Econometrics 2 - Lecture 5

Multi-equation Models

Contents

- Systems of Regression Equations
- Multivariate Regression Models
- SUR Model: Estimation
- Simultaneous Equation Models
- Simultaneous Equation Models: Identification
- Identification: Criteria
- Single Equation Estimation Methods
- System Estimation Methods

Multiple Dependent Variables

In general, economic processes involve a multiple set of variables which show a simultaneous and interrelated development

Examples:

- Households consume a set of commodities (food, durables, etc.); the demanded quantities depend on the prices of commodities, the household income, the number of persons living in the household, etc. A consumption model includes a set of dependent variables and a common set of explanatory variables.
- The market of a product is characterized by (a) the demanded and supplied quantity and (b) the price of the product; a model for the market consists of equations representing the development and interdependencies of these variables.
- An economy consists of markets for commodities, labor, finances, etc. A model for a sector or the full economy contains descriptions of the development of the relevant variables and their interactions.

Systems of Regression Equations

Economic processes involve the simultaneous developments as well as interrelations of a set of dependent variables

- For modeling an economic process a system of relations, typically in the form of regression equations: multi-equation model

Example: Two dependent variables y_{t1} and y_{t2} are modeled as

$$y_{t1} = x'_{t1}\beta_1 + \varepsilon_{t1}$$
$$y_{t2} = x'_{t2}\beta_2 + \varepsilon_{t2}$$

with $V\{\varepsilon_{ti}\} = \sigma_i^2$ for $i = 1, 2$, $Cov\{\varepsilon_{t1}, \varepsilon_{t2}\} = \sigma_{12} \neq 0$

Typical situations:

1. The set of regressors x_{t1} and x_{t2} coincide
2. The set of regressors x_{t1} and x_{t2} differ, may overlap
3. Regressors contain one or both dependent variables
4. Regressors contain lagged variables

Types of Multi-equation Models

Multivariate regression or multivariate multi-equation model

- A set of regression equations, each explaining one of the dependent variables
  - Possibly common explanatory variables
  - Seemingly unrelated regression (SUR) model: each equation is a valid specification of a linear regression, related to other equations only by the error terms
  - See cases 1 and 2 of “typical situations” (slide 4)

Simultaneous equation models

- Describe the relations within the system of economic variables
  - in form of model equations
  - See cases 3 and 4 of “typical situations” (slide 4)

Error terms: dependence structure is specified by means of second moments or as joint probability distribution

Capital Asset Pricing Model

Capital asset pricing (CAP) model: describes the return R_i of asset *i*

$$R_i - R_f = \beta_i(E\{R_m\} - R_f) + \varepsilon_i$$

with

- R_f: return of a risk-free asset
- R_m: return of the market portfolio

- β_i: indicates how strong fluctuations of the returns of asset *i* are determined by fluctuations of the market as a whole
- Knowledge of the return difference $R_i - R_f$ will give information on the return difference $R_j - R_f$ of asset *j* , at least for some assets
- Analysis of a set of assets $i = 1, \ldots, s$
  - The error terms ε_i, $i = 1, \ldots, s$, represent common factors, have a common dependence structure
  - Efficient use of information: simultaneous analysis

A Model for Investment

Grunfeld investment data [Greene, (2003), Chpt.13; Grunfeld & Griliches (1960)]: Panel data set on gross investments I_{it} of firms over 20 years and related data

- Investment decisions are assumed to be determined by

$$I_{it} = \beta_{i1} + \beta_{i2}F_{it} + \beta_{i3}C_{it} + \varepsilon_{it}$$

  with
  - F_{it}: market value of firm at the end of year t-1
  - C_{it}: value of stock of plant and equipment at the end of year t-1
- Simultaneous analysis of equations for the various firms: efficient use of information
  - Error terms for the firms include common factors such as economic climate
  - Coefficients may be the same for the firms

The Hog Market

Model equations:

$$Q^d = \alpha_1 + \alpha_2 P + \alpha_3 Y + \varepsilon_1 \quad \text{(demand equation)}$$
$$Q^s = \beta_1 + \beta_2 P + \beta_3 Z + \varepsilon_2 \quad \text{(supply equation)}$$
$$Q^d = Q^s \text{ (equilibrium condition)}$$

with Q^d: demanded quantity, Q^s: supplied quantity, P: price, Y: income, and Z: costs of production, or

$$Q = \alpha_1 + \alpha_2 P + \alpha_3 Y + \varepsilon_1 \quad \text{(demand equation)}$$
$$Q = \beta_1 + \beta_2 P + \beta_3 Z + \varepsilon_2 \quad \text{(supply equation)}$$

- Model describes the equilibrium transaction quantity and price
- Model determines simultaneously Q and P, given Y and Z
- Error terms
  - May be correlated: $\mathrm{Cov}\{\varepsilon_1, \varepsilon_2\} \neq 0$
- Simultaneous analysis necessary for efficient use of information

Klein‘s Model I

1. $C_t = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \alpha_4 (W_t^p + W_t^g) + \varepsilon_{t1}$ (consumption)
2. $I_t = \beta_1 + \beta_2 P_t + \beta_3 P_{t-1} + \beta_4 K_{t-1} + \varepsilon_{t2}$ (investment)
3. $W_t^p = \gamma_1 + \gamma_2 X_t + \gamma_3 X_{t-1} + \gamma_4 t + \varepsilon_{t3}$ (wages)
4. $X_t = C_t + I_t + G_t$
5. $K_t = I_t + K_{t-1}$
6. $P_t = X_t - W_t^p - T_t$

with C (consumption), P (profits), W^p (private wages), W^g (governmental wages), I (investment), K_{-1} (capital stock), X (national product), G (governmental demand), T (taxes) and t [time (year-1936)]

- Model determines simultaneously C, I, W^p, X, K, and P
- Simultaneous analysis necessary in order to take dependence structure of error terms into account: efficient use of information

Examples of Multi-equation Models

Multivariate regression models

- Capital asset pricing (CAP) model: for all assets, return R_i is a function of E{R_m} – R_f; dependence structure of the error terms caused by common variables
- Model for investment: firm-specific regressors, dependence structure of the error terms like in CAP model
- Seemingly unrelated regression (SUR) models

