U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 303N /MASTAT – 303N

Econometrics

Block: 1 Linear Model and its Gerneralization

Unit – 1 : Linear Regression Models

Unit – 2 : Multicollinearity

Unit – 3 : Estimation of Parameters and Prediction

Unit – 4 : Model with Qualitative Independent Variables

Unit – 5 : Non-Spherical Disturbances

Block: 2 Simultaneous Equation Models and Forecasting

Unit – 6 : Structural and Reduced from of the Model and Identification Problem

Unit – 7 : Estimators in Simultaneous Equation Models - I

Unit – 8 : Estimators in Simultaneous Equation Models - I

Unit – 9 : Forecasting

Unit – 10 : Instrumental Variable Estimation

Block: 3 Advance Econometrics

Unit – 11 : Autoregressive Process

Unit – 11 : Vector Autoregressive Process

Unit – 11 : Granger Causality

Unit – 11 : Cointegration

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University, Gorakhpur.

Prof. Shruti **Member-Secretary**
Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Sandeep Mishra **Writer**
Department of Statistics (***Unit 1, 2, 6, 7, 8, 9, 10, 11, 12, 13 & 14***)
Shahid Rajguru College of Applied Sciences for Women
University of Delhi, New Dehi

Dr. Anupma Singh **Writer**
Department of Statistics (***Unit 3, 4 & 5***)
Ewing Christian College
University of Allahabad, Prayagraj

Prof. Anoop Chaturvedi **Editor**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Shruti **Course Coordinator**
School of Sciences,
U. P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 303N/ MASTAT – 303N **ECONOMETRICS**
©UPRTOU
First Edition: ***July 2024***
ISBN :

 Printed and Published by Col. Vinay Kumar, Registrar, Uttar Pradesh Rajarshi Tandon Open University, 2024.
Printed By:

Blocks & Units Introduction

The present SLM on ***Econometrics*** consists of fourteen units with three blocks.

The ***Block - 1 –Linear Model and its generalizations,*** is the first block, which is divided into five units.

The ***Unit - 1 – Linear Regression Models,*** is the first unit of present self-learning material, which describes Linear regression model, Assumptions, estimation of parameters by least squares and maximum likelihood methods. LOGIT, PROBIT, TOBIT and multinomial choice models, passion regression models.

The ***Unit – 2 - Multicollinearity,*** deals with Multicollinearity, problem of multicollinearity, consequences and solutions, regression, and LASSO estimators.

The ***Unit – 3 - Estimation of Parameters and Prediction*** deals with Testing of hypotheses and confidence estimation for regression coefficients, R^2 and adjusted R^2, point and interval predictors.

The ***Unit – 4 - Model with qualitative independent variables*** deals with Models with dummy independent variables, discreet and limited dependent variables. Use of dummy variables, model with non-spherical disturbances, estimation of parametric by generalized equation.

The ***Unit – 5 - Non-Spherical Disturbances,*** seemingly unrelated regression equations (SURE) model and its estimation, Panel data models, estimation in random effect and fixed effect models.

The ***Block – 2 - Simultaneous Equations Models and Forecasting,*** is the second block, which is divided into five units.

The ***Unit – 6 - Structural and reduced form of the model and identification problem,*** deals with the Simultaneous equations model, concept of structural and reduced forms, problem of identification, rank and order conditions of identifiability.

The ***Unit – 7 - Estimators in Simultaneous Equation Models – I,*** deals with the Limited and full information estimators, indirect least squares estimators, two stage least squares estimators, three stage least squares estimators and k class estimator.

The ***Unit – 8 - Estimators in Simultaneous Equation Models – I,*** deals with the Limited information maximum likelihood estimation, full information maximum likelihood estimation, prediction, and simultaneous confidence interval.

The ***Unit – 9 - Forecasting,*** deals with the Forecasting, exponential and adaptive smoothing methods, periodogram and correlogram analysis.

The ***Unit – 10 - Instrumental Variable Estimation,*** deals with the Review of GLM, analysis of GLM and generalized leased square estimation, Instrumental variables, estimation, consistency properties, asymptotic variance of instrumental variable estimators.

The ***Block - 3 – Advance Econometrics,*** is the third block, which is divided into four units.

The ***Unit – 11 - Autoregressive Process***, deals with the Moving average (MA), Auto regressive (AR), ARMA and ARMA models, Box-Jenkins models, estimation of ARIMA model parameters, auto covariance and auto correlation function.

The ***Unit – 12 - Vector Autoregressive Process***, deals with the Multivariate time series process and their properties, vector autoregressive (VAR), Vector moving average (VMA) and vector autoregressive moving average (VARMA) process.

The ***Unit – 13- Granger Causality***, deals with the Granger causality, instantaneous Granger causality and feedback, characterization of casual relations in bivariate models, Granger causality tests, Haugh-Pierce test, Hsiao test.

The ***Unit – 14- Cointegration***, deals with the Cointegration, Granger representation theorem, Bivariate cointegration and cointegration tests in static model.

At the end of every block/unit the summary, self assessment questions and further readings are given.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 303N/ MASTAT – 303N

Econometrics

Block: 1 Linear Model and its Gerneralization

Unit – 1 : Linear Regression Models

Unit – 2 : Multicollinearity

Unit – 3 : Estimation of Parameters and Prediction

Unit – 4 : Model with Qualitative Independent Variables

Unit – 5 : Non-Spherical Disturbances

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University, Gorakhpur.

Prof. Shruti **Member-Secretary**
Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Sandeep Mishra **Writer**
Department of Statistics (***Unit 1& 2***)
Shahid Rajguru College of Applied Sciences for Women
University of Delhi, New Dehi

Dr. Anupma Singh **Writer**
Department of Statistics (***Unit 3, 4 & 5***)
Ewing Christian College
University of Allahabad, Prayagraj

Prof. Anoop Chaturvedi **Editor**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Shruti **Course Coordinator**
School of Sciences,
U. P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 303N/ MASTAT – 303N **ECONOMETRICS**

First Edition: ***July 2024***
ISBN :
 Printed and Published by Col. Vinay Kumar, Registrar, Uttar Pradesh Rajarshi Tandon Open University, 2024.
Printed By:

Block & Units Introduction

The present SLM on ***Econometrics*** consists of fourteen units with three blocks.

The ***Block - 1 –Linear Model and its generalizations,*** is the first block, which is divided into five units.

The ***Unit - 1 – Linear Regression Models,*** is the first unit of present self-learning material, which describes Linear regression model, Assumptions, estimation of parameters by least squares and maximum likelihood methods. LOGIT, PROBIT, TOBIT and multinomial choice models, passion regression models.

The ***Unit – 2 - Multicollinearity,*** deals with Multicollinearity, problem of multicollinearity, consequences and solutions, regression, and LASSO estimators.

The ***Unit – 3 - Estimation of Parameters and Prediction*** deals with Testing of hypotheses and confidence estimation for regression coefficients, R^2 and adjusted R^2, point and interval predictors.

The ***Unit – 4 - Model with qualitative independent variables*** deals with Models with dummy independent variables, discreet and limited dependent variables. Use of dummy variables, model with non-spherical disturbances, estimation of parametric by generalized equation.

The ***Unit – 5 - Non-Spherical Disturbances,*** seemingly unrelated regression equations (SURE) model and its estimation, Panel data models, estimation in random effect and fixed effect models.

At the end of every unit the summary, self-assessment questions and further readings are given.

Blocks and Units Introduction

Block 1: Linear Model and its generalizations

Unit 1: Linear regression models:

Linear regression model. Assumptions, estimation of parameters by least squares and maximum likelihood methods.

Unit 2: Multicollinearity:

Multicollinearity, problem of multicollinearity, consequences and solutions, regression and LASSO estimators.

Unit 3: Estimation of parameters and prediction

Testing of hypotheses and confidence estimation for regression coefficients, R^2 and adjusted R^2, point and interval predictors.

Unit 4: Model with qualitative independent variables:

Models with dummy independent variables, discreet and limited dependent variables. Use of dummy variables, LOGIT, PROBIT, TOBIT and multinomial choice models, Poisson regression models.

Unit 5: Non-spherical disturbances

Model with non-spherical disturbances, estimation of parametric by generalized equation., Seemingly unrelated regression equations (SURE) model and its estimation, Panel data models, estimation in random effect and fixed effect models.

UNIT 1 LINEAR REGRESSION MODELS

Structure

1.1 Introduction

Econometrics may be defined as the application of statistical and mathematical methods to the analysis of economic data. Aims to give empirical content to economic relations for testing economic theories, forecasting, decision making, and policy evaluation. Econometrics may be considered as the combination of Economic Theory, Mathematical Economics and Statistics.

For example, the microeconomic theory states that the demand of a commodity is expected to increase as the price of that commodity decreases, provided the other things remain constant. How much the demand will go up or down because of certain change in the price of the commodity? Econometrician Job is to provide empirical content to the economic theory.

- Mathematical Economics, Economic Statistics and Econometrics:

Mathematical Economics

Mathematical economics involves the application of mathematical methods to represent theories and analyze problems in economics. It uses mathematical symbols and equations to model economic phenomena and relationships.

Economic Models are simplified mathematical representations of economic processes. An economic model describes the relationships between different economic variables and is used to explain how economies function or predict future economic behaviors. For example, supply and demand models, cost functions, and utility maximization problems.

Economic Statistics

Economic statistics focuses on the collection, processing, and presentation of economic data. It provides the quantitative basis for economic analysis, aiding in the visualization and understanding of economic trends and patterns.

Data Collection: Gathering data from various sources such as surveys, censuses, and administrative records.

Data Processing: Cleaning and organizing raw data to make it suitable for analysis. This includes handling missing values, correcting errors, and standardizing formats.

Data Presentation: Displaying data in an accessible format, often using charts, diagrams, tables, and graphs to facilitate interpretation and decision-making.

Econometrics

Econometrics combines economic theory, mathematical economics, and economic statistics to empirically test economic theories and quantify economic relationships. It uses statistical methods to estimate and test hypotheses about economic models.

Mathematical economics expresses economic theory in mathematical form. The mathematical description of relationship between different economic variables (causes and effects) describing the behavior of an economy is called an economic model.

Objective of the econometrician is to put economic model in such a form that allows empirical testing and empirical verification of economic theory.

1.1.1 How Econometric Analysis Proceeds?

Steps involved

- Statement of Economic Theory or Hypothesis
- Specification of Mathematical Model
- Specification of Statistical or Econometric Model

- Collection of data on relevant variables
- Estimation of parameters of chosen econometric model
- Tests of the hypothesis derived from the model
- Forecasting or Prediction

Statement of Economic Theory or Hypothesis:

Law of demand states that as the price of a commodity increases, the demand decreases provided the other things held constant.

Specification of Mathematical Model:

An inverse relationship exists between the price and demand. It does not tell the precise form of the relationship. For this purpose, we must express the statement in mathematical form.

q: Quantity demanded, p: Price.

We can write

$$q = \beta_1 + \beta_2 p, \beta_2 < 0$$

(1)

$$\text{or } q = Ap^{\beta};\ \beta < 0$$

(2)

In both relationships, q has an inverse relationship with p. Economic theory does not provide much information about the functional form of the relationship. For this purpose, we require statistical tools.

Specification of Statistical or Econometric Model:

Economic relationships are usually stochastic in nature. There are variables, other than main dominant variable p, affecting q. Let u be a random variable including the effect of all other variables. We can write (1) as

$$q = \beta_1 + \beta_2 p + u, \beta_2 < 0$$

(3)

u is called the random error term or disturbance term. Equation (3) is a statistical model or econometric model.

Econometric Model may have more than one equation. For example, consider the following model:

$$\text{Wage Equation: } W = \alpha_0 + \alpha_1 U + \alpha_2 P + u_1,$$
$$\text{Price Equation: } P = \beta_0 + \beta_1 W + \beta_2 R + \beta_3 M + u_2$$

$W =$ Rate of change in money wage
$P =$ Rate of change in prices

Variables explained. $U =$ % Unemployment rate,

$M =$ Supply of money; $R =$ Rate of change in cost of capital

Collection of data on relevant variables:

Three types of data usually available

Time series data: Time series data is collected over a time period. For example, data on unemployment rate of a country for 10 consecutive years.

Cross-section data: Cross section data is collected on one or more variables at a single point of time. For example, data on unemployment rate of 20 countries at a particular time point.

Pooled or Panel data: Panel data is the combination of time series and cross section data. For example, data on unemployment rate of 20 countries for 10 consecutive years.

Estimation of parameters:

Law of demand states that $\beta_2 < 0$. Statistics provide us methods to estimate the parameters based on given observations on p and q.

Tests of the hypothesis derived from the model:

Does the estimated model support the economic theory? For instance, is $\beta_2 < 0$?

Forecasting or Prediction:

Estimated demand function can be used to predict the value of demand for a specific value of price.

Econometrics Applications

Econometrics is a powerful tool used across various fields to analyze and interpret data, uncover relationships between variables, and make informed predictions. It is widely applied in domains such as business, economics, government, and finance to analyze relationships between variables, test hypotheses, and make predictions. Its ability to transform data into actionable insights makes it an invaluable tool for strategic planning, policy evaluation, risk management, and more.

Here are some of the key applications:

1. Business

Strategic Planning: Econometric models help businesses forecast future sales, determine optimal pricing strategies, and allocate resources efficiently.

Investment Decisions: Firms use econometric analysis to assess the potential returns on investments, analyze market trends, and make data-driven investment choices.

Marketing and Advertising: Econometrics helps in evaluating the effectiveness of advertising campaigns, understanding consumer behavior, and optimizing marketing strategies.

Budgeting and Revenue Forecasting: Companies employ econometric techniques to predict future revenues and plan budgets accordingly.

2. Economics

Macroeconomic Analysis: Economists use econometrics to study economic growth, inflation, unemployment, and other macroeconomic variables. This helps in understanding the broader economic environment and policy impacts.

Microeconomic Analysis: Econometrics is used to analyze individual and firm behavior, market structures, and the effects of regulations on industries.

3. Government and Policy Organizations

Policy Evaluation: Governments use econometric models to evaluate the impact of policies such as tax changes, subsidies, and social programs on the economy and society.

Economic Forecasting: Econometric models help in predicting economic indicators like GDP growth, inflation rates, and employment trends, aiding in policy formulation and planning.

4. Central Banks

Monetary Policy: Central banks utilize econometrics to analyze the effects of interest rates, money supply, and other monetary policies on the economy. This helps in maintaining economic stability and achieving policy targets.

Financial Stability: Econometrics assists in assessing the health of financial systems, identifying potential risks, and devising strategies to mitigate financial crises.

5. Financial Services

Risk Management: Financial institutions use econometric models to measure and manage risks associated with investments, loans, and market fluctuations.

Asset Pricing: Econometrics is employed to develop models for pricing financial assets and derivatives, helping in investment decision-making and portfolio management.

6. Economic Consulting Firms

Economic Impact Studies: Consulting firms use econometrics to conduct studies on the economic impact of projects, policies, and market changes, providing valuable insights for clients.

Market Analysis: Firms analyze market trends, consumer behavior, and competitive dynamics to offer strategic advice to businesses and governments.

The core of econometrics lies in analyzing causal relationships between variables and making predictions based on empirical data.

Explanatory and response variables

An explanatory variable is the expected cause, and it explains the results. A response variable is the expected effect, and it responds to changes in explanatory variables.

Example: Researcher has five brands of coffee and believes that different brands used to make a cup of coffee affect hyperactivity differently. The explanatory variable is coffee brand. The response variable is hyperactivity level.

Exogeneous and Endogenous Variables

Exogenous variable is determined outside the model and is imposed on the model. An exogenous change is a change in an exogenous variable.

Endogenous variable is the variable whose measure is determined by the model. An endogenous change is a change in an endogenous variable in response to an exogenous change that is imposed upon the model. An endogenous random variable is correlated with the error term while an exogenous variable is not.

Example: Amount of wheat produced may depend on weather variables, farmer skill, pests, price of seeds, price of diesel etc.

These are exogenous to crop production. The amount of wheat produced is an endogenous variable. Are other variables exogeneous? If we consider the entire system, then

insects depend upon weather variable, price of seeds depend upon "price of diesel". Hence these are endogenous variables.

1.2 Objectives

After completing this Block, students should have developed a clear understanding of:

- Regression analysis relevant for analysing economic data.
- The fundamental concepts of econometrics.
- Multiple Linear Regression Model

1.3 Multiple Linear Regression Model

Simple regression model involves a dependent and one independent variable. In Multiple Regression Model, the study or dependent variable depends on more than one explanatory or independent variables.

