An extremum estimator is the minimizer / maximizer of a sample criterion function $Q_n(\theta)$, where $\theta \in \Theta \subset \mathbb{R}^K$.

We assume that F.O.C. are differentiable and allow a mean value expansion around the true value θ_0:

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta}$$

⇒ By Mean Value Expansion around θ_0

$$o_p(1) = \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_n(\theta_n^*)}{\partial \theta\, \partial \theta'} \sqrt{n}\,(\hat{\theta}_n - \theta_0)$$

Assume
- $\sup_{\theta \in \Theta} \left\| \frac{\partial^2 Q_n(\theta)}{\partial \theta\, \partial \theta'} - B(\theta) \right\| = o_p(1)$
- $B(\theta)$ is continuous at θ_0.

⇒ Uniform LLN holds

⇒ By ULLN

$$o_p(1) = \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial \theta} + \left(B(\theta_0) + o_p(1)\right) \sqrt{n}\,(\hat{\theta}_n - \theta_0)$$

⇒ Rearrange

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = -B(\theta_0)^{-1} \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p(1),$$

where $\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \Omega_0)$.

Therefore,

This is called the influence function of $\hat{\theta}_n$.

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i + o_p(1) \xrightarrow{d} N(0, B_0^{-1}\Omega_0 B_0^{-1}).$$

$\Omega_0 = E\,\xi_i \xi_i'$

A) Ordinary Least Squares

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} (y_i - x_i'\theta)^2 / 2$$

- $$\frac{\partial Q_n(\theta)}{\partial \theta} = -\frac{1}{n}\sum_{i=1}^{n} x_i \, (y_i - x_i'\theta)$$

- $$\frac{\partial^2 Q_n(\theta)}{\partial\theta\,\partial\theta'} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{p} E x_i x_i'$$

at true parameter

- $$\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i \xrightarrow{d} N(0, E u_i^2 x_i x_i')$$

- $$\sqrt{n}(\hat\theta_n - \theta_0) \xrightarrow{d} N(0, (E x_i x_i')^{-1} E u_i^2 x_i x_i' (E x_i x_i')^{-1})$$

- $$\xi_i = (E x_i x_i')^{-1} x_i u_i$$

B) Linear GMM

$$Q_n(\theta) = \left(\frac{1}{n}\sum_{i=1}^{n} z_i (y_i - x_i'\theta)\right)' A_n' A_n \left(\frac{1}{n}\sum_{i=1}^{n} z_i (y_i - x_i'\theta)\right) / 2$$

- $$\frac{\partial Q_n(\theta)}{\partial\theta} = -\frac{1}{n}\sum_{i=1}^{n} x_i z_i' \, A_n' A_n \left(\frac{1}{n}\sum_{i=1}^{n} z_i (y_i - x_i'\theta)\right)$$

- $$\frac{\partial^2 Q_n(\theta)}{\partial\theta\,\partial\theta'} = \frac{1}{n}\sum_{i=1}^{n} x_i z_i' \, A_n' A_n \, \frac{1}{n}\sum_{i=1}^{n} z_i x_i' \xrightarrow{p} \Gamma_0' A' A \Gamma_0$$

- $$\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta} = -\frac{1}{n}\sum_{i=1}^{n} x_i z_i' \, A_n' A_n \, \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i \xrightarrow{d} N(0, \Gamma_0' A' A \, E u_i^2 z_i z_i' \, A' A \Gamma_0)$$

- $$\sqrt{n}(\hat\theta_n - \theta_0) \xrightarrow{d} N(0, (\Gamma_0' A' A \Gamma_0)^{-1} \Gamma_0' A' A \, E u_i^2 z_i z_i' \, A' A \Gamma_0 (\Gamma_0' A' A \Gamma_0)^{-1})$$

- $$\xi_i = (\Gamma_0' A' A \Gamma_0)^{-1} \Gamma_0' A' A \, z_i u_i$$

C) Maximum Likelihood

$$Q_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log f(w_i,\theta)$$

- $$\frac{\partial Q_n(\theta)}{\partial\theta} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(w_i,\theta)$$

- $$\frac{\partial^2 Q_n(\theta)}{\partial\theta\,\partial\theta'} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta'}\log f(w_i,\theta) \xrightarrow[p]{\text{at } \theta_0} \underbrace{E\frac{\partial^2}{\partial\theta\partial\theta'}\log f(w_i,\theta_0)}_{B_0}$$

- $$\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial \log f(w_i,\theta_0)}{\partial\theta} \xrightarrow{d} N\Big(0,\ \underbrace{E\,\frac{\partial \log f(w_i,\theta_0)}{\partial\theta}\,\frac{\partial \log f(w_i,\theta_0)}{\partial\theta}'}_{\Omega_0}\Big)$$

- $$\sqrt{n}(\hat\theta_n-\theta_0) \xrightarrow{d} N(0,\ B_0^{-1}\Omega_0 B_0^{-1}) \overset{\text{under correct specification}}{=} N(0,\ \Omega_0^{-1})$$

