Topic 19 Header Slide

- Topic 19. Estimation of non-linear regression models

  Estimators 8.etc. Non-Linear Least Squares (NLLS), Non-Linear GMM, Non-Linear IVE, MLE

Estimator	Notation	Under Complications	Extends to
1.	$\hat{\beta}_{OLS}$	$A2 Nonlinear.additive Error$	$\rightarrow \hat{\beta}_{NLLS}$
2.	$\hat{\beta}_{LAD}$		
3.	$\hat{\beta}_{Lstar}$		
4.	$\hat{\beta}_{GMM}$	$A3 Endogenous X/Exogenous W$	$\rightarrow \hat{\beta}_{IVE}$
$5a.$	$\hat{\beta}_{IGLS}$		
$5b.$	$\hat{\beta}_{FGLS}$		
6.	$\hat{\beta}_{MLE}$	$Multiple$	$\rightarrow \hat{\beta}_{FIMLE}$
7.	$\hat{\beta}_{IVE}$	$A2 Nonlinear.Additive Error$	$\rightarrow \hat{\beta}_{NLIVE}$
$8.etc.$		Nonlinear extensions (NLIV, NLGMM, etc)	

1. *Ostensibly* non-linear models
   - A suitable/judiciously-chosen transformation restores $A2Linear$ (partially/fully)
   - Leading example: Cobb-Douglas (C/D) production function. Consider

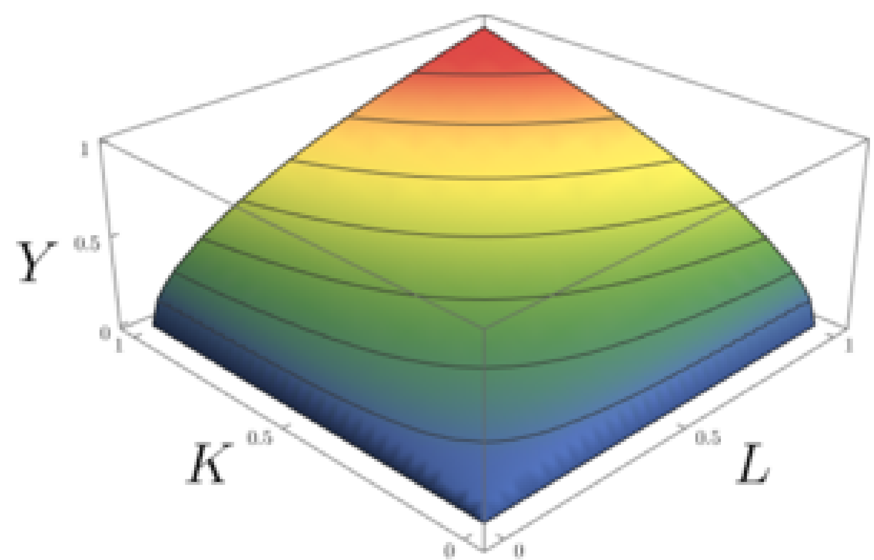
$$Y_s = \alpha L_s^{\beta_1} K_s^{\beta_2} \nu_s,$$

which exhibits log-linearity (for at least the main parameters) since we can readily obtain

$$y_s = \log(Y_s) = \beta_0 + \beta_1 \log(L_s) + \beta_2 \log(K_s) + \varepsilon_s,$$

where $\beta_0 \equiv \log(\alpha)$ and $\varepsilon_s \equiv \log(\nu_s)$ – i.e., linear in $\beta = (\beta_0, \beta_1, \beta_2)'$ and additive ε_s.

   - Techniques: OLS and similar methods are suitable



Introduction (2 of 2)

2. *Inherently* non-linear models
   - There exists no obvious transformation to restore $A2Linear$
   - Simple example: Constant Elasticity of Substitution (CES) production function
   - Leading example: Limited dependent variable (LDV) models (which we focus on below)
   - Techniques: NLLS, Non-linear GMM, Non-linear IVE, MLE

The inherent non-linearity of LDV models

- Models for LDVs represent a leading example of inherently non-linear regressions.