Simultaneous equation models

- Hog market model: endogenous regressors, dependence structure of error terms
- Klein's model I: endogenous regressors, dynamic model, dependence of error terms from different equations and possibly over time

Single- vs. Multi-equation Models

Complications for estimation of parameters of multi-equation models:

- Dependence structure of error terms
- Violation of exogeneity of regressors

Example: Hog market model, demand equation

$$Q = \alpha_1 + \alpha_2 P + \alpha_3 Y + \varepsilon_1$$

- P is not exogenous: $\mathrm{Cov}\{P, \varepsilon_1\} = (\sigma_1^2 - \sigma_{12})/(\beta_2 - \alpha_2) \neq 0$
- Covariance matrix of $\varepsilon = (\varepsilon_1, \varepsilon_2)'$

$$\mathrm{Cov}\{\varepsilon\} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Statistical analysis of multi-equation models requires methods adapted to these features

Analysis of Multi-equation Models

Issues of interest:

- Estimation of parameters
- Interpretation of model characteristics, prediction, etc.

Estimation procedures

- Multivariate regression models
  - GLS , FGLS, ML
- Simultaneous equation models
  - Single equation methods: indirect least squares (ILS), two stage least squares (TSLS), limited information ML (LIML)
  - System methods of estimation: three stage least squares (3SLS), full information ML (FIML)
  - Dynamic models: estimation methods for vector autoregressive (VAR) and vector error correction (VEC) models

Contents

- Systems of Regression Equations
- Multivariate Regression Models
- SUR Model: Estimation
- Simultaneous Equation Models
- Simultaneous Equation Models: Identification
- Identification: Criteria
- Single Equation Estimation Methods
- System Estimation Methods

The SUR Model

Seemingly unrelated regression model

Multivariate regression model: the general case, *m* equations

$$y_{t1} = x'_{t1}\beta_1 + \varepsilon_{t1}$$

$$\ldots$$

$$y_{tm} = x'_{tm}\beta_m + \varepsilon_{tm}$$

with $V\{\varepsilon_{ti}\} = \sigma_i^2$ for $i = 1,\ldots,m$; $Cov\{\varepsilon_{ti}, \varepsilon_{tj}\} = \sigma_{ij} \neq 0$ for $i \neq j$, $i,j = 1,\ldots,m$, and $t = 1,\ldots,T$, i.e., contemporaneously correlated error terms

Regressors

- Can be specific for each equation
- Multivariate regression with common regressors

$$x_{ti} = x_t \text{ for } i = 1,\ldots,m$$

e.g., the CAP model

Example: Investment Model

Investment model based on the Grunfeld data set [Greene, (2003), Chpt.13; Grunfeld & Griliches (1960)]

$$I_{it} = \beta_{i1} + \beta_{i2}F_{it} + \beta_{i3}C_{it} + \varepsilon_{it}$$

with

- I_{it}: gross investment I_{it} of firm *i*
- F_{it}: market value of firm at the end of year *t*-1
- C_{it}: value of stock of plant and equipment at the end of year *t*-1

- Explanatory variables observed for firm *i* may affect other firms due to the dependence structure of the error terms
- Estimation methods
  - Each equation separately using OLS
  - Generalized least squares estimation may take dependence structure of the error terms into account

The SUR Model: Notation

For $m = 2$

- With T-vectors y_i, ε_i, and $(T \times K)$-matrix X_i

  $y_i = X_i \beta_i + \varepsilon_i, \quad i = 1, 2$

  and $V\{\varepsilon_{ti}\} = \sigma_i^2$, $Cov\{\varepsilon_{t1}, \varepsilon_{t2}\} = \sigma_{12} \neq 0$ for $t = 1,\ldots,T$,

$$V\{\varepsilon_t\} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \Sigma$$

- With $2T$-vectors

$$\bar{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \bar{X}\bar{\beta} + \bar{\varepsilon}$$

  and $V = V\{\bar{\varepsilon}\} = \Sigma \otimes I_T = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \otimes I_T$

The Kronecker Product

Definition of the Kronecker product of matrices A (order $n \times m$) and B (order $p \times q$)

$$A = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix}, B = \begin{pmatrix} b_{11} & \dots & b_{1q} \\ \vdots & \ddots & \vdots \\ b_{p1} & \dots & b_{pq} \end{pmatrix}$$

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1m}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \dots & a_{nm}B \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & \dots & a_{1m}b_{1q} \\ \vdots & \ddots & \vdots \\ a_{n1}b_{p1} & \dots & a_{nm}b_{pq} \end{pmatrix}$$

The product matrix has the order $np \times mq$

Some rules: $(i)\ (A \otimes B)(C \otimes D) = AC \otimes BD,$ suitable A, B, C, D

$(ii)\ (A \otimes B)' = A' \otimes B'$

$(iii)\ (A \otimes B)^{-1} = A^{-1} \otimes B^{-1},$ square A, B

Contents

- Systems of Regression Equations
- Multivariate Regression Models
- SUR Model: Estimation
- Simultaneous Equation Models
- Simultaneous Equation Models: Identification
- Identification: Criteria
- Single Equation Estimation Methods
- System Estimation Methods

Parameter Estimation

For $m = 2$

OLS estimators for β_1 and β_2 on basis of $y_i = X_i \beta_i + \varepsilon_i$, $i = 1, 2$

- $b_i = (X_i'X_i)^{-1} X_i'y_i$, $i = 1, 2$
- Or $\bar{b} = (b_1', b_2')' = (\bar{X}'\bar{X})^{-1} \bar{X}'\bar{y}$
- With $V\{b_i\} = \sigma_i^2(X_i'X_i)^{-1}$
- Ignores $\Sigma \neq I$, i.e., the contemporaneous correlation of error terms

GLS estimators

$$\tilde{\bar{\beta}} = (\tilde{\beta}_1', \tilde{\beta}_2')' = (\bar{X}'V^{-1}\bar{X})^{-1} \bar{X}'V^{-1}\bar{y}$$