- $$\xi_i = -B_0^{-1}\frac{\partial}{\partial\theta}\log f(w_i,\theta_0)$$

D) Non-linear Least Squares

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} (y_i - g(x_i,\theta))^2/2$$

- $$\frac{\partial Q_n(\theta)}{\partial\theta} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(x_i,\theta)}{\partial\theta}(y_i - g(x_i,\theta))$$

- $$\frac{\partial^2 Q_n(\theta)}{\partial\theta\,\partial\theta'} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial g(x_i,\theta)}{\partial\theta}\,\frac{\partial g(x_i,\theta)}{\partial\theta'} - (y_i-g(x_i,\theta))\,\frac{\partial^2 g(x_i,\theta)}{\partial\theta\,\partial\theta'}\right\}$$

$$\xrightarrow[p]{\text{at } \theta_0} E\,\frac{\partial g(x_i,\theta_0)}{\partial\theta}\,\frac{\partial g(x_i,\theta_0)}{\partial\theta'} - \underbrace{E\,u_i\,\frac{\partial^2 g(x_i,\theta_0)}{\partial\theta\partial\theta'}}_{=0 \text{ if correctly specified by UE}:\ E[u_i|X_i]=0} := B_0^{-1}$$

- $$\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(x_i,\theta_0)}{\partial\theta}u_i \xrightarrow{d} N\Big(0,\ E u_i^2\,\frac{\partial g(x_i,\theta_0)}{\partial\theta}\,\frac{\partial g(x_i,\theta_0)}{\partial\theta'}\Big)$$

- $$\sqrt{n}(\hat\theta_n-\theta_0) \xrightarrow{d} N\Big(0,\ B_0^{-1}\,E u_i^2\,\frac{\partial g(x_i,\theta_0)}{\partial\theta}\,\frac{\partial g(x_i,\theta_0)}{\partial\theta'}\Big)$$

- $$\xi_i = -B_0^{-1}\,\frac{\partial g(x_i,\theta_0)}{\partial\theta}\,u_i$$

E) <u>Non-linear GMM</u>

$$Q_n(\theta) = \left(\frac{1}{n}\sum_{i=1}^{n} g(w_i,\theta)\right)' A_n' A_n \left(\frac{1}{n}\sum_{i=1}^{n} g(w_i,\theta)\right)$$

- $$\frac{\partial Q_n(\theta)}{\partial\theta} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(w_i,\theta)}{\partial\theta'}\right)' A_n' A_n \left(\frac{1}{n}\sum_{i=1}^{n} g(w_i,\theta)\right)$$

- $$\frac{\partial^2 Q_n(\theta)}{\partial\theta\,\partial\theta'} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(w_i,\theta)}{\partial\theta'}\right)' A_n' A_n \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(w_i,\theta)}{\partial\theta'}\right) +$$

$$\left[I_k \otimes \left(\frac{1}{n}\sum_{i=1}^{n} g(w_i,\theta)' A_n' A_n\right)\right]\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\operatorname{vec}\left(\frac{\partial g(w_i,\theta)}{\partial\theta'}\right)\right]$$

$$\xrightarrow[p]{\text{at } \theta_0} \Gamma_0' A' A \Gamma_0 + [I_k \otimes \underbrace{E g(w_i,\theta_0)'}_{=0 \text{ under correct specification.}} A'A]\; E\frac{\partial}{\partial\theta}\operatorname{vec}\left(\frac{\partial g(w_i,\theta_0)}{\partial\theta'}\right)$$

- $$\sqrt{n}\,\frac{Q_n(\theta_0)}{\partial\theta} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(w_i,\theta_0)}{\partial\theta'}\right)' A_n' A_n \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(w_i,\theta_0)$$

$$\xrightarrow{d} N(0,\ \Gamma_0' A' A\, E g(w_i,\theta_0) g(w_i,\theta_0)' A' A \Gamma_0)$$

- $$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0,\ B_0^{-1}\Gamma_0' A'A\, E g(w_i,\theta_0) g(w_i,\theta_0)' A'A \Gamma_0 B_0^{-1}).$$

- $$\xi_i = B_0^{-1}\Gamma_0' A' A\, g(w_i,\theta_0)$$

If there's no misspecification then $A^{*\prime} A^{*} = [E g(w_i,\theta_0) g(w_i,\theta_0)']^{-1}$. Otherwise, θ_0 depend on $A'A$ (pseudo-true parameter) and there cannot be an $\hat{\text{Avar}}$ robust to misspecification.

F) <u>Minimum Distance (MD)</u>

$$Q_n(\theta) = (\hat{\pi}_n - g(\theta))' A_n' A_n (\hat{\pi}_n - g(\theta))/2,$$

where $\sqrt{n}(\hat{\pi}_n - \pi_0) \xrightarrow{d} N(0, V_0)$ is a first-step estimator.

