BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title : Econometrics

Course Code: 23ST03DEC

Unit-V

Official Statistics

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. E. Devi
Guest Faculty
Department of Statistics

UNIT - 5

K - Class Estimators

The K-class estimator is a family of econometric techniques that encompasses Ordinary Least Squares (OLS), Two-Stage Least Squares (2SLS), and Limited Information Maximum Likelihood (LIML) by varying a parameter k. It provides a flexible framework for handling endogeneity in simultaneous equation models, where the k value determines the estimator's position on a spectrum between the biased but consistent OLS and the more complex but asymptotically efficient 2SLS and LIML methods.

A Range of Estimators:

The K-class estimator is defined by a parameter, k, which can be a deterministic value or a stochastic (data-dependent) one.

Special Cases:

OLS (k=0) : When k equals zero, the K-class estimator reduces to Ordinary Least Squares.

2SLS (k=1) : When k equals one, it becomes the Two-Stage Least Squares (2SLS) estimator.

LIML (k < 1 but > 0) : The Limited Information Maximum Likelihood (LIML) estimator is also a member of the K-class, with a k value typically less than one.

Handling Endogeneity:

In the presence of endogeneity (where endogenous variables are correlated with the error term), K-class estimators provide a way to obtain consistent estimates by using instrumental variables. The choice of k affects the bias-variance trade-off, with values between 0 and 1 balancing the properties of OLS and 2SLS.

The formula for the k-class estimator can be expressed as:

- $\hat{\beta}_k = ((1-k)X'X + kX'\hat{X})^{-1}((1-k)X'y + kX'\hat{y})$
- Where:
  - y is the vector of the dependent variable
  - X is the matrix of independent variables
  - $\hat{X}$ is the matrix of first-stage predicted values from the reduced form
  - k is the k-class parameter

Full information estimator

A "full information estimator" is a method of statistical estimation that uses all available data points, including those with missing values, to produce the best possible parameter estimates for a statistical model

Introduction

In econometrics and statistical estimation, researchers often deal with complex models involving multiple equations that are interdependent. In such cases, estimating each equation separately may lead to inefficient or biased results because it ignores the interconnections among the equations. To overcome this limitation, Full Information Estimators (FIE) are used. These estimators consider the entire system of equations simultaneously, making use of all available information in the model.

The concept of full information estimation is particularly relevant in simultaneous equation models (SEMs), where endogenous variables appear as explanatory variables in other equations. Unlike limited information methods that estimate one equation at a time, full information methods treat the model as a whole.

Concept and Theoretical Framework

A Full Information Estimator refers to any estimation method that utilizes all available information from the entire system of equations in a model. The best-known example is the Full Information Maximum Likelihood (FIML) estimator.

In general terms, suppose we have a system of simultaneous equations:

$$[Y = XB + U]$$

where,

(Y) is a matrix of endogenous variables,

(X) is a matrix of exogenous variables,

(B) is a matrix of parameters,

(U) is the error term.

Full information estimation involves specifying the joint probability distribution of all the error terms and using it to estimate all parameters in the system simultaneously. This differs from Limited Information Estimation (LIML), which focuses only on one structural equation at a time.

Types of Full Information Estimators

Full Information Maximum Likelihood (FIML) :_The most common full information method. FIML estimates all equations together by maximizing the likelihood function of the joint distribution of the disturbance terms. It assumes that the error terms are normally distributed and that the model is correctly specified.

Three-Stage Least Squares (3SLS) : An extension of the Two-Stage Least Squares (2SLS) estimator, 3SLS combines the efficiency of Generalized Least Squares (GLS) with system-wide information, leading to more efficient parameter estimates when the error terms are correlated across equations.

Advantages of Full Information Estimators

- They are asymptotically efficient, meaning they make optimal use of available data as the sample size increases.
- They account for correlation among error terms across equations, improving estimation accuracy.
- They provide system-wide consistency in estimated parameters, ensuring that all equations fit together logically.
- Useful for policy analysis where simultaneous effects between variables are significant (e.g., supply and demand models, macroeconomic systems).

Limitations

- Full information methods require correct model specification. If even one equation is misspecified, the entire system's estimates may become inconsistent.
- They are computationally intensive, especially for large systems.
- The assumption of normal distribution of errors may not always hold in practice.
- Sensitive to outliers and data irregularities.

Comparison with Limited Information Estimators

Aspect	Full Information (FIML/3SLS)	Limited Information (LIML/2SLS)
Information used	All equations jointly	One equation at a time
Efficiency	High (if correctly specified)	Lower
Computational complexity	High	Moderate
Sensitivity to specification	Very sensitive	Less sensitive

.