- with $V\left\{\tilde{\bar{\beta}}\right\} = (\bar{X}'V^{-1}\bar{X})^{-1}$

Analogous for any number m of equations

Investment Model

Investment models

$$I_{it} = \beta_1 + \beta_2 F_{it} + \beta_3 C_{it} + \varepsilon_{it}$$

for General Motors, Chrysler, and General Electric

	β_1		β_2		β_3		R^2
	OLS	GLS	OLS	GLS	OLS	GLS	
GM	-149.8	-133.6	0.119	0.115	0.371	0.376	0.92
p-val.	0.175	0.178	0.0002	0.0001	1.5E-8	3.1E-9	
Crysler	-6.190	-3.266	0.078	0.073	0.316	0.320	0.91
p-val.	0.653	0.794	0.0011	0.0009	4.0E-9	8.6E-10	
GE	-9.96	-11.96	0.027	0.028	0.152	0.152	0.71
p-val.	0.755	0.680	0.106	0.067	1.7E-5	6.1E-6	

F: market value, *C*: value of stock

The General SUR Model

SUR model with *m* equations

- The *i*-the equation: *T*-vectors y_i, ε_i, and (TxK_i)-matrix X_i

  $y_i = X_i \beta_i + \varepsilon_i, \quad i = 1, \ldots, m$

  $V\{\varepsilon_{ti}\} = \sigma_i^2$, $Cov\{\varepsilon_{ti}, \varepsilon_{tj}\} = \sigma_{ij} \neq 0$ for $i, j = 1, \ldots, m$, $t = 1, \ldots, T$,
- Full model

$$\bar{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix} = \bar{X}\bar{\beta} + \bar{\varepsilon}$$

with $\mathrm{V} = \mathrm{V}\{\bar{\varepsilon}\} = \Sigma \otimes I_n = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{1m} & \cdots & \sigma_m^2 \end{pmatrix} \otimes I_n$

FGLS Estimator

2-step procedure: estimation of GLS estimators

$$\tilde{\tilde{\beta}} = (\tilde{\beta}_1', \tilde{\beta}_2')' = (\bar{X}' V^{-1} \bar{X})^{-1} \bar{X}' V^{-1} \bar{y}$$

requires knowledge of covariance matrix V of the error terms

1. OLS estimation of each of the m equations; calculation of the m-squared matrix $S = E'E/T$, estimator of Σ, using the Txm matrix $E = (e_1, \ldots, e_m)$ of OLS residuals, with elements s_{ij} of S

   $$s_{ij} = (\Sigma_t e_{ti} e_{tj})/T$$

   for $i,j = 1, \ldots, m$; suitable degree of freedom correction
2. GLS estimation using instead of V the estimated matrix $\tilde{V} = S \otimes I_m$ i.e., V with s_{ij} substituted for σ_{ij} for $i, j = 1, \ldots, m$

   $$\tilde{\tilde{b}} = (\tilde{b}_1', \tilde{b}_2')' = (\bar{X}' \tilde{V}^{-1} \bar{X})^{-1} \bar{X}' \tilde{V}^{-1} \bar{y}$$

GLS and OLS Estimators

GLS estimators

$$\tilde{\tilde{\beta}} = (\tilde{\beta}_1{}', \tilde{\beta}_2{}')' = (\bar{X}'\tilde{V}^{-1}\bar{X})^{-1}\bar{X}'\tilde{V}^{-1}\bar{y}$$

- More efficient than OLS estimators
- Efficiency gain increases
  - with growing correlation of error terms
  - with shrinking correlation of regressors
- GLS estimator for β_i coincides with OLS estimator b_i if
  - matrix X_i of regressors is the same for all equations: $X_i = X$
  - ε_{ti} is uncorrelated with all ε_{tj}, $j \neq i$
- FGLS estimates are consistent, asymptotically efficient

Goodness of Fit Measures

Definition of an R^2-like measure:

$$R_I^2 = 1 - \frac{S_g(\tilde{\bar{\beta}})}{S_g(0)} = 1 - \frac{\mathrm{tr}(S^{-1}\tilde{\Sigma})}{\mathrm{tr}(S^{-1}S_{yy})}$$

with $S_g(b) = (\bar{y}\text{-}\bar{X}\tilde{\bar{b}})'\tilde{V}^{-1}(\bar{y}\text{-}\bar{X}\tilde{\bar{b}}) = T\,\mathrm{tr}(S^{-1}\tilde{\Sigma}),\ \tilde{\Sigma}=(\tilde{E}'\tilde{E})/T$
using the Txm matrix $\tilde{E} = (\tilde{e}_1,\ldots,\tilde{e}_m)$ of FGLS residuals and analogously the mxm matrix S_{yy} of sample covariances of the y_{ti}

An alternative measure is

$$R_*^2 = 1 - \frac{m}{\mathrm{tr}(S^{-1}S_{yy})}$$

SUR Model in GRETL

Model > simultaneous equation …

- to be specified
  - the simultaneous equations
  - option for estimator: Seemingly Unrelated Regression (sur)
- FGLS estimation

Contents

- Systems of Regression Equations
- Multivariate Regression Models
- SUR Model: Estimation
- Simultaneous Equation Models
- Simultaneous Equation Models: Identification
- Identification: Criteria
- Single Equation Estimation Methods
- System Estimation Methods

Example: Two-equation Model

Dependent variables Y_1, Y_2; the model is

$Y_1 = \alpha_1 + \alpha_2 Y_2 + \alpha_3 X_1 + \varepsilon_1$ (equation A)

$Y_2 = \beta_1 + \beta_2 Y_1 + \beta_3 X_2 + \varepsilon_2$ (equation B)

1. Violation of assumption A2 (exogeneity of regressors): a positive ε_1
   - results in an increase of Y_1 (see equation A)
   - Consequence of this (see equation B) is, given a positive β_2, an increase of Y_2
   - i.e., ε_1 and Y_2 are correlated
2. Biased OLS estimators
   - Large values of Y_1 are observed – due to positive values of ε_1 – together with large values of Y_2
   - Overestimated α_2

Example: Market Model

Describes the transactions in the market of a commodity, e.g., hogs

$$Q^d = \alpha_1 + \alpha_2 P + \alpha_3 Y + \varepsilon_1 \quad \text{(demand equation)}$$
$$Q^s = \beta_1 + \beta_2 P + \beta_3 Z + \varepsilon_2 \quad \text{(supply equation)}$$
$$Q^d = Q^s \text{ (equilibrium condition, market clearing assumption)}$$

with Q^d: demanded quantity, Q^s: supplied quantity, P: price, Y: income, and Z: costs of production, or

$$Q = \alpha_1 + \alpha_2 P + \alpha_3 Y + \varepsilon_1 \quad \text{(demand equation)}$$
$$Q = \beta_1 + \beta_2 P + \beta_3 Z + \varepsilon_2 \quad \text{(supply equation)}$$