- The key feature of LDV models is the distinction between an underlying latent variable, LV, (denoted by y_s^*) and an LDV (denoted by y_s), whereby y_s^* and y_s are linked by a "partial observability" or "information filtering" function given, say, by:

$$y_s = \tau(y_s^*).$$

Fact 1. The LV y_s^* is governed by assumptions $A1^*, A2^*, A3^*, A4^*, A5^*$

Fact 2. The LDV y_s is governed both by assumptions $A1^*, A2^*, A3^*, A4^*, A5^*$ and the $y_s = \tau(y_s^*)$ mapping

Fact 3. Recall that $A1, A2, A3$ are assumptions that are critical in ensuring consistency (i.e., our first-order concern, as opposed to efficiency, which is typically a second-order objective). These assumptions will be determined by the complete set, $A1^*, A2^*, A3^*, A4^*, A5^*$

- Leading Examples:
  1. Example 1a and 1b – Probit and Logit models for binary outcomes
  2. Example 2 – Tobit model for censored and truncated outcomes
  3. Example 3 – Selectivity and Discrete/Continuous models for censored and truncated simultaneous outcomes

Example 1a – Probit model for binary outcomes

LDV is binary, LV is Gaussian $\Longrightarrow$ Probit model

$$
\begin{aligned}
A1^* &: rank(X) = k \\
A2^* linear &: y_s^* = x_s'\beta + \varepsilon_s \quad \mathbb{E}(\varepsilon_s) = 0 \\
A3^* &: \mathbb{E}(\varepsilon_s|X) = \mathbb{E}(\varepsilon_s) \\
A4^* GMiid &: \varepsilon_s|X \sim iid \\
A5^* &: \varepsilon_s|X \sim \mathcal{N}(0,1), \text{ so that } y_s^*|X \sim \mathcal{N}(x_s'\beta, 1) \\
pdf_\varepsilon(z) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \equiv \varphi(z) \\
cdf_\varepsilon(z) &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}w^2\right) \mathrm{d}w \equiv \Phi(z) \\
\tau(\cdot) &: y_s = \begin{cases} 1, & y_s^* > 0 \\ 0, & y_s^* \leq 0 \end{cases} \\
A2nonlinear &: y_s = \mathbb{E}(y_s|X) + u_s = \Phi(x_s'\beta) + u_s \\
A3 &: \mathbb{E}(u_s|X) = 0 \\
y_s|X &\sim \text{Bernoulli}(p) \text{ with } p = \Pr(y_s = 1|X) = \Phi(x_s'\beta) = \mathbb{E}(y_s|X)
\end{aligned}
$$

Example 1b – Logit model for binary outcomes

LDV is binary, LV is Logistic $\implies$ Logit model

$$
\begin{aligned}
A1^* &: rank(X) = k \\
A2^* linear &: y_s^* = x_s'\beta + \varepsilon_s \quad \mathbb{E}(\varepsilon_s) = 0 \\
A3^* &: E(\varepsilon_s|X) = \mathbb{E}(\varepsilon_s) \\
A4^* GMiid &: \varepsilon_s|X \sim iid \\
A5^* &: \varepsilon_s|X \sim \text{Logistic}(0, \pi^2/3), \text{ so that } y_s^*|X \sim \text{Logistic}(x_s'\beta, \pi^2/3) \\
pdf_\varepsilon(z) &= \frac{\exp(-z)}{(1+\exp(-z))^2} \equiv \lambda(z) \\
cdf_\varepsilon(z) &= \frac{1}{1+\exp(-z)} = \frac{\exp(z)}{1+\exp(z)} \equiv \Lambda(z) \\
\tau(\cdot) &: y_s = \begin{cases} 1, & y_s^* > 0 \\ 0, & y_s^* \le 0 \end{cases} \\
A2nonlinear &: y_s = \mathbb{E}(y_s|X) + u_s = \Lambda(x_s'\beta) + u_s \\
A3 &: E(u_s|X) = 0 \\
y_s|X &\sim \text{Bernoulli}(p) \text{ with } p = \Pr(y_s = 1|X) = \Lambda(x_s'\beta) = \mathbb{E}(y_s|X)
\end{aligned}
$$

Example 2 – Tobit model for censored outcomes (1 of 2)