Focus of Attention:

Our main emphasis is on studying

(i) What is causing variation in dependent variable?

(ii) Which variables are mainly responsible for variation in dependent variable?

Examples:

(i) Scientists might be interested in observing the effect of different amounts of fertilizer, different levels of irrigation, different types of soil on crop yield.

(ii) Selling price of a house depends upon its location, House area, House has sea facing or not, Number of bedrooms, Number of bathrooms, how old the house is, etc.

Let y be a dependent variable, and $x_1, x_2, \dots, x_k$ are k independent or explanatory variables.

We assume

$$E(y) = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k$$

(4)

$$\text{or } y = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k + u \quad (5)$$

Usually, $x_1 = 1$ to allow for the intercept term.

Here u is the random error or disturbance term and gives the difference between actual value of dependent variable and its expected value or its value predicted by the multiple regression.

$\beta_1, \beta_2 \ldots, \beta_k$ are unknown (constants) regression coefficients

$\beta_j = \dfrac{\partial E(y)}{\partial x_j}$ gives the rate of change in y with respect to x_j

In (4), if we change x_j by one unit, *i.e.*, to $x_j + 1$, then $E(y)$ changes by amount β_j.

Interpretation of regression coefficient as Elasticity

In economics and engineering applications, Elasticity is measured as a percentage change/response. For instance, the price elasticity of demand of a commodity is the percentage change in quantity of demand resulting from unit change in price.

For example, if a 10% increase in price of petroleum results in a 2 percent decrease in demand, then price elasticity is .02/.10 = 0.2. Corresponding regression coefficient is -0.2. If 20% increase in price of mango results in 40% decrease in its demand, then its price elasticity is 0.4/0.2=2.0. Corresponding regression coefficient -2.0.

Model Setup: Consider the set of n observations on dependent and independent variables arranged in the following table:

Sample no.	Dependent Variable	k independent variables

1	y_1	$x_{11}\ x_{12}\ \dots\ x_{1k}$
2	y_2	$x_{21}\ x_{22}\ \dots\ x_{2k}$
$\vdots$	$\vdots$	$\vdots$
n	y_n	$x_{n1}\ x_{n2}\ \dots\ x_{nk}$

$x_{ij}: i^{th}$ observation on j^{th} independent variable, $i = 1,2, \dots, n; j = 1,2, \dots, k$

1.3.1 Assumptions

We consider the following assumptions for the multiple

Assumption 1:

The following linear relationship exists between y and $x's$

$$y_1 = x_{11}\beta_1 + x_{12}\beta_2 + \cdots + x_{1k}\beta_k + u_1$$
$$y_2 = x_{21}\beta_1 + x_{22}\beta_2 + \cdots + x_{2k}\beta_k + u_2$$
$$\vdots$$
$$y_n = x_{n1}\beta_1 + x_{n2}\beta_2 + \cdots + x_{nk}\beta_k + u_n \quad (6)$$

$u_1, \dots, u_n$ are the disturbances or error terms.

Let us write

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; x_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix}; \ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}; \ u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

Then, using these vector notations, we can write (6) as

$$y = x_1\beta_1 + \cdots + x_k\beta_k + u. \quad (7)$$

Further, we write

$$X = (x_1 \dots x_k) = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}$$

Then, in matrix notations, we can write (6) as

$$y = X\beta + u$$

(8)

Usually, the first column of X consists of all elements equal to 1 to allow for the intercept term. Thus $x_{11} = \cdots = x_{n1} = 1,$ and

$$X = \begin{pmatrix} 1 & x_{12} & \dots & x_{1k} \\ 1 & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

Assumption 2:

$$E(u_i) = 0 \; \forall \; i = 1,2, \dots, n$$

$$\text{or } E(u) = 0$$

(9)

Assumption 3:

$$E(u_i^2) = \sigma_u^2 \; \forall \; i = 1,2, \dots, n$$

$$E(u_i u_{i'}) = 0 \; \forall \; i \neq i'$$

or

$$E(uu') = \sigma_u^2 I_n$$

(10)

Thus, $u_i's$ have same variance and pairwise uncorrelated. The disturbances are said to be homoscedastic if all the $u_i's$ have the same variances.

Assumption 4:

Rank of $X = \rho(X)$ is $k\ (\leq n)$.

Thus $x_1, \dots, x_k$ are linearly independent.

Note: If $\rho(X) < k,$ some of the linear combinations of $\beta_1, \dots, \beta_k,$ say, $\lambda_1\beta_1 + \cdots + \lambda_k\beta_k = \lambda'\beta$ can be estimated unbiasedly but all such linear combinations cannot be estimated unbiasedly.

Assumption 5:

X is a non-stochastic matrix. Even if X is stochastic, it is uncorrelated with u, i.e.,

$E(X'u) = 0$.

Using assumptions 2 and 5, we have

$$E(y) = X\beta$$

or $E(y_i) = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ik}\beta_k \ \ \forall i = 1,2, \dots, n$

Assumption 6:

Sometimes we assume that u follows a normal distribution.

We may combine assumptions 2,3 and 6 as $u \sim N(0, \sigma_u^2 I_n)$.

Assumption 7:

Sometimes, for studying the asymptotic properties such as consistency of the estimator of β, we assume that

$$\underset{n\to\infty}{plim}\left(\frac{X'X}{n}\right)=Q$$

exists and Q is a non-stochastic and positive definite matrix with finite elements.

1.3.2 Estimation of parameters by least squares

1.3.2.1 Ordinary Least Squares (OLS) Estimator of β

Let β be estimated by $b=(b_1,\dots,b_k)'$.

In method of least squares, b is obtained by minimizing the residual sum of squares

$$S=\sum_{i=1}^{n}(y_i-x_{i1}b_1-x_{i2}b_2-\cdots-x_{ik}b_k)^2=(y-Xb)'(y-Xb)$$

The resulting estimator is called the ordinary least squares (OLS) estimator.

Result 1.3.1: When X is of full column rank, the OLS estimator of β is given by

$$b=(X'X)^{-1}X'y.$$

Proof: We can write S as

$$\begin{aligned}S&=y'y-2b'X'y+b'X'Xb\\&=b'X'Xb-2b'X'X(X'X)^{-1}X'y+y'X(X'X)^{-1}X'y+y'y-y'X(X'X)^{-1}X'y\\&=(b-(X'X)^{-1}X'y)'X'X(b-(X'X)^{-1}X'y)+v\end{aligned}\qquad(11)$$

where

$$\begin{aligned}v&=y'y-y'X(X'X)^{-1}X'y\\&=y'My\end{aligned}$$

where $M=I_n-X(X'X)^{-1}X'$

In (11), S is minimum when the term

$$(b-(X'X)^{-1}X'y)\,'X'X(b-(X'X)^{-1}X'y)$$

(12)

is minimum.

For any $p\times 1$ vector $c, c'X'Xc \geq 0$ and $c'X'Xc = 0$ iff $c = 0$. For showing this, let us write $d = Xc$. Since X is of full column rank, $d = Xc$ is zero if and only if $c = 0$. Hence $c'X'Xc = d'd = 0$, iff $c = 0$. Thus, $X'X$ is positive definite and the minimum value of (12) is zero. This value is attained for

$$b = (X'X)^{-1}X'y$$

(13)

Alternative Derivation: We have

$$S = y'y - 2b'X'y + b'X'Xb$$

$$\frac{\partial S}{\partial b} = -2X'y + 2X'Xb = 0$$

$$\Rightarrow b = (X'X)^{-1}X'y$$

Further

$$\frac{\partial^2 S}{\partial b \partial b'} = 2X'X$$

Since $X'X$ is positive definite, S is minimum for $b = (X'X)^{-1}X'y$.■

b is known as the ordinary least squares (OLS) estimator of β.

Result 1.3.2: *The OLS estimator b is an unbiased estimator of β.*

Proof: We can write b as

$$b = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u)$$

$$= \beta + (X'X)^{-1}X'u$$

Taking expectation and observing that

$E[(X'X)^{-1}X'u] = 0$ (assumption 5),

we have

$E(b) = \beta \quad \forall\beta$■

For $b = (X'X)^{-1}X'y$, the error sum of squares is

$v = e'e = (y - Xb)'(y - Xb) = y'My$.

1.3.2.2 Ordinary Least Squares (OLS) Estimator of σ_u^2

Let us write

$v = (y - Xb)'(y - Xb) = y'My$

We also observe that

$v = y'y - b'X'Xb = y'y - b'X'y$

Result 1.3.3: *An unbiased estimator of* σ_u^2 *is*

$s^2 = \frac{v}{n-k}$.

Proof: OLS residual vector is given by

$e = y - Xb = My$

where $M = I_n - X(X'X)^{-1}X'$ is a symmetric, idempotent matrix and $MX = 0$. Hence $e = My = Mu$. Thus

$e'e = u'Mu = v$

Further

$$tr(M) = (n-k).$$

Taking expectation, we get

$$E(v) = E(u'Mu)$$
$$= tr[E(uu'M)]$$
$$= \sigma_u^2 tr(M)$$
$$= \sigma_u^2 (n-k)$$

Hence

$$E(s^2) = \sigma_u^2 \blacksquare$$

Result 1.3.4: *The variance covariance matrix of b is given by*

$$E(b-\beta)(b-\beta)' = \sigma_u^2 (X'X)^{-1}.$$

Proof: *We have*

$$b - \beta = (X'X)^{-1} X'u$$

Hence

$$E(b-\beta)(b-\beta)' = (X'X)^{-1} X' E(uu') X (X'X)^{-1}$$
$$= \sigma_u^2 (X'X)^{-1} \blacksquare$$

Since b is a linear function of y, it is said to be a linear unbiased estimator of β.

1.4 Best Linear Unbiased Estimator (BLUE) Property of $\mathbf{b}$: Gauss Markov Theorem

Result 1.4.1: *Let* $\lambda = (\lambda_1, \dots, \lambda_k)'$ *and* $\lambda'\beta = \lambda_1\beta_1 + \cdots + \lambda_k\beta_k$. *Then* $\lambda' b$ *is a Best Linear Unbiased Estimator (BLUE) of* $\lambda'\beta$ *in the sense that (i) it is an unbiased estimator of* $\lambda'\beta$, *(ii) for any linear unbiased estimator* $a'y$ *of* $\lambda'\beta$, $Var(\lambda' b) \leq Var(a'y)\ \forall\ \beta$.

Proof: We assume that X is of full column rank, *i.e.*, $rank(X) = k.$ Then

$(i)\quad E(\lambda' b) = \lambda' E(b) = \lambda' \beta, \forall\, \beta.$

(ii) Let $a'y$ be any other unbiased estimator of $\lambda'\beta$ then,

$$E(a'y) = a'X\beta = \lambda'\beta$$

$$\Rightarrow \lambda = X'a$$

$$V(\lambda' b) = \sigma_u^2 \lambda' (X'X)^{-1}\lambda$$

$$V(a'y) - V(\lambda' b) = \sigma_u^2 a'(I - X(X'X)^{-1}X')a = \sigma^2 a'Ma$$

Since $MM = M,$ writing $\delta = Ma$, we have

$$V(a'y) - V(\lambda' b) = \sigma_u^2 \delta'\delta \geq 0$$

or $V(a'y) \geq V(\lambda' b)$■

If $\lambda_i = 1$ and all other elements of λ are zero, then $\lambda'\beta = \beta_i$. Thus $\lambda' b = b_i$ is BLUE of β_i. In this sense, b is a BLUE of β.

1.4.1 Alternative form of Gauss-Markov Theorem

Result 1.4.2: *Let* $\tilde{\beta} = Cy$ be any linear unbiased estimator of β where C is a $k \times n$ matrix and $\mathrm{V}(\tilde{\beta}) = E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'$. Then $V(\tilde{\beta}) - V(b)$ is positive semi definite.

Proof: We write

$$C = (X'X)^{-1}X' + D$$

Then

$$\tilde{\beta} = (X'X)^{-1}X'y + Dy$$

$$= \beta + (X'X)^{-1}X'u + DX\beta + Du$$

$$E(\tilde{\beta}) = \beta + DX\beta = \beta \Rightarrow DX = 0$$

Hence

$$V(\tilde{\beta}) = E((X'X)^{-1}X'u + Du)((X'X)^{-1}X'u + Du)'$$

$$= \sigma_u^2(X'X)^{-1} + \sigma_u^2 DD'$$

$$= V(b) + \sigma_u^2 DD'$$

Therefore

$$V(\tilde{\beta}) - V(b) = \sigma_u^2 DD'$$

which is positive semi-definite■

1.4.2 Maximum Likelihood Estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma_u^2}$

Result 1.4.3: If $u \sim N(0, \sigma_u^2 I_n)$, then maximum likelihood estimators of β and σ_u^2 are

$$b = (X'X)^{-1}X'y \hat{\sigma}_u^2 = \frac{v}{n}.$$

Proof: Since $u \sim N(0, \sigma_u^2 I_n)$, the likelihood function is given by

$$L(\beta, \sigma_u^2) = \prod_{j=1}^{n} f(u_j)$$

$$= \frac{1}{(2\pi\sigma_u^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma_u^2}\sum_{j=1}^{n} u_j^2\right]$$

$$= \frac{1}{(2\pi\sigma_u^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma_u^2} u'u\right]$$

$$= \frac{1}{(2\pi\sigma_u^2)^{n/2}} \exp\left[-\frac{1}{2\sigma_u^2}(y - X\beta)'(y - X\beta)\right].$$

Since the log transformation is monotonic, the maximization of likelihood function is equivalent to the maximization of log likelihood function.

Further

$$\ln L(\beta, \sigma_u^2) = -\frac{n}{2}\ln(2\pi\sigma_u^2) - \frac{1}{2\sigma_u^2}(y - X\beta)'(y - X\beta).$$

Hence

$$\frac{\partial \ln L(\beta, \sigma_u^2)}{\partial \beta} = \frac{1}{2\sigma_u^2} 2X'(y - X\beta) = 0$$

$$\frac{\partial \ln L(\beta, \sigma_u^2)}{\partial \sigma_u^2} = -\frac{n}{2\sigma_u^2} + \frac{1}{2(\sigma_u^2)^2}(y - X\beta)'(y - X\beta) = 0$$

Let $\hat{\beta}$ and $\hat{\sigma}_u^2$ denote the MLEs of β and σ_u^2 respectively. Then, we obtain $\hat{\beta}$ and $\hat{\sigma}_u^2$ by solving

$$\frac{1}{2\hat{\sigma}_u^2} 2X'\left(y - X\hat{\beta}\right) = 0$$

$$-\frac{n}{2\hat{\sigma}_u^2} + \frac{1}{2(\hat{\sigma}_u^2)^2}\left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right) = 0$$

This gives

$$\hat{\beta} = (X'X)^{-1}X'y = b$$

$$\hat{\sigma}_u^2 = \frac{1}{n}(y - Xb)'(y - Xb)$$

$$= \frac{v}{n} = \frac{n-k}{n}s^2$$

Further

$$\left.\frac{\partial^2 \ln L(\beta, \sigma_u^2)}{\partial \beta \partial \beta'}\right|_{\beta=b, \sigma_u^2=\hat{\sigma}_u^2}$$

$$= -\frac{1}{\hat{\sigma}_u^2}X'X\left.\frac{\partial^2 \ln L(\beta, \sigma^2)}{\partial^2(\sigma_u^2)^2}\right|_{\beta=b, \sigma_u^2=\hat{\sigma}_u^2}$$

$$= \frac{n}{2\hat{\sigma}_u^4} - \frac{1}{\hat{\sigma}_u^6}(y - Xb)'(y - Xb)$$

$$= \frac{n}{2\hat{\sigma}_u^4} - \frac{n}{\hat{\sigma}_u^4}$$

$$= -\frac{n}{2\hat{\sigma}_u^4}\frac{\partial^2 \ln L(\beta,\sigma^2)}{\partial\beta\partial\sigma_u^2}\Big|_{\beta=b,\sigma_u^2=\hat{\sigma}_u^2}$$

$$= -\frac{1}{\hat{\sigma}_u^4}X'(y-Xb) = 0.$$

The Hessian matrix of log likelihood is

$$H(\beta,\sigma_u^2)|_{\beta=b,\sigma_u^2=\hat{\sigma}_u^2}$$

$$= \begin{pmatrix} \frac{\partial^2 \ln L(\beta,\sigma_u^2)}{\partial\beta\partial\beta'} & \frac{\partial^2 \ln L(\beta,\sigma_u^2)}{\partial\beta\partial\sigma_u^2} \\ \frac{\partial^2 \ln L(\beta,\sigma_u^2)}{\partial\sigma_u^2\partial\beta} & \frac{\partial^2 \ln L(\beta,\sigma_u^2)}{\partial^2(\sigma_u^2)^2} \end{pmatrix}\Bigg|_{\beta=b,\sigma_u^2=\hat{\sigma}_u^2}$$

$$= \begin{pmatrix} -\frac{1}{\hat{\sigma}_u^2}X'X & 0 \\ 0 & -\frac{n}{2\hat{\sigma}_u^4} \end{pmatrix}$$

The Hessian matrix is negative definite for $\beta = b, \sigma_u^2 = \hat{\sigma}_u^2$. This ensures that the likelihood function is maximized at these values.

