Structural Equation Modeling

Mgmt 290
Lecture 5 – Model Estimation
& Modeling Process
Oct. 24, 2005

About Estimation

When sample size big as infinite, est close to true value

- 2 criteria
- 1) Unbiased ~ E(est) = **true value**
- (biased, inconsistent)
- 2) Efficient ~ Var(est) small
- (not reliable)

One method more efficient than another – one dominates another

Estimation Methods Offered by LISREL

- Instrumental Variables (IV)
- Two-Stage Least Squares (2SLS) ← **small sample**
- Unweighted Least Squares (ULS) - below all iterative
- Generalized Least Squares (GLS)
- **Maximum Likelihood (ML)** ← **Large sample**
- Generally Weighted Least Squares (WLS)
- Diagonally Weighted Least Squares (DWLS) least squares

Weighted LS targets at the violation of Homoskedasticity

OLS Estimation

BLUE if assumptions good

- if all assumptions are met
- if recursive

For multiple regression,
If all the assumptions are
Valid, OLS and ML will give
the same results.

$E(e_j)=0$ --- the mean value of the error term is 0
$Var(e_j) = \sigma^2$ --- the variance of the error term is constant - Homoskedasticity
$Cov(e_i, e_j) = 0$, no autocorrelation
No serious collinearity
E_j is normally distributed
Additive and Linearity

When OLS Does Not Work

- For OLS, the disturbance must not be correlated with each causal variable. There are three reasons why such a correlation might exist:
- **1) Spuriousness (Third Variable Causation): A variable causes two or more causal variables and one or more of that variables are not included in the model.**
- **2) Reverse Causation (Feedback Model): The endogenous variable causes, either directly or indirectly, one of its causes.**
- **3) Measurement Error: There is measurement error in a causal variable.**

IV Estimation for Simple Regression

Special case
Of SEM

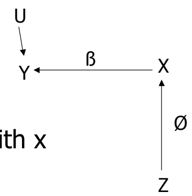
- Y = BX + U (X U, and Y are Standardized x, u, and y)
- YX = B XX + UX
- E(YX) = B E(XX) (E(UX) = 0)
- So, B = Cov(x, y) / var(x)

- YZ = BXZ + UZ
- E(YZ) = B E(XZ) (E(ZU) = 0)
- So, B = Cov(y, z) / cov(x, z)

Get a Z that
E(ZU)=0
E(x,z) not 0
Z is
Standardized z

Why IV Estimation Works

- unbiased
-
- efficient if
- highly correlated with x



- E(B) = Cov (yz) / cov(xz)
- = ßØ / Ø = ß

Z does not affect y directly

IV Estimation

- Conditions for instrumental variable I estimation:
- 1) The variable I must not be correlated with the error U.
- 2) For a given structural equation, there must be as many or more I variables as there are variables needing an instrument.
- 3) The variable I must be associated with the variable that needs an instrument, and does not affect Y directly.

2SLS for regression -- application of IV method

- Multiple regression :
- Y = cX1 + … + dXn + U
- Z1 … Zn are IVs
- Step 1: run OLS regression of Xi on Zi (or on all Zs) to get predicted X'i
- Step 2: run OLS regression of Y on X'1 ~ X'n

Why?

- as Zi uncorrelated with U
-
- among Xi = a+ bZi + ei
- a+bZi also uncorrelated with U
- the correlated part gets isolated

2SLS for SEM -- application of IV method

- Structural Equations:
- $Z = aX + bY + U$
  $Y = cQ + dZ + V$
- Note that the notation has changed. For this example, variable Q serves as an instrumental variable for Y in the Z equation, and X serves as an instrumental variable for Z in the Y equation.

Model generated IVs.

2SLS

- For the Z equation:
  Stage 1: Regress Y on Q.
  Stage 2: Regress Z on the stage 1 predicted score for Y and X.
- For the Y equation:
  Stage 1: Regress Z on X.
  Stage 2: Regress Y on the stage 1 predicted score for Z and Q.

Implement IV Estimation in SPSS and LISREL

- in SPSS
- Step 1: click on File, then Read Text Data to read in your data file
- Step 2: click on Analyze, then Regression, then 2-Stage Least Squares…
- A 2-Stage Least Squares box will open that you should (1) move your dependent variable to the box with Dependent: above it, then (2) move your instrumental variables AND your other independent variables not needing instrumental variables to the box with Instrumental: on the top, and (3) move all your independent variables (not IVs) to the box with Explanatory: on the top.
- Click on OK to get your results.
- For more, see http://www.researchmethods.org/instru-var.htm

Review of ML

- Maximize a Likelihood Value
- Iterative B0 -> B1 -> …. Stop when the improvement is not significant

2SLS over ML

- Does not require any distributional assumptions
- do not require numerical optimization algorithms (simple computing)
- permit using routine diagnostic procedures
- perform better in small samples

ML over 2SLS

- ML gives simultaneous estimation & use full info
- if assumptions are valid and the model specification is correct, ML is more efficient
- especially for sufficiently large sample
- 2SLS results depend on the choice of IVs