LDV is censored, LV is Gaussian $\Longrightarrow$ Tobit model

$$
\begin{aligned}
A1^* &: rank(X) = k \\
A2^* linear &: y_s^* = x_s'\beta + \varepsilon_s \quad \mathbb{E}(\varepsilon_s) = 0 \\
A3^* &: E(\varepsilon_s|X) = \mathbb{E}(\varepsilon_s) \\
A4^* GMiid &: \varepsilon_s|X \sim iid \\
A5^* &: \varepsilon_s|X \sim \mathcal{N}(0, \sigma^2), \text{ so that } y_s^*|X \sim \mathcal{N}(x_s'\beta, \sigma^2) \\
pdf_\varepsilon(z) &= \frac{1}{\sigma}\varphi\left(\frac{z}{\sigma}\right) \\
cdf_\varepsilon(z) &= \Phi\left(\frac{z}{\sigma}\right) \\
\tau(\cdot) &: y_s = \begin{cases} y_s^*, & y_s^* > 0 \\ -999, & y_s^* \leq 0 \end{cases} \\
A2nonlinear &: y_s = \mathbb{E}(y_s|X) + u_s = x_s'\beta + \frac{1}{\sigma}\left(\frac{\varphi\left(\frac{x_s'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{x_s'\beta}{\sigma}\right)}\right) + u_s \text{ if } y_s \neq -999 \\
A3 &: E(u_s|X) = 0
\end{aligned}
$$

(Note: -999 is simply an arbitrary code for missing data, or "MDC".)

Example 2 – Tobit model for censored outcomes (2 of 2)

$y_s|X \sim$ mixed distribution with PDF as given by $\begin{cases} \frac{1}{\sigma}\varphi\left(\frac{y_s - x_s'\beta}{\sigma}\right), & y_s \neq -999 \\ \Phi\left(\frac{x_s'\beta}{\sigma}\right), & y_s = -999 \end{cases}$

Example 3 – Selectivity model with Gaussianity

$$
\begin{aligned}
A1^* &: \; rank(X) = k_x, \quad rank(Z) = k_z \\
A2^* linear &: \; y_s^* = x_s'\beta + \varepsilon_s, \quad \mathbb{E}(\varepsilon_s) = 0, \quad R_s^* = z_s'\gamma + \zeta_s, \quad \mathbb{E}(\zeta_s) = 0 \\
A3^* &: \; E(\varepsilon_s|X,Z) = \mathbb{E}(\varepsilon_s), \quad E(\zeta_s|X,Z) = \mathbb{E}(\zeta_s) \\
A4^* GMiid &: \; \varepsilon_s|X,Z \sim iid, \quad \zeta_s|X,Z \sim iid \\
A5^* &: \; \begin{pmatrix} \varepsilon_s \\ \zeta_s \end{pmatrix} \Bigg| X,Z \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right), \text{ so that} \\
& \quad \begin{pmatrix} y_s^* \\ R_s^* \end{pmatrix} |X,Z \sim \mathcal{N}\left(\begin{pmatrix} x_s'\beta \\ z_s'\gamma \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right) \\
pdf_\varepsilon(z) &= \frac{1}{\sigma}\varphi\left(\frac{z}{\sigma}\right), \quad cdf_\varepsilon(z) = \Phi\left(\frac{z}{\sigma}\right), \quad etc. \\
\tau(\cdot) &: \; y_s = \begin{cases} y_s^*, & d_s = 1 \\ -999, & d_s = 0 \end{cases} \text{ where participation, } d_s \text{, is indicated by} \\
& \quad d_s = \begin{cases} 1, & R_s^* > 0 \\ 0, & R_s^* \leq 0 \end{cases} \\
A2nonlinear &: \; y_s = \mathbb{E}(y_s|X,Z) + u_s = x_s'\beta + \rho\sigma\left(\frac{\varphi(z_s'\beta)}{1-\Phi(z_s'\beta)}\right) + u_s \text{ if } d_s = 1 \\
A3 &: \; E(u_s|X) = 0 \\
y_s|X,Z &\sim \text{ mixed distribution; } d_s|Z \sim \text{Bernoulli}(p) \text{ where } p = \Pr(d_s = 1|Z) = \Phi(z_s'\beta)
\end{aligned}
$$

Key estimation strategies for LDV models

Quick summary of available approaches –

1. Full MLE – consistent and asymptotically efficient
2. GMM-type non-linear estimation – consistent but not asymptotically efficient
3. PHP (pretend, hope, and pray!) – inconsistent due to endogeneity (cause 6 or cause 7)

Dealing with Regressor Endogeneity

The fundamental point is the following:
The appropriate estimation strategy in the presence of one or more endogenous regressors depends on the particular endogeneity cause that applies in the particular setting.
Hence, we must review the seven leading causes and identify the one that pertains.