Note: *The MLE of* β *is the same as OLS estimator* b.

$\hat{\sigma}_u^2$ is not an unbiased estimator of σ_u^2. The bias of $\hat{\sigma}_u^2$ is given by

$$E[\hat{\sigma}_u^2 - \sigma_u^2] = -\frac{k}{n}\sigma_u^2.$$

1.4.3 Distributions of b and s^2

Result 1.4.4: *If* $u \sim N(0, \sigma_u^2 I_n)$, *then* $b \sim N(\beta, \sigma_u^2(X'X)^{-1}), v = \left(\frac{e'e}{\sigma_u^2}\right) \sim \chi^2(n-k)$. *Further* b *and* v *are independently distributed.*

Proof: We can write

$$b = (X'X)^{-1}X'y = \beta + Cu$$

where $C = (X'X)^{-1}X'$.

Since $u \sim N(0, \sigma_u^2 I_n), Cu \sim N(0, \sigma_u^2 CC')$. Further

$CC' = (X'X)^{-1}$. Hence $b \sim N(\beta, \sigma_u^2 (X'X)^{-1})$.

Further $e'e = u'Mu$, where $M = I_n - X(X'X)^{-1}X'$ is an idempotent matrix. The rank of M is

$$rank(M) = tr(M) = n - k.$$

Hence $n - k$ eigenvalues of M are 1 and remaining k eigen values are 0. Thus, there exists an orthogonal matrix P such that

$$P'MP = diag(1, \dots, 1, 0, \dots, 0) = \Lambda \text{ (say)}.$$

Let $w = \frac{1}{\sigma_u} P'u$. Then $w \sim N(0, I_n)$. Therefore, the elements of $w,\ i.e.,\ w_1, w_2, \dots, w_n$ are *iid* standard normal variates. Hence

$$v = \frac{1}{\sigma_u^2} u'Mu = w'\Lambda w = \sum_{j=1}^{n-k} w_j^2 \sim \chi^2(n-k).$$

For proving that b and v are independently distributed, we have

$$E(b - \beta)(Mu)'$$
$$= (X'X)^{-1}X'E(uu')M$$
$$= \sigma_u^2 (X'X)^{-1}X'M = 0$$

Hence b and Mu are uncorrelated and both follow normal distribution. Thus, b and Mu are independently distributed. Further, $v = \frac{1}{\sigma_u^2} u'Mu = \frac{1}{\sigma_u^2} (Mu)'(Mu)$ is a function of Mu. Hence, b and v are independently distributed■

We observe that

$$Var(s^2) = \frac{\sigma_u^4}{(n-k)^2} Var\left(\frac{e'e}{\sigma_u^2}\right)$$

$$= \frac{\sigma_u^4}{(n-k)^2} 2(n-k)$$

$$= \frac{2\sigma_u^4}{n-k}$$

1.4.4 Cramer-Rao lower bound

We observed that

$$\mathrm{E}\left[\frac{\partial^2 \ln L\,(\beta, \sigma_u^2)}{\partial\beta\partial\beta'}\right] = -\frac{1}{\sigma_u^2} X'X$$

$$E\left[\frac{\partial^2 \ln L\,(\beta, \sigma_u^2)}{\partial^2 (\sigma_u^2)^2}\right]$$

$$= E\left[\frac{n}{2\sigma_u^4} - \frac{1}{\sigma_u^6}(y - X\beta)'(y - X\beta)\right]$$

$$= \frac{n}{2\sigma_u^4} - \frac{1}{\sigma_u^6} n\sigma_u^2$$

$$= -\frac{n}{2\sigma_u^4}$$

$$E\left[\frac{\partial^2 \ln L\,(\beta, \sigma_u^2)}{\partial\beta\partial\sigma_u^2}\right]$$

$$= E\left[-\frac{1}{\sigma_u^4} X'(y - X\beta)\right] = 0$$

Cramer-Rao lower bound for the variance of unbiased estimators of (β, σ_u^2) is

$$[\mathrm{I}(\beta,\sigma_u^2)]^{-1} = [-E\{H(\beta,\sigma_u^2)\}]^{-1} = \begin{pmatrix} \sigma_u^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma_u^4}{n} \end{pmatrix}$$

Cramer-Rao lower bound is attained for the variance-covariance matrix of b but not for variance of s^2.

1.4.5 Large Sample Properties

Consistency of the Least Squares Estimator:

We assume that

$(i)\ (x_j, u_j), j = 1, \dots, n$ is a sequence of *iid* random variables

$$(ii)\ \underset{n\to\infty}{plim}\left(\frac{X'X}{n}\right) = Q$$

exists and is a non-stochastic, positive definite matrix with finite elements.

Result 1.4.5: *(i) The OLS estimator b is a consistent estimator of β, (ii) s^2 is a consistent estimator of σ_u^2.*

Proof: We can write

$$b = \beta + \left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'u\right)$$

Now

$$\frac{1}{n}X'u = \frac{1}{n}\sum_{j=1}^{n} x_j u_j = \frac{1}{n}\sum_{j=1}^{n} w_j = \bar{w}$$

where $w_j = x_j u_j$. Hence

$$plim\ b = \beta + Q^{-1} plim(\bar{w})$$

Further

$$E(w_j|X) = x_j E(u_j|X) = 0$$

This implies that $E(\bar{w}) = 0$. Again, variance covariance matrix of $\bar{w}$ is

$$Var(\bar{w}|X) = E(\bar{w}\bar{w}'|X)$$

$$= \frac{\sigma_u^2}{n}\left(\frac{1}{n}X'X\right)$$

so that

$$Var(\bar{w}) = \frac{\sigma_u^2}{n}E\left(\frac{1}{n}X'X\right)$$

Hence

$$\lim_{n\to\infty} Var(\bar{w}) = \lim_{n\to\infty}\left\{\frac{\sigma_u^2}{n}E\left(\frac{1}{n}X'X\right)\right\}$$

Since $\text{plim}\left(\frac{1}{n}X'X\right) = Q$ is a finite positive definite matrix and $\lim_{n\to\infty}\left(\frac{\sigma_u^2}{n}\right) = 0$, we have $\lim_{n\to\infty} Var(\bar{w}) = 0.$

Hence $plim\left(\frac{1}{n}X'u\right) = 0$ implying that $plim\ b = \beta.$

(ii) We can write s^2 as

$$s^2 = \frac{1}{n-k}u'Mu$$

$$= \frac{n}{n-k}\left[\frac{1}{n}u'u - \left(\frac{1}{n}u'X\right)\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'u\right)\right]$$

Now, as $n \to \infty, n/(n-k) \to 1$. Further

$$plim\left(\frac{1}{n}u'X\right)\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'u\right) = 0.$$

and

$$plim\left(\frac{1}{n}u'u\right) = plim\left(\frac{1}{n}\sum_{j=1}^{n} u_j^2\right)$$

Since $u_j^2\ (j = 1, \dots, n)$ are *iid* with common mean σ_u^2,

$$plim\left(\frac{1}{n}u'u\right) = \sigma_u^2.$$

Hence $plim(s^2) = \sigma_u^2$■

Obviously, asymptotic variance covariance matrix of b is

$$n.plim\left(\sigma_u^2\left(\frac{1}{n}X'X\right)^{-1}\right) = n\sigma_u^2 Q^{-1}$$

Further

$$plim\left(s^2\left(\frac{1}{n}X'X\right)^{-1}\right) = \sigma_u^2 Q^{-1}.$$

Thus, an estimator of the asymptotic variance covariance matrix of b is $s^2(X'X)^{-1}$.

1.5 Self-Assessment Exercise

1. Discuss the role and applications of econometrics.
2. What is the general form of the linear regression model?
3. Why is linear regression considered a fundamental statistical tool?
4. What are the main assumptions of the linear regression model?
5. How does the assumption of linearity influence the model's application?
6. Why is it important to check for homoscedasticity in a regression model?
7. What is the objective of the **Ordinary Least Squares (OLS)** method in linear regression?

8. How are the regression coefficients estimated using the OLS method?
9. How does MLE differ from OLS in terms of methodology and assumptions?
10. What assumptions are made about the error term (ϵ\epsilonϵ) when using MLE?
11. How is the likelihood function maximized to estimate regression parameters?
12. Under what conditions would OLS and MLE provide identical parameter estimates?
13. When is MLE preferred over OLS in regression modeling?
14. What are the practical implications if the assumptions of linear regression are violated?
15. How would you interpret the regression coefficients in a multiple linear regression model?
16. Define a multiple linear regression model and give its assumptions.
17. Derive the least squares estimator of coefficients vector in a linear regression model and show that it is unbiased.
18. State and prove the Gauss Markov theorem.

1.6 Summary

This unit provides a detailed exploration of regression modeling techniques, focusing on **multiple linear regression**. It begins with an in-depth study of the linear regression model, covering:

1. **Model Structure**: Establishing a linear relationship between a dependent variable and multiple independent variables.
2. **Assumptions**:
    - Linearity of the relationship between yyy and predictors.
    - Independence of observations.

- Homoscedasticity (constant variance of errors).
- Normality of residuals.
- Absence of multicollinearity among predictors.

3. **Parameter Estimation Methods**:

Least Squares (OLS): Minimizes the sum of squared residuals to derive parameter estimates.

Maximum Likelihood Estimation (MLE): Optimizes the likelihood of the observed data, providing flexible parameter estimates under normality assumptions.

1.7.1 References

Gujarati, D. and Guneshker, S. (2007). *Basic Econometrics*, 4th Ed., McGraw Hill Companies.

Johnston, J. and J. Dinardo (1997). *Econometric Methods*, 2nd Ed., McGraw Hill International.

1.7.2 Further Readings

- Judge, G.G., W.E. Griffiths, R.C. Hill Lütkepohl and T. C. Lee (1985). The theory and practice of econometrics, Wiley.
- Koutsoyiannis, A. (2004). *Theory of Econometrics*, 2 Ed., Palgrave Macmillan Limited.
- Maddala, G.S. and Lahiri, K. (2009). *Introduction to Econometrics*, 4 Ed., John Wiley & Sons.
- Ullah, A. and Vinod, H.D. (1981). Recent advances in Regression Methods, Marcel Dekker.

Block 1: Linear Model and its generalizations

Unit 2: Multicollinearity:

Multicollinearity, problem of multicollinearity, consequences and solutions, regression, and LASSO estimators.

UNIT 1 MULTICOLLINEARITY

Structure

1.1 Introduction

Multicollinearity in regression analysis refers to the situation where two or more predictor variables in a model are highly correlated. This correlation can lead to unreliable estimates of the coefficients, reduced statistical power, interpretation difficulties, and instability of model coefficients. It occurs when predictor variables contain redundant information about the response variable, complicating the accurate estimation of their effects. Addressing multicollinearity is crucial for improving the robustness and reliability of regression models.

***Key Points about Multicollinearity*:**

1. Identification:

- Variance Inflation Factor (VIF): A common metric used to detect multicollinearity. VIF values greater than 10 (some sources use a threshold of 5) indicate significant multicollinearity.
- Correlation Matrix: By examining the correlation coefficients between pairs of predictor variables. High absolute values (close to 1 or -1) suggest multicollinearity.
- Condition Index: Values above 30 indicate strong multicollinearity.

2. Problems Caused:

- Unstable Estimates: Coefficients become very sensitive to changes in the model.
- Reduced Precision: Confidence intervals for coefficients can become very wide.

- Misleading Significance Tests: The p-values for predictors can be misleading, showing some predictors as non-significant when they contribute to the model.

3. Solutions:

- Removing Predictors: Eliminating one or more correlated predictors can help reduce multicollinearity.

- Combining Predictors: Creating composite variables or using techniques like principal component analysis (PCA) to combine correlated variables into a single predictor.

- Regularization Techniques: Methods like Ridge Regression (L2 regularization) and Lasso Regression (L1 regularization) can help manage multicollinearity by adding penalties to the size of the coefficients.

Multicollinearity does not affect the predictive accuracy of the model per se, but it affects the interpretability and stability of the model coefficients. Techniques such as variance inflation factor (VIF) and principal component analysis (PCA) can be used to detect and mitigate multicollinearity in regression analysis.

1.2 Objectives

After completing this course, there should be a clear understanding of:

- Multicollinearity
- Its problem, consequences, and solution
- LASSO Estimator

1.3 Multicollinearity

The multicollinearity exists when two or more explanatory variables have high correlation. The high correlation means one predictor variable can be used to predict the other. This creates unnecessary information, adversely affecting the results in a regression model. Some examples of multicollinear predictors are:

(i) a person’s height and weight,

(ii) age and sales price of a car,

(iii) years in university teaching jobs and annual salary.

Let us write

$$y = X\beta + u.$$

(1)

where, $y: n \times 1;\quad X: n \times k;\ u: n \times 1;$

and $E(u) = 0;\ E(uu') = \sigma_u^2 I_n.$

One of the assumptions is that the matrix X is of full column rank, i.e., $\rho(X) = k$. When $\rho(X) < k$, we face the problem of (exact) multicollinearity.

What are the implications of exact multicollinearity?

Let $x_1, \cdots, x_k$ be the columns of X.

Exact Multicollinearity

In case of exact multicollinearity, there exists a relation of the form $c_1 x_1 + \cdots + c_k x_k = 0$; where $c_1, \ldots, c_k$ are the constants, not all equal to 0.

A linear parametric function $w'\beta = w_1\beta_1 + \cdots + w_k\beta_k$ *is estimable iff* w' *can be expressed as a linear combination of rows of* X*, i.e.,* $\rho(X' \quad w) = \rho(X')$. *In other words, if* w *belongs to the row space of* X*, then* $w'\beta$ *is estimable.*

If above condition of estimability is satisfied then a BLUE of $w'\beta$ is $w'b^*$, where b^* is a solution of $X'Xb^* = X'y$.

Equivalently $w'\beta$ is estimable *iff* $\rho(X'X \ \ w) = \rho(X'X)$, i.e., w can be expressed as a linear combination of rows or columns of $X'X$.

Result 1.3.1: *The linear parametric function* $w'\beta$ *is estimable iff w can be expressed as a linear combination of the eigen vectors of* $X'X$ *corresponding to non-zero eigen values of* $X'X$.

Proof: Suppose P is an orthogonal matrix consists of orthonormal eigen vectors of $X'X$, so that $PP' = I_k$.

Let us write $XP = Z$. We have

$$P'X'XP = Z'Z = diag(\lambda_1, \dots, \lambda_k),$$

where $\lambda_1, \dots, \lambda_k$ are eigen values of $X'X$.

We can write the equation (1) as

$$y = XPP'\beta + u = Z\theta + u$$

where $\theta = P'\beta$. If $\rho(X) = J < k$. Then $k - J$ columns of $Z = XP$ are zero. Without loss of generality, we assume that last $k - J$ columns of Z are zero and then last $k - J$ components of θ disappear from the model.

Hence $\theta_{J+1}, \dots \theta_k$ cannot be estimated. In other words, $\theta_1, \dots \theta_J$ or any linear combination of them can be estimated from the model. Then we have

$$w'\beta = w'PP'\beta = (P'w)'\theta$$

(2)

Hence, we can estimate $w'\beta$ *iff* last $k - J$ components of $P'w$ are zero. This gives the required result■

Let p be a normalized eigen vector corresponding to non-zero eigen value λ of X'X, so that $X'Xp = \lambda p$. Then $p'\beta$ is estimable and its BLUE is $p'b^*$, where b^* is a solution of $X'Xb^* = X'y$.