- Model describes the equilibrium transaction quantity Q and price P
- Exogeneity assumption for variables Y, Z
  - Model determines simultaneously Q and P, given Y and Z
  - Endogenous variables: Q, P

Market Model

Structural form of the model

$$Q = \alpha_1 + \alpha_2 P + \alpha_3 Y + \varepsilon_1 \quad \text{(demand equation)}$$

$$Q = \beta_1 + \beta_2 P + \beta_3 Z + \varepsilon_2 \quad \text{(supply equation)}$$

Corresponding reduced form

$$Q = \pi_{11} + \pi_{12} Y + \pi_{13} Z + u_1$$

$$P = \pi_{21} + \pi_{22} Y + \pi_{23} Z + u_2$$

with $\pi_{11} = (\alpha_1 \beta_2 - \alpha_2 \beta_1)/(\beta_2 - \alpha_2)$, $u_1 = (\beta_2 \varepsilon_1 - \alpha_2 \varepsilon_2)/(\beta_2 - \alpha_2)$, etc.

- Given values for Y and Z, values for Q and P can be calculated
- Parameter estimation:
  - Estimates from structural form parameters are biased, inconsistent; see above
  - Reduced form equations are a SUR model; FGLS estimates are consistent, asymptotically efficient

Simultaneous Equation Models: Estimation

Issues:

- Estimation problem: What methods can be applied to multi-equation model, what properties will the estimates have?
- Identification problem: Given estimates of reduced form parameters, can from them structural parameters be derived?

Types of Variables

Endogenous variables

- Determined by the model

Exogenous variables

- Determined from outside the model
- Types of exogenous variables
  - Strictly exogenous variables: uncorrelated with past, actual, and future error terms
  - Predetermined variables: uncorrelated with actual and future error terms, e.g., lagged endogenous variables

Complete system of equations: number of equations equals the number of endogenous variables

Structural and Reduced Form

Structural form: represents relations between endogenous variables and exogenous (and predetermined) variables according to economic theory

Reduced form: describes the dependence of endogenous variables upon exogenous or predetermined variables

Coefficients of

- Structural form: Interpretation as structural parameters corresponding to economic theory
- Reduced form: Interpretation as impact multiplicator, indicating the effects of changes of the predetermined variables on endogenous variables

Market Model: Structural Form

Structural form of the 2-equation model

$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{t1}$ (demand equation)

$Q_t = \beta_1 + \beta_2 P_t + \beta_3 Z_t + \varepsilon_{t2}$ (supply equation)

with $\varepsilon_t = (\varepsilon_{t1}, \varepsilon_{t2})'$: bivariate white noise with

$$V\{\varepsilon_t\} = \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Structural form in matrix notation:

$Ay_t = \Gamma z_t + \varepsilon_t$

with

$y_t = (Q_t, P_t)'$, $z_t = (1, Y_t, Z_t)'$

$$A = \begin{pmatrix} 1 & -\alpha_2 \\ 1 & -\beta_2 \end{pmatrix}, \Gamma = \begin{pmatrix} \alpha_1 & \alpha_3 & 0 \\ \beta_1 & 0 & \beta_3 \end{pmatrix}$$

Market Model: Reduced Form

Reduced form in matrix notation

$$y_t = A^{-1}\,\Gamma z_t + A^{-1}\varepsilon_t = \Pi z_t + u_t$$

with

$$\Pi = \begin{pmatrix} \dfrac{\alpha_1\beta_2 - \alpha_2\beta_1}{\beta_2 - \alpha_2} & \dfrac{\alpha_3\beta_2}{\beta_2 - \alpha_2} & \dfrac{-\alpha_2\beta_3}{\beta_2 - \alpha_2} \\ \dfrac{\alpha_1 - \beta_1}{\beta_2 - \alpha_2} & \dfrac{\alpha_3}{\beta_2 - \alpha_2} & \dfrac{-\beta_3}{\beta_2 - \alpha_2} \end{pmatrix}$$

and $\Omega = V\{u_t\} = A^{-1}\,\Sigma(A^{-1})'$

Reduced form equations:

$$Q_t = \pi_{11} + \pi_{12}Y_t + \pi_{13}Z_t + u_{t1}$$

$$P_t = \pi_{21} + \pi_{22}Y_t + \pi_{23}Z_t + u_{t2}$$

Multi-equation Model: General Structural Form

Model with m endogenous variables (and equations), K regressors

$$Ay_t = \Gamma z_t + \varepsilon_t$$

with m-vectors y_t and ε_t, K-vector z_t, (mxm)-matrix A, (mxK)-matrix Γ, and (mxm)-matrix $\Sigma = V\{\varepsilon_t\}$

Structure of the multi-equation model: (A, Γ, Σ)

Structural parameters: Elements of A and Γ

Normalized matrix A: $\alpha_{ii} = 1$ for all i

Complete multi-equation model: A is a square matrix with full rank, i.e., A is invertible

Recursive multi-equation model: A is a lower triangular matrix, i.e., y_{ti} can be regressor in equations j with $j > i$ only

Contents

- Systems of Regression Equations
- Multivariate Regression Models
- SUR Model: Estimation
- Simultaneous Equation Models
- Simultaneous Equation Models: Identification
- Identification: Criteria
- Single Equation Estimation Methods
- System Estimation Methods

Example: A Simple Market Model

Structural form of the 2-equation model

$$Q_t = \alpha_1 + \alpha_2 P_t + \varepsilon_{t1} \quad \text{(demand equation)}$$

$$Q_t = \beta_1 + \beta_2 P_t + \varepsilon_{t2} \quad \text{(supply equation)}$$

- Observations (Q_t, P_t), $t = 1, \ldots, T$,
  - A cloud of points in the scatter diagram
  - OLS estimation gives slope and intercept: not clear whether these parameters correspond to demand or supply equation
  - Demand and supply equations are "not identified"
- Reduced form equations

$$Q_t = \pi_{11} + u_{t1},\ P_t = \pi_{21} + u_{t2}$$

with $\pi_{11} = (\alpha_1\beta_2 - \alpha_2\beta_1)/(\beta_2 - \alpha_2)$, $\pi_{21} = (\alpha_1 - \beta_1)/(\beta_2 - \alpha_2)$