- $$\frac{\partial Q_n(\theta)}{\partial \theta} = -\left(\frac{\partial g(\theta)}{\partial \theta'}\right)' A_n' A_n \left(\hat{\pi}_n - g(\theta)\right)$$

- $$\frac{\partial^2 Q_n(\theta)}{\partial \theta \, \partial \theta'} = \left(\frac{\partial g(\theta)}{\partial \theta'}\right)' A_n' A_n \left(\frac{\partial g(\theta)}{\partial \theta'}\right) + \left[I_K \otimes (\hat{\pi}_n - g(\theta))' A_n' A_n\right] \frac{\partial}{\partial \theta} \operatorname{vec}\left(\frac{\partial g(\theta)}{\partial \theta'}\right)$$

$$\xrightarrow[p]{\text{at } \theta_0} \Gamma_0' A' A \Gamma_0 + \underbrace{\left[I_K \otimes (\pi_0 - g(\theta_0)) A' A\right] \frac{\partial}{\partial \theta} \operatorname{vec}\left(\frac{\partial g(\theta_0)}{\partial \theta'}\right)}_{=0 \text{ if } \pi_0 = g(\theta_0) \text{ or } g(\cdot) \text{ is linear in } \theta.}$$

- $$\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} = -\left(\frac{\partial g(\theta_0)}{\partial \theta'}\right)' A_n' A_n \sqrt{n}(\hat{\pi}_n - \underbrace{g(\theta_0)}_{\pi_0})$$

$$\xrightarrow{d} N(0, \Gamma_0' A' A V_0 A' A \Gamma_0)$$

- $$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, B_0^{-1} \Gamma_0' A' A V_0 A' A \Gamma_0 B_0^{-1})$$

- $$\xi_i = B_0^{-1} \Gamma_0' A' A \xi_i^{\pi}, \quad \text{where} \quad V_0 = E\, \xi_i^{\pi} \xi_i^{\pi\prime}.$$

↳ influence function of first step estimator.

If there's no misspecification then $A^{*\prime} A^* = V_0^{-1}$.
Otherwise, θ_0 depend on $A'A$ (pseudo-true parameter) and there cannot be an $\hat{\text{Avar}}$ robust to misspecification.

Before introducing two-step estimators more generally, let's introduce "stacking" of influence functions.

Suppose with the same n observations we can estimate two objects

$$\sqrt{n}\left(\theta_n^A - \theta_0^A\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i^A + o_p(1) \; ; \; \sqrt{n}\left(\theta_n^B - \theta_0^B\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i^B + o_p(1)$$

We can test, for example $\hat{\theta}_n^A = \hat{\theta}_n^B$. How?

$$\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\xi_i^A \\ \xi_i^B\end{pmatrix}}_{\text{We stack them and "pretend" they come from the same estimation}} \xrightarrow{d} \mathcal{N}\underbrace{\begin{pmatrix} E\,\xi_i^A \xi_i^{A\prime} & E\,\xi_i^A \xi_i^{B\prime} \\ E\,\xi_i^B \xi_i^{A\prime} & E\,\xi_i^A \xi_i^{B\prime}\end{pmatrix}}_{\text{We get the covariance structure of the estimators by co-varying their influence functions.}}$$

Fun Little Problem

- Suppose you have a random sample of three variables, (x_i, y_i, z_i), $\quad i = 1, \ldots, N$.
- x_i and y_i are mean zero, but z_i is not.
- Suppose you want to estimate the mean, μ of z_i
- Suppose you want to estimate the slope, β, in a simple linear regression

$$y_i = x_i\beta + u_i$$

- Suppose you then want to know the covariance between μ and β?

35

1 6

Toni Whited @toniwhited · Feb 12

6/7

Stacking

- You could just bootstrap.
- OR! . . . you could stack the influence functions for $\hat{\mu}$ and $\hat{\beta}$ and covary them.
- That's what Tim Erickson and I did in Erickson and Whited (2002, Econometric Theory).
- The next slide shows you how to do it in Julia. Super easy.
- You can do this for most estimators because most estimators we use are asymptotically linear.