We have

$$p'X'y = p'X'Xb^* = \lambda p'b^*,$$

$$Var(p'X'y) = \sigma_u^2 p'X'Xp = \lambda\sigma_u^2 p'p = \lambda\sigma_u^2.$$

Hence

$$Var(p'b^*) = \frac{1}{\lambda^2} Var(p'X'y) = \frac{\sigma_u^2}{\lambda}.$$

(3)

If p_1 and p_2 are eigen vectors corresponding to two non-zero eigen values λ_1, λ_2 of $X'X$, then

$$Cov(p_1'b^*, p_2'b^*) = \frac{1}{\lambda_1\lambda_2} Cov(p_1'X'y, p_2'X'y)$$

$$= \frac{\sigma_u^2}{\lambda_1\lambda_2} p_1'X'Xp_2$$

$$= \frac{\sigma_u^2}{\lambda_1\lambda_2} \lambda_2 p_1'p_2 = 0. \qquad (4)$$

Let $w'\beta$ be an estimable linear parametric function. Then

$$w = \delta_1 p_1 + \cdots + \delta_J p_J$$

where,

$\delta_1, \ldots, \delta_J$ are constants and p_i is eigen vector corresponding to non zero eigen value λ_i. Then

$$Var(w'b^*) = \sigma_u^2 \sum_{i=1}^{J} \frac{\delta_i^2}{\lambda_i} \qquad (5)$$

Thus, precision of $w'b^*$ depends upon σ_u^2, $\delta_i's$ and $\lambda_i's$.

Comparatively precise estimates can be obtained in the directions of eigen vectors of $X'X$ *corresponding to large eigen values. An estimable linear parametric function* $w'\beta$ *will be comparatively estimated with less precision if w, as written in (5), has large weights* δ_i *attached to small eigen values* λ_i.

1.3.1 Case of near Multicollinearity

Let us suppose that $\rho(X'X) = k$.

Then all linear parametric functions $w'\beta = (\sum_{i=1}^{k} \delta_i p_i)'\beta$ are estimable and BLUE is

$w'b = \left(\sum_{i=1}^{k} \delta_i p_i\right)' b$, where $b = (X'X)^{-1}X'y$.

Then

$$Var(w'b) = \sigma_u^2 w'(X'X)^{-1}w$$

$$= \sigma_u^2 w' \left(\sum_{i=1}^{k} \lambda_i^{-1} p_i p_i'\right) w$$

$$= \sigma_u^2 \sum_{i=1}^{k} \frac{\delta_i^2}{\lambda_i}.$$

In the case of near multicollinearity $w'\beta$ is estimated with less precision if *w* has large weights attached to small eigen values.

1.3.2 Sources of Multicollinearity

Data-based multicollinearity

1) Poorly designed experiments: If only a subspace of regressors have been sampled, i.e., there are k regressors but sample is collected from a lower dimensional space.

2) Insufficient data: If number of observations is less than the number of regressors, then collecting more data can resolve the issue but this is not possible in all the cases. For example, gene expression data.

3) Variables may be highly correlated while collecting data from purely observational studies.

Structural multicollinearity

1) Constraints on the model or in the population: when two or more regressors are related with some kind of linear relationship.

2) Model specification: if range of x is small, including x^2 in the model may cause multicollinearity.

3) An over defined model: if in a model there are more regressors than number of observations. OLS estimator cannot be obtained. This problem is often faced in gene expression data. Then one of the usual approaches is to eliminate some of the regressors.

For variables selection we cannot apply test of significance for regression coefficients as it involves OLS estimator. Then the question arises "How to select regressors to eliminate?"

Some other Causes of multicollinearity

1) Incorrectly using Dummy variables is also a cause of multicollinearity. For example, if we add a dummy variable for every category.

2) Including a regressor, which is a combination of two other regressors can also cause the problem of multicollinearity. For example, interest rates of various terms to maturity influence amount of fixed investment. But various terms interest rates are usually highly correlated.

1.3.3 Consequences of Multicollinearity

1) Let us consider a model with two explanatory variables in deviation form (observations are deviations from mean):

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + u_i; i = 1,2, \dots, n.$$

We also assume that the observations are scaled to unit length.

Let $r_{12} = \sum_{j=1}^{n} x_{1i} x_{2i}$ be the correlation coefficient between X_1 and X_2, r_{1y}, and r_{2y} be the correlation coefficients of y with X_1 and X_2 respectively.

Then

$$X'X = \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix}$$

The OLS estimator of $\beta = (\beta_1 \ \ \beta_2)'$ is a solution of

$$\begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} r_{1y} \\ r_{2y} \end{pmatrix}$$

Now

$$(X'X)^{-1} = \frac{1}{1 - r_{12}^2} \begin{pmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{pmatrix}$$

Then, the OLS estimator of $\beta = (\beta_1 \ \ \beta_2)'$ is

$$b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \frac{1}{1 - r_{12}^2} \begin{pmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{pmatrix} \begin{pmatrix} r_{1y} \\ r_{2y} \end{pmatrix}$$

$$\text{or } b_1 = \frac{r_{1y} - r_{12} r_{2y}}{1 - r_{12}^2}, \qquad b_2 = \frac{r_{2y} - r_{12} r_{1y}}{1 - r_{12}^2}. \tag{6}$$

The covariance matrix of b is

$$Var(b) = \frac{\sigma_u^2}{1 - r_{12}^2} \begin{pmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{pmatrix}$$

(7)

The strong multicollinearity between X_1 and X_2 results in r_{12} close to 1 or -1. This leads to

(i) Large variances and covariances between the OLS estimators of regression coefficients.

(ii) The regression coefficients are large in magnitude.

2) Let us write the model in the following canonical form:

$$y = Z\theta + u$$

where $Z'Z = \Lambda = diag(\lambda_1, \dots, \lambda_k)$.

The OLS estimator of θ is

$$\hat{\theta} = \Lambda^{-1} Z' y$$

Then $\hat{\theta}_i = \frac{1}{\lambda_i} z_i' y$

$$Var(\hat{\theta}_i) = \frac{\sigma^2}{\lambda_i} \Rightarrow E(\hat{\theta}_i^2) = \theta_i^2 + \frac{\sigma^2}{\lambda_i}$$

The small λ_i results in estimates large in magnitude $(E(\hat{\theta}_i^2) \gg \theta_i^2)$ and large variance.

Main consequences of presence of multicollinearity are

1. For exact multicollinearity, OLS estimators cannot be defined.

2. Leads to estimators with large variances and covariances and hence imprecise estimators of regression coefficients.

3. Since variances of individual coefficients are large, in testing significance of regression parameters, the null hypothesis of insignificant regression parameter is often accepted.

4. Even when the coefficients are jointly significant and R^2 is high, the individual coefficients are insignificant.

5. Because of large variances of coefficients estimates, the confidence intervals tend to be much wider.

6. It leads to estimates which are large in magnitude and often having *wrong* signs.

7. OLS estimators and their variances become very sensitive to small changes in data. Thus, results are not very robust.

1.3.4 Detection of Multicollinearity

The measures are:

(i) *Condition number and condition index*:

Let $\kappa(X) = \left(\frac{\lambda_1}{\lambda_k}\right)^{\frac{1}{2}}$

(8)

where,

λ_1: maximum eigen value of Λ,

λ_k: minimum eigen value of Λ, and

$\kappa(X)$ is a measure of sensitivity of b to changes in $X'y$ or $X'X$.

If the condition number is around 5 to 10, then it shows weak dependence. If condition number is around 30 to 100 then it shows strong relations.

(ii) *Multicollinearity Index*:

Multicollinearity index (mci) is defined as

$MCI = \sum_{j=1}^{k} \left(\frac{\lambda_k}{\lambda_j}\right)^2$

(9)

If $MCI \approx 1$ it indicates high multicollinearity, and

if $MCI > 2$ it indicates little or no multicollinearity.

(iii) *Variance Decomposition Proportions:*

Let us write the Spectral decomposition of $X'X$ as

$$X'X = \sum_{i=1}^{k} \lambda_i p_i p_i',$$

$$(X'X)^{-1} = \sum_{i=1}^{k} \frac{1}{\lambda_i} p_i p_i'$$

Here, p_{li}^2 is the l^{th} diagonal element of $p_i p_i'$.

Let $\sum_{i=1}^{k} \frac{p_{li}^2}{\lambda_i}$ be the l^{th} diagonal element of $(X'X)^{-1}$, so that

$$Var(b_l) = \sigma_u^2 \sum_{i=1}^{k} \frac{p_{li}^2}{\lambda_i}.$$

Define,

$$\phi_{li} = \frac{\frac{p_{li}^2}{\lambda_i}}{\sum_{j=1}^{k} \frac{p_{lj}^2}{\lambda_j}} \qquad (10)$$

where, ϕ_{li} is the proportion of $Var(b_l)$ associated with λ_i

Table: Variance Decomposition Proportions

Eigen Values	$Var(b_1)$	$Var(b_2)$		$Var(b_k)$
λ_1	ϕ_{11}	ϕ_{21}	…	ϕ_{k1}

λ_2	ϕ_{12}	ϕ_{22}	…	ϕ_{k2}
⋮	⋮	⋮	⋮	⋮
λ_k	ϕ_{1k}	ϕ_{2k}		ϕ_{kk}

Here, sum of diagonal elements is 1.

Two or more large values of ϕ_{ij} *in a row indicate that multicollinearity is adversely affecting the precision of estimate of the associated coefficient.*

(iv) *Variance Inflation Factor (VIF):*

Let R_j^2 is the multiple correlation coefficient between j^{th} regressor and the remaining $k-1$ regressors. Then VIF is defined as

$$VIF_j = \frac{1}{1-R_j^2} \quad (11)$$

i) If VIF is close to 1, it means there is no correlation between j^{th} predictor and remaining predictors.

ii) VIF exceeding 4 warrant further investigation.

iii) VIFs exceeding 10 are signs of serious multicollinearity.

1.3.5 Solutions to Multicollinearity Problem

Suppose $X'X$ has a small root λ corresponding to eigen vector p. An additional observation y_{n+1} is taken corresponding to $x_{n+1} = l.p$, where l is a scalar. The model for complete set of observations is

$$\begin{pmatrix} y \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} X \\ x'_{n+1} \end{pmatrix} \beta + u$$

$$\text{or } y^* = X^* \beta + u$$

(12)

$$\text{Then } X^{*\prime} X^* p = (X'X + x_{n+1} x'_{n+1}) p = (X'X + l^2 pp')p = (\lambda + l^2)p$$

where, p is the eigen vector of $X^{*\prime} X^*$ corresponding to the eigen value $(\lambda + l^2)$. Choosing an additional observation in the direction of p can improve the precision of estimator.

***Some other methods to overcome multicollinearity problem*:**

***Exact Linear Constraint*:**

It leads to restricted regression estimator. The restrictions imposed presumably describe some physical constraint on the variables involved and are the product of a theory relating the variables. One effect of using exact parameter restrictions is to reduce the sampling variability of the estimators, a desirable end given multicollinear data. The imposition of binding constraints, even if incorrect, may reduce the mean square error of the estimator although incorrect restrictions produce biased parameter estimators. Exact linear restrictions may be employed in case of both extreme and near-extreme multicollinearity. Exact restrictions "work" by reducing the dimensionality of the parameter space, one dimension for each independent linear constraint.

***Stochastic Linear Restrictions*:**

Here, you get mixed regression estimator. It arises from prior statistical information, usually in the form of previous estimates of parameters that are also included in a current model

***Linear Inequality restrictions*:**

Linear inequality restrictions can indeed be useful in addressing multicollinearity in regression analysis. By imposing linear inequality restrictions on the coefficients of the

regression model, we can restrict the possible values that the coefficients can take. This restriction can potentially reduce the variance of the estimates and improve the precision of estimation, especially in cases of moderate to high multicollinearity. In cases of extreme multicollinearity, where predictors are nearly perfectly correlated, even imposing linear inequality restrictions may not resolve the problem. The estimates of the coefficients can become highly unstable and may not yield meaningful results regardless of the restrictions imposed.

1.4 Principal Component Regression

Principal Component Regression (PCR) is a technique that combines Principal Component Analysis (PCA) and linear regression. It is used when there are high levels of multicollinearity among the predictors in a regression model, leading to unstable estimates of regression coefficients.

Let us consider the model $y = X\beta + u$. Suppose $P = (p_1, \dots, p_k)$ is $(k \times k)$ matrix of orthogonal eigen vectors of $X'X$.

$z_i = Xp_i$ is the i^{th} principal component. Then $z_i'z_i = \lambda_i$, where λ_i is the i^{th} largest eigen value of $X'X$.

$Z = XP = (z_1, \dots, z_k)$ is $n \times k$ matrix of principal components.

We can rewrite the model as

$$y = XPP'\beta + u = Z\theta + u;\ (\theta = P'\beta)$$

(13)

1.4.1 Steps for obtaining the principal component estimator

(i) Delete some of the principal components $z_i's$ corresponding to small eigen values.

(ii) Partition $Z = X(P_1\ P_2) = (Z_1\ Z_2)$

where,

Z_1 be the matrix of principal components to be retained.

Z_2 be the Matrix of principal components to be deleted.

and Z_1 and Z_2 are orthogonal to each other.

Then the model (13) become

$$y = Z_1\theta_1 + Z_2\theta_2 + u \quad (14)$$

(iii) OLS estimator of θ_1 is $\hat{\theta}_1 = (Z_1'Z_1)^{-1}Z_1'y$, with covariance matrix $\sigma_u^2(Z_1'Z_1)^{-1}$.

(iv) We have $\beta = P\theta = P_1\theta_1 + P_2\theta_2$.

Omitting the components of Z_2 means setting $\theta_2 = 0$. Hence the principal component estimator of β is

$$\hat{\beta} = P_1\hat{\theta}_1 = P\hat{\theta}^*,$$

where $\hat{\theta}^* = \left(\hat{\theta}_1' \quad 0'\right)'$.

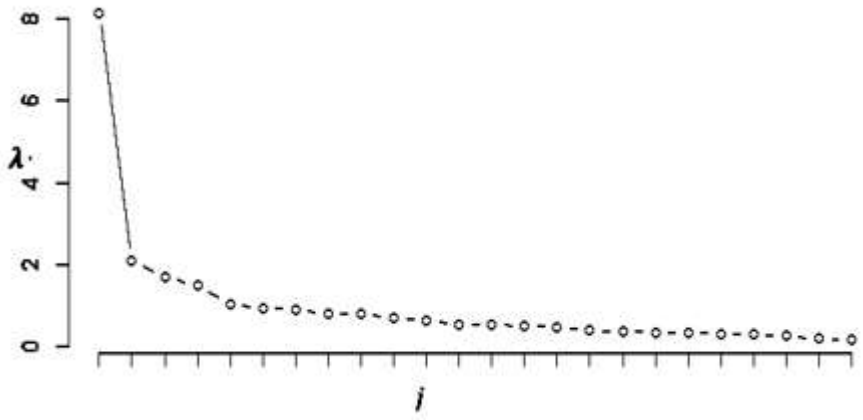
The principal component estimator has lower variance than OLS estimator but biased unless the restriction $P_2\theta_2 = 0$ is satisfied.

1.4.2 How to select number of principal components to be omitted?

Visual Examination

The principal component matrix is of the same size as the original data matrix. However, fewer principal components are usually needed because many of them might not be meaningful. Examining the amount of variance explained by each new principal component vector is helpful in achieving this. For this purpose, one can use scree plot. Scree plot shows the eigenvalues in decreasing order.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ be the eigen values in decreasing order. We plot $\lambda_j' s$ against $j = 1,2, \ldots, k$. Then select the index of the last component before the plot flattens.



Variance explained criteria:

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$. Then the trace of covariance matrix with J components deleted is equal to $\sigma_u^2 \sum_{i=1}^{k-J} \lambda_i^{-1}$. Percentage reduction in trace of covariance matrix obtainable from using a least squares estimator with J independent linear restrictions is:

$$V_J = \frac{\sum_{i=k-J+1}^{k} \lambda_i^{-1}}{\sum_{i=1}^{k} \lambda_i^{-1}} \times 100\%; J = 1,2, \ldots, k;$$

where V_J is the benchmark function of principal component regression. This is the variance of the component which we have retained.

The total % variation that is explained by the first $\mathrm{k} - J$ loadings is

$$\frac{\sum_{i=1}^{k-J} \lambda_i^{-1}}{\sum_{i=1}^{k} \lambda_i^{-1}} \times 100\%;$$

We may select J so that, say, the above percentage explained variation is greater than 80%.

1.5 Ordinary Ridge Regression (ORR) Estimator

Ordinary Ridge Regression (ORR), often simply referred to as Ridge Regression, is a technique used in linear regression to mitigate the problem of multicollinearity among predictor variables. It extends the ordinary least squares (OLS) method by adding a regularization term to the regression objective function.

The Ordinary Ridge Regression is given by

$$b^*(c) = (X'X + cI)^{-1}X'y; \;\; c > 0$$

(15)

For $c = 0$, we get the OLS estimator b. We have

$$E(b'b) = \beta'\beta + \sigma_u^2 tr(X'X)^{-1} > \beta'\beta + \frac{\sigma_u^2}{\lambda_k}$$

(16)

where λ_k= minimum eigen value of $X'X$ and, $tr(X'X)^{-1} = \sum_{i=1}^{k} \frac{1}{\lambda_i} > \frac{1}{\lambda_k}$.