A Simple Market Model, cont'd

- Given estimates for π_{11} and π_{21}, the two equations

  $$\pi_{11} = (\alpha_1\beta_2 - \alpha_2\beta_1)/(\beta_2 - \alpha_2)$$
  $$\pi_{21} = (\alpha_1 - \beta_1)/(\beta_2 - \alpha_2)$$

  have no unique solution for four structural parameters α_1, α_2, β_1, and β_2
- "Identifying" the demand or supply equation from the data is linked to a unique solution for the structural parameters
- Both the demand and supply equations are not identified or unidentified

A Modified Market Model

$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{t1}$ (demand equation)

$Q_t = \beta_1 + \beta_2 P_t + \varepsilon_{t2}$ (supply equation)

Coefficients of the reduced form

$\pi_{11} = (\alpha_1\beta_2 - \alpha_2\beta_1)/(\beta_2 - \alpha_2)$, $\pi_{12} = \alpha_3\beta_2/(\beta_2 - \alpha_2)$

$\pi_{21} = (\alpha_1 - \beta_1)/(\beta_2 - \alpha_2)$, $\pi_{22} = \alpha_3/(\beta_2 - \alpha_2)$

with OLS estimates p_{ij}, i, j =1, 2, j=1, 2

- Supply equation: estimates for β_1, β_2 are uniquely determined

  $b_2 = p_{12}/p_{22}$, $b_1 = p_{11} - p_{21}b_2$

- Demand equation: only two equations for $\alpha_1, \ldots, \alpha_3$

  $a_1 = p_{11} - p_{21}a_2$, $a_3 = p_{22}(b_2 - a_2)$

  No unique solutions

- The supply equation is identified, the demand equation is unidentified

One more Market Model

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{t1} \quad \text{(demand equation)}$$
$$Q_t = \beta_1 + \beta_2 P_t + \beta_3 Z_t + \varepsilon_{t2} \quad \text{(supply equation)}$$

OLS estimates for reduced form parameters p_{ij}, i=1, 2, j=1, ..., 3, give estimates

$$a_2 = p_{13}/p_{23},\ b_2 = p_{12}/p_{22}$$

- The parameters of the demand equation are uniquely determined:

$$a_3 = p_{22}(b_2 - a_2),\ a_1 = p_{11} - p_{21}a_2$$

- The supply equation parameters are uniquely determined:

$$b_3 = -p_{23}(b_2 - a_2),\ b_1 = p_{11} - p_{21}b_2$$

- Both the supply equation and the demand equation are identified

Contents

- Systems of Regression Equations
- Multivariate Regression Models
- SUR Model: Estimation
- Simultaneous Equation Models
- Simultaneous Equation Models: Identification
- Identification: Criteria
- Single Equation Estimation Methods
- System Estimation Methods

Counting the Parameters

- Number of structural parameters:
  - A: $m \times m$ non-singular matrix, i.e., m^2 parameters
  - Γ: $m \times K$ matrix, i.e., mK parameters
  - Σ: $m \times m$ symmetric, positive definite matrix, i.e., $m(m+1)/2$ parameters
- Number of reduced form parameters:
  - Π: $m \times K$ matrix, i.e., mK parameters
  - Ω: $m \times m$ symmetric, positive definite matrix , i.e., $m(m+1)/2$ parameters

Number of structural parameters exceeds that of reduced form parameters by m^2

- Identification requires further information such as restrictions for parameters

Identification: Parameter Restrictions

Restrictions on structural parameters: reduce the number of parameters to be estimated, so that equations are identified

- Normalization: in each structural equation, one coefficient is a “1”
- Exclusion: the omission of a regressor in an equation results in a zero in A or Γ, i.e., reduces the number of structural parameters
- Identities, like equations 4 through 6 in Klein’s I model, reduce the number of structural parameters to be estimated
- Linear – or non-linear – restrictions on structural parameters, restrictions on the elements of Σ also, reduce the number of structural parameters to be estimated

Check of identification

- Order condition
- Rank condition

Order Condition

Model with m endogenous variables (and equations), K regressors

$$\mathrm{A}y_t = \Gamma z_t + \varepsilon_t$$

- Equation j:
  - m_j: number of explanatory endogenous variables
  - m_j^*: number of excluded endogenous variables ($m_j^* = m - m_j - 1$)
  - K_j^*: number of excluded exogenous variables ($K_j^* = K - K_j$)
- Order Condition: Equation j is identified if

  $$K_j^* \geq m_j$$

  i.e., the number of exogenous variables excluded from equation j is at least as large as the number of explanatory endogenous variables included in the equation
- The order condition is a necessary but not sufficient condition for identification

Market Model

Model:

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{t1} \quad \text{(demand equation)}$$

$$Q_t = \beta_1 + \beta_2 P_t + \varepsilon_{t2} \quad \text{(supply equation)}$$

$m = 2$ (Q, P), $K = 2$ (1 for the intercept, Y)

- Supply equation ($j = 2$):

  $$m_2^* = 0,\ m_2 = 1,\ K_2^* = 1,\ K_2 = 1$$

  Order condition is fulfilled: $K_2^* = 1 = m_2 = 1$; the supply equation is identified

- Demand equation ($j = 1$):

  $$m_1^* = 0,\ m_1 = 1,\ K_1^* = 0,\ K_1 = 2$$

  Order condition is not fulfilled: $K_1^* = 0 < m_1 = 1$; the demand equation is not identified

Rank Condition

Model with m endogenous variables (and equations), K regressors

$$Ay_t = \Gamma z_t + \varepsilon_t$$

- Equation j:
  - A*: obtained by deleting from A the j-th row and all column with a non-zero element in the j-th row
  - Γ*: obtained by deleting from Γ the j-th row and all column with a non-zero element in the j-th row
- Rank Condition: Equation j is identified if

  $$r(A^* | \Gamma^*) \geq m - 1$$

  i.e., the rank of the matrix $(A^* | \Gamma^*)$ is at least as large as the number of endogenous variables minus 1
- The order condition is a sufficient condition for identification of equation j