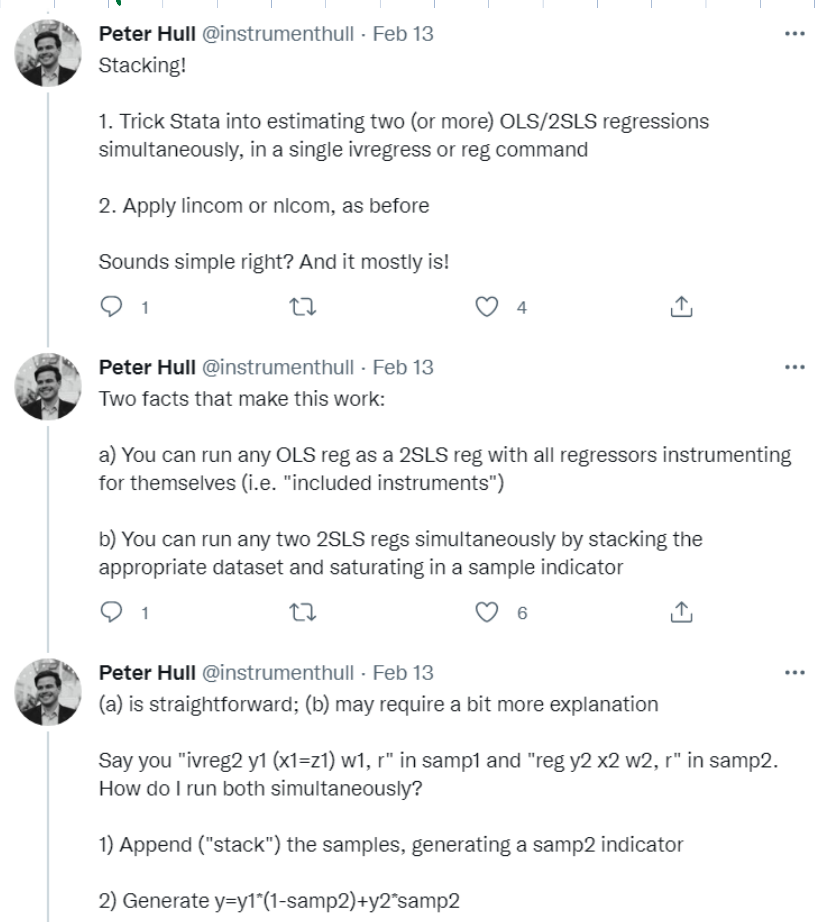
In practice, the observations in the sample used for each estimator can differ, but you can still create sample dummies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \xi_i^A \times \text{sample } A_{i\,(1\times 1)} \\ \xi_i^B \times \text{sample } B_{i\,(1\times 1)} \end{pmatrix} \xrightarrow[d]{} \mathcal{N}\left(\begin{matrix} E\, \xi_i^A \xi_i^{A'} s_i^A & E\, \xi_i^A \xi_i^{B'} s_i^A s_i^B \\ E\, \xi_i^B \xi_i^{A'} \underbrace{s_i^A s_i^B} & E\, \xi_i^B \xi_i^{B'} s_i^B \end{matrix} \right)$$

all observations such that sample $A_i = 1$ or sample $B_i = 1$

if the two samples do not overlap at all, we can assume there's no covariance!

One of the most important applied econometricians right now!


Peter Hull @instrumenthull · Feb 13
Stacking!

1. Trick Stata into estimating two (or more) OLS/2SLS regressions simultaneously, in a single ivregress or reg command

2. Apply lincom or nlcom, as before

Sounds simple right? And it mostly is!

1 4

Peter Hull @instrumenthull · Feb 13
Two facts that make this work:

a) You can run any OLS reg as a 2SLS reg with all regressors instrumenting for themselves (i.e. "included instruments")

b) You can run any two 2SLS regs simultaneously by stacking the appropriate dataset and saturating in a sample indicator

1 6

Peter Hull @instrumenthull · Feb 13
(a) is straightforward; (b) may require a bit more explanation

Say you "ivreg2 y1 (x1=z1) w1, r" in samp1 and "reg y2 x2 w2, r" in samp2. How do I run both simultaneously?

1) Append ("stack") the samples, generating a samp2 indicator

2) Generate y=y1*(1-samp2)+y2*samp2

But this is more of a practical concern, we will assume that any estimator uses all the observations in the data.

Two-Step Estimation (GMM)

$$Q_n(\theta,\hat{\tau}_n) = G_n(\theta,\hat{\tau}_n)' A_n' A_n \, G_n(\theta,\hat{\tau}_n)/2$$

- $$\frac{\partial Q_n(\theta,\hat{\tau}_n)}{\partial \theta} = \left(\frac{\partial G_n(\theta,\hat{\tau}_n)}{\partial \theta}\right)' A_n' A_n \, G_n(\theta,\hat{\tau}_n)$$

- $$\frac{\partial^2 Q_n(\theta,\hat{\tau}_n)}{\partial \theta \, \partial \theta'} = \left(\frac{\partial G_n(\theta,\hat{\tau}_n)}{\partial \theta}\right)' A_n' A_n \left(\frac{\partial G_n(\theta,\hat{\tau}_n)}{\partial \theta}\right) +$$

$$\left[I_k \otimes G_n(\theta,\hat{\tau}_n)' A_n' A_n \right] \frac{\partial}{\partial \theta} \mathrm{vec}\left(\frac{\partial G_n(\theta,\hat{\tau}_n)}{\partial \theta'}\right)$$

$$\xrightarrow[p]{\text{at } \theta_0} \Gamma_0' A' A \, \Gamma_0 + [I_k \otimes \underbrace{G(\theta_0,\tau_0)'}_{=0} A' A] \frac{\partial}{\partial \theta} \mathrm{vec}\left(\frac{\partial G(\theta_0,\tau_0)}{\partial \theta'}\right)$$

Uniform convergence under θ and τ. (Lemma 12.2)

under correct specification

(think of GMM or MD)

- $$\sqrt{n}\,\frac{\partial Q_n(\theta_0,\hat{\tau}_n)}{\partial \theta} = \sqrt{n}\left(\frac{\partial G_n(\theta_0,\hat{\tau}_n)}{\partial \theta}\right)' A_n' A_n \, G_n(\theta_0,\hat{\tau}_n)$$