Full information maximum likelihood (FIML)

- Full Information Maximum Likelihood (FIML) is a statistical method for handling missing data in a dataset by maximizing the likelihood function across all observed data, even when some data points are missing.
- FIML is commonly used in structural equation modeling (SEM) and assumes multivariate normality for accurate results

Full information definition

In statistics, "full information" refers to complete and comprehensive data that is collected, organized, and analyzed to provide a deep understanding of a phenomenon or population, allowing for robust conclusions and reliable decision-making

Maximum likelihood definition

- Maximum Likelihood (ML) is a statistical method for estimating the parameters of a distribution by finding the parameter values that maximize the probability of observing the given data.
- It does this by treating the probability of the observed data as a function of the unknown parameters, known as the Likelihood function.
- The parameter values that make this function's output the highest are chosen as the maximum likelihood estimates.

The principle of fiml

- **Maximizing the Likelihood:**
  FIML finds the parameter values that make the observed data as likely as possible under the assumed statistical model.
- **Utilizing All Information:**

Instead of excluding observations with missing values (like listwise delection), FIML uses all available data, calculating likelihoods for each observation based on the observed subset of its variables.

- **Multivariate Normality:**
  The method generally assumes that the data are multivariate normal and linearly related.

Full Information Maximum Likelihood (FIML)

- **Data Utilization:**

  Uses all available information, including partially missing data, to estimate the model's parameters.

- **Method:**

  Constructs a likelihood function for the entire dataset and finds the parameter values that maximize this function.

- **Advantages:**

  More efficient and accurate estimates, as it leverages all data.
  Produces consistent and asymptotically unbiased estimates when data is missing at random (MAR).
  Can provide more accurate fit statistics.

- **Disadvantages:**

  More computationally intensive and complex.
  Assumes multivariate normality for its theoretical properties; if data is not normal, other robust standard errors may be necessary.

FORMULA

$$l(\theta) = \sum_{i}^{1} \ln f(w_i\ ;\theta)$$

PROBLEM

A study collects values for two variables, **X** and **Y**, from 6 individuals. Some values are missing, as shown in the table:

Person	X	Y
1	4	6
2	5	7
3	6	–
4	–	8
5	7	9
6	6	–

Assume **(X, Y)** follow a bivariate normal distribution with unknown means $\boldsymbol{\mu_x}$**,** $\boldsymbol{\mu_Y}$ and variances $\boldsymbol{\sigma_x^2}$**,** $\boldsymbol{\sigma_Y^2}$.

Using Full Information Maximum Likelihood (FIML):

1. Find the maximum likelihood estimates for $\boldsymbol{\mu_x}$**,** $\boldsymbol{\mu_Y}$**,** $\boldsymbol{\sigma_x^2}$**,** and $\boldsymbol{\sigma_Y^2}$**.**

Solution

Maximum Likelihood Estimates

The bivariate normal density for complete cases and marginal normal for univariate cases is used.
For parameter estimation:

Step-by-step calculations:

- **X values observed:** 4, 5, 6, 7, 6
- **Y values observed:** 6, 7, 8, 9

Means:

$\hat{\mu}X \;=\; 4+5+6+7+6 / 5 = 5.6$

$\hat{\mu}Y \;=\; 6+7+8+9 / 4 = 7.5$

Variances:

For X:

$\hat{\sigma}X^2 \;=\; 1/5\ [(4-5.6)^2+(5-5.6)^2+(6-5.6)^2+(7-5.6)^2+(6-5.6)^2]$

CALCULATION

$(4-5.6)^2 = 2.56$

$(5.56)^2 = 0.36$

$(6-5.6)^2 = 0.16$

$(7-5.6)^2 = 1.96$

$(6-5.6)^2 = 0.16$

Sum = 2.56+0.36+0.16+1.96+0.16

= 5.2

$\hat{\sigma}X^2 = 5.2 / 5$

= 1.04

For Y:

$\hat{\sigma}Y^2 = 1/4\ [(6-7.5)^2 + (7-7.5)^2 + (8-7.5)^2 + (9-7.5)^2\ ]$

CALCULATION

$(6-7.5)^2 = 2.25$

$(7-7.5)^2 = 0.25$

$(8-7.5)^2 = 0.25$

$(9-7.5)^2 = 2.25$

Sum = 2.25 + 0.25 + 0.25 + 2.25

= 5

$\hat{\sigma}Y^2 = 5/4$

= 1.25

MLEs

$\hat{\mu}X = 5.6$

$\hat{\mu}Y = 7.5$

$\hat{\sigma}X^2 = 1.04$

$\hat{\sigma}Y^2 = 1.2$

Three Stage Least Squares [3SLS]: An Efficient Estimator:

The 2SLS method does not exploit the correlation of the disturbances across 2SLS equations. If the disturbances in the structural equations of the disturbances across 2SLS estimator is not even consistent. Zellner and Theil developed a method which considers all the equation of a model known as Three Stage Least Squares estimation method. This method has greater efficiency than 2SLS method under certain circumstances. This method was developed as a logical extension of Theil's 2SLS.