More About ML

- A "large sample" method
- 100 observations as minimum
- 200 or more for moderate complexity in structure model
- Or 5:1 ~ 10:1 as sample size to parameters ratio

Starting Values & Converge

- Software generated starting values
- Sometimes they do not lead to the convergence of iterative estimation
- We need to come up some good starting values

Data in c:\program files\lisrel87s\lis87ex
student version

Example

- The data set, **klein.dat**, consists of the following 15 variables:
- Ct = Aggregate Consumption
  Pt_1 = Total Profits, previous year
  Wt_s = Private Wage Bill
  It = Net Investment
  Kt_1 = Capital Stock, previous year
  Et_1 = Total Production of Private Industry, previous year
  Wt** = Government Wage Bill
  Tt = Taxes
  At = Time in Years from 1931
  Pt = Total Profits
  Kt = End-of-year Capital Stock
  Et = Total Production of Private Industry
  Wt = Total Wage Bill
  Yt = Total Income
  Gt = Government Non-Wage Expenditure

Need to take IT off.

Example

- Ct <- Pt, Pt_1, wt, It

Correlations

		PT	PT_1	WT	IT
PT	Pearson Correlation	1	.769**	.634**	.923**
	Sig. (2-tailed)	.	.000	.002	.000
	N	21	21	21	21
PT_1	Pearson Correlation	.769**	1	.579**	.768**
	Sig. (2-tailed)	.000	.	.006	.000
	N	21	21	21	21
WT	Pearson Correlation	.634**	.579**	1	.407
	Sig. (2-tailed)	.002	.006	.	.067
	N	21	21	21	21
IT	Pearson Correlation	.923**	.768**	.407	1
	Sig. (2-tailed)	.000	.000	.067	.
	N	21	21	21	21

**. Correlation is significant at the 0.01 level (2-tailed).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	14.200	2.517		5.643	.000
	PT	.375	.213	.230	1.763	.097
	PT_1	.132	.101	.078	1.305	.210
	WT	.761	.054	.839	14.003	.000
	IT	-.223	.235	-.115	-.947	.358

a. Dependent Variable: CT

OLS Estimation

- Ct = 16.237 + .193 Pt + .0899 Pt_1 + .796 wt

From LISREL

- t ratio - (2.115, .992, 19.933)
- R^2 = .981

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	16.237	1.303		12.464	.000
	PT	.193	.091	.119	2.115	.049
	PT_1	.090	.091	.053	.992	.335
	WT	.796	.040	.877	19.933	.000

a. Dependent Variable: CT

From SPSS

2SLS Estimation

LISREL

- Ct = 16.15 + .0565 Pt + .206 Pt_1 + .808 wt
- only wt's effect significant
- R^2 = .977
- IVs ~ Wt_s, Tt, Gt, At, Kt_1, Et_1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	16.150	1.432		11.278	.000
	Unstandardized Predicted Value	5.646E-02	.114	.034	.497	.626
	Unstandardized Predicted Value	.206	.112	.118	1.846	.082
	Unstandardized Predicted Value	.808	.044	.890	18.208	.000

a. Dependent Variable: CT

SPSS

General Modeling Process

- 1) Model specification
- 2) Identification
- 3) Estimation and Fit
- 4) Model Modification
- 5) Estimation and Fit

Data Preparation

Use fit indexes

Ref: Kelloway's book

A Step by Step Approach

- Generals one adopted by almost everyone
- See Kelloway's book
- Very helpful to make all your steps explicit!
- See
- http://www.researchmethods.org/step-by-step1.pdf

Our Step by Step Approach (1)

- 1) Proposal ← Assignment 1
- 2) Initial Model Specification ← Assignment 2
- 3) Prepare Data ← Assignment 2
- 4) Estimate Models ← Assignment 3
- 5) Evaluate Models ← Assignment 3, Assignment 4
- 6) Diagnostics and Modify Models ← Assignment 4
- 7) Final Results ← Assignment 4, Presentation

Our Step by Step Approach (2)

- 1) Proposal – foundation for specifying models and evaluating results
- 2) Initial Model Specification – language -> math, narrow down step, concerns strategies
- 3) Prepare Data – take care of data problems and foundation for selecting estimation methods & diagnostics
- 4) Estimate Models – work with software packages and try to use more info (IVs & starting values)
- 5) Evaluate Models – 15 or more fit indexes
- 6) Diagnostics and Modify Models – check assumptions again and again
- 7) Final Results – a final smile ☺

Main Advantages of Our Step by Step Approach

- A clear relationship among assumptions, model representation, estimation and model evaluation (modification) & interpretation
- Not strictly confirmative
- Search for "best" models
- Search for "true" values
- (a better results than others or something closer to the TRUTH)

Prepare to Set Up LISREL

- Use LISREL

Ready for LISREL???

- Import External Data in Other Format …
- .psf for LISREL
- (looks similar to any other table formats)

- Can import datasets in almost any format
- SPSS, SAS, Stata, Excel, …

Can export as well
ASCII & SPSS formats