Attempting to Overcome the Seven Leading Causes of Regressor Endogeneity

3.1 *Cause 1*

Combination of lagged dependent variables among the regressors and autocorrelated errors.

- Depending on Order of Lags in A2Linear and Type of Autocorrelation:

  If A4 autocorrelation is $MA(q)$, Best approach: IVE with older lags of A2linear specification

  If A4 autocorrelation is $AR(p)$, IVE *does* *not* work.

3.2 *Cause 2*

Omitted regressors from the specification.

Recall Paul Broca historical example with ~200 autopsies.

- Best approach: Think what the Omitted Important Variables are. Find them and Include them.
- Poor approach: IVE would be impossible or counterproductive.

3.3 *Cause 3*

Measurement errors in one or more regressor variables(s)

A2linear in terms of X^*:

$$y = X^* \beta^{true} + \varepsilon^{true}$$

vs.

A2linear in terms of X:

$$\begin{aligned} y &= (X - V)\,\beta^{true} + \varepsilon^{true} \\ &= X\beta^{true} + \varepsilon^{true} - V\beta^{true} \\ &= X\beta^{true} + \underbrace{\mathcal{U}}_{\text{composite true error}} \end{aligned}$$

using the fact that, by assumption, $X = X^* + V$

Suggested IVE idea: Use the *ranking* of mismeasured variable as Instrument. Idea/Hope: the ranking of a variable (assigning 1 to its smallest value, 2 to its next largest, ..., and S to its largest value) should be correlated to the original variable (i.e., Relevant Instrument) while it is hoped that it is *less* correlated to the source of endogeneity in the composite error U, which is the $-\beta^{true}V$ term involving the measurement error.

NB: this is only an *approximate* idea/hope, since the Ranking is clearly not fully Valid, nor the most Relevant one.

3.4 *Cause 4*

Functional form misspecification of A2 part of the model.

- Best approach: Think what the correct Functional Form of $E(y|X)$ is and re-model A2Nonlinear to reflect the true nonlinear regression relation.
- Poor approach: IVE would be impossible or counterproductive.

3.5 *Cause 5*

Simultaneity - System of simultaneous equations determines LHS and RHS variables simultaneously.

- Best approach: IVE of the SFEs, using X^{Bad} and $X^{Good} = Z^I$ as regressors, and all the Good (=Exogenous) variables throughout the model as Instruments. Clearly, the complete set of Exogenous variables in the whole System of Equations is: $Union[Z^I, Z^{Elsewhere}]$

3.5.1 *Simple illustration:*

Consider modeling the market for ice cream, assume perfect markets with flexible prices

- SF1: $y = X^B\beta^B + Z^I\beta^G + \varepsilon^{true}$
  - y: quantity of ice cream demanded — dependent variable (endogenous)
  - X^B: price of ice cream — endogenous
  - Z^I: consumer income, weather — exogenous
- SF2: $X^B = y\gamma + Z^E\gamma^E + u^{true}$
  - X^B: price of ice cream — dependent variable (endogenous)
  - y: quantity of ice cream supplied — endogenous
  - Z^E: transportation in refrigerated lorries etc. — exogenous
- Note: we could always consider also the reduced form (RF) equation for price of ice cream y on the LHS as a function of only Z_s (Z^I and Z^E) on the RHS

3.6 *Cause 6*

[Related to cause 4] + LDV (limited development variable) model

3.7 *Cause 7*

[Related to cause 5] + LDV (limited development variable) model with Selectivity/Sample Selection/Discrete-Continuous Switching models

3.7.1 BEST parametric CUAN: LDV Approach 1: Full MLE of LDV Models

.

3.7.2 CUAN but NOT BEST: LDV Approach 2: Consistent but Inefficient GMM-type Nonlinear Estimation of LDV Models

.

3.7.3 TERRIBLE: Inconsistent in General: LDV Approach 3: PHP (Pretend Hope and Pray) is Inconsistent because Classic Endogeneity (either Cause 6 and/or Cause 7)

-