IS-LM Model

$$C_t = \gamma_{11} - \alpha_{14} Y_t + \varepsilon_{t1}$$
$$I_t = \gamma_{21} - \alpha_{23} R_t + \varepsilon_{t2}$$
$$R_t = -\alpha_{34} Y_t + \gamma_{32} M_t + \varepsilon_{t3}$$
$$Y_t = C_t + I_t + Z_t$$

C: consumption, *I*: investments, *R*: interest rate, *Y*: production/income, *M*: money, *Z*: autonomous expenditures

endo.: C,I,R,Y (m=4); exo.: $1,M,Z$ (K=3)

- Equation 1:
  - Order condition: $K_1^* = 2 \geq m_1 = 1$
  - Rank condition: $r(A^*|\,\Gamma^*) = 3 \geq m - 1 = 3$

$$\begin{pmatrix} A & \Gamma \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \alpha_{14} & \gamma_{11} & 0 & 0 \\ 0 & 1 & \alpha_{23} & 0 & \gamma_{21} & 0 & 0 \\ 0 & 0 & 1 & \alpha_{34} & 0 & \gamma_{32} & 0 \\ -1 & -1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} A^* & \Gamma^* \end{pmatrix} = \begin{pmatrix} 1 & \alpha_{23} & 0 & 0 \\ 0 & 1 & \gamma_{32} & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

Both conditions are fulfilled, equation 1 is identified

Identification Checking: The Practice

1. A multi-equation model is identified if each equation is identified
2. Most equations which fulfill the order condition also fulfill the rank condition
3. Identification checking for small models is usually easy; equations of large models usually are identified (large models contain large numbers of predetermined variables)
4. Addition of an equation to an identified model: the resulting model is identified if the new equation contains at least one additional variable

Identification: More Notation

Equation j is

1. Exactly identified: $K_j^* = m_j$ and rank condition is met
2. Overidentified: $K_j^* > m_j$ and rank condition is met
3. Underidentified: $K_j^* < m_j$ or rank condition fails

Contents

- Systems of Regression Equations
- Multivariate Regression Models
- SUR Model: Estimation
- Simultaneous Equation Models
- Simultaneous Equation Models: Identification
- Identification: Criteria
- Single Equation Estimation Methods
- System Estimation Methods

Simultaneous Equation Models: Estimation Methods

1. Single equation methods, also limited information methods
   - Indirect least squares estimation (ILS)
   - Two stage least squares estimation (2SLS or TSLS)
   - Limited information ML estimation (LIML)
2. (Complete) system methods, also full information methods
   - Three stage least squares estimation (3SLS)
   - Full information ML estimation (FIML)

The Modified Market Model

Estimator for β_2 from

$$Q_t = \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{t1} \quad \text{(demand equation)}$$

$$Q_t = \beta_2 P_t + \varepsilon_{t2} \quad \text{(supply equation)}$$

with contemporaneously correlated error terms

T-vectors *p* and *q;*

- OLS estimate for β_2 from the supply equation: $b_2 = (p'p)^{-1}p'q$; is biased
- IV estimate for β_2 with instrumental variable *Y*: $b_2^{IV} = (y'p)^{-1}y'q$; is consistent
- ILS estimate: $b_2^{ILS} = p_2/p_1 = (y'p)^{-1}y'q$

  with OLS estimates p_1 and p_2 for π_1 and π_2 from the reduced form

$$P = \pi_1 Y + u_1$$

$$Q = \pi_2 Y + u_2$$

Modified Market Model, cont'd

4. 2SLS estimate for β_2 of the supply equation
   - Step 1: Regression of the explanatory variable *P* on the instrumental variable *Y*, calculation of fitted values

     $\hat{p}$ = $[(y'y)^{-1}y'p]\, y$
   - Step 2: OLS estimation of β_2 from $Q_t = \beta_2 \hat{P}_t + v_t$

     $$b_2^{2SLS} = (\hat{p}'\hat{p})^{-1}\hat{p}'q$$

OLS Estimation

OLS estimators of structural parameters: in general

- biased
- not consistent

- But often a feasible alternative
  - OLS estimator is efficient, i.e., has minimal variance; may be a good estimator in spite of unbiasedness
  - Tends to be robust against not fulfilled assumptions
  - May be advantageous for small or moderate sample sizes; not depending upon asymptotics
- OLS estimators for parameters of recursive simultaneous equation models: asymptotically unbiased
- OLS technique: important procedure in all estimation methods for simultaneous equation models

Indirect Least Squares (ILS) Estimation

Model with *m* endogenous variables (and equations), *K* regressors

- Structural form

  $$Ay_t = \Gamma z_t + \varepsilon_t,\ V\{\varepsilon_t\} = \Sigma$$

- Reduced form

  $$y_t = A^{-1}\Gamma z_t + A^{-1}\varepsilon_t = \Pi z_t + u_t,\ V\{u_t\} = \Omega$$

Estimation of the structural parameters of equation *j*:

  - Step 1: OLS estimation of reduced form parameters Π
  - Step 2: Calculation of estimates for structural parameters, solving $A\Pi = \Gamma$ for the structural parameters of equation *j*

- Estimation of structural parameters of equation *j*: equation *j* needs to be identified

Two Stage Least Squares (2SLS) Estimation

Estimation of structural parameters of equation j

$$y_j = X_j\beta_j + \varepsilon_j = Y_j\alpha_i + Z_j\gamma_j + \varepsilon_j$$

with X_j: $[T\text{x}(m_j\text{-}1+K_j)]$-matrix of explanatory variables, Y_j: $[T\text{x}(m_j\text{-}1)]$-matrix of explanatory endogenous variables, Z_j: $(T\text{x}K_j)$-matrix of exogenous variables

2SLS (or TSLS) estimation in two steps:

- Step 1: OLS estimation of reduced form parameters Π, calculation of predictions $\hat{Y}_j$
- Step 2: OLS estimation of structural parameters β_j, using