Mean value around τ_0

$$= \left(\frac{\partial G_n(\theta_0,\tau_0)}{\partial \theta}\right)' A_n' A_n \sqrt{n}\, G_n(\theta_0,\tau_0) +$$

$$\frac{\partial}{\partial \tau'}\left\{\left(\frac{\partial G_n(\theta_0,\tau_n^*)}{\partial \theta}\right)' A_n' A_n \, G_n(\theta_0,\tau_n^*)\right\} \sqrt{n}\,(\hat{\tau}_n - \tau_0)$$

$$= \Gamma_0' A' A \sqrt{n}\, G_n(\theta_0,\tau_0) + \Lambda_0 \sqrt{n}\,(\hat{\tau}_n - \tau_0) + o_p(1)$$

$$= \left[\Gamma_0' A' A \;\vdots\; \Lambda_0\right] \underbrace{\sqrt{n}\begin{pmatrix} G_n(\theta_0,\tau_0) \\ \hat{\tau}_n - \tau_0 \end{pmatrix}}_{\text{stacking trick!}} + o_p(1)$$

$$\xrightarrow[d]{} N\left(0, \left[\Gamma_0' A' A \;\vdots\; \Lambda_0\right] \begin{pmatrix} V_{10} & V_{20} \\ V_{20}' & V_{30} \end{pmatrix} \left[\Gamma_0' A' A \;\vdots\; \Lambda_0\right]' \right),$$

where $$\sqrt{n}\begin{pmatrix} G_n(\theta_0,\tau_0) \\ \hat{\tau}_n - \tau_0 \end{pmatrix} \xrightarrow[d]{} N\left(0, \begin{pmatrix} V_{10} & V_{20} \\ V_{20}' & V_{30} \end{pmatrix}\right)$$

Finally,

- $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, B_0^{-1}\left[\Gamma_0' A' A \;\vdots\; \Delta_0\right]\begin{pmatrix} V_{10} & V_{20} \\ V_{20}' & V_{30} \end{pmatrix}\left[\Gamma_0' A' A \;\vdots\; \Delta_0\right]' B_0^{-1}\right)$

- $\xi_i = B_0^{-1}\left[\Gamma_0' A' A \;\vdots\; \Delta_0\right]\begin{pmatrix} \xi_i^G \\ \xi_i^T \end{pmatrix}$, where ξ_i^G is the influence function of $\sqrt{n}\, G_n(\theta_0, \tau_0)$, and ξ_i^T is the influence function of $\sqrt{n}(\hat{\tau}_n - \tau_0)$.

Final 2019

2. **(Logit Binary Choice model)** Let iid data $\{(X_i', Y_i)' : i = 1, \ldots, n\}$, where $Y_i \in \{0, 1\}$ is a binary variable, be generated according to the model

$$E(Y_i \mid X_i) = P(Y_i = 1 \mid X_i) = \Lambda(X_i'\beta_0),$$

where $\Lambda(u) = e^u/(1 + e^u)$ is the CDF of the Logistic distribution, X_i is the k-vector of regressors, and $\beta_0 \in \mathbb{R}^k$ is the unknown vector of parameters. Note that the conditional distribution of Y_i conditional on X_i can be described as

$$P(Y_i = y \mid X_i) = \left(\Lambda(X_i'\beta_0)\right)^y \left(1 - \Lambda(X_i'\beta_0)\right)^{1-y}, \quad y \in \{0, 1\}.$$

Let $\hat{\beta}_n$ denote the maximum likelihood estimator (MLE) of β_0:

$$\hat{\beta}_n = \arg\max_{\beta \in \mathbb{R}^k} Q_n(\beta), \text{ where}$$

$$Q_n(\beta) = n^{-1} \sum_{i=1}^{n} \left\{Y_i \ln \Lambda(X_i'\beta) + (1 - Y_i) \ln\left(1 - \Lambda(X_i'\beta)\right)\right\}.$$

Let $\tilde{\beta}_n$ denote the nonlinear least squares (NLS) estimator of β_0:

$$\tilde{\beta}_n = \arg\min_{\beta \in \mathbb{R}^k} R_n(\beta), \text{ where}$$

$$R_n(\beta) = n^{-1} \sum_{i=1}^{n} \left(Y_i - \Lambda(X_i'\beta)\right)^2 / 2.$$

(a) **(3 points)** What assumption does one need to impose on X_i to ensure identification of β_0?

solution:

$(*)$ $1 - \frac{1}{x} \le \log x \le x - 1$

First, consider the case of MLE

$$Q(\beta) = E\{ y_i \ln \Lambda(x_i'\beta) + (1-y_i) \ln (1 - \Lambda(x_i'\beta))\}$$

Notice that we need to show that $Q(\hat{\beta_n}) - Q(\beta_0) < 0$, i.e. $Q(\beta_0)$ is the unique maximizer of MLE.

