EE, IV and Sample Selection

Empirical Analysis II, Econ 311: Topic 2

Prof. Harald Uhlig[1]

[1]University of Chicago
Department of Economics
huhlig@uchicago.edu

Winter 2019

Outline

Outline

1. Extremum Estimators (EE)
   - Extremum Estimators (EE)
2. Instrumental Variables (IV)
   - Instrumental Variables (IV)
3. Sample Selection
   - Sample Selection

Extremum Estimators

- Some material: Hayashi, Econometrics, Princeton Univ. Press (2000).
- $\hat{\theta}_n$ is called an extremum estimator, if it solves

$$\hat{\theta}_n \in \text{argmax}_{\theta \in \Theta \subseteq \mathbb{R}^m} Q_n(\theta)$$

  for some objective $Q_n(\theta)$.

Examples for Extremum Estimators

- Maximum Likelihood (ML), Nonlinear Least Squares (NLLS), M-Estimator / Moment-Estimator (M).

$$\text{ML:} \quad Q_n(\theta) = \frac{1}{n}\sum_{j=1}^{n} \ell(\theta \mid y_j)$$

$$\text{NLLS:} \quad Q_n(\theta) = -\frac{1}{n}\sum_{j=1}^{n} (y_j - h(X_j;\theta))^2$$

$$\text{M:} \quad Q_n(\theta) = \frac{1}{n}\sum_{j=1}^{n} m(y_j;\theta)$$

- Generalized Method of Moments (GMM).

$$\text{GMM:} \quad Q_n(\theta) = -\frac{1}{2} g_n(\theta)' \hat{W}_n g_n(\theta)$$

$$\text{where} \quad g_n(\theta) = \frac{1}{n}\sum_{j=1}^{n} g(y_j;\theta)$$

Consistency

Theorem

Suppose, $Q_n, n = 0, 1, 2 \ldots$ are continuous functions of $\theta \in \Theta$, Θ compact. Suppose

- *Identification:* $\theta_0 = \text{argmax}_{\theta\in\Theta} Q_0(\theta)$ *is unique.*
- *Uniform convergence:* $Q_n(\cdot)$ *converges uniformly in probability to* $Q_0(\cdot)$, *i.e.*

$$\sup_{\theta\in\Theta} \mid Q_n(\theta) - Q_0(\theta) \mid \xrightarrow{P} 0 \text{ as } n \to \infty$$

Then $\hat{\theta}_n \xrightarrow{P} \theta_0$ *(consistency).*

M-estimators: differentiating.

- Assume twice differentiability. Truth: $\theta = \theta_0$.
- Define

$$\text{Score: } s(y_j;\theta) = \frac{\partial m(y_j;\theta)}{\partial \theta}, \quad \text{Hessian: } H(y_j;\theta) = \frac{\partial^2 m(y_j;\theta)}{\partial \theta \, \partial \theta'}$$

- Similar to MLE calculations,

$$\begin{aligned}
Q_n(\theta) &= \tfrac{1}{n}\textstyle\sum_{j=1}^n m(y_j;\theta) &\xrightarrow{P} E_{\theta_0}[m(y;\theta)] = Q_0(\theta) \\
s_n(\theta) = \frac{\partial Q_n(\theta)}{\partial \theta} &= \tfrac{1}{n}\textstyle\sum_{j=1}^n s(y_j;\theta) &\xrightarrow{P} E_{\theta_0}[s(y;\theta)] \\
H_n(\theta) = \frac{\partial^2 Q_n(\theta)}{\partial \theta \, \partial \theta'} &= \tfrac{1}{n}\textstyle\sum_{j=1}^n H(y_j;\theta) &\xrightarrow{P} E_{\theta_0}[H(y;\theta_0)] = -\Psi \text{ at } \theta = \theta_0
\end{aligned}$$