It involves the method of least square application in three successive stages. 3 SLS are more efficient than 2 SLS, since in their estimation we use more information than in 2 SLS. The 3 SLS method gives consistent estimators, though they are biased. However, this method requires better specification of the model and more data, since all parameters are estimated simultaneously. It is also very sensitive to specification errors, and is computationally more cumbersome. Hence, if there is no specific reasons for using the method, 2 SLS should be preferred than 3 SLS. Several other methods such as limited information maximum likelihood and full information maximum likelihood exist for estimating the parameters of a simultaneous equation system, but a discussion of these is beyond the scope of this book. In order to apply 3SLS method the following points should be considered while making calculations.

- Definitional equations (identities) should be left out.
- Unidentified equations should also be omitted.
- Just and or over identified equations should be included. It is computationally efficient to apply 3SLS to each of these groups separately.

Structural Model is

$$M = \alpha_0+\alpha_1 P+\alpha_2 N+\epsilon$$

$$P = \beta_0+\beta_1 M+\beta_2 R+\mu$$

Where,

M = Money supply

P = Price level

N = Rate of interest

R = Income

M and P are endogenous variables

R and N are exogenous variables

If we knew μ was high, we would know that ε is probably also high, and hence M is higher than predicted from P and N only. Similarly for μ and P.

The reduced form equations are:

$$M=\Pi_0+\Pi_1 R+\Pi_2 N+\nu$$
$$P=\Pi_3+\Pi_4 R+\Pi_5 N+u$$

Estimate reduced form equations. Obtain $\widehat{M}$,$\hat{P}$

Regress structural (original) equations, replacing endogenous explanatory variables M and P with predicted values $\widehat{M}$ and $\hat{P}$and save residuals from these regressions, labeled $\hat{\varepsilon}$ and $\hat{\mu}$

Reestimate structural equations with $\hat{\varepsilon}$ and $\hat{\mu}$ included as explanatory variables.

$$M=\alpha_0+\alpha_1\hat{P}+\alpha_2 N+\delta\hat{\mu}+\varepsilon{*}{*}$$

$$P=\beta_0+\beta_1\widehat{M}+\beta_2 R+\rho\hat{\varepsilon}+\mu{*}{*}$$

Because ε and μ are correlated, $\hat{\mu}$ provides information for explaining M and $\hat{\varepsilon}$ provides information for explaining P. Including this information makes the estimates better.

In simultaneous equation models (SEMs), multiple equations are estimated together because endogenous variables appear on both sides of the equations.

Ordinary Least Squares (OLS) fails here because the repressors are correlated with the error term — causing biased and inconsistent estimates.

Therefore, several special estimation methods have been developed to handle such systems accurately.

These include:

- K-Class Estimators
- Full Information Estimators
- Full Information Maximum Likelihood (FIML)
- Three-Stage Least Squares (3SLS)

K-Class Estimators:

The K-Class Estimator is a generalization of the Two-Stage Least Squares (2SLS) estimator, which is commonly used in the context of instrumental variables (IV) regression. The K-Class Estimator allows for a more flexible approach depending on the choice of the parameter k.

The general formula for a K-Class estimator is:

$$\hat{\beta}_K = (X'X + KZ'Z\,)'X'Y$$

Where:

X is the matrix of explanatory variables (repressors),

Y is the vector of dependent variables (outcome),

Z is an instrumental variable matrix (optional, used for IV estimation),

K is a scalar parameter that determines the type of estimator.

When K=0K = 0K=0, you get the Ordinary Least Squares (OLS) estimator.

Full Information Estimator (FIE)

A Full Information Estimator is an econometric method that estimates all parameters of a system of simultaneous equations by using all available data from the entire system, considering the interrelationships among all endogenous variables. Unlike limited information methods, which focus on individual equations, full information estimators analyze the system as a whole.

ORDINARY LEAST SQUARES (OLS)

- **Method**: OLS is the simplest and most widely used estimation method. It minimizes the sum of squared residuals (the differences between observed and predicted values) to estimate the parameters of the model.
- **Formula:**

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

- **Assumption:** No endogeneity — explanatory variables are uncorrelated with the error term.

TWO-STAGE LEAST SQUARES (2SLS)

Method: Used when explanatory variables are endogenous.

1. **First Stage:** Regress endogenous variables on instruments Z to get fitted values:

$$\hat{X}_{IV} = P_Z X = Z(Z'Z)^{-1}Z'X.$$

2. **Second Stage:** Regress dependent variable on fitted values:

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y$$

- **Assumption:** Instruments Z must be correlated with endogenous repressors but uncorrelated with the error term.