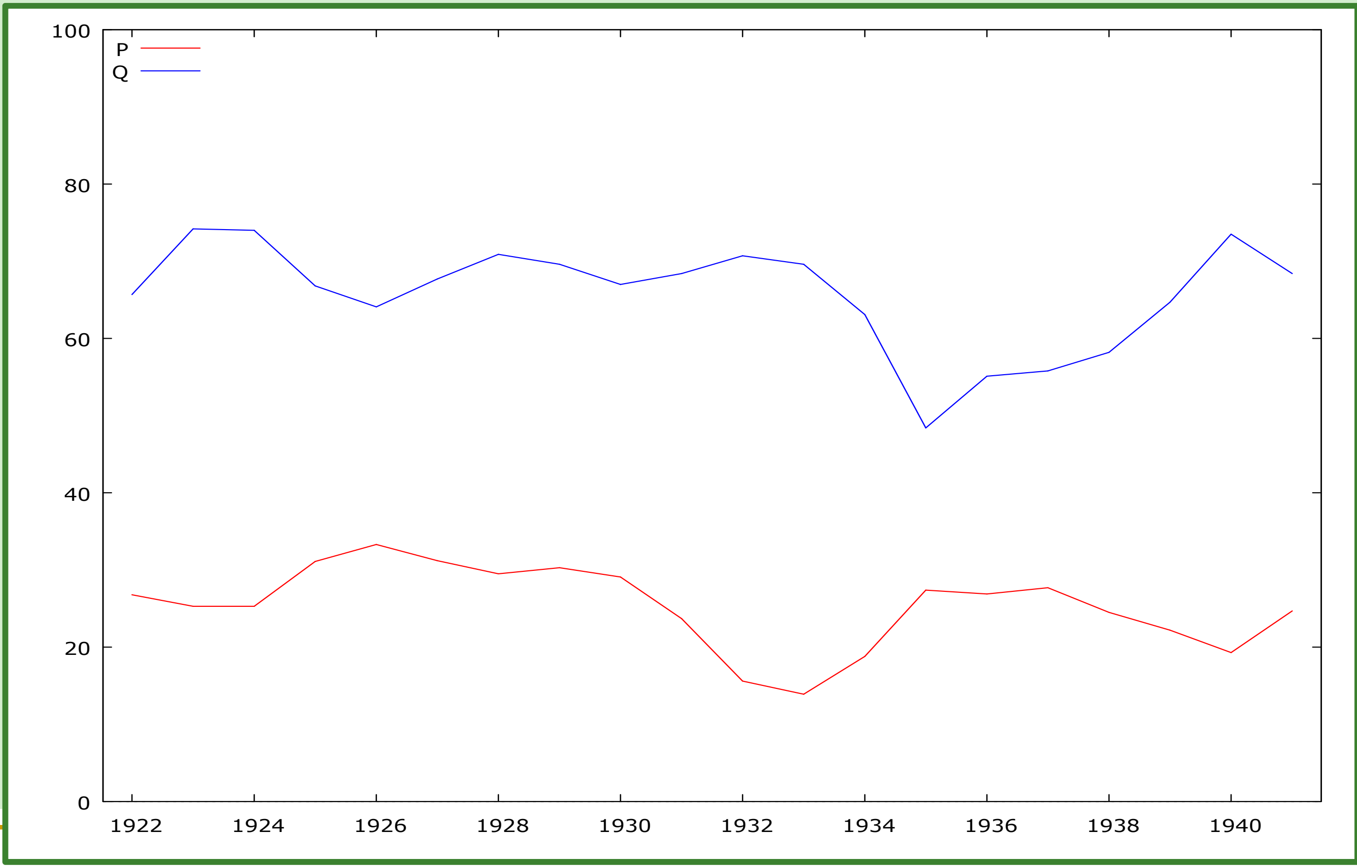
$$y_j = \hat{X}_j\beta_j + v_j$$

with $\hat{X}_j = (\hat{Y}_j\ Z_j)$

2SLS (or TSLS) estimation of structural parameters of equation j requires the equation to be identified

Example: Hog Market

US hog market 1922-1941 (Merill & Fox, 1971): *P*: retail price for hog (US cents p.lb.), *Q*: hog-consumption p.c., *Y*: income p.c. (USD), *Z*: exogenous production factor



Hog Market: The Model

Model with endogenous variables Q, P, exogenous variables Y, Z

$$Q_t = \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \varepsilon_{t1} \quad \text{(demand equation)}$$

$$Q_t = \beta_1 + \beta_2 P_t + \beta_3 Z_t + \varepsilon_{t2} \quad \text{(supply equation)}$$

both equations are exactly identified

Coefficients of demand and supply equation estimated by three single equation methods

- Separate OLS estimation of both equations
- ILS estimation
- 2SLS estimation

Example: Hog Market

Comparison of three single equation estimation methods

- Strong coincidence of ILS and 2SLS estimates
- OLS estimates of demand equation deviates substantially from ILS and 2SLS estimates

		Demand			**Supply**		
		const	*P*	*Y*	const	*P*	*Z*
OLS	*coeff*	56.962	**-1.410**	**0.080**	15.355	-0.030	**0.744**
	p-val	8.7e-9	0.002	0.003	0.004	0.701	1.4e-10
2SLS	coeff	60.885	**-3.088**	**0.149**	8.318	0.177	**0.770**
	p-val	4.2e-9	0.001	0.000	0.241	0.222	2.9e-24
ILS	*coeff*	60.885	**-3.088**	**0.149**	8.318	0.177	**0.770**

2SLS Estimator: Properties

2SLS (or TSLS) estimation of structural parameters of equation *j* requires the equation to be identified

- Order condition $K_j^* \geq m_j$: number of excluded exogenous variables (K_j^*) is at least the number of explanatory endogenous variables (m_j)
- i.e., the number of potential instrumental variables is at least the number of variables to be substituted by predictions

Properties: 2SLS estimators are

- Consistent
- Asymptotically normally distributed

LIML Estimator

Limited information ML (LIML) estimation:

- Maximization of the likelihood function derived from the system of
  - one structural equation
  - reduced form equations for the remaining endogenous variables
  - Assumes normally distributed error terms

Application:

- Asymptotic distribution of LIML estimator equivalent to that of the 2SLS estimator
- Wrt computational effort, 2SLS estimation is much easier to use
- In practical applications, LIML estimation is hardly used

Contents

- Systems of Regression Equations
- Multivariate Regression Models
- SUR Model: Estimation
- Simultaneous Equation Models
- Simultaneous Equation Models: Identification
- Identification: Criteria
- Single Equation Estimation Methods
- System Estimation Methods

Why System Estimation Methods?