- Assume: y_j correlated across "nearby" j' ("ergodicity"). Then, $\sqrt{n} s_n(\theta_0) \xrightarrow{d} \mathcal{N}(0,\Sigma)$, where long-run variance Σ of $s(y_j;\theta_0)$ is

$$\Sigma = \sum_{k=-\infty}^{\infty} \Gamma_k \text{ , where } \Gamma_k = E[s(y_j;\theta_0)\, s(y_{j+k};\theta_0)']$$

Asymptotics for the M-Estimator: Delta method

- The M-Estimator $\hat{\theta}_n$ solves $\frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta} = s_n(\hat{\theta}_n) = 0$.
- First-order expansion around θ_0:

$$0 = s_n(\hat{\theta}_n) \approx s_n(\theta_0) + H_n(\theta_0)(\hat{\theta}_n - \theta_0)$$

- Assume Ψ is invertible (hence: positive definite).

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx -\sqrt{n}\,\Psi^{-1} H_n(\theta_0)(\hat{\theta}_n - \theta_0) \approx \sqrt{n}\,\Psi^{-1} s_n(\theta_0)$$

- Take the limit.

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

- If $\Sigma = \Psi$: $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \Psi^{-1}\right)$.
- Keep in mind: $\Psi = \Psi(\theta_0)$, $\Sigma = \Sigma(\theta_0)$ depend on θ_0.

GMM-estimators: differentiating.

- Assume twice differentiability, ergodicity. Further,

  Assume: $\hat{W}_n \xrightarrow{P} \mathcal{W}$

  Define: $\mathcal{S} = \sum_{k=-\infty}^{\infty} \Gamma_k$ (long-run variance of $g(y_j;\theta_0)$)

  where: $\Gamma_k = E[g(y_j;\theta_0)\,g(y_{j+k};\theta_0)']$ (ass.: $E[g(y,\theta_0)] = 0$)

  Define: $G = E[\frac{\partial g(y;\theta_0)}{\partial \theta'}]$

- Differentiate:

$$Q_n(\theta) = -\frac{1}{2} g_n(\theta)'\,\hat{W}_n\,g_n(\theta) \quad \text{with } g_n(\theta) = \frac{1}{n}\sum_{j=1}^n g(y_j;\theta)$$

$$G_n(\theta) = \frac{\partial g_n(\theta)}{\partial \theta'} = \frac{1}{n}\sum_{j=1}^n \frac{\partial g(y_j;\theta)}{\partial \theta'} \xrightarrow{P} G \text{ at } \theta = \theta_0$$

$$s_n(\theta) = \frac{\partial Q_n(\theta)}{\partial \theta} = -G_n(\theta)'\,\hat{W}_n\,g_n(\theta)$$

$$\text{so: } \sqrt{n} s_n(\theta_0) \xrightarrow{d} \mathcal{N}(0,\Sigma) \quad \text{where } \Sigma = G'\mathcal{W}\mathcal{S}\mathcal{W}G$$

Asymptotics for the GMM-Estimator: Delta method

- The GMM-Estimator $\hat{\theta}_n$ solves $\frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta} = s_n(\hat{\theta}_n) = 0$.
- First-order expansion of $g_n(\theta)$ around θ_0:

$$0 = s_n(\hat{\theta}_n) = -G_n(\hat{\theta}_n)' \hat{W}_n \, g_n(\hat{\theta}_n) \approx s_n(\theta_0) - G_n(\hat{\theta}_n)' \hat{W}_n \, G_n(\hat{\theta}_0)(\hat{\theta}_n - \theta_0)$$

exploiting $-G_n(\hat{\theta}_n)' \hat{W}_n \, g_n(\theta_0) \approx -G_n(\theta_0)' \hat{W}_n \, g_n(\theta_0) = s_n(\theta_0)$.

- Note: $G_n(\hat{\theta}_n)' \hat{W}_n \, G_n(\hat{\theta}_0) \xrightarrow{P} G' \mathcal{W} G = \Psi$.
- Assume Ψ is invertible (hence: positive definite).

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx \sqrt{n}\, \Psi^{-1} s_n(\theta_0)$$

- Take the limit.

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

- A good choice: $\mathcal{W} = \mathcal{S}^{-1}$. Then $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \Psi^{-1}\right)$.

Summary and Testing

1. $\sqrt{n}s_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$**.**
2. $\sqrt{n}(\hat{\theta}_n - \theta_0) = \Psi^{-1}\sqrt{n}s_n(\theta_0) + o_P \xrightarrow{d} \mathcal{N}(0, \Psi^{-1}\Sigma\Psi^{-1})$
3. For MLE, for GMM with $\mathcal{W} = \mathcal{S}^{-1}$**:** **(**$*$**)** $\Psi = \Sigma$

Constrained estimation: $\hat{\theta}_{c,n} = \text{argmax}_{\theta\in\Theta} Q_n$ s.t. $a(\theta) = 0$, where $\partial a(\theta_0)/\partial\theta$ has rank k. Assume ($*$). Assume $\hat{\Psi}_n \xrightarrow{P} \Psi$. Three tests:

1. Likelihood-ratio test: (note: abuse of language, if not MLE)

$$LR = 2n * (Q_n(\hat{\theta}_n) - Q_n(\hat{\theta}_{c,n})) \xrightarrow{d} \chi^2_k$$

2. Score test or Lagrange multiplier test or Rao test:

$$LM = n\, s_n(\hat{\theta}_{c,n})'\, \hat{\Psi}_n^{-1}\, s_n(\hat{\theta}_{c,n}) \xrightarrow{d} \chi^2_k$$

3. Wald test: Define $A_n = \partial a(\hat{\theta}_n)/\partial\theta \xrightarrow{P} A = \partial a(\theta_0)/\partial\theta$

$$W = n\, a(\hat{\theta}_n)' \left(A_n\, \hat{\Psi}_n^{-1}\, A_n'\right)^{-1} a(\hat{\theta}_n) \xrightarrow{d} \chi^2_k$$

Summary and Testing

1. $\sqrt{n} s_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$.
2. $\sqrt{n}(\hat{\theta}_n - \theta_0) = \Psi^{-1}\sqrt{n} s_n(\theta_0) + o_P \xrightarrow{d} \mathcal{N}(0, \Psi^{-1}\Sigma\Psi^{-1})$
3. **For MLE, for GMM with $\mathcal{W} = \mathcal{S}^{-1}$: (*) $\Psi = \Sigma$**

Constrained estimation: $\hat{\theta}_{c,n} = \text{argmax}_{\theta \in \Theta} Q_n$ s.t. $a(\theta) = 0$, where $\partial a(\theta_0)/\partial\theta$ has rank k. Assume (*). Assume $\hat{\Psi}_n \xrightarrow{P} \Psi$. Three tests:

1. Likelihood-ratio test: (note: abuse of language, if not MLE)

$$LR = 2n * (Q_n(\hat{\theta}_n) - Q_n(\hat{\theta}_{c,n})) \xrightarrow{d} \chi^2_k$$

2. Score test or Lagrange multiplier test or Rao test:

$$LM = n\, s_n(\hat{\theta}_{c,n})'\, \hat{\Psi}_n^{-1}\, s_n(\hat{\theta}_{c,n}) \xrightarrow{d} \chi^2_k$$

3. Wald test: Define $A_n = \partial a(\hat{\theta}_n)/\partial\theta \xrightarrow{P} A = \partial a(\theta_0)/\partial\theta$

$$W = n\, a(\hat{\theta}_n)' \left(A_n\, \hat{\Psi}_n^{-1}\, A_n'\right)^{-1} a(\hat{\theta}_n) \xrightarrow{d} \chi^2_k$$

Outline

The Instrumental Variables (IV) Estimator

- Linear regression. Suppose X and ϵ are correlated in

$$Y = X\beta + \epsilon$$

- OLSE: biased. $E[\hat{\beta}] = \beta + E[(X'X)^{-1}X'\epsilon]$.
- Instruments: Z, correlated with X, but uncorrelated with ϵ. Suppose: same number of variables as X.
- Calculate

$$\begin{aligned} Z'Y &= Z'X\beta + Z'\epsilon \\ \hat{\beta}_{IV} &= (Z'X)^{-1}Z'Y \end{aligned}$$

- For estimating the variance σ^2 of ϵ, use $\hat{\epsilon} = Y - X\hat{\beta}_{IV}$
- For iid observations:

$$\begin{aligned} \widehat{\sigma^2} &= \frac{1}{n}\hat{\epsilon}'\hat{\epsilon} \\ \widehat{\text{Var}}(\beta_{IV}) &= \widehat{\sigma^2}(Z'X)^{-1}Z'Z(X'Z)^{-1} \end{aligned}$$

The Instrumental Variables (IV) Estimator

- Linear regression. Suppose X and ϵ are correlated in

$$Y = X\beta + \epsilon$$

- OLSE: biased. $E[\hat{\beta}] = \beta + E[(X'X)^{-1}X'\epsilon]$.
- Instruments: Z, correlated with X, but uncorrelated with ϵ. Suppose: same number of variables as X.
- Calculate

$$\begin{aligned} Z'Y &= Z'X\beta + Z'\epsilon \\ \hat{\beta}_{IV} &= (Z'X)^{-1}Z'Y \end{aligned}$$

- For estimating the variance σ^2 of ϵ, use $\hat{\epsilon} = Y - X\hat{\beta}_{IV}$
- For iid observations:

$$\begin{aligned} \widehat{\sigma^2} &= \frac{1}{n}\hat{\epsilon}'\hat{\epsilon} \\ \widehat{\text{Var}}(\beta_{IV}) &= \widehat{\sigma^2}(Z'X)^{-1}Z'Z(X'Z)^{-1} \end{aligned}$$

The Instrumental Variables (IV) Estimator

- Linear regression. Suppose X and ϵ are correlated in

$$Y = X\beta + \epsilon$$

- OLSE: biased. $E[\hat{\beta}] = \beta + E[(X'X)^{-1}X'\epsilon]$.
- Instruments: Z, correlated with X, but uncorrelated with ϵ. Suppose: same number of variables as X.
- Calculate

$$\begin{aligned} Z'Y &= Z'X\beta + Z'\epsilon \\ \hat{\beta}_{IV} &= (Z'X)^{-1}Z'Y \end{aligned}$$

- For estimating the variance σ^2 of ϵ, use $\hat{\epsilon} = Y - X\hat{\beta}_{IV}$
- For iid observations:

$$\begin{aligned} \widehat{\sigma^2} &= \frac{1}{n}\hat{\epsilon}'\hat{\epsilon} \\ \widehat{\text{Var}}(\beta_{IV}) &= \widehat{\sigma^2}(Z'X)^{-1}Z'Z(X'Z)^{-1} \end{aligned}$$

The Instrumental Variables (IV) Estimator

- Linear regression. Suppose X and ϵ are correlated in

$$Y = X\beta + \epsilon$$

- OLSE: biased. $E[\hat{\beta}] = \beta + E[(X'X)^{-1}X'\epsilon]$.
- Instruments: Z, correlated with X, but uncorrelated with ϵ. Suppose: same number of variables as X.
- Calculate

$$\begin{aligned} Z'Y &= Z'X\beta + Z'\epsilon \\ \hat{\beta}_{IV} &= (Z'X)^{-1}Z'Y \end{aligned}$$

- For estimating the variance σ^2 of ϵ, use $\hat{\epsilon} = Y - X\hat{\beta}_{IV}$
- For iid observations:

$$\begin{aligned} \widehat{\sigma^2} &= \frac{1}{n}\hat{\epsilon}'\hat{\epsilon} \\ \widehat{\mathrm{Var}}(\beta_{IV}) &= \widehat{\sigma^2}(Z'X)^{-1}Z'Z(X'Z)^{-1} \end{aligned}$$

The Two-Stage Least Squares (2SLS) Estimator

- Linear regression. Suppose X and ϵ are correlated in

$$Y = X\beta + \epsilon$$

- Instruments: Z, correl. with X, uncorrel. with ϵ, $\dim(Z) \geq \dim(X)$.
- Replace X with $\tilde{X} = E[X \mid Z]$, i.e. run two regressions:

$$\begin{aligned} 1. \quad X &= Z\alpha + \nu \\ \text{thus:} \quad \tilde{X} &= Z\hat{\alpha} \\ 2. \quad Y &= \tilde{X}\beta + \tilde{\epsilon} \end{aligned}$$

- OLSE: [Remark: for "same number of variables", same as above!]

$$\begin{aligned} \tilde{X} &= Z(Z'Z)^{-1}Z'X = P_Z X \\ \hat{\beta}_{IV} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y \end{aligned}$$

- But: for estimating the variance of ϵ,

use: $\hat{\epsilon} = Y - X\hat{\beta}_{IV}$, not: $\hat{\tilde{\epsilon}} = Y - \tilde{X}\hat{\beta}_{IV}$

The Two-Stage Least Squares (2SLS) Estimator

- Linear regression. Suppose X and ϵ are correlated in

$$Y = X\beta + \epsilon$$

- Instruments: Z, correl. with X, uncorrel. with ϵ, $\dim(Z) \geq \dim(X)$.
- Replace X with $\tilde{X} = E[X \mid Z]$, i.e. run two regressions:

$$\begin{aligned} \text{1.} \quad X &= Z\alpha + \nu \\ \text{thus:} \quad \tilde{X} &= Z\hat{\alpha} \\ \text{2.} \quad Y &= \tilde{X}\beta + \tilde{\epsilon} \end{aligned}$$

- OLSE: [Remark: for "same number of variables", same as above!]

$$\begin{aligned} \tilde{X} &= Z(Z'Z)^{-1}Z'X = P_Z X \\ \hat{\beta}_{IV} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y \end{aligned}$$

- But: for estimating the variance of ϵ,

$$\text{use: } \hat{\epsilon} = Y - X\hat{\beta}_{IV}, \quad \text{not: } \hat{\tilde{\epsilon}} = Y - \tilde{X}\hat{\beta}_{IV}$$

The Two-Stage Least Squares (2SLS) Estimator

- Linear regression. Suppose X and ϵ are correlated in

$$Y = X\beta + \epsilon$$

- Instruments: Z, correl. with X, uncorrel. with ϵ, $\dim(Z) \geq \dim(X)$.
- Replace X with $\tilde{X} = E[X \mid Z]$, i.e. run two regressions:

$$\begin{aligned} \text{1.} \quad X &= Z\alpha + \nu \\ \text{thus:} \quad \tilde{X} &= Z\hat{\alpha} \\ \text{2.} \quad Y &= \tilde{X}\beta + \tilde{\epsilon} \end{aligned}$$

- OLSE: [Remark: for "same number of variables", same as above!]

$$\begin{aligned} \tilde{X} &= Z(Z'Z)^{-1}Z'X = P_Z X \\ \hat{\beta}_{IV} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y \end{aligned}$$

- But: for estimating the variance of ϵ,

$$\text{use: } \hat{\epsilon} = Y - X\hat{\beta}_{IV}, \quad \text{not: } \hat{\tilde{\epsilon}} = Y - \tilde{X}\hat{\beta}_{IV}$$

IV and GMM

- Consider iid case. Z_t, t=1,...,T are uncorrelated with ϵ_t:

$$E[Z_t(y_t - X_t\beta)] = 0$$

- "Moment condition".
- This can be generalized:

$$g([X_t, Z_t]; \theta) = E[Z_t' f(X_t; \theta)] = 0$$

- Example:
  - Asset pricing: $0 = E_t[\beta \left(\frac{C_t}{C_{t+1}}\right)^\eta R_{t+1} - 1]$
  - Let $X_t = [C_{t+1}, C_t, R_{t+1}]$. Let $\theta = [\beta, \eta]$.
  - Use data known in $t-1$ and earlier as instruments Z_t.
- Find $\hat{\theta}$ per GMM for some suitable weighting matrix $\hat{W}_t$.

Outline

Type-2-Tobit

- some material: Cameron-Trivedi, *Microeconometrics*, Cambridge University Press, 2005
- Example: hours worked depend on wage, provided the agent chooses to work at all. That decision ("probit") depends on something else.
- Standard specification: linear.

$$\begin{aligned} y_1^* &= X_1\beta_1 + \epsilon_1 \\ y_2^* &= X_2\beta_2 + \epsilon_2 \end{aligned}$$

  Observe: $X = [X_1, X_2]$. Observe $y_2 = y_2^*$, iff $y_1^* > 0$.
- Assume: $E[\epsilon_2 \mid \epsilon_1] = \rho\epsilon_1$.
- OLSE of y_2 on X_2 has a sample selection bias, since

$$\begin{aligned} E[y_2 \mid X, y_1^* > 0] &= X_2\beta_2 + E[\epsilon_2 \mid y_1^* > 0] \\ &= X_2\beta_2 + \rho E[\epsilon_1 \mid \epsilon_1 > -X_1\beta_1] \end{aligned}$$

The Heckit estimator. Heckman's two-step procedure

- Special case: $\epsilon_1 \sim \mathcal{N}(0,1)$. Let ϕ: pdf, Φ: cdf. Calculate

$$\int_c^\infty \frac{\epsilon_1}{\sqrt{2\pi}} e^{-\epsilon_1^2/2} d\epsilon_1 = \int_c^\infty \frac{d}{d\epsilon_1}\left(-\frac{1}{\sqrt{2\pi}} e^{-\epsilon_1^2/2}\right) d\epsilon_1 = \phi(c)$$

  Therefore,

$$E[\epsilon_1 \mid \epsilon_1 > -X_1\beta] = \frac{\phi(-X_1\beta)}{1-\Phi(-X_1\beta)} = \frac{\phi(X_1\beta)}{\Phi(X_1\beta)} = \lambda(X_1\beta_1)$$

  $\lambda(X_1\beta_1)$ is the inverse Mills ratio. [Remark: slightly different from definition in "Topic 1"].

- Therefore:

$$E[y_2 \mid X, y_1^* > 0] = X_2\beta_2 + \rho\,\lambda(X_1\beta_1)$$

- The Heckit estimator per Heckman's two-step procedure:
  1. Probit: estimate $\hat{\beta}_1$.
  2. OLSE of $y_2 = X_2\beta_2 + \rho\,\lambda(X_1\hat{\beta}_1) + \nu_2$

The Heckit estimator. Heckman's two-step procedure

- Special case: $\epsilon_1 \sim \mathcal{N}(0,1)$. Let ϕ: pdf, Φ: cdf. Calculate

$$\int_c^\infty \frac{\epsilon_1}{\sqrt{2\pi}} e^{-\epsilon_1^2/2} d\epsilon_1 = \int_c^\infty \frac{d}{d\epsilon_1}\left(-\frac{1}{\sqrt{2\pi}} e^{-\epsilon_1^2/2}\right) d\epsilon_1 = \phi(c)$$

  Therefore,

$$E[\epsilon_1 \mid \epsilon_1 > -X_1\beta] = \frac{\phi(-X_1\beta)}{1-\Phi(-X_1\beta)} = \frac{\phi(X_1\beta)}{\Phi(X_1\beta)} = \lambda(X_1\beta_1)$$

  $\lambda(X_1\beta_1)$ is the inverse Mills ratio. [Remark: slightly different from definition in "Topic 1"].
- Therefore:

$$E[y_2 \mid X, y_1^* > 0] = X_2\beta_2 + \rho\,\lambda(X_1\beta_1)$$

- The Heckit estimator per Heckman's two-step procedure:
  1. Probit: estimate $\hat{\beta}_1$.
  2. OLSE of $y_2 = X_2\beta_2 + \rho\,\lambda(X_1\hat{\beta}_1) + \nu_2$