For small λ_k, squared length of OLS estimator is much larger than squared length of coefficients vector β. The squared length of ORR estimator in less than that of OLS estimator.

Result 1.5.1: The bias and MSE of ORR estimator $b^*(c)$ are given by

$$E[b^*(c) - \beta] = -c(X'X + cI)^{-1}\beta$$

and

$$E[b^*(c) - \beta][b^*(c) - \beta]'$$
$$= \sigma_u^2(X'X + cI)^{-1}X'X(X'X + cI)^{-1} + c^2(X'X + cI)^{-1}\beta\beta'(X'X + cI)^{-1}.$$

Proof: We consider that,

$$b^*(c) = (X'X + cI)^{-1}X'y$$
$$= (X'X + cI)^{-1}X'Xb$$

$$= (X'X + cI)^{-1}(X'X + cI - cI)b$$

$$= b - c(X'X + cI)^{-1}b$$

or

$$E\left(b^*(c)\right) = \beta - c(X'X + cI)^{-1}\beta$$

or, the bias of $b^*(c)$ is

$$E\left(b^*(c)\right) - \beta = -c(X'X + cI)^{-1}\beta$$

(17)

again, we have

$$b^*(c) - E\left(b^*(c)\right) = (X'X + cI)^{-1}X'X(b - \beta)$$

or

$$\left[b^*(c) - E\left(b^*(c)\right)\right]\left[b^*(c) - E\left(b^*(c)\right)\right]'$$
$$= (X'X + cI)^{-1}X'X(b - \beta)(b - \beta)'X'X(X'X + cI)^{-1})$$

Taking expectation on both sides, we get

$$E\left[b^*(c) - E\left(b^*(c)\right)\right]\left[b^*(c) - E\left(b^*(c)\right)\right]'$$
$$= E[(X'X + cI)^{-1}X'X(b - \beta)(b - \beta)'X'X(X'X + cI)^{-1})]$$
$$= \sigma_u^2(X'X + cI)^{-1}X'X(X'X + cI)^{-1}$$
$$= V\left(b^*(c)\right).$$

Then the MSE of $b^*(c)$ is

$$MSE\ [b^*(c)] = V\left(b^*(c)\right) + Bias[b^*(c)]\ Bias\ [b^*(c)]'$$

$$= \sigma_u^2(X'X + cI)^{-1}X'X(X'X + cI)^{-1}$$
$$+ c^2(X'X + cI)^{-1}\beta\beta'(X'X + cI)^{-1}\blacksquare \qquad (18)$$

Result 1.5.2: The mean squared error of ORR estimator is given by

$$E[b^*(c)-\beta]'[b^*(c)-\beta] = \sigma_u^2 \sum_{i=1}^{k} \frac{\lambda_i}{(\lambda_i + c)^2} + c^2\beta'(X'X + cI)^{-2}\beta.$$

Proof: The mean square error of ORR estimator is obtained by taking the trace of mean square error of $b^*(c)$. From equation (18);

$$E[b^*(c)-\beta]'[b^*(c)-\beta]$$
$$= tr[\sigma_u^2(X'X + cI)^{-1}X'X(X'X + cI)^{-1} + c^2(X'X + cI)^{-1}\beta\beta'(X'X + cI)^{-1}]$$
$$= tr[\sigma_u^2(X'X + cI)^{-1}X'X(X'X + cI)^{-1}] + c^2\beta'(X'X + cI)^{-2}\beta$$
$$= \sigma_u^2 \sum_{i=1}^{k} \frac{\lambda_i}{(\lambda_i + c)^2} + c^2\beta'(X'X + cI)^{-2}\beta \blacksquare$$

How to select c?

The estimation of ridge regression estimator depends upon the value of *c*. Various approaches have been suggested in the literature to determine the value of *c*. The value of *c* can be chosen on the bias of criteria like

- the stability of estimators with respect to *c*.
- reasonable signs.
- the magnitude of residual sum of squares etc.

We consider here the determination of *c* by the inspection of ridge trace.

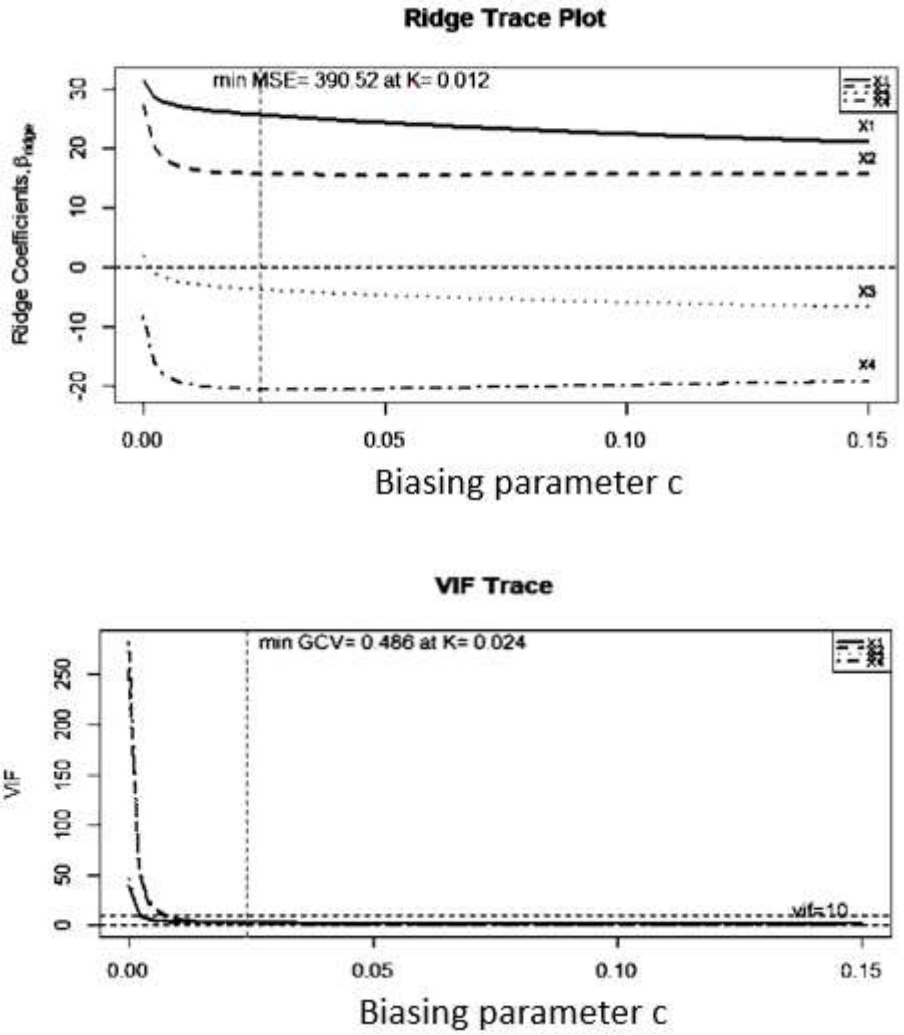
Ridge Trace

Choosing an appropriate value for c is one of the main difficulties in using ridge regression. The inventors of ridge regression, Hoerl and Kennard (1970), recommended utilising a diagram known as the ridge trace. The ridge regression coefficients as a function of c are displayed in this graphic. The analyst selects a value for c for which the regression

coefficients have stabilised when seeing the ridge trace. For modest values of c, the regression coefficients frequently fluctuate greatly before stabilising. Select the minimum value of c resulting in the least amount of bias, beyond which the regression coefficients appear to stay constant.

Ridge Trace is a two-dimensional plot of $b_i^*(c)$ and residual sum of squares against *c*. Select that value of *c* for which the estimated coefficients stabilize with increasing *c*.

Figure: Univariate ridge trace and VIF trace plots for the coefficients



From the above two graphs we observe that:

1) As c increases the coefficients shrink toward 0.

2) VIF decreases rapidly as c gets bigger than 0.

3) The VIF values begin to change slowly as c increases.

4) We choose the smallest value of c where the regression coefficients become stable in the ridge trace and the VIF values become sufficiently small.

Some operational choices of c are

1) *Hoerl, Kennard and Baldwin (1975):*

$$c_{HKB} = \frac{ks^2}{b'b}$$

$$s^2 = \frac{(y - Xb)'(y - Xb)}{n - k}$$

2) *Lawless and Wang (1976):*

$$c_{LW} = \frac{ks^2}{b'X'Xb}$$

1.6 Generalized Ridge Regression Estimator

Generalized Ridge Regression (GRR) is an extension of the classical Ridge Regression method, which is used to handle multicollinearity and improve the stability of regression estimates when there are correlated predictors in a linear regression model. Like Ridge Regression, the goal of GRR is to stabilize the parameter estimates by shrinking them towards zero, especially when multicollinearity is present.

Consider the model in canonical form

$$y = Z\theta + u$$

$$Z'Z = diag(\lambda_1, \dots, \lambda_k) = \Lambda \text{ (say)}$$

Let $C = diag(c_1, \dots, c_k)$. Then GRR estimator of θ is defined as

$$\hat{\theta}_R = (\Lambda + C)^{-1} Z' y = (\Lambda + C)^{-1} \Lambda \hat{\theta}$$

where

$Z' y = Z' Z (Z' Z)^{-1} Z' y = \Lambda \hat{\theta}, \hat{\theta}$ is the OLS estimator of θ.

The i^{th} component of $\hat{\theta}_R$ is given by

$$\hat{\theta}_{R_i} = \frac{\lambda_i}{\lambda_i + c_i} \hat{\theta}_i$$

(19)

Result 1.6.1: the expressions for bias and MSE of $\hat{\theta}_{R_i}$ are given by

$$E\left[\hat{\theta}_{R_i} - \theta_i\right] = -\frac{c_i}{\lambda_i + c_i} \theta_i$$

$$E\left[\hat{\theta}_{R_i} - \theta_i\right]^2 = \frac{\lambda_i \sigma_u^2 + c_i^2 \theta_i^2}{(\lambda_i + c_i)^2}$$

The MSE of $\hat{\theta}_{R_i}$ is minimum when $c_i = \frac{\sigma_u^2}{\theta_i^2}$.

Proof: We have,

$$\hat{\theta}_{R_i} = \frac{\lambda_i}{\lambda_i + c_i} \hat{\theta}_i = \hat{\theta}_{R_i} = \frac{\lambda_i + c_i - c_i}{\lambda_i + c_i} \hat{\theta}_i = \hat{\theta}_i - \frac{c_i}{\lambda_i + c_i} \hat{\theta}_i$$

$$\Rightarrow E\left[\hat{\theta}_{R_i}\right] = \theta_i - \frac{c_i}{\lambda_i + c_i} \theta$$

This is the expression for Bias of the $\hat{\theta}_{R_i}$.

Now, for the expression of MSE of $\hat{\theta}_{R_i}$, let

$$\hat{\theta}_{R_i} - \theta_i = \frac{\lambda_i}{\lambda_i + c_i} (\hat{\theta}_i - \theta_i) + \frac{\lambda_i}{\lambda_i + c_i} \theta_i - \theta_i$$

$$= \frac{\lambda_i}{\lambda_i + c_i} (\hat{\theta}_i - \theta_i) - \frac{c_i}{\lambda_i + c_i} \theta_i$$

$$\mathrm{E}\left[\hat{\theta}_{\mathrm{R_i}} - \theta_{\mathrm{i}}\right]^2 = E\left[\frac{\lambda_{\mathrm{i}}}{\lambda_{\mathrm{i}} + \mathrm{c_i}}\left(\hat{\theta}_{\mathrm{i}} - \theta_i\right) - \frac{c_{\mathrm{i}}}{\lambda_{\mathrm{i}} + \mathrm{c_i}}\theta_i\right]^2$$

$$= E\left[\frac{\lambda_{\mathrm{i}}}{\lambda_{\mathrm{i}} + \mathrm{c_i}}\left(\hat{\theta}_{\mathrm{i}} - \theta_i\right)\right]^2 + E\left[\frac{c_{\mathrm{i}}}{\lambda_{\mathrm{i}} + \mathrm{c_i}}\theta_i\right]^2 - 2E\left[\frac{\lambda_{\mathrm{i}}}{\lambda_{\mathrm{i}} + \mathrm{c_i}}\left(\hat{\theta}_{\mathrm{i}} - \theta_i\right)\right]\left[\frac{c_{\mathrm{i}}}{\lambda_{\mathrm{i}} + \mathrm{c_i}}\theta_i\right]$$

$$= \frac{\lambda_i^2}{(\lambda_{\mathrm{i}} + \mathrm{c_i})^2} E\left(\hat{\theta}_{\mathrm{i}} - \theta_i\right)^2 + \frac{C_i^2}{(\lambda_{\mathrm{i}} + \mathrm{c_i})^2}\theta_i^2$$

$$= \frac{\lambda_i^2}{(\lambda_{\mathrm{i}} + \mathrm{c_i})^2}\frac{\sigma_{\mathrm{u}}^2}{\lambda_{\mathrm{i}}} + \frac{C_i^2}{(\lambda_{\mathrm{i}} + \mathrm{c_i})^2}\theta_i^2$$

$$= \frac{\lambda_{\mathrm{i}}\sigma_{\mathrm{u}}^2 + \mathrm{c_i}^2\theta_{\mathrm{i}}^2}{(\lambda_{\mathrm{i}} + \mathrm{c_i})^2} \qquad (20)$$

Differentiate equation (20) with respect to c_{i} and equate it to zero gives

$$\frac{\partial\left(\mathrm{E}\left[\hat{\theta}_{\mathrm{R_i}} - \theta_{\mathrm{i}}\right]^2\right)}{\partial c_{\mathrm{i}}}$$

$$= \frac{(\lambda_{\mathrm{i}} + \mathrm{c_i})^2(2c_{\mathrm{i}}\theta_i^2) - (\lambda_{\mathrm{i}}\sigma_{\mathrm{u}}^2 + \mathrm{c_i}^2\theta_{\mathrm{i}}^2)2.(\lambda_{\mathrm{i}} + \mathrm{c_i})}{(\lambda_{\mathrm{i}} + \mathrm{c_i})^4} = 0$$

$$\Rightarrow (\lambda_{\mathrm{i}} + \mathrm{c_i})(c_{\mathrm{i}}\theta_i^2) - (\lambda_{\mathrm{i}}\sigma_{\mathrm{u}}^2 + \mathrm{c_i}^2\theta_{\mathrm{i}}^2) = 0$$

$$\Rightarrow c_{\mathrm{i}} = \frac{\sigma_{\mathrm{u}}^2}{\theta_i^2}.$$

Now,

$$\left[\frac{\partial^2\left(\mathrm{E}\left[\hat{\theta}_{\mathrm{R_i}} - \theta_{\mathrm{i}}\right]^2\right)}{\partial {c_{\mathrm{i}}}^2}\right]_{c_{\mathrm{i}} = \frac{\sigma_{\mathrm{u}}^2}{\theta_i^2}} > 0$$

Thus the MSE of $\hat{\theta}_{R\,i}$ is minimum when $c_i = \frac{\sigma_u^2}{\theta_i^2}$.

1.7 Shrinkage Estimator

A shrinkage estimator in statistics is a method that combines information from a sample with some form of prior knowledge or assumptions to produce more stable and

sometimes more accurate estimates of parameters or predictions. In the context of regression analysis, particularly when dealing with multicollinearity or high-dimensional data, shrinkage estimators like penalized regression and stein-rule estimators are commonly used.

For the problem of multicollinearity, the selection of a subset of variables (among a large number of variables) are required. In the variable selection procedure variables are either retained or discarded. Variable selection procedure often leads to high variance and prediction error so it does not work well.

The other method which are considered for variable selection is Shrinkage Methods, which is more continuous and do not suffer as much from high variability. Here we consider some shrinkage methods for estimating the regression.

1.7.1 Penalized Regression Estimators

A penalized regression estimator, also known as regularized regression, refers to a class of regression techniques that introduce a penalty term into the ordinary least squares (OLS) objective function. These penalties are designed to Shrink the estimators by imposing a penalty on their size, thereby mitigating issues like multicollinearity and overfitting. The two main types of penalized regression estimators are: Ridge Regression and LASSO (Least Absolute Shrinkage and Selection Operator) Regression.

Sparsity: Large number of predictors recorded but a relatively small number (proportion) of strong effects. This corresponds to sparsity.