Single equation (limited information) estimation methods ignore the contemporaneous correlation of error terms

System (full information, complete system) estimation methods take contemporaneous correlation of error terms into account

- Estimation of equation parameters is more efficient: estimation of coefficients of equation *j* makes use information contained in other equations
- Estimation methods
  - 3SLS estimation
  - Iterative 3SLS estimation
  - Full information ML (FIML) estimation

3SLS Estimator

The *m* equations of the full model in matrix notation

$$\bar{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix} = \bar{X}\bar{\beta} + \bar{\varepsilon}$$

with $\mathrm{V} = \mathrm{V}\{\bar{\varepsilon}\} = \Sigma \otimes I_n = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{1m} & \cdots & \sigma_m^2 \end{pmatrix} \otimes I_n$

3SLS estimation: FGLS estimation based on

- 2SLS residuals for each of the *m* equations
- estimate for Σ obtained from these residuals

3SLS Estimator: Three Steps

The steps of the 3SLS estimation are

1. Based on the reduced form equations, calculation of predicted values for all explanatory endogenous variables; cf. the first stage of 2SLS estimation
2. For the *j*-th equation, $j = 1, \ldots, m$,
   - Calculation of 2SLS estimators b_j and
   - 2SLS residuals $e_j = y_j - X_j b_j$

   Estimation of elements $\sigma_{ij} = \mathrm{Cov}\{\varepsilon_{ti}, \varepsilon_{tj}\}$ of Σ: $s_{ij} = (e_j' e_j)/T$
3. Calculation of the 3SLS estimator

$$b^{3SLS} = [\bar{X}'(S^{-1} \otimes P_z)\bar{X}]^{-1} \bar{X}'(S^{-1} \otimes P_z)\bar{y}$$

with the projection matrix $P_z = Z(Z'Z)^{-1}Z'$, S the estimated matrix Σ

3SLS Estimator: Properties

3SLS estimation requires all equations of the system to be identified

Properties: 3SLS estimators are

- Consistent
- Asymptotically normally distributed

3SLS estimates coincide with 2SLS estimates if

- All equations are exactly identified
- The error terms are contemporaneously uncorrelated, i.e., Σ is diagonal

Example: Hog Market

Comparison of 2SLS and 3SLS estimation

- Strong coincidence of 3SLS and 2SLS estimates
- Smaller *p*-values of most 3SLS estimate indicate higher efficiency

		Demand			**Supply**		
		const	*P*	*Y*	const	*P*	*Z*
3SLS	*coeff*	60.885	**-3.088**	**0.149**	8.318	0.177	**0.770**
	p-val	1.8e-10	0.002	6.9e-5	0.204	0.186	2.9e-28
2SLS	coeff	60.885	**-3.088**	**0.149**	8.318	0.177	**0.770**
	p-val	4.2e-9	0.001	0.000	0.241	0.222	2.9e-24

More System Estimators

Iterative 3SLS estimator

- 3SLS estimates b^{3SLS} of structural parameters (or A^{3SLS} and Γ^{3SLS}) give
  - revised reduced form parameters ($A^{-1}\Gamma = \Pi$) and
  - predictions of the explanatory endogenous variables;
- Iterative 3SLS estimator: starting with an initial 3SLS estimator, the following iterations are repeatedly executed until convergence is reached
  - Outer iteration: step 1 of 3SLS estimation resulting in improved predictions of the predetermined variables,
  - Inner iteration: step 2 of 3SLS estimation, resulting in improved 2SLS residuals and estimate *S* for Σ, and step 3, resulting in improved 3SLS estimators b^{3SLS}
  - The inner iteration can be repeated using 3SLS residuals for estimate *S*

More System Estimators, cont'd

Full information ML (FIML) estimator:

- Assumes normally distributed error terms
- Maximizes likelihood function with respect to structural parameters

Simultaneous Equation Models in GRETL

Model > Simultaneous Equations …

- choice of estimator
  - SUR
  - 2SLS
  - LIML
  - 3SLS
  - FIML
- Specification of equations, instrumental variables, endogenous variables, and identities

Your Homework

Klein's model I consists of the following equations (see the GRETL data file "klein"):

$C_t = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \alpha_4 W_t + \varepsilon_{t1}$ (consumption)

$I_t = \beta_1 + \beta_2 P_t + \beta_3 P_{t-1} + \beta_4 K_{t-1} + \varepsilon_{t2}$ (investment)

$W_t^p = \gamma_1 + \gamma_2 X_t + \gamma_3 X_{t-1} + \gamma_4 t + \varepsilon_{t3}$ (wages)

$W_t = W_t^p + W_t^g$

$X_t = C_t + I_t + G_t$

$K_t = I_t + K_{t-1}$

$P_t = X_t - W_t^p - T_t$

Endogenous variables are: *C I Wp X W K P*

1. Which of the equations are identified? Use (a) order and (b) rank conditions in answering the question.
2. Estimate the structural parameters using (a) OLS, (b) SUR, (c) 2SLS, (d) 3SLS, and (e) FIML; compare the results and explain pros and cons of the methods.

Your Homework, cont'd

3. The goodness of fit measure

$$R_I^2 = 1 - \frac{S_g(\tilde{\bar{\beta}})}{S_g(0)} = 1 - \frac{\mathrm{tr}(S^{-1}\tilde{\Sigma})}{\mathrm{tr}(S^{-1}S_{yy})}$$

makes use of

$$S_g(b) = (\bar{y} - \bar{X}\tilde{\bar{b}})'\tilde{V}^{-1}(\bar{y} - \bar{X}\tilde{\bar{b}}) = T\,\mathrm{tr}(S^{-1}\tilde{\Sigma}),\ \tilde{\Sigma} = (\tilde{E}'\tilde{E})/T$$

with the Txm matrix $\tilde{E} = (\tilde{e}_1, \dots, \tilde{e}_m)$ of FGLS residuals and analogously the mxm matrix S_{yy} of sample covariances of the y_{ti}. Show for $T = m = 2$ that the two numerators in the definition of R_I^2, $S_g(.)$ and tr(.), coincide.