FULL INFORMATION METHODS (LIKE FIML)

These methods analyse all the variables and their relationships simultaneously. It fits the entire proposed model to the entire raw dataset at once, making the most complete use of the available information.

HOW FIML WORKS

FIML operates based on the principles of Maximum Likelihood Estimation (MLE).

1. The Goal of Maximum Likelihood

The goal of any MLE is to find the parameter values (e.g., regression coefficients, means, variances) that make the observed data "most probable" or "most likely."

2. The Likelihood Function for Complete Data

If you had a complete dataset with no missing values, you would compute a single likelihood value for the entire dataset, given a set of parameters.

3. The FIML Adjustment for Missing Data

FIML modifies this for incomplete data. It doesn't have one single, neat dataset. Instead, it constructs the overall likelihood case-by-case.

- The variables that are present for that individual
- It does this by using the model's assumptions (e.g., that the data are multivariate normal) to "marginalize" over the missing variables. In simpler terms, it focuses the probability calculation on the dimensions where data exists.

4. Combining the Information

The individual likelihoods for all cases are then multiplied together (or their log-likelihoods are summed) to form the overall likelihood function for the entire sample.

5. Estimation

The software then iteratively searches for the parameter values that maximize this overall likelihood function. The result is a set of parameter estimates that are the "best guess" given every single observed data point.

THREE-STAGE LEAST SQUARES (3SLS)

The Three-Stage Least Squares (3SLS) method is a system estimation technique used in simultaneous equation models (SEMs). It is an extension of the Two-Stage Least Squares (2SLS) method that also accounts for correlations among the error terms of different equations.

STEPS INVOLVED IN THE 3SLS

STEP 01: Reduced Form Estimation

- Estimate each endogenous variable as a function of all exogenous variables (like in 2SLS).
- Get predicted values for endogenous variables ($\hat{Y}_1$, $\hat{Y}_2$).

$$Y_1 = \pi_{10} + \pi_{11}X_1 + \pi_{12}X_2 + v_1$$

$$Y_2 = \pi_{20} + \pi_{21}X_1 + \pi_{22}X_2 + v_2$$

STEP 02: Structural Equation Estimation

- Substitute the predicted values ($\hat{Y}_1$, $\hat{Y}_2$) into the structural equations.
- Estimate each equation using OLS or 2SLS to obtain consistent estimates

STEP 03: Generalized least square (GLS) adjustment

- Since disturbances $\mathcal{U}_1$ *and* $\mathcal{U}_2$ may be correlated, we use GLS to make the estimates efficient.
- Use the estimated covariance matrix of residuals from Stage 2 to perform GLS across all equations.

$$\hat{\Sigma}_u = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Then estimate the full system jointly using this covariance information.

COMPARISON TABLE:

Criteria	K-Class Estimator	Full Information Maximum Likelihood (FIML)	Full Information Least Squares (FILS)	Three-Stage Least Squares (3SLS)
Basic Idea	A general class of estimators including OLS, 2SLS, and LIML as special cases	Estimates all equations in the system jointly by maximizing the likelihood function	Estimates all equations jointly using least squares principle considering full system	Combines 2SLS with SUR (Seemingly Unrelated Regression) to account for correlation across equations

Information Used	Uses sample data from relevant equation(s)	Uses full information from the entire system of equations	Uses full information (like FIML) but minimizes squared errors	Uses all equations' information including cross-equation error correlations
Type of Estimation	Limited information (focuses on single equation at a time)	Full information (joint estimation)	Full information (joint estimation)	Limited information but system-based
Estimator Type	Linear (depends on choice of K)	Nonlinear (requires iterative estimation)	Linear (joint least squares)	Linear (system GLS-type)
Computation Complexity	Simple to moderate	High (requires iteration and convergence)	Moderate to high	Moderate
Efficiency	Depends on choice of K (LIML more efficient than 2SLS)	Most efficient under correct model specification	Less efficient than FIML	More efficient than 2SLS; close to FIML
Bias	Consistent if instruments valid	Consistent and asymptotically efficient	Consistent	Consistent and asymptotically efficient
When to Use	When you need flexibility or to compare with 2SLS/LIML	When full model and all equations are correctly specified	When FIML is complex, but system info is needed	When equations have correlated disturbances
Assumptions	Valid instruments	Correct specification of	System identified and	Valid instruments and

	and identification	entire system, normality of errors	errors homoscedastic	correlated error terms
Examples of Special Cases	OLS (K=0), 2SLS (K=1), LIML (K=λ)	–	–	2SLS + SUR combined
Advantages	Flexible, includes several estimators as special cases	Statistically most efficient	Easier than FIML	Accounts for cross-equation correlations
Disadvantages	Efficiency depends on K choice	Computationally demanding	Less efficient	Sensitive to instrument quality

*****oooooooooo*****