Now,

- $$Q(\hat{\beta_n}) - Q(\beta_0) = E\left\{ y_i \ln \frac{\Lambda(x_i'\hat{\beta_n})}{\Lambda(x_i'\beta_0)} + (1-y_i) \ln \frac{1-\Lambda(x_i'\hat{\beta_n})}{1-\Lambda(x_i'\beta_0)} \right\}$$

$$\le E\left\{ y_i \left(\frac{\Lambda(x_i'\hat{\beta_n}) - \Lambda(x_i'\beta_0)}{\Lambda(x_i'\beta_0)} \right) + (1-y_i) \frac{\Lambda(x_i'\beta_0) - \Lambda(x_i'\hat{\beta_n})}{1-\Lambda(x_i'\beta_0)} \right\}$$

$\overset{\text{LIE}}{=} 0.$ (by str. concavity it achieves max at 1 point)

So when is it zero? i.e. it achieves the upper bound

- $Q(\hat{\beta}_n) - Q(\beta_0) = E\left\{ Y_i \ln \frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} + (1-Y_i) \ln \frac{1-\Lambda(X_i'\hat{\beta}_n)}{1-\Lambda(X_i'\beta_0)} \right\}$

$$= \underbrace{E\left\{ Y_i \ln \frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} + (1-Y_i) \ln \frac{1-\Lambda(X_i'\hat{\beta}_n)}{1-\Lambda(X_i'\beta_0)} \,\middle|\, X_i'\hat{\beta}_n = X_i'\beta_0 \right\}}_{=0} P(X_i'\hat{\beta}_n = X_i'\beta_0)$$

$$+$$

$$\underbrace{E\left\{ Y_i \ln \frac{\Lambda(X_i'\hat{\beta}_n)}{\Lambda(X_i'\beta_0)} + (1-Y_i) \ln \frac{1-\Lambda(X_i'\hat{\beta}_n)}{1-\Lambda(X_i'\beta_0)} \,\middle|\, X_i'\hat{\beta}_n \neq X_i'\beta_0 \right\}}_{\neq 0 \text{ otherwise there would be set identification (which contradicts strict concavity of the obj. function)}} \underbrace{P(X_i'\hat{\beta}_n \neq X_i'\beta_0)}_{\text{if this } = 0}$$

Then, our condition is $P(X_i'\hat{\beta}_n \neq X_i'\beta_0) > 0$.

Next, for NLS we have

$$R(\beta) = \frac{1}{2} E\,(Y_i - \Lambda(X_i'\beta))^2$$

$$R_n(\beta) = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} (Y_i - \Lambda(X_i'\beta))^2$$

We want to show that $R(\hat{\beta}_n) - R(\beta_0) > 0$.

$$R(\hat{\beta}_n) - R(\beta_0) = \frac{1}{2}\left\{ E(\underbrace{Y_i - \Lambda(X_i'\beta_0)}_{u_i} + \Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n))^2 - E u_i^2 \right\}$$

$$= \frac{1}{2}\left\{ E(\cancel{u_i^2} + 2u_i(\Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n)) + (\Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n))^2) - \cancel{E u_i^2} \right\}$$

$$= \frac{1}{2}\left\{ E[\underbrace{E(u_i|X_i)}_{=0}(\Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n))] + \underbrace{E(\Lambda(X_i'\beta_0) - \Lambda(X_i'\hat{\beta}_n))^2}_{\text{can only be zero when } P(\Lambda(X_i'\beta_0) = \Lambda(X_i'\hat{\beta}_n)) = 1} \right\} \geq 0$$

so the condition is the same.

(b) **(12 points)** Find the asymptotic variance of the MLE $\hat{\beta}_n$. Assume that EX_iX_i' is positive definite and finite. Hints: (i) Use the property

$$\frac{d\Lambda(u)}{du} = \Lambda(u)\left(1-\Lambda(u)\right)$$

to show that

$$\frac{\partial Q_n(\beta_0)}{\partial \beta} = n^{-1}\sum_{i=1}^{n}\left(Y_i - \Lambda(X_i'\beta_0)\right)X_i.$$

(ii) Show that

$$Var(Y_i \mid X_i) = \Lambda(X_i'\beta_0)\left(1-\Lambda(X_i'\beta_0)\right).$$

Solution:

(i)
$$\frac{\partial Q_n(\beta_0)}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n}\left\{ Y_i \frac{\partial \Lambda(X_i'\beta_0)/\partial\beta}{\Lambda(X_i'\beta_0)} - (1-Y_i)\frac{\partial\Lambda(X_i'\beta_0)/\partial\beta}{1-\Lambda(X_i'\beta_0)}\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{ Y_i \frac{\cancel{\Lambda(X_i'\beta_0)}(1-\Lambda(X_i'\beta_0))X_i}{\cancel{\Lambda(X_i'\beta_0)}} - (1-Y_i)\frac{\Lambda(X_i'\beta_0)\cancel{(1-\Lambda(X_i'\beta_0))}X_i}{\cancel{1-\Lambda(X_i'\beta_0)}}\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{(Y_i - \Lambda(X_i'\beta_0))X_i\right\}$$