Bias-Variance Trade-off

General regression relationship: $y = f(x) + u$

Let $\hat{f}(x)$ be the regression prediction. The effectiveness of prediction is measured through squared error of prediction $[y - \hat{f}(x)]^2$. Then

$$MSE = E[y - \hat{f}(x)]^2$$

$$= E[\{y - f(x)\} - \{\hat{f}(x) - f(x)\}]$$

$$= [Bias\{\hat{f}(x)\}]^2 + Var[\hat{f}(x)] + \sigma_u^2$$

This shows that

(i) the estimators having lower bias have higher variance.

(ii) the estimators having lower variance have higher bias.

In terms of higher bias there is a trade-off between bias and variance then tune the estimator and find the best possible trade-off.

1.7.1.1 Ridge Regression

In ridge regression, a penalty term proportional to the sum of the squares of the coefficients is added to the OLS objective function. The ridge penalty is parameterized by a tuning parameter, which controls the amount of shrinkage applied to the coefficients. Ridge regression is effective in reducing the impact of multicollinearity by shrinking the coefficients of correlated predictors towards each other.

Ridge estimator is obtained by minimizing the penalized residual sum of squares

$$RSS(c) = (y - X\beta)'(y - X\beta) + c\sum_{j=1}^{k}\beta_j^2$$

$$= (y - X\beta)'(y - X\beta) + c\beta'\beta$$

(21)

Thus, $\hat{\beta}_{Ridge} = \underset{\beta}{\operatorname{argmin}}\{RSS(c)\}$

Equivalently

$$\hat{\beta}_{Ridge} = \underset{\beta}{\operatorname{argmin}}(y - X\beta)'(y - X\beta) \text{ subject to } \sum_{j=1}^{k}\beta_j^2 \leq \delta$$

(22)

There is a one-to-one correspondence between (21) and (22). Differentiating $RSS(c)$ with respect to β and substituting it equal to zero leads to

$$\hat{\beta}_{Ridge} = (X'X + cI_k)^{-1}X'y$$

The solution adds a positive constant in the diagonal elements of $X'X$. The term c is the Penalty parameter controlling the amount of shrinkage.

Consider singular value decomposition of X:

$X = UDV'U$ is $n \times k$ matrix

V is $k \times k$ orthogonal matrix

$D = diag(d_1, \dots, d_k);\ d_1 \geq \cdots \geq d_k$ are the singular values of X.

X is singular if any one of the d_j is zero.

$$X(X'X)^{-1}X' = UDV'(VDU'UDV')^{-1}VDU' = UU'$$

Then

$$Xb = X(X'X)^{-1}X'y$$

$$= UU'y$$

where, $U'y$ is Co-ordinates of y with respect to the orthonormal basis U.

Further

$$X\hat{\beta}_{Ridge} = X(X'X + cI_k)^{-1}X'y$$

$$= UD(D^2 + cI_k)^{-1}DU'y$$

$$= \sum_{j=1}^{k} u_j \frac{d_j^2}{(d_j^2 + c)} u_j'y \ \left(u_j : j^{th} \text{ column of } U\right) \quad (23)$$

The ridge regression shrinks the co-ordinates of y with respect to the orthonormal basis U by the factor $\frac{d_j^2}{(d_j^2+c)}$. Greater shrinkage is applied to co-ordinates corresponding to smaller values of $d_j^2 = \lambda_j$.

$\lambda_j\ (j = 1, \dots, k)$ are the eigen values of $X'X$.

1.7.1.2 Least Absolute Shrinkage and Selection Operator (LASSO)

In LASSO regression, a penalty term proportional to the sum of the absolute values of the coefficients is added to the OLS objective function. LASSO can shrink some coefficients exactly to zero, performing automatic variable selection and providing sparse solutions.

The LASSO estimator is defined as

$$\hat{\beta}_{lasso} = \underset{\beta}{\operatorname{argmin}}\{(y - X\beta)'(y - X\beta) + c\sum_{j=1}^{k}|\beta_j|\}$$

(24)

We have,

Ridge Estimator is with $L_2 - \text{penalty}$ and LASSO estimator is with $L_1 - \text{penalty}$. Equivalently

$$\hat{\beta}_{lasso} = \underset{\beta}{\operatorname{argmin}}(y - X\beta)'(y - X\beta) \text{ subject to} \sum_{j=1}^{k}|\beta_j| \leq \delta$$

(25)

LASSO solution is a quadratic programming.

If δ is larger than $\delta_0 = \sum_{j=1}^{k}|\beta_j|$, $\hat{\beta}_{lasso}$ is equivalent to b. If δ is equal to $\delta_0/2$, the amount of shrinkage is 50%. It makes some of the coefficients zero.

Choice of regularization parameter c

K-fold cross validation:

Data are randomly split in K groups. Model constructed for $K-1$ groups and validated on K^{th} group. Sum of predictive errors for each model is an estimate of squared prediction error.

We select that value of c for which estimate of squared prediction error is minimum.

NOTE: Both ridge regression and lasso regression are forms of penalized regression that offer different trade-offs between bias and variance:

- Ridge regression generally reduces variance more effectively than LASSO but does not perform variable selection.
- LASSO can perform both variable selection and shrinkage but may have higher bias in estimation compared to ridge regression.

The choice between ridge regression and LASSO (or a combination known as elastic net) depends on the specific characteristics of the data, including the presence of multicollinearity, the desired interpretability of the model, and the importance of variable selection in the analysis. These techniques are widely used in machine learning and statistical modeling to improve the performance and interpretability of regression models, especially when dealing with high-dimensional data or correlated predictors.

1.8 Self-Assessment Exercise

1. Discuss the problem of multicollinearity and its various consequences.
2. Give various measures of multicollinearity.
3. Describe the method of principal components to overcome the multicollinearity problem.
4. Describe the ordinary ridge regression (ORR) estimator to overcome the multicollinearity problem.
5. How can we obtain characterizing scalar of ordinary ridge regression estimator using ridge trace?
6. Gie interpretations of ridge regression estimator and LASSO as penalized regression estimators.
7. Describe the ordinary ridge regression (ORR) estimator to overcome the problem of multicollinearity and derive its bias and MSE. Discuss various methods for selecting the characterizing scalar in ORR estimator.

1.9 Summary

Multicollinearity is a common issue in regression analysis where predictor variables exhibit high correlations among themselves. It leads to to inflated standard errors of OLS estimators of regression coefficients and ambiguous interpretation of the relationships between predictors and the dependent variable. We have discussed the consequences of multicollinearity, which include unreliable coefficient estimates, difficulty in interpreting variable importance, and potential for misleading conclusions. Different measures such as Variance Inflation Factors (VIF) can be used to detect and address highly correlated predictors.

We have explored several strategies to overcome multicollinearity, such as increasing sample size, careful variable selection, using regularization techniques like Ridge regression and LASSO, applying Principal Component regression.

1.10 References

- Gujarati, D. and Guneshker, S. (2007). *Basic Econometrics*, 4th Ed., McGraw Hill Companies.
- Johnston, J. and J. Dinardo (1997). *Econometric Methods*, 2nd Ed., McGraw Hill International.
- Judge, G.G., W.E. Griffiths, R.C. Hill Lütkepohl and T. C. Lee (1985). The theory and practice of econometrics, Wiley.

1.11 Further Readings

- Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression. Applications to non-orthogonal problems. Technometrics 12.
- Koutsoyiannis, A. (2004). *Theory of Econometrics*, 2 Ed., Palgrave Macmillan Limited.

- Maddala, G.S. and Lahiri, K. (2009). *Introduction to Econometrics*, 4 Ed., John Wiley & Sons.

- Ullah, A. and Vinod, H.D. (1981). Recent advances in Regression Methods, Marcel Dekker.

Unit 3. Estimation of parameters and prediction

3.1. Introduction

3.2. Objectives

3.3. Estimation under Linear Restrictions

3.3.1. Restricted Least Squares Estimation

3.3.2. Properties of b_R:

3.3.3. Maximum Likelihood Estimator under Exact Restrictions:

3.4. Tests of Linear Hypothesis

3.4.1. Likelihood Ratio Test for Set of Linear Hypothesis

3.5. Model in Deviation Form and ANOVA

3.5.1. Analysis of Variance

3.6. Performance of a regression model: R^2 and adjusted R^2

3.6.1. Relation between F-ratio for testing the significance of the regression and Coefficient of Determination:

3.7. Confidence interval estimation

3.7.1. Joint Confidence Set for all the Coefficients:

3.8. Point and Interval Prediction

3.9. Self-Assessment Exercise

3.10. Summary

3.11. References

3.12. Further Reading

3.1. Introduction

This unit focuses on advanced techniques in multiple linear regression analysis, designed to refine estimation methods, test model assumptions, and enhance interpretability and prediction accuracy. By integrating statistical tools with theoretical insights, this unit equips learners to handle complex regression scenarios with confidence. The key topics covered include:

Restricted Regression Estimation in Multiple Linear Regression Models

Explore how to incorporate constraints on regression coefficients to improve model estimation and reflect specific theoretical or practical considerations.

Tests for Linear Restrictions

Learn hypothesis testing techniques to evaluate whether specific linear relationships among coefficients hold, providing insights into model structure and variable importance.

Model in Deviation Form

Understand how re-expressing a model in deviation form (mean-centered variables) simplifies interpretation and calculation, particularly in the presence of interaction terms.

Analysis of Variance (ANOVA) in Regression

Examine how ANOVA techniques decompose variability in regression models, offering a deeper understanding of the contribution of explanatory variables.

R-Square and Adjusted R-Squar

Gain proficiency in assessing the goodness-of-fit of a regression model, balancing the trade-off between explanatory power and model complexity.

Interval Estimation of Regression Coefficients

Learn how to calculate confidence intervals for regression coefficients, providing a measure of the precision and reliability of the estimates.

Point and Interval Prediction

Develop skills to make accurate predictions from regression models, including specific value forecasts (point predictions) and range-based predictions (interval predictions) with confidence levels.

By the end of this unit, learners will have a comprehensive understanding of these advanced regression topics, enabling them to evaluate and enhance regression models effectively for robust analysis and decision-making.

3.2. Objective

After completing this Block, students should have developed a clear understanding of:

- Estimation under Linear Restrictions
- Tests of Linear Hypothesis
- Model in Deviation Form and ANOVA
- Performance of a regression model: R^2 and adjusted R^2
- Confidence interval estimation
- Point and Interval Prediction

3.3. Estimation under Linear Restrictions

Example:

Cobb-Douglas production function

$$Y = AL^{\beta_1}K^{\beta_2}$$

$$logY = logA + \beta_1 logL + \beta_2 logK + u$$

Y: Total production, L: Labour input, K: Capital input

Assumption of constant returns to scale means that doubling the usage of capital K and labor L will also double output Y.

This implies that $\beta_1 + \beta_2 = 1$.

i.e.

$$2Y = A(2L)^{\beta_1}(2K)^{\beta_2} = (2)^{\beta_1+\beta_2}Y$$

$$\Rightarrow 2 = (2)^{\beta_1+\beta_2}$$

Diminishing returns to scale means output increases in lesser proportion than increase in factor inputs.

This implies that $\beta_1 + \beta_2 < 1$.

Increasing returns to scale means output increases in bigger proportion than increase in factor inputs.

This implies that $\beta_1 + \beta_2 > 1$.

Sometimes we have prior information about the regression coefficients in the form of linear constraints binding the coefficients.

Such prior information may be available from

(a) Some theoretical considerations.

(b) Past experience.

(c) Empirical investigations.

(d) Some extraneous sources etc.

Forms of linear restrictions:

(i) Exact linear restrictions. i.e. $\beta_1 + \beta_2 = 1$

(ii) Stochastic linear restrictions. i.e. $\beta_1 + \beta_2 + v = 1$; v is a random variable

(iii)Inequality restrictions. i.e. $\beta_1 + \beta_2 < 1$

$$\beta_1 + \beta_2 > 1$$

We are interested in:

(i) Estimation under set of exact linear restrictions (point estimation or confidence set/interval estimation)

(ii) Test of set of linear restrictions

Let us consider a set of q (<k) linearly independent restrictions on β:

$$R_{11}\beta_1 + \cdots + R_{1k}\beta_k = r_1$$

$$\vdots$$

$$R_{q1}\beta_1 + \cdots + R_{qk}\beta_k = r_q$$

Define

$$R = \begin{pmatrix} R_{11} & \cdots & R_{1k} \\ \vdots & \ddots & \vdots \\ R_{q1} & \cdots & R_{qk} \end{pmatrix} : q \times k, \; r = \begin{pmatrix} r_1 \\ \vdots \\ r_q \end{pmatrix} : q \times 1$$

Then the set of linear restrictions can be written as

$$R\beta = r$$

The rank of matrix R is q (<k).

We require the method of Lagrange multiplier, which is defined as follows:

For minimizing (or maximizing) a function $f(\beta)$ subject to the restriction $g(\beta) = 0$, we define a Lagrangian function

$$h(\beta, \lambda) = f(\beta) - \lambda g(\beta)$$

and then minimizing (or maximizing) $h(\beta, \lambda)$.

λ is the Lagrange's multiplier.

3.3.1. Restricted Least Squares Estimation:

We estimate β utilizing sample information

$$y = X\beta + u$$

and prior information

$$R\beta = r \qquad (3.1)$$

Result 3.3.: The restricted regression estimator obtained by minimizing the residual sum of squares

$$(y - X\beta)'(y - X\beta)$$

subject to the restriction $R\beta = r$ *is*

$$b_R = b + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - Rb) \quad (3.2)$$

Proof: Consider the Lagrange's function

$$S(\beta, \lambda) = (y - X\beta)'(y - X\beta) - 2\lambda'(R\beta - r)$$

Differentiating with respect to β and λ, we obtain

$$\frac{\partial S(\beta, \lambda)}{\partial \beta} = 2X'X\beta - 2X'y - 2R'\lambda = 0 \qquad (3.3)$$

$$\frac{\partial S(\beta, \lambda)}{\partial \lambda} = R\beta - r = 0 \qquad (3.4)$$

Writing the resulting solution for β as b_R, from (3.3) we get

$$b_R = (X'X)^{-1}(X'y + R'\lambda) = b + (X'X)^{-1}R'\lambda \qquad (3.5)$$

From (3.5), substituting b_R in (3.4), we get

$$Rb_R - r = 0 \Rightarrow \lambda = (R(X'X)^{-1}R')^{-1}(r - Rb)$$

so that

$$b_R = b + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - Rb)$$

Hence the result follows■

3.3.2. Properties of b_R:

(i) b_R satisfies Linear Restrictions: We have

$$Rb_R = Rb + R(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - Rb) = r$$

Hence b_R satisfies the linear restrictions (3.1).

(ii) Unbiasedness: We have

$$E(b_R) = E(b) + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}E(r - Rb)$$

$$= \beta + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta)$$

When the restriction $R\beta = r$ is correct, $E(b_R) = \beta$ and b_R is an unbiased estimator of β.

