{interior}(\Theta)$
2. Q_n is twice differentiable in a neighbourhood $\mathcal{N}$ of θ_0
3. $\sqrt{n}\nabla Q_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$
4. there is a function $H(\theta)$ and $H = H(\theta_0)$ such that
   - $\sup_{\theta \in \mathcal{N}} \|\nabla_{\theta\theta} Q_n(\theta) - H(\theta)\| \xrightarrow{p} 0$
   - that is continuous and invertible at θ_0

then

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

AN (2)

To show the above, note that

$$\begin{aligned} 0 &= \nabla_\theta Q_n(\hat{\theta}_n) \\ &= \nabla_\theta Q_n(\theta_0) + \nabla_{\theta\theta} Q_n(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0) \end{aligned}$$

where the

- first equality sets θ_n as FOC solution
- the second uses assumption 2
- with a mean value of $\tilde{\theta}_n$

AN (3)

- pick n large enough so that $\hat{\theta}_n$ (and therefore $\tilde{\theta}_n$) is in $\mathcal{N}$
- by continuity, $\nabla_{\theta\theta}$ is invertible
- then:

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta_0) &= -(\nabla_{\theta\theta} Q_n(\tilde{\theta}_n))^{-1} \nabla_\theta Q_n(\theta_0) \\ &\xrightarrow{d} H^{-1} \mathcal{N}(0, \Sigma) \\ &= \mathcal{N}(0, H^{-1} \Sigma H^{-1})\end{aligned}$$

where

- the first equality rewrites the result of the previous slide, assuming invertibility of the Hessian
- the limit follows from
  - the ULLN (condition 4)
  - the CLT (condition 2)

Two-step estimators

Setup

- specialize extremum estimation to M-estimation
- two parameters, (θ, γ).
- Consider the following two-step estimator:

1. In a first stage, estimate nuisance parameter γ from

$$\frac{1}{n}\sum_{i=1}^{n} m(z_i, \gamma) = 0$$

2. The second step estimation is from

$$\frac{1}{n}\sum_{i=1}^{n} g(z_i, \theta, \hat{\gamma}) = 0$$

One-step

Two-step estimation can be written as a one-step:

- moments: $\tilde{g} = (g, m)$
- parameters: $\tilde{\theta} = (\theta, \gamma)$

Previous results go through: if $\tilde{g}$ satisfies the AN conditions, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

where

$$V = G_\theta^{-1} E\left[(g(z) + G_\gamma \psi(z))(g(z) + G_\gamma \psi(z))'\right] G_\theta^{-1}$$

Conclusion

1. Asymptotic normality goes through
2. Some notation:

$$\begin{aligned} G_\theta &= E[\nabla_\theta g(z, \theta_0, \gamma_0)] \\ G_\gamma &= E[\nabla_\gamma g(z, \theta_0, \gamma_0)] \\ g(z) &= g(z, \theta_0, \gamma_0) \\ M &= E[\nabla_\gamma m(z, \gamma_0)] \\ \psi(z) &= -M^{-1} m(z, \gamma_0) \end{aligned}$$

3. If $G_\gamma = 0$ then V is equal to that for known-γ_0
4. $G_\gamma \psi(z) = -G_\gamma M^{-1} m$ is an adjust term that accounts for variance due to first step estimation

Semiparametric two-step estimators

Setup

Changing notation to Newey (1994), whose exposition (Section 5) we will follow, specialized to the just-identified case.

1. In a first stage, estimate nuisance parameter h by a **nonparametric** estimator $\hat{h}$
2. The second step estimation is from

$$\hat{m}_n(z_i, \beta, \hat{h}) = 0$$

where $\hat{m}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} m(z_i, \beta, \hat{h})$.

Results

I follow the exposition in Newey (1994), Section 5, specialized to the just-identified case.

Informally speaking, if we have:

1. uniform convergence of $\hat{m}_n(\beta_0)$ and $\partial \hat{m}_n(\beta_0)/\partial \beta$
2. asymptotic normality of $\sqrt{n}\hat{m}_n(\beta_0) \xrightarrow{d} \mathcal{N}(0, \Omega)$
3. consistency of $\hat{\beta}_n$

then we can do the usual mean-value expansion and

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(0, V),$$

where

$$V = M^{-1}\Omega(M')^{-1}$$
$$M = E[\partial m(z, \beta_0, h_0)/\partial \beta]$$

Asymptotic normality

- Start with asymptotic normality of $\sqrt{n}\hat{m}_n(\beta_0) \xrightarrow{d} \mathcal{N}(0, \Omega)$
- We need to control:
  - smoothness of m
  - estimation error in $\hat{h}$

Linearization

Let $\|\|$ be an appropriate norm, such as the Sobolev norm.

Assumption 1. For a small enough $\|h - h_0\|^2$, we have (i) $\|m(z,h) - m(z,h_0) - D(z,h-h_0)\| \leq b(z)\|h - h_0\|^2$ and (ii) $E(b(z))\sqrt{n}\|\hat{h} - h_0\|^2 \xrightarrow{p} 0$.

- implies Frechet-differentiability of m in h: the objective function depends smoothly on h
- think of $D(z, h - h_0)$ as the derivative ("$D(z)\|\hat{h} - h_0\|$")
- convergence rates for h should be faster than $n^{-1/4}$
  - choose bandwidth appropriately
  - or: require enough smoothness (increasing in first step's dimension)

Stochastic equicontinuity

Assumption 2. Stochastic equicontinuity:

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(D(z_i, \hat{h} - h_0) - \int D(z, \hat{h} - h_0)dF_0) \xrightarrow{p} 0$$

- imposes smoothness on m

Uniform convergence

Stochastic equicontinuity is important in econometrics. If the parameter space is compact, then

pointwise convergence and stochastic equicontinuity $\Leftrightarrow$ uniform converger

- replaces Jennrich's ULLN (Lemma 2.4 in NM'94)
- used in NM'94 for nonsmooth objective functions
- primitive conditions in Andrews (94) - chapter after NM
- literature on **empirical processes** provides larger toolbox for uniform convergence (e.g. vdVaart, Ch 18+19+20+25)

Continuity

Assumption 3. (i) there exists an $a(z)$ and a measure $\hat{F}$ such that $E(a(z)) = 0$ and $Var(a(z)) < \infty$ such that for small enough $\|\hat{h} - h_0\|$:

$$\int D(z, \hat{h} - h_0) dF_0 = \int a(z) d\hat{F}$$

(ii) for the EDC $\tilde{F}$, $\sqrt{n}(\int a(z) d\hat{F} - \int a(z) d\tilde{F}) \xrightarrow{p} 0$

- think $\hat{F} = \tilde{F}$ for part (i)
- $a(z)$ is a pathwise (Gateaux-like) derivative in the Hilbert space of functions that is the first step parameter space
- if $a(z) = 0$ (which happens when $E(D(z, h - h_0)) = 0$) then first step estimation does not affect $\hat{\beta}$
  - interpretation: linearity of m in h
    - Imbens, Newey, Ridder (2003)
    - Chernozhukov, Escanciano, Ichimura, Newey (2016)

Asymptotic normality

Lemma 1. Under Assumptions 1-3,

$$\sqrt{n}\hat{m}_n(\beta_0) \xrightarrow{d} \mathcal{N}(0, \Omega)$$

with

$$\Omega = E[(m(z, h_0) + a(z))(m(z, h_0) + a(z)'].$$

- Done: asymptotic normality
- To do:
  - consistency of $\hat{\beta}$
  - uniform convergence

Consistency

Assumptions 4 and 5. There exists ϵ, $\|h\|$, $b(z)$ and $\tilde{b}(z)$ such that: (i) w.p. 1, m is continuous in β on the parameter space $\mathcal{B}$; (ii) $\|m(z, \beta, h_0)\| \leq b(z)$; (iii) $\|m(z, \beta, h) - m(z, \beta, h_0)\| \leq \tilde{b}(z)\|h - h_0\|^{\epsilon}$; (iv) $E(m(z, \beta, h_0))$ has a unique solution at $\beta_0 \in \mathcal{B}$; $\mathcal{B}$ is compact.

Lemma 2. If Assumptions 4 and 5 hold, and $\|\hat{h} - h_0\| \xrightarrow{p} 0$, then $\hat{\beta} - \beta_0 \xrightarrow{p} 0$.

- Conditions as in one-step
  - id / compactness / continuity / boundedness
  - strengthened by consistency of $\hat{h}$
  - bound effect that $\hat{h}$ can have on m

Overview

- Done:
  - consistency of $\hat{\beta}$
  - asymptotic normality of $\sqrt{n}\hat{m}_n(\beta_0)$
- To do:
  - uniform convergence

Uniform convergence

Assumption 6. (i) β_0 is in the interior of $\mathcal{B}$; (ii) there is $\|h\|, \epsilon > 0$ and a neighbourhood $\mathcal{N}$ such that for all $\|h - h_0\| < \epsilon$, $m(z, \beta, h)$ is differentiable in β on $\mathcal{N}$; (iii) M is invertible; (iv) $E[\|m(z, \beta_0, h_0)\|^2] < \infty$; (v) Assumption 4 (i)-(iii) applies to $\partial m / \partial \beta$.

Lemma 2. If Assumptions 1-6 hold, and $\|\hat{h} - h_0\| \xrightarrow{p} 0$ then

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

Inference

Inference

- How to construct standard errors, do hypothesis testing?
- Focus on semiparametric two-step estimation (nests parametric and one-step)
- Define

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(z_i, \hat{\beta}, \hat{h})}{\partial \beta}$$

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}([m(z_i, \hat{\beta}, \hat{h}) + \hat{a}(z_i)][m(z_i, \hat{\beta}, \hat{h}) + \hat{a}(z_i)]')$$

- Newey, Lemma 5.4 provides conditions under which $\hat{V} \xrightarrow{p} V$
- Then use

$$\hat{V}^{-1/2}\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, I)$$

for inference

Inference (2)

Problem: $a(z_i)$ can be hard to obtain.

- it results from projection in an infinite-dimensional Hilbert space

Solutions:

1. use insights in Newey (1994), Sections 2-4, special cases in 6-7
2. use Ichimura and Newey (2015), who propose a very clever use of the Gateuaux derivative
3. use Ackerberg, Chen, and Hahn (2012), who show that - if you use sieves - you can use standard errors from parametric two-step
4. use the **bootstrap**

Bootstrap

Will do in an application.