(ii)
$$Var(Y_i \mid X_i) = E\left\{(Y_i - E(Y_i\mid X_i))^2 \mid X_i\right\}$$

$$= E\left\{(Y_i - \Lambda(X_i'\beta_0))^2 \mid X_i\right\}$$

$$= E\left\{\underbrace{Y_i^2}_{\text{binary}} - 2Y_i\Lambda(X_i'\beta_0) + \Lambda(X_i'\beta_0)^2 \mid X_i\right\}$$

$$= E\left\{Y_i - 2Y_i\Lambda(X_i'\beta_0) + \Lambda(X_i'\beta_0)^2 \mid X_i\right\}$$

$$= \Lambda(X_i'\beta_0) - 2\Lambda(X_i'\beta_0)^2 + \Lambda(X_i'\beta_0)^2$$

$$= \Lambda(X_i'\beta_0)\left(1-\Lambda(X_i'\beta_0)\right).$$

Then
$$\sqrt{n}\,\frac{\partial Q_n(\beta_0)}{\partial\beta} \to N\left(0, E(Y_i - \Lambda(X_i'\beta_0))^2 X_iX_i'\right)$$
$$= N\left(0, E\,\Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))X_iX_i'\right)$$
$$= N(0, \Omega_0)$$

Recall that

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{\partial Q_n(\hat{\beta}_n)}{\partial\beta} = \frac{\partial Q_n(\beta_0)}{\partial\beta} + \frac{\partial Q_n(\beta^*)}{\partial\beta\,\partial\beta'}(\hat{\beta}_n - \beta_0)$$

where
$$\frac{\partial Q_n(\beta_0)}{\partial\beta\,\partial\beta'} = -\frac{1}{n}\sum_{i=1}^{n} \Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))\, X_i X_i'$$
$$\xrightarrow[p]{} -E\left[\Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))\, X_i X_i'\right] = \Omega_0$$

Finally,
$$\sqrt{n}(\hat{\beta}_n - \beta_0) = -\Omega_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{(Y_i - \Lambda(X_i'\beta_0))X_i\} + o_p(1)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \underbrace{-\Omega_0^{-1}\{(Y_i - \Lambda(X_i'\beta_0))X_i\}}_{\xi_i^{MLE}} + o_p(1)$$
$$\xrightarrow[d]{} N(0, \Omega_0^{-1}).$$

(c) **(12 points)** Find the asymptotic variance of the NLS estimator $\tilde{\beta}_n$.

solution:

Again, we require to analyze the following objects
$$\frac{\partial R_n(\beta_0)}{\partial\beta} = -\frac{1}{n}\sum_{i=1}^{n}(Y_i - \Lambda(X_i'\beta_0))\frac{\partial\Lambda(X_i'\beta_0)}{\partial\beta}$$
$$= -\frac{1}{n}\sum_{i=1}^{n}(Y_i - \Lambda(X_i'\beta_0))\Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))X_i$$

Hence,
$$\sqrt{n}\frac{\partial R_n(\beta_0)}{\partial\beta} \xrightarrow[d]{} N(0, E[(Y_i-\Lambda(X_i'\beta_0))^2\Lambda(X_i'\beta_0)^2(1-\Lambda(X_i'\beta_0))^2X_iX_i'])$$
$$\underset{LIE}{=} N(0, \underbrace{E[\Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))]^3 X_iX_i'}_{\Sigma_0})$$

Next,
$$\frac{\partial^2 R_n(\beta_0)}{\partial\beta\,\partial\beta'} = \frac{1}{n}\sum_{i=1}^{n}\left\{\underbrace{-(Y_i-\Lambda(X_i'\beta_0))}_{=o_p(1)}\frac{\partial^2\Lambda(X_i'\beta_0)}{\partial\beta\,\partial\beta'} + \frac{\partial\Lambda(X_i'\beta_0)}{\partial\beta}\frac{\partial\Lambda(X_i'\beta_0)}{\partial\beta'}\right\}$$
$$= E\,\frac{\partial\Lambda(X_i'\beta_0)}{\partial\beta}\frac{\partial\Lambda(X_i'\beta_0)}{\partial\beta'} + o_p(1)$$

$$= \underbrace{E\{[\Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))]^2 X_i X_i'\}}_{B_0} + o_p(1)$$

Finally,

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) = -B_0^{-1}\sqrt{n}\,\frac{\partial R_n(\beta_0)}{\partial \beta} + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \underbrace{-B_0^{-1}(Y_i - \Lambda(X_i'\beta_0))\Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))X_i}_{\xi_i^{NLS}} + o_p(1)$$

(d) **(7 points)** Find the asymptotic covariance between the MLE and the NLS estimator. Using the asymptotic covariance, show that the MLE estimator is more efficient than the NLS estimator by comparing their asymptotic variances.

Remember, we stack

we only want this!