However, if $r - R\beta = \gamma \neq 0$, then b_R is a biased estimator and its bias is given by

$$E(b_R - \beta) = (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}\gamma$$

(iii)MSE Matrix: Let us write

$$C = (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}$$

The MSE Matrix of b_R is given by

$$E(b_R - \beta)(b_R - \beta)' = E[b - \beta + C(r - Rb)][b - \beta + C(r - Rb)]'$$

$$= E[\{(I_k - CR)(b - \beta) + C(r - R\beta)\} \times \{(I_k - CR)(b - \beta) + C(r - R\beta)\}']$$

$$= E[(I_k - CR)(b-\beta)(b-\beta)'(I_k - CR)'] + E[C(r-R\beta)(b-\beta)'(I_k - CR)']$$
$$+ E[C(r-R\beta)(r-R\beta)'C'] + E[(I_k - CR)(b-\beta)(b-\beta)'(I_k - CR)']$$

$$= \sigma_u^2(I_k - CR)(X'X)^{-1}(I_k - R'C') + C(r-R\beta)(r-R\beta)'C'$$

We observe that

$$CR(X'X)^{-1}R'C' = CR(X'X)^{-1} = (X'X)^{-1}R'C' = (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}$$

So that

$$(I_k - CR)(X'X)^{-1}(I_k - R'C') = (X'X)^{-1} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}$$

Hence

$$\mathrm{E}(b_R-\beta)(b_R-\beta)'$$
$$= \sigma_u^2[(X'X)^{-1} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}]$$
$$+ (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r-R\beta)(r-R\beta)'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}$$

When the restrictions $R\beta = r$ are correct, the variance-covariance matrix of b_R is

$$\mathrm{E}(b_R-\beta)(b_R-\beta)' = \sigma_u^2[(X'X)^{-1} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}]$$

3.3.3. Maximum Likelihood Estimator under Exact Restrictions:

Let $u \sim N(0, \sigma_u^2 I_n)$. The LF is given by

$$L(\beta, \sigma_u^2) = \left(\frac{1}{2\pi\sigma_u^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{\frac{(y-X\beta)'(y-X\beta)}{\sigma_u^2}\right\}\right]$$

We maximize the following log of LF combined with the Lagrange's multiplier term:

$$\ln L(\beta, \sigma_u^2, \lambda) = -\frac{n}{2}\ln\sigma_u^2 - \frac{1}{2\sigma_u^2}(y-X\beta)'(y-X\beta) + \lambda'(R\beta - r)$$

Partially differentiating $lnL(\beta, \sigma_u^2, \lambda)$ with respect to β, σ_u^2 and λ, substituting the derivatives equal to zero and denoting the MLE of β and σ_u^2 as $\hat{\beta}_R$ and $\hat{\sigma}_{uR}^2$ respectively, we have

$$\left.\frac{\partial \ln L(\beta, \sigma_u^2, \lambda)}{\partial \beta}\right|_{\beta=\hat{\beta}_R, \sigma_u^2=\hat{\sigma}_{uR}^2} = -\frac{1}{\hat{\sigma}_{uR}^2}\left(X'X\hat{\beta}_R - X'y\right) + 2R'\lambda = 0$$

$$\left.\frac{\partial \ln L(\beta, \sigma_u^2, \lambda)}{\partial \lambda}\right|_{\beta=\hat{\beta}_R, \sigma_u^2=\hat{\sigma}_{uR}^2} = 2\left(R\hat{\beta}_R - r\right) = 0$$

$$\left.\frac{\partial \ln L(\beta, \sigma^2, \lambda)}{\partial \sigma_u^2}\right|_{\beta=\hat{\beta}_R, \sigma_u^2=\hat{\sigma}_{uR}^2} = -\frac{2n}{\hat{\sigma}_{uR}^2} + \frac{2\left(y - X\hat{\beta}_R\right)'\left(y - X\hat{\beta}_R\right)}{\hat{\sigma}_{uR}^4} = 0.$$

Solving these normal equations, we get

$$\tilde{\lambda} = \frac{[R(X'X)^{-1}R']^{-1}(r - Rb)}{\hat{\sigma}_u^2}$$

$$\hat{\beta}_R = b + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - Rb) = b_R$$

$$\hat{\sigma}_{uR}^2 = \frac{(y - Xb_R)'(y - Xb_R)}{n}.$$

The Hessian matrix of second order partial derivatives is positive definite at $\beta = \hat{\beta}_R, \sigma_u^2 = \hat{\sigma}_{uR}^2$. The restricted least squares and restricted maximum likelihood estimators of β are the same whereas for σ_u^2, the two estimators are different.

3.4. Tests of Linear Hypothesis

Suppose we want to test the null hypothesis H_0:

$$R_{11}\beta_1 + \cdots + R_{1k}\beta_k = r_1 \vdots R_{q1}\beta_1 + \cdots + R_{qk}\beta_k = r_q$$

against $H_1: not\ H_0$

or $H_0: R\beta = r$ against $H_1: R\beta \neq r$.

$$\mathrm{R} = \begin{pmatrix} R_{11} & \cdots & R_{1k} \\ \vdots & \ddots & \vdots \\ R_{q1} & \cdots & R_{qk} \end{pmatrix}: q \times k, \qquad r = \begin{pmatrix} r_1 \\ \vdots \\ r_q \end{pmatrix}: q \times 1$$

Cobb-Douglas production function

$$logY = \beta_0 + \beta_1 logL + \beta_2 logK, \beta_0 = logA$$

Y: Total production, L: Labour input, K: Capital input

Assumption of constant returns to scale implies that

$H_0: \beta_1 + \beta_2 = 1$. We take

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, R = (0 \quad 1 \quad 1), r = 1, \qquad q = 1$$

$$R\beta = r \Rightarrow \beta_1 + \beta_2 = 1$$

Examples:

(i) Significance of a Regression Coefficient:

$q = 1$, and $R: 1 \times k$ vector with all elements equal to zero except the i^{th} element which is 1.

Thus $R = (0\ 0\ \ldots 0\ 1\ 0\ \ldots 0)$

Further $r = \beta_{io}$. Then, we get the hypothesis $\beta_i = \beta_{io}$.

For $r = 0$, we get the hypothesis $\beta_i = 0$.

(ii) Significance of Complete Regression:

$q = k - 1,$

$r = (0\ 0\ \ldots\ 0)' : (k-1) \times 1$

$$R = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = (0 \quad I_{k-1})$$

Then the hypothesis becomes

$$\begin{pmatrix}\beta_2\\ \vdots\\ \beta_k\end{pmatrix}=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix} \text{ or } \beta_2=\beta_3=\cdots=\beta_k=0$$

We are not interested in the hypothesis $\beta_1=0$ as it is corresponding to the intercept term and taking $\beta_1=0$ means additional assumption that mean level is also zero. We are mainly interested in the significance of entire regression not in the mean level of Y.

(iii) Equality of Two Coefficients:

$q=1, r=0$

$R=(0\ 0\ \ldots 0\ 1\ 0\ \ldots -1\ \ldots 0)$(1 at ith place and – 1 at jth place)

This leads to the hypothesis $\beta_i=\beta_j$.

3.4.1. Likelihood Ratio Test for Set of Linear Hypothesis:

We assume that $u\sim N(0,\sigma_u^2 In)$. The Likelihood Function is

$$L(\beta,\sigma_u^2)=\left(\frac{1}{2\pi\sigma_u^2}\right)^{\frac{n}{2}}\exp\left[-\frac{1}{2}\left\{\frac{(y-X\beta)'(y-X\beta)}{\sigma_u^2}\right\}\right]$$

The likelihood ratio test statistic for testing $H_0: R\beta=r$ against $H_1: R\beta\neq r$

$$\lambda_{LR}=\frac{\max L\,(\beta,\sigma_u^2)}{\max L\,(\beta,\sigma_u^2|R\beta=r)}$$

Now $\max L\,(\beta,\sigma_u^2)$ occurs when $\beta=b=(X'X)^{-1}X'y$ and

$$\sigma_u^2=\hat{\sigma}_u^2=\frac{1}{n}(y-Xb)'(y-Xb)$$

For these values

$$\max L\,(\beta,\sigma_u^2)=\left(\frac{1}{2\pi\hat{\sigma}_u^2}\right)^{\frac{n}{2}}\exp\left[-\frac{1}{2}\left\{\frac{(y-Xb)'(y-Xb)}{\hat{\sigma}_u^2}\right\}\right]$$

$$= \left(\frac{1}{2\pi\hat{\sigma}_u^2}\right)^{\frac{n}{2}} \exp\left[-\frac{n}{2}\right]$$

Further $\max L\,(\beta, \sigma_u^2 | R\beta = r)$ occurs when $\beta = b_R$ and

$$\sigma_u^2 = \hat{\sigma}_{uR}^2 = \frac{1}{n}(y - Xb_R)'(y - Xb_R)$$

For these values

$$\max L\,(\beta, \sigma_u^2 | R\beta = r) = \left(\frac{1}{2\pi\hat{\sigma}_{uR}^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{\frac{(y - Xb_R)'(y - Xb_R)}{\hat{\sigma}_{uR}^2}\right\}\right]$$

$$= \left(\frac{1}{2\pi\hat{\sigma}_{uR}^2}\right)^{\frac{n}{2}} \exp\left[-\frac{n}{2}\right]$$

Hence

$$\lambda_{LR}^{\frac{2}{n}} = \frac{\hat{\sigma}_{uR}^2}{\hat{\sigma}_u^2} = \frac{(y - Xb_R)'(y - Xb_R)}{(y - Xb)'(y - Xb)}$$

Further

$$(y - Xb_R)'(y - Xb_R)$$

$$= [y - Xb - X(X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}(r - Rb)]'[y - Xb - X(X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}(r - Rb)]$$

$$\because\ (y - Xb)'X(X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}(r - Rb) = 0$$

Therefore

$$(y - Xb_R)'(y - Xb_R) = (y - Xb)'(y - Xb) + (r - Rb)'(R(X'X)^{-1}R')^{-1}(r - Rb)$$

Therefore

$$\lambda_{LR}^{\frac{2}{n}} - 1 = \frac{(r - Rb)'(R(X'X)^{-1}R')^{-1}(r - Rb)}{(y - Xb)'(y - Xb)}$$

Distribution of test statistic

Since $b \sim N(\beta, \sigma_u^2 (X'X)^{-1})$, we have

$Rb \sim N(R\beta, \sigma_u^2 R(X'X)^{-1}R')$

Under H_0 $Rb \sim N(r, \sigma_u^2 R(X'X)^{-1}R')$.

Hence under H_0

$$Q_1 = \frac{1}{\sigma_u^2}(r - Rb)'\{R(X'X)^{-1}R'\}^{-1}(r - Rb) \sim \chi^2(q)$$

Further

$$Q_2 = \frac{1}{\sigma_u^2} e'e = \frac{1}{\sigma_u^2}(y - Xb)'(y - Xb) \sim \chi^2(n - k),$$

independently of b, and hence, independently of Q_1.

Therefore, under H_0

$$F_{cal} = \frac{\frac{Q_1}{q}}{\frac{Q_2}{n-k}} = \frac{(r - Rb)'\{R(X'X)^{-1}R'\}^{-1}(r - Rb)}{\frac{e'e}{n-k}} \sim F(q, n-k)$$

We reject $H0$ when $F_{cal} > F(\alpha; q, n-k)$.

$F(\alpha; q, n-k)$: tabulated value of $F(q, n-k)$ at α level of significance.

(i) <u>Significance of a Regression Coefficient:</u>

$q = 1, R = (0\ 0\ \ldots 0\ 1\ 0\ \ldots 0), r = \beta_{io}$

Hypothesis $H_0: \beta_i = \beta_{io}$, against $H_1: \beta_i \neq \beta_{io}$

$Rb - r = b_i - \beta_{io}$

$R(X'X)^{-1}R' = c_{ii}$

c_{ii}: i^{th} diagonal element of $(X'X)^{-1}$

$$(Rb - r)'\{R(X'X)^{-1}R'\}^{-1}(Rb - r) = \frac{(b_i - \beta_{io})^2}{c_{ii}}$$

$$F_{cal} = \frac{(b_i - \beta_{io})^2/c_{ii}}{s^2} \sim F(1, n-k) (\text{under } H_0)$$

$$s^2 = \frac{e'e}{n-k}$$

Hence, under H_0

$$\frac{b_i - \beta_{io}}{s\sqrt{c_{ii}}} \sim t(n-k)$$

We reject the null hypothesis^{e} H_0: $\beta_i = \beta_{io}$ at α level of significance if

$$\left|\frac{b_i - \beta_{io}}{s\sqrt{c_{ii}}}\right| > t(\alpha; n-k)$$

(ii) <u>Testing Significance of Complete Regression:</u>

$q = k-1, r = (0\ 0\ \ldots\ 0)'$,

$R = (0 \quad I_{k-1})$

$Rb - r = b_{(2)}$ $\left(\because b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right)$

$R(X'X)^{-1}R' : (k-1) \times (k-1)$

sub matrix formed by last $k-1$ rows and columns in $(X'X)^{-1}$

$X = (l_n \quad X_{(2)})$

$$X'X = \begin{pmatrix} l_n' \\ X_{(2)}' \end{pmatrix}(l_n \quad X_{(2)}) = \begin{pmatrix} n & l_n'X_{(2)} \\ X_{(2)}'l_n & X_{(2)}'X_{(2)} \end{pmatrix} \qquad (\because\ l_n'l_n = n)$$

Write

$$(X'X)^{-1} = \begin{pmatrix} \lambda^{11} & \lambda' \\ \lambda & \Lambda \end{pmatrix}$$

Then

$$\begin{pmatrix} n & l_n'X_{(2)} \\ X_{(2)}'l_n & X_{(2)}'X_{(2)} \end{pmatrix}\begin{pmatrix} \lambda^{11} & \lambda' \\ \lambda & \Lambda \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{I_{k-1}} \end{pmatrix} \qquad (\because X'X(X'X)^{-1} = I_k)$$

$$\Rightarrow \Lambda = \left(X_{(2)}'AX_{(2)}\right)^{-1} \qquad \left(\because I_n - \frac{1}{n}l_nl_n'\right)$$

Thus F-statistic becomes

$$F = \frac{\frac{b_{(2)}'X_{(2)}'AX_{(2)}b_{(2)}}{(k-1)}}{\frac{e'e}{(n-k)}} = \frac{\frac{ESS}{(k-1)}}{\frac{RSS}{(n-k)}}$$

In terms of R^2

$$F = \frac{(n-k)}{(k-1)}\frac{R^2}{(1-R^2)}.$$

3.5. Model in Deviation Form and ANOVA

The OLS Regression equation is

$$y = Xb + e \qquad (3.6)$$

Let us define

$$A = I_n - \left(\frac{1}{n}\right)l_nl_n'$$

$l_n : n \times 1$ vector with all elements equal to 1.

For any $n \times 1$ vector y

$\frac{1}{n} l_n' y = \bar{y}$,

Further

$Ay = y - \bar{y} l_n$

Hence Ay gives the observation vector y as deviation from mean. Further A is a symmetric idempotent matrix.

$$Al_n = \left\{ I_n - \left(\frac{1}{n}\right) l_n l_n' \right\} l_n$$

$$= l_n - \frac{1}{n} l_n l_n' l_n$$

$$= l_n - l_n = 0, \quad (l_n' l_n = n)$$

$X'e = 0,$

$Ae = e$

The first column of X matrix is l_n. Partition X matrix as

$X = [l_n \ X_{(2)}],$

Here $X_{(2)}$ is $n \times (k-1)$ matrix of observations on all explanatory variables, except first column corresponding to intercept term. We have

$$\frac{1}{n} l_n' X_{(2)} = (\bar{X}_2 \ \ldots \bar{X}_k)$$

Further, $\forall j = 2, \ldots, k$

$\bar{X}_j = \frac{1}{n} \sum_{i=1}^{n} X_{ij}$ is the mean of j^{th} explanatory variable

Then, we can write (3.6) as

$$y = l_n b_1 + X_{(2)} b_{(2)} + e \tag{3.7}$$

where

$$b = \begin{pmatrix} b_1 \\ b_{(2)} \end{pmatrix}.$$

b_1 is the intercept term and $b_{(2)}$ is $(k-1 \times 1)$ vector of slope coefficients. Pre multiplying (3.7) by A, and observing that $Al_n = 0, Ae = e$, we obtain

$$Ay = AX_{(2)} b_{(2)} + e \tag{3.8}$$

Since $X'e = 0$ implies that $X'_{(2)} e = 0,$ we have

$$X'_{(2)} Ay = X'_{(2)} AX_{(2)} b_{(2)} \left(AX_{(2)}\right)' Ay = \left(AX_{(2)}\right)' \left(AX_{(2)}\right) b_{(2)} \tag{3.9}$$

Here Ay is y vector as deviation from mean and $AX_{(2)}$ is the matrix of explanatory variables as deviation from mean.

Equation (3.9) is a set of normal equations in terms of deviations, whose solution leads to OLS estimators of slope coefficients, *i.e.*, $b_{(2)}$.

Pre multiplying

$$y = l_n b_1 + X_{(2)} b_{(2)} + e$$

by $\frac{1}{n} l'_n$ leads to

$$\bar{y} = b_1 + (\bar{X}_2 \ \ldots \bar{X}_k) \begin{pmatrix} b_2 \\ \vdots \\ b_k \end{pmatrix}$$

or $b_1 = \bar{y} - b_2 \bar{X}_2 \ - \cdots - b_k \bar{X}_k$

3.5.1. Analysis of Variance

We have

$y'Ay = (Ay)'(Ay)$

$= \left(AX_{(2)}b_{(2)} + e\right)'\left(AX_{(2)}b_{(2)} + e\right)$

$= b'_{(2)}X'_{(2)}AX_{(2)}b_{(2)} + e'e$

$y'Ay = \sum_{j=1}^{n}\left(y_j - \bar{y}\right)^2$ denotes the Total Sum of Squares (TSS) having $(n-1)$ d.f.

$b'_{(2)}X'_{(2)}AX_{(2)}b_{(2)}$ is the Explained Sum of Squares (ESS) having $(k-1)$ d.f.

$e'e$ is the Residual Sum of Squares (RSS) having $(n-k)$ d.f.

Hence TSS=ESS+RSS

Sum of squares and corresponding d.f. are additive.

Analysis of Variance (ANOVA) Table

S.S.	d.f.	M.S.S.	F ratio
ESS	$k-1$	$ESS/(k-1)$	$\frac{ESS/(k-1)}{RSS/(n-k)}$
RSS	$n-k$	$RSS/(n-k)$	
TSS	$n-1$		

3.6. Performance of a regression model: R^2 and adjusted R^2

The error sum of squares is

$e'e = (y - Xb)'(y - Xb)$

In two variables simple linear regression, the model is good if r^2 (square of correlation coefficient) is high. Here

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

R^2 is the coefficient of determination or square of multiple correlation coefficient.

It gives the proportion of variation in the dependent variable that is explained by the independent variables. Further

$$0 \leq R^2 \leq 1$$

If $R^2 \approx 0$, it indicates the poor fit of the model. Further, $R^2 \approx 1$ indicates the best fit of the model.

Suppose $R^2 = 0.95$, then it indicates that 95% of the variation in *y* is explained by the explanatory variables. In simple words, the model is 95% good. Similarly, we can interpret any other value of R^2 between 0 and 1. Thus, R^2 indicates the adequacy of fitted model. Whenever we add an explanatory variable to the model, the value of R^2 always increases. In case the included variable is irrelevant, the increase in R^2 gives a misleading picture. We may also face the problem of multicollinearity if too many irrelevant variables are included. Keeping in view this problem, adjusted R^2, denoted as $\bar{R}^2$ or $adj\ R^2$ is used.

The adjusted R^2 is defined as

$$\bar{R}^2 = 1 - \frac{\frac{RSS}{n-k}}{\frac{TSS}{n-1}}$$

$$= 1 - \left(\frac{n-1}{n-k}\right)(1 - R^2) \qquad \left(R^2 = 1 - \frac{RSS}{TSS}\right)$$

$RSS/(n-k)$: Unbiased estimator of variance of residual e

$TSS/(n-1)$: Unbiased estimator of variance of y

If adding a variable produces a too small reduction in $(1 - R^2)$ to compensate for the increase in $(n-1)/(n-k)$, $\bar{R}^2$ may decrease. A problem with $\bar{R}^2$ is that it may take negative value, which is difficult to interpret. For instance, suppose, $k = 5,\ n = 15,\ R^2 = 0.16$, then

$$\bar{R}^2 = 1 - \frac{15-1}{15-5}(1-0.16) = -0.176$$

Some limitations of R^2

(i) For model with intercept term

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2}$$

For model without intercept term

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{n} y_j^2}$$

R^2 for model without intercept term (regression forced through origin) is usually greater than R^2 for model with intercept term. However, R^2 of two different linear models cannot be compared.

(ii) R^2 is sensitive to extreme values, so R^2 lacks robustness.

(iii) Suppose we have following two alternative models:

(a) $y_i = \beta_1 + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i, i = 1,2,..,n$

(b) $\log y_i = \gamma_1 + \gamma_2 X_{i2} + \cdots + \gamma_k X_{ik} + v_i.$

Then coefficients of determination for the two models are

$R_1^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$: For model (a)

$R_2^2 = 1 - \frac{\sum_{i=1}^{n}(\log y_i - \widehat{\log y_i})^2}{\sum_{i=1}^{n}(\log y_i - \overline{\log y})^2}$: For model (b)

Then R_1^2 and R_2^2 are not comparable. If we define

$$R_3^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \text{antilog}\,\hat{z}_i)}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \hat{z}_i = \widehat{logy}_i$$

then R_1^2 and R_3^2 provide a better option for the comparison of two models.

3.6.1. Relation between F-ratio for testing the significance of the regression and Coefficient of Determination:

For testing $H_0: \beta_2 = \beta_3 = \cdots = \beta_k = 0$, the *F*-test statistic is

$$\text{F} = \frac{ESS/(k-1)}{RSS/(n-k)}$$

Further R^2 is

$$R^2 = 1 - \frac{RSS}{TSS}$$

Hence

$$\text{F} = \frac{n-k}{k-1}\frac{R^2}{1-R^2}$$

$R^2 = 0 \Rightarrow F = 0,$

$$R^2 = 1 \Rightarrow F = \infty$$

In fact, F is a monotone increasing function of R^2.

3.7. Confidence interval estimation

Confidence interval for the individual regression coefficient:

Consider the model

$$y = X\beta + u; u \sim N(0, \sigma_u^2 I_n)$$

Then $b \sim N(\beta, \sigma_u^2 (X'X)^{-1})$.

Let $c_{jj} = j^{th}$ diagonal element of $(X'X)^{-1}$. Then $b_j \sim N(\beta_j, \sigma_u^2 c_{jj})$.

$$s^2 = \frac{1}{n-k}(y - Xb)'(y - Xb)$$

Then

$$\frac{b_j - \beta_j}{\sigma_u \sqrt{c_{jj}}} \sim N(0,1)$$

$$\frac{(n-k)s^2}{\sigma_u^2} \sim \chi^2(n-k)$$

Hence

$$\frac{b_j - \beta_j}{\sqrt{s^2 c_{jj}}} \sim t(n-k)$$

$(1-\alpha)100\%$ confidence interval is obtained as

$$P\left[-t_{\frac{\alpha}{2},n-k} \leq \frac{b_j - \beta_j}{\sqrt{s^2 c_{jj}}} \leq t_{\frac{\alpha}{2},n-k}\right]$$

$$= 1 - \alpha P\left[b_j - t_{\frac{\alpha}{2},n-k}\sqrt{s^2 c_{jj}} \leq \beta_j \leq b_j + t_{\frac{\alpha}{2},n-k}\sqrt{s^2 c_{jj}}\right]$$

$$= 1 - \alpha. \text{Confidence Interval is } \left(b_j - t_{\frac{\alpha}{2},n-k}\sqrt{s^2 c_{jj}}, b_j + t_{\frac{\alpha}{2},n-k}\sqrt{s^2 c_{jj}}\right)$$

3.7.1. Joint Confidence Set for all the Coefficients:

We have

$$\frac{\frac{(b-\beta)'X'X(b-\beta)}{k}}{\frac{RSS}{n-k}} \sim F(k, n-k)$$

$$\Rightarrow P\left[\frac{(b-\beta)'X'X(b-\beta)}{ks^2} \leq F_\alpha(k, n-k)\right] = 1 - \alpha.$$

Then $(1-\alpha)100\%$ joint confidence set for β is

$$\left\{\beta: \frac{(b-\beta)'X'X(b-\beta)}{ks^2} \leq F_\alpha(k, n-k)\right\}.$$

3.8. Point and Interval Prediction

Prediction or Forecasting When X variables are Uncertain Model:

$$y = X\beta + u$$

$x_f' = (1\ x_{2f}\ \ldots x_{kf})$: True value of the explanatory variable for the forecast period *f*

$\hat{x}_f' = (1\ \hat{x}_{2f}\ \ldots \hat{x}_{kf})$: Estimated values of the explanatory variables for the forecast period *f*

The true value of y_f is given by

$$y_f = x_f'\beta + u_f$$

The point prediction is

$$\hat{y}_f = \hat{x}_f' b$$

$b = (X'X)^{-1}X'y$: OLS Estimator of β $\qquad (E(uu') = \sigma_u^2 I_n)$

Forecast error is

$$e_f = y_f - \hat{y}_f$$

$$= u_f + x_f'\beta - \hat{x}_f' b$$

$$= u_f - (\hat{x}_f' b - \hat{x}_f'\beta) - \hat{x}_f'\beta + x_f'\beta$$

$$= u_f - \hat{x}_f'(b - \beta) - \beta'(\hat{x}_f - x_f)$$

We assume that

(i) $E(\hat{x}_f) = x_f$, *i.e.*, forecaster makes unbiased forecasts of the x values.

(ii) Covariance between $\hat{x}_f$ and OLS estimator *b* is zero, i.e., $\mathrm{E}\left(\hat{x}_f'(b - \beta)\right) = 0$.

Then $E(e_f) = E[u_f - \hat{x}_f'(b - \beta) - \beta'(\hat{x}_f - x_f)] = 0.$

Thus $\hat{y}_f = \hat{x}_f' b$ is an unbiased forecast of y_f.

i.e. $E(\hat{y}_f) = E(\hat{x}_f' b) = x_f'\beta = E(y_f)$

Let $V(\hat{x}_f) = E\left[(\hat{x}_f - x_f)(\hat{x}_f - x_f)'\right]$.

The variance of forecast error is given by

$$E(e_f^2) = \sigma_{\hat{y}_f}^2$$

$$= E\left[u_f - \hat{x}_f'(b-\beta) - \beta'(\hat{x}_f - x_f)\right]^2$$

$$= E\left[u_f^2 + \hat{x}_f'(b-\beta)(b-\beta)'\hat{x}_f + \beta'(\hat{x}_f - x_f)(\hat{x}_f - x_f)'\beta\right] \qquad (3.10)$$

The cross-product terms have expectation zero. i.e.

$$E\left[u_f\ \hat{x}_f'(b-\beta)\right] = 0$$

$$E\left[u_f\ \beta'(\hat{x}_f - x_f)\right] = 0$$

$$E\left[\hat{x}_f'(b-\beta)\beta'(\hat{x}_f - x_f)\right] = 0$$

Now

$$E(u_f^2) = \sigma_u^2 \qquad (3.11)$$

$$E\left[\hat{x}_f'(b-\beta)(b-\beta)'\hat{x}_f\right] = E\left[tr\ (b-\beta)(b-\beta)'\hat{x}_f\hat{x}_f'\right]$$

$$= \sigma_u^2\left\{x_f'(X'X)^{-1}x_f + tr\left[(X'X)^{-1}V(\hat{x}_f)\right]\right\} \qquad (3.12) \qquad (\because\ E(\hat{x}_f\hat{x}_f') = V(\hat{x}_f) + x_f x_f')$$

$$\beta' E(\hat{x}_f - x_f)(\hat{x}_f - x_f)'\beta = \beta' V(\hat{x}_f)\beta \qquad (3.13)$$

Utilizing (3.10), (3.11), (3.12), (3.13), we obtain

$$\sigma_{\hat{y}_f}^2 = \sigma_u^2\left\{1 + x_f'(X'X)^{-1}x_f + tr\left[(X'X)^{-1}V(\hat{x}_f)\right]\right\} + \beta' V(\hat{x}_f)\beta$$

If $\hat{x}_f = x_f$, i.e., x_f is exactly known, so that $V(\hat{x}_f) = 0$, then

$$E(e_f^2) = \sigma_u^2\{1 + x_f'(X'X)^{-1}x_f\}.$$

An unbiased estimator of σ_u^2 is

$$s^2 = \frac{1}{n-k}(y - Xb)'(y - Xb).$$

Result 3.8.1.: If an unbiased estimator of $V(\hat{x}_f)$, say $\widehat{V(\hat{x}_f)}$ is available, then an unbiased estimator of $\sigma_{\hat{y}_f}^2$ is given by

$$\hat{\sigma}_{\hat{y}_f}^2 = s^2\left[1 + \hat{x}_f'(X'X)^{-1}\hat{x}_f\right] + tr\left\{(X'X)^{-1}\widehat{V(\hat{x}_f)}\right\} + b'\widehat{V(\hat{x}_f)}b \qquad (3.14).$$

Proof: The expression for $\sigma_{\hat{y}_f}^2$ is

$$\sigma_{\hat{y}_f}^2 = \sigma_u^2\{1 + x_f'(X'X)^{-1}x_f + tr\left[(X'X)^{-1}V(\hat{x}_f)\right]\} + \beta'V(\hat{x}_f)\beta. \qquad (3.15)$$

An unbiased estimator of σ_u^2 is s^2. Further

$$E(\hat{x}_f\hat{x}_f') = x_f x_f' + V(\hat{x}_f).$$

Hence, an unbiased estimator of $x_f x_f' + V(\hat{x}_f)$ is $\hat{x}_f\hat{x}_f'$, so that an unbiased estimator of

$x_f'(X'X)^{-1}x_f + tr\left[(X'X)^{-1}V(\hat{x}_f)\right]$ is

$$\hat{x}_f'(X'X)^{-1}\hat{x}_f \qquad (3.16)$$

Further

$$E\left\{b'\widehat{V(\hat{x}_f)}b\right\} = E\left\{tr\left(\widehat{V(\hat{x}_f)}bb'\right)\right\}$$

$$= tr\{V(\hat{x}_f)E(bb')\}$$

$$= \sigma_u^2 tr\{V(\hat{x}_f)(X'X)^{-1}\} + \beta'V(\hat{x}_f)\beta.$$

Thus, an unbiased estimator of $\beta' V(\hat{x}_f)\beta$ is

$$b'\widehat{V(\hat{x}_f)}b - s^2 tr\left\{\widehat{V(\hat{x}_f)}(X'X)^{-1}\right\} \quad (3.17)$$

Combining (3.15), (3.16), and (3.17), we obtain the result (3.14)■

When x_f is exactly known,

$$\widehat{\sigma^2_{\hat{y}_f}} = s^2\left[1 + x_f'(X'X)^{-1}x_f\right].$$

Prediction for the Model with Non-spherical disturbances

• In the linear model with non-spherical disturbances, the disturbances have interdependence.

• The pattern of sample residuals contain information which is useful in prediction of post-sample observations.

• This information can be utilized to obtain the best linear unbiased predictor.

• This leads to the gain in efficiency over the usual expected value estimator.

Best Linear Unbiased Prediction

Consider the model

$$y = X\beta + u; E(u) = 0,\ E(uu') = \sigma_u^2\Omega.$$

Consider the problem of predicting y_f for given $x_f\ (k \times 1)$ for the forecast period f. Let u_f be the disturbance term for the forecast period so that

$$y_f = x_f'\beta + u_f,\ E(u_f) = 0,\ E(u_f^2) = \sigma_u^2$$

$$E(u_f u) = \sigma_u^2\omega\omega: n \times 1 \text{ vector of correlations between } u_f \text{ and elements of } u$$

Result 3.8.2.: The best linear unbiased predictor (BLUP) of y_f is

$\hat{y}_f = X'b = x_f'\hat{\beta} + \omega'\Omega^{-1}(y - X\hat{\beta})$.

$\hat{\beta} = (X'\Omega^{-1}X)^{-1}\Omega^{-1}y$: GLS estimator of β.

Proof: Let $c\ (n \times 1)$ be a vector and $\hat{y}_f = c'y$ is the linear predictor of y_f.

For the BLUP, we obtain $c'y$ such that

$\sigma_p^2 = E(c'y - y_f)^2$ is minimum subject to $E(c'y - y_f) = 0$.

Now $p = c'y = c'X\beta + c'uc'y - y_f = c'X\beta + c'u - x_f'\beta - u_f$ $(\because y_f = x_f'\beta - u_f)$

$E(c'y - y_f) = c'X\beta - x_f'\beta = 0$ requires that

$X'c = x_f$ (3.18)

Further, using (3.18), we have

$\sigma_p^2 = E(c'y - y_f)^2$

$= E(c'u - u_f)^2$

$= E(c'u)^2 + E(u_f)^2 - 2E(u_f c'u)$

$= \sigma_u^2 c'\Omega c + \sigma_u^2 - 2\sigma_u^2 c'\omega$

$= \sigma_u^2(1 - 2c'\omega + c'\Omega c$

3.9. Self-Assessment Exercise

1. Explain the concept of restricted regression estimation. Why might it be useful in a multiple linear regression model?

2. Given a regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$, impose the restriction $\beta_1 = 2\beta_2$ and derive the restricted estimator.

3. What is the null hypothesis in a test for linear restrictions? Provide an example.

4. For the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$, test the restriction $\beta_1 + \beta_2 = 1$ using the F-test. Describe the steps and interpret the results.

5. Rewrite the regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$ in deviation form. Explain how this transformation can simplify computations.

6. Discuss the test for a set of linear hypotheses in a linear regression model. How can we test (i) the significance of complete regression, (ii) the significance of a single coefficient? Derive the bias and MSE matrix of restricted regression estimator when linear restrictions are not correct.

7. Explain how ANOVA is used in the context of regression analysis.

8. For a given regression model, decompose the total sum of squares (SST) into explained sum of squares (SSR) and residual sum of squares (SSE). Verify that SST=SSR+SSE.

9. Write the linear model in deviation form and give the ANOVA. Define R^2 and adjusted R^2 and discuss their merits/demerits.

10. Describe the difference between point prediction and interval prediction. When would you use each?

11. Obtain an unbiased forecast for value of dependent variable of out of sample forecast period. Derive the variance of the forecