Generalized Method of Moments

These transparencies were prepared in 2002.

They can be used as an adjunct to

Chapter 6 of our subsequent book

Microeconometrics: Methods and Applications

Cambridge University Press, 2005.

Original version of slides: October 2002

OUTLINE

- Introduction.
- Examples: OLS, IV
- Theory:
- Linear IV: GMM and 2SLS
- IV in practice (weak instruments)
- Nonlinear IV: NL2SLS
- Linear and Nonlinear Sets of Equations: SUR, 3SLS, panel
- Two-Step Estimators and Empirical Likelihood

1 Introduction

- GMM is generalization of method of moments

- Example is estimation of μ for y i.i.d.

- Population moment condition

$$\mathsf{E}[y-\mu]=0.$$

- Sample moment condition:

$$\frac{1}{N}\sum_{i=1}^{N}(y_i-\mu)=0.$$

- Solving yields MM estimator

$$\widehat{\mu}=\bar{y}.$$

Introduction (continued)

- More generally
  a **population moment condition** for $\boldsymbol{\theta}$
  leads to
  a corresponding **sample moment condition** for $\boldsymbol{\theta}$
  which we **solve** for $\boldsymbol{\theta}$.

- What if nonlinear in $\boldsymbol{\theta}$? Nonlinear MM.

- What if more moment conditions than components of $\boldsymbol{\theta}$? GMM.

- What is the best moment condition to start with? Optimal GMM.

2 GMM Examples: OLS

- Population **conditional** moment condition

$$\mathsf{E}[u_i|\mathbf{x}_i] = \mathsf{E}[y_i - \mathbf{x}_i'\boldsymbol{\beta}|\mathbf{x}_i] = 0.$$

- Population **unconditional** moment condition

$$\mathsf{E}[\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{x}_i] = \mathbf{0}.$$

- Sample moment condition

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{x}_i = \mathbf{0}.$$

- Solving yields OLS estimator

$$\widehat{\boldsymbol{\beta}}_{\mathsf{OLS}} = \left(\sum_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\sum_i \mathbf{x}_i y_i$$

Regression with Symmetric Errors

- Population **conditional** moment condition

$$\begin{aligned}\mathsf{E}[u_i|\mathbf{x}_i] &= \mathsf{E}[y_i - \mathbf{x}_i'\boldsymbol{\beta}|\mathbf{x}_i] = 0 \\ \mathsf{E}[u_i^3|\mathbf{x}_i] &= \mathsf{E}[\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^3|\mathbf{x}_i] = 0.\end{aligned}$$

- Population **unconditional** moment condition

$$\mathsf{E}\left[\begin{array}{c}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{x}_i \\ \left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^3\mathbf{x}_i\end{array}\right] = \left[\begin{array}{c}\mathbf{0} \\ \mathbf{0}\end{array}\right].$$

- There are $2K$ moment conditions and only K parameters, so cannot solve for $\boldsymbol{\beta}$.

- Instead GMM minimizes quadratic form

$$\begin{aligned}&\left[\begin{array}{cc}\frac{1}{N}\sum_{i=1}^N\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{x}_i' & \frac{1}{N}\sum_{i=1}^N\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^3\mathbf{x}_i'\end{array}\right] \\ &\times\mathbf{W}_N\times\left[\begin{array}{c}\frac{1}{N}\sum_{i=1}^N\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{x}_i \\ \frac{1}{N}\sum_{i=1}^N\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^3\mathbf{x}_i\end{array}\right].\end{aligned}$$

Maximum Likelihood

- Population moment condition

$$\mathsf{E}\left[\frac{\partial \ln f(y|\mathbf{x},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}.$$

- Sample moment condition

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln f(y_i|\mathbf{x}_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}.$$

- Solving yields the MLE.

Instrumental Variables (IV)

- Population **conditional** moment condition

$$\mathsf{E}[u_i|\mathbf{z}_i] = 0 \Rightarrow \mathsf{E}[y_i - \mathbf{x}_i'\boldsymbol{\beta}|\mathbf{z}_i] = 0.$$

- Population **unconditional** moment condition

$$\mathsf{E}[\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{z}_i] = \mathbf{0}.$$

- Sample moment condition

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{z}_i = \mathbf{0}.$$

- If dim($\mathbf{z}$)=dim($\mathbf{x}$) can solve to obtain IV estimator

$$\widehat{\boldsymbol{\beta}}_{\mathsf{IV}} = \left(\sum_i \mathbf{z}_i\mathbf{x}_i'\right)^{-1}\sum_i \mathbf{z}_i y_i.$$

Two-Stage Least Squares

- If dim($\mathbf{z}$)>dim($\mathbf{x}$) cannot solve for $\boldsymbol{\beta}$.

- Instead GMM minimizes quadratic form

$$\left[\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{z}_i'\right]\times\mathbf{W}_N\times\left[\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{z}_i'\right].$$

- The choice $\mathbf{W}_N = \left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right]^{-1}$ is optimal if errors are independent and homoskedastic.

- This is generalized IV or two-stage least squares (though no "two-stage" motivation here).

Structural Models (Hansen)

- Maximize expected PDV of lifetime utility $\mathsf{E}\left[\sum_{t=0}^{\infty} \beta^t U(C_t) \quad | \mathcal{I}_0\right]$ with budget constraint with labor and asset income.

- Euler equation with constant RR aversion utility

$$\mathsf{E}\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{\alpha} \frac{(P_{t+1}+D_{t+1})}{P_t} - 1 \middle| \mathcal{I}_t \quad \right] = 0,$$

where $\mathcal{I}_t$ is information set at time t.

- GMM estimator using time series data

$$\left[\frac{1}{T}\sum_{t=1}^{T}\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{\alpha} \frac{(P_{t+1}+D_{t+1})}{P_t} - 1\right)\mathbf{z}_t'\right]$$
$$\mathbf{W}_T\left[\frac{1}{T}\sum_{t=1}^{T}\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{\alpha} \frac{(P_{t+1}+D_{t+1})}{P_t} - 1\right)\mathbf{z}_t\right]$$

where $\mathbf{z}_t \in \mathcal{I}_t$. e.g. C_t/C_{t-1}, $(P_t + D_t)/P_{t-1}$.

3 Theory

- Population **unconditional** moment condition

$$\mathsf{E}[\mathbf{h}(\mathbf{w}, \boldsymbol{\theta}_0)] = \mathbf{0},$$

where $\mathbf{w} = (\mathbf{y}, \mathbf{x}, \mathbf{z})$ is all observables.

- Sample moment condition

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{h}(\mathbf{w}_{ii}, \widehat{\boldsymbol{\theta}}) = \mathbf{0}.$$

- If dim($\mathbf{h}$)=dim($\boldsymbol{\theta}$) can solve (numerically if not analytically) to obtain **method of moments estimator.**

- If dim($\mathbf{h}$)>dim($\boldsymbol{\theta}$) then do GMM.

Theory (continued)

- GMM minimizes

$$Q_N(\theta) = \left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{h}(\mathbf{w}_i,\boldsymbol{\theta})\right]' \mathbf{W}_N \left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{h}(\mathbf{w}_i,\boldsymbol{\theta})\right]$$

- Equivalently where $r = \mathbf{dim}(\mathbf{h})$

$$\sum_{j=1}^{r}\sum_{k=1}^{r} W_{N,jk}\left(\frac{1}{N}\sum_{i=1}^{N} h_j(\mathbf{w}_i,\boldsymbol{\theta})\right)\left(\frac{1}{N}\sum_{i=1}^{N} h_j(\mathbf{w}_i,\boldsymbol{\theta})\right)$$

- Equivalently when $\mathbf{W}_N$ is an identity matrix

$$\left(\frac{1}{N}\sum_{i=1}^{N} h_1(\mathbf{w}_i,\boldsymbol{\theta})\right)^2 + \cdots + \left(\frac{1}{N}\sum_{i=1}^{N} h_r(\mathbf{w}_i,\boldsymbol{\theta})\right)^2.$$

Theory (continued)

Similar issues as for weighted LS in the linear model.

- Model choice entails specification of moment conditions that are basis for estimation.

- Estimator choice entails specification of a weighting function.

- Statistical inference is based on robust standard errors that do not assume the weighting function to be the optimal weighting function.

- Leads to expression for variance of GMM estimator qualitatively similar to that for the WLS estimator.

Asymptotic Distribution (continued)

- First-order conditions re-scaled

$$\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \mathbf{h}_i(\widehat{\boldsymbol{\theta}})'}{\partial \boldsymbol{\theta}}\bigg|_{\widehat{\boldsymbol{\theta}}}\right]\mathbf{W}_N\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{h}_i(\widehat{\boldsymbol{\theta}})\right]=\mathbf{0}.$$

- First-order Taylor series of third term

$$\widehat{\mathbf{G}}'\mathbf{W}_N\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{h}_i(\boldsymbol{\theta}_0)+\mathbf{G}(\boldsymbol{\theta}^+)'\sqrt{N}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right]=\mathbf{0}.$$

- Solving

$$\begin{aligned}
&\sqrt{N}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\\
=\;&\left[\widehat{\mathbf{G}}'\mathbf{W}_N\mathbf{G}(\boldsymbol{\theta}^+)\right]^{-1}\widehat{\mathbf{G}}'\mathbf{W}_N\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{h}_i(\boldsymbol{\theta}_0)\\
\to\;&\left[\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0)\right]^{-1}\mathbf{G}_0'\mathbf{W}_0\times \mathsf{N}\left[\mathbf{0},\mathbf{S}_0\right]
\end{aligned}$$

Asymptotic Distribution (continued)

· So

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{\mathsf{GMM}} - \boldsymbol{\theta}_0) \xrightarrow{d}$$
$$\mathcal{N}[\mathbf{0}, (\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0)^{-1}(\mathbf{G}_0'\mathbf{W}_0\mathbf{S}_0\mathbf{W}_0\mathbf{G}_0)(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0)^{-1}$$

- where $\mathbf{W}_0 =$plim$\mathbf{W}_N$ and

$$\begin{aligned}\mathbf{G}_0 &= \lim \frac{1}{N}\sum_{i=1}^{N} \mathsf{E}\left[\left.\frac{\partial \mathbf{h}_i}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0}\right] \\ \mathbf{S}_0 &= \lim \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathsf{E}\left[\left.\mathbf{h}_i\mathbf{h}_j'\right|_{\boldsymbol{\theta}_0}\right]\end{aligned}$$

- and to implement use $\mathbf{W}_N$ and for i.i.d. case

$$\begin{aligned}\widehat{\mathbf{G}} &= \frac{1}{N}\sum_{i=1}^{N} \left.\frac{\partial \mathbf{h}_i}{\partial \boldsymbol{\theta}'}\right|_{\widehat{\boldsymbol{\theta}}} \\ \widehat{\mathbf{S}} &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{h}_i(\widehat{\boldsymbol{\theta}})\mathbf{h}_i(\widehat{\boldsymbol{\theta}})'.\end{aligned}$$

Optimal GMM

- (1) Optimal weighting matrix (for $r > q$).
  For given choice of $\mathbf{h}(\mathbf{w}, \boldsymbol{\theta}_0)$ use
  $\mathbf{W}_N = \widehat{\mathbf{S}}^{-1}$ where $\widehat{\mathbf{S}}$ is consistent for $\mathbf{S}_0$. Then

  $$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{\mathsf{GMM}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}[\mathbf{0}, (\mathbf{G}_0' \mathbf{W}_0 \mathbf{G}_0)^{-1}].$$

  This is usually what people call optimal GMM.

- Result (1) is routinely used.
  One step GMM uses $\mathbf{W}_N = \mathbf{I}$.
  Two step GMM uses $\mathbf{W}_N = \widehat{\mathbf{S}}^{-1}$.

  Often one-step does better - see Ziliak (1997).

Optimal GMM (continued)

- (2) Optimal moment condition.
  The best choice of $\mathbf{h}(\mathbf{w}, \boldsymbol{\theta}_0)$ is that corresponding to the MLE so
  $$\mathbf{h}(\mathbf{w}, \boldsymbol{\theta}_0) = \frac{\partial \ln f(y|\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$
  Requires specification of conditional density.

Optimal GMM (continued)

- (3) Optimal moment condition for given choice of conditional moment condition. For
$$\mathsf{E}[\boldsymbol{\rho}(y, \mathbf{x}, \boldsymbol{\theta}_0)|\mathbf{z}] = \mathbf{0}$$
the best unconditional moment is
$$\mathsf{E}[\mathbf{D}^*(\mathbf{z}, \boldsymbol{\theta}_0)\boldsymbol{\rho}(y, \mathbf{x}, \boldsymbol{\theta}_0)] = \mathbf{0},$$
where
$$\begin{aligned}\mathbf{D}^*(\mathbf{z}, \boldsymbol{\theta}) &= \mathsf{E}\left[\frac{\partial \boldsymbol{\rho}(y, \mathbf{x}, \boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}|\mathbf{z}\right] \\ &\quad \left\{\mathsf{E}\left[\boldsymbol{\rho}(y, \mathbf{x}, \boldsymbol{\theta})\boldsymbol{\rho}(y, \mathbf{x}, \boldsymbol{\theta})'|\mathbf{z}\right]\right\}^{-1}.\end{aligned}$$
Requires specification of cond. variance of $\boldsymbol{\rho}(y, \mathbf{x}, \boldsymbol{\theta}_0)$.

- Example is that efficient LS based on $\mathsf{E}[u|\mathbf{x}] = \mathbf{0}$ but with heteroskedastic error is GLS with
$$\mathsf{E}\left[\frac{\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{x}_i}{\mathsf{Var}[y_i|\mathbf{x}_i]}\right] = 0.$$

Test of Overidentifying Restrictions

- Test H_0 :E$[\mathbf{h}(\mathbf{w}, \boldsymbol{\theta}_0)] = \mathbf{0}$.

- Obvious is to test if $N^{-1}\sum_i \mathbf{h}_i(\mathbf{w}_i, \widehat{\boldsymbol{\theta}}) \simeq \mathbf{0}$.

- When $r = q$, estimation imposes $N^{-1}\sum_i \mathbf{h}_i(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$ and no test is possible.

- When $r > q$, if optimal weighting matrix $\widehat{\mathbf{S}}^{-1}$ is used then use

$$\tau_N = \left(\frac{1}{\sqrt{N}} \sum_i \mathbf{h}_i \left(\widehat{\boldsymbol{\theta}}^{\text{opt}} \right) \right)' \widehat{\mathbf{S}}^{-1} \left(\frac{1}{\sqrt{N}} \sum_i \mathbf{h}_i \left(\widehat{\boldsymbol{\theta}}^{\text{opt}} \right) \right),$$

Reject H_0 if $\tau_N > \chi^2_\alpha(r - q)$.

4 Linear IV: GMM

- Leading GMM example where
  $\#$ moment conditions $>$ $\#$ parameters.
  But algebra is very lengthy.

- For $u_i = y_i - \mathbf{x}_i\boldsymbol{\beta}$ minimize

$$Q_N(\boldsymbol{\beta}) = \left[\frac{1}{N}\sum_{i=1}^{N} u_i\mathbf{z}_i'\right] \mathbf{W_N} \left[\frac{1}{N}\sum_{i=1}^{N} u_i\mathbf{z}_i\right]$$

- In matrix algebra minimize

$$\begin{aligned} Q_N(\boldsymbol{\beta}) &= \mathbf{u}'\mathbf{Z}\mathbf{W_N}\mathbf{Z}'\mathbf{u} \\ &= (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{Z}\mathbf{W_N}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

- Can solve f.o.c. (not given) to get

$$\widehat{\boldsymbol{\beta}}_{\mathsf{GMM}} = \left[\mathbf{X}'\mathbf{Z}\mathbf{W_N}\mathbf{Z}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W_N}\mathbf{Z}'\mathbf{y}.$$

Linear IV: GMM (continued)

- Then

$$\widehat{\mathsf{V}}\left[\widehat{\beta}_{\mathsf{GMM}}\right] = N\left[\mathbf{X}'\mathbf{Z}\mathbf{W}_{\mathbf{N}}\mathbf{Z}'\mathbf{X}\right]^{-1} \times \mathbf{X}'\mathbf{Z}\mathbf{W}_{\mathbf{N}}\widehat{\mathbf{S}}\mathbf{W}_{\mathbf{N}}\mathbf{Z}'\mathbf{X} \times \left[\mathbf{X}'\mathbf{Z}\mathbf{W}_{\mathbf{N}}\mathbf{Z}'\mathbf{X}\right]^{-1}.$$

- For heteroskedastic error

$$\widehat{\mathbf{S}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{z}_i\mathbf{z}_i'$$

- For homoskedastic error

$$\widehat{\mathbf{S}} = \frac{1}{N}s^2\sum_{i=1}^{N}\widehat{u}_i^2\mathbf{z}_i\mathbf{z}_i' = s^2\mathbf{Z}'\mathbf{Z}/N.$$

Linear IV: 2SLS

- Two-stage least squares uses $\mathbf{W}_N = (N^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$.

- This is optimal weighting matrix if errors are homoskedastic.

- Best to be robust and assume heteroskedastic errors.

- But if assume homoskedastic variance simplifies to familiar
$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_{\mathrm{2SLS}}] = s^2 \left[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right]^{-1}.$$

- Test for endogenity using over-identifying restrictions test.

Linear IV: 2SLS (continued)

The 2SLS estimator can be motivated in several ways.

1. Optimal GMM if errors are homoskedastic.

2. GLS estimation in transformed regression

   $\mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{X} + \mathbf{Z}'\mathbf{u}$ if errors are homoskedastic.

3. OLS regression of $\mathbf{y}$ on $\widehat{\mathbf{X}} = \mathbf{P_Z X}$ rather than of $\mathbf{y}$ on $\mathbf{X}$.

   The two-stage interpretation. Does not generalize to nonlinear.

4. IV estimation of $\mathbf{y}$ on $\mathbf{X}$ with instrument $\widehat{\mathbf{Z}} = \mathbf{P_Z X}$ rather than $\mathbf{Z}$.

5 IV in Practice

- Simple single equation linear model

$$y = \mathbf{x}'\boldsymbol{\beta} + u,$$

where part of $\mathbf{x}$ is endogenous.

- Estimate by IV where instrument $\mathbf{z}$ is such that

1. Valid: $\mathbf{z}$ uncorrelated with error u.

2. Relevant: $\mathbf{z}$ correlated with regressor $\mathbf{x}$.

3. Strong: $\mathbf{z}$ strongly correlated, rather than weakly correlated, with the regressor $\mathbf{x}$.

Weak Instrument

- For scalar regressor and instrument, a weak instrument is one for which $\mathsf{Cor}[x, z]$ is small.

- Equivalently $R^2_{x,z}$ (from regress x on z) is small.

- Large standard errors as
$$\mathsf{V}[\widehat{\beta}_{\mathsf{IV}}] = \mathsf{V}[\widehat{\beta}_{\mathsf{OLS}}]/\mathsf{Cor}^2[\mathbf{z}, \mathbf{x}].$$

- IV could be more inconsistent than OLS as
$$\frac{\textbf{plim}\,\widehat{\beta}_{\mathsf{IV}} - \beta}{\textbf{plim}\,\widehat{\beta}_{\mathsf{OLS}} - \beta} = \frac{\mathsf{Cor}[\mathbf{z}, \mathbf{u}]}{\mathsf{Cor}[\mathbf{x}, \mathbf{u}]} \times \frac{1}{\mathsf{Cor}[\mathbf{z}, \mathbf{x}]}.$$

- Poor finite sample performance as $\widehat{\beta}_{\mathsf{IV}}$ not centered around β. [Note that $\mathsf{E}[\widehat{\beta}_{\mathsf{IV}}]$ does not exist in just-identified case].

- Alternative estimators such as split-sample IV.

6 Nonlinear IV

- Nonlinear regression model with additive error term

$$y = g(\mathbf{x}, \boldsymbol{\beta}) + u$$

- In matrix notation

$$\mathbf{y} = \mathbf{g} + \mathbf{u}.$$

- Assume existence of instruments that satisfy $\mathsf{E}[\mathbf{u}|\mathbf{Z}] = \mathbf{0}$, so $\mathsf{E}[\mathbf{Z}'\mathbf{u}] = \mathbf{0}$, or

$$\mathsf{E}[\mathbf{Z}'(\mathbf{y} - \mathbf{g})] = \mathbf{0}.$$

- The GMM estimator minimizes

$$Q_N(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{g})'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'(\mathbf{y} - \mathbf{g}).$$

Nonlinear 2SLS

- Nonlinear 2SLS minimizes
$$Q_N(\boldsymbol{\beta}) = (\mathbf{y}-\mathbf{g})'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{y}-\mathbf{g}).$$

- Optimal choice if error is homoskedastic.