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\xi_i^{MLE}\\ \xi_i^{NLS}\end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} E\xi_i^{MLE}\xi_i^{MLE'} & E\xi_i^{MLE}\xi_i^{NLS'} \\ E\xi_i^{NLS}\xi_i^{MLE'} & E\xi_i^{NLS}\xi_i^{NLS'}\end{pmatrix}\right)$$

$$E\xi_i^{MLE}\xi_i^{NLS'} = E\{-\Omega_0^{-1}(Y_i - \Lambda(X_i'\beta_0))X_i\,(Y_i - \Lambda(X_i'\beta_0))\Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))X_i'(-B_0^{-1})\}$$

$$= \Omega_0^{-1}\,E\{(Y_i - \Lambda(X_i'\beta_0))^2\Lambda(X_i'\beta_0)(1-\Lambda(X_i'\beta_0))X_iX_i'\}\,B_0^{-1}$$

$$= \Omega_0^{-1}\,B_0\,B_0^{-1} = \Omega_0^{-1}.$$

⊛ If the covariance = Variance of $\hat{\theta}^A$, then $\hat{\theta}^A$ is the most efficient.

$$E\{(\xi_i^{MLE} - \xi_i^{NLS})(\xi_i^{MLE} - \xi_i^{NLS})'\} = E\xi_i^{MLE}\xi_i^{MLE'} - E\xi_i^{NLS}\xi_i^{MLE'} - E\xi_i^{MLE}\xi_i^{NLS'} + E\xi_i^{NLS}\xi_i^{NLS'}$$

$$= \Omega_0^{-1} - 2\Omega_0^{-1} + \Sigma_0 \leftarrow \text{Avar}(\hat{\beta}_n^{NLS})$$

$$= \Sigma_0 - \Omega_0^{-1} \geq 0 \Leftrightarrow \Sigma_0 \geq \Omega_0^{-1}.$$

(e) **(6 points)** The model can also be viewed as a conditional moment restriction model: for some unique β_0,

$$E\left(Y_i - \Lambda(X_i'\beta_0) \mid X_i\right) = 0.$$

Let $\mathcal{G}$ be the set of measurable k-vector valued functions of X_i. Consider a class of estimators $\{\hat{\beta}_n^g : g \in \mathcal{G}\}$, where $\hat{\beta}_n^g$ is defined as a solution to the following sample moment condition:

$$n^{-1}\sum_{i=1}^{n}\left(Y_i - \Lambda(X_i'\hat{\beta}_n^g)\right) g(X_i) = 0.$$

Show that $\hat{\beta}_n^g$ corresponding to the optimal choice of g is the MLE.

solution:

Recall the efficient IV for a conditional moment restriction

$$E\left(m(y_i, x_i, \beta_0) \mid z_i\right) = 0$$

$$g^*(z_i) = \frac{1}{E\left(m^2(y_i, x_i, \beta_0) \mid z_i\right)} E\left(\frac{\partial m(y_i, x_i, \beta_0)}{\partial \beta} \mid z_i\right)$$

And in this case : $m(y_i, x_i, \beta_0) = y_i - \Lambda(x_i'\beta_0)$
$z_i = x_i$

Then ,

- $E\left(\frac{\partial m(y_i, x_i, \beta_0)}{\partial \beta} \mid x_i\right) = -\Lambda(x_i'\beta_0)\left(1 - \Lambda(x_i'\beta_0)\right) x_i$

- $E\left((y_i - \Lambda(x_i'\beta_0))^2 \mid x_i\right) = \Lambda(x_i'\beta_0)\left(1 - \Lambda(x_i'\beta_0)\right)$

$\Rightarrow \quad g^*(x_i) = x_i$.

Therefore, the efficient IV estimator satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \Lambda(x_i'\hat{\beta}_n^*)\right) x_i = 0$$

which corresponds to the MLE estimator.

⊛ Notice that if $\Lambda(x_i'\beta) = x_i'\beta$ then OLS is doing the efficient IV and it does the same as MLE (attains the efficiency bound).

Jeffrey Wooldridge @jmwooldridge

One of the remarkable features of Bruce's result, and why I never could have discovered it, is that the "asymptotic" analog doesn't seem to hold. Suppose we assume random sampling and in the population specify

A1. E(y|x) = x*b0
A2. Var(y|x) = (s0)^2

#metricstotheface

$E[y_i|x_i] = x_i'\beta_0$

Then efficient IV is x_i,
and optimal F.O.C.:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\hat{\beta}_n)x_i = 0$$

which is OLS.

Jeffrey Wooldridge @jmwooldridge · Feb 12
Replying to @jmwooldridge
Also assume rank E(x'x) = k so no perfect collinearity in the population. Then OLS is asymptotically efficient among estimators that only use A1 for consistency. But OLS is not asymp effic among estimators that use A1 and A2 for consistency.

1 1 10

Jeffrey Wooldridge @jmwooldridge · Feb 12
A2 adds many extra moment conditions that, generally, are useful for estimating b0 -- for example, if D(y|x) is asymmetric with third central moment depending on x. So there are GMM estimators more asymp efficient than OLS under A1 and A2.

1 1 9

Jeffrey Wooldridge @jmwooldridge · Feb 12
With asymp analysis, we see a clear tradeoff between robustness and effic.

Perhaps my choice of asymptotic analogy isn't a good one, or the optimal IVs for b0 collapse to OLS (harder for me to believe). In any case, I still have more to learn ....

1 8

Next time, more about 2-step estimation and over-id testing.