- Amemiya (1976) called this NL2SLS as handles endogeneity.

- But not really two-stage. In particular if regress $\mathbf{x}$ on $\mathbf{z}$ to get $\widehat{\mathbf{x}}$ and then do GLS of y on $g(\widehat{\mathbf{x}}, \boldsymbol{\beta})$ will generally get inconsistent estimator.

7 Two-Step Estimators

- Sequential two-step estimator $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ jointly solves the equations

$$
\begin{aligned}
N^{-1} \sum_{i=1}^{N} \mathbf{h}_{1i}(y_i, \mathbf{x}_i, \widehat{\boldsymbol{\theta}}_1) &= \mathbf{0} \\
N^{-1} \sum_{i=1}^{N} \mathbf{h}_{2i}(y_i, \mathbf{x}_i, \widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) &= \mathbf{0}.
\end{aligned}
$$

- Defining $\boldsymbol{\theta} = (\boldsymbol{\theta}_1' \quad \boldsymbol{\theta}_2')'$ and $\mathbf{h}_i = (\mathbf{h}_{1i}' \quad \mathbf{h}_{2i}')'$ this is just GMM with f.o.c. in case where $\dim(\mathbf{h}_i) = \dim(\boldsymbol{\theta})$

$$
N^{-1} \sum_{i=1}^{N} \mathbf{h}_i(y_i, \mathbf{x}_i, \widehat{\boldsymbol{\theta}}) = \mathbf{0}.
$$

- Apply general result. Simplification occurs as $\partial \mathbf{h}_{1i} / \partial \boldsymbol{\theta}_2 = \mathbf{0}$. But still messy (like delta method).

8 Empirical Likelihood

- Maximize empirical likelihood function $N^{-1}\sum_i \ln p_i$ subject to constraint $\sum_i p_i = 1$ and additional constraint from $\mathrm{E}[\mathbf{h}(y_i, \mathbf{x}_i, \boldsymbol{\theta})] = \mathbf{0}$ that
$$\sum_i p_i \mathbf{h}(y_i, \mathbf{x}_i, \boldsymbol{\theta}) = 0.$$

- Thus maximize the Lagrangian
$$\begin{aligned}\mathcal{L}(\mathbf{p}, \eta, \boldsymbol{\lambda}) &= \frac{1}{N}\sum_{i=1}^{N} \ln p_i - \eta\left(\sum_{i=1}^{N} p_i - 1\right) \\ &\quad -\boldsymbol{\lambda}'\sum_i p_i \mathbf{h}(y_i, \mathbf{x}_i, \boldsymbol{\theta}),\end{aligned}$$
w.r.t. $p_1, ..., p_N$ and the Lagrangian multipliers η and $\boldsymbol{\lambda}$.

- Same asymptotic distribution as GMM.
But different in finite samples.

9 Linear Sets of Equations

- Systems OLS same as equation by equation OLS
- Systems GLS more efficient usually.
- Examples:
  - Seemingly unrelated regressions
  - Three-stage least squares
  - panel data

10 Nonlinear Sets of Equations

- Stack equations to give vector $\mathbf{h}(\mathbf{w}_i, \boldsymbol{\theta})$.

- Systems GMM

11 References

Amemiya, T. (1974), "The Nonlinear Two-Stage Least Squares Estimator," Journal of Econometrics, 2, 105-110.

Bound, J., D.A. Jaeger and R.M. Baker (1995), "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak", Journal of the American Statistical Association, 90, 443-50.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Methods of Moments Estimators", Econometrica, 1029-1054.

Newey, W.K. (1984), "A Methods of Moments Interpretation of Sequential Estimators," Economics Letters, 14, 201-206.

Newey, W. K. (1990a), "Efficient Instrumental Variable Estimation of Nonlinear Models," Econometrica, 58, 809-838.

Qin, J. and J. Lawless (1994), "Empirical Likelihood and General Estimating Equations," Annals of Statistics, 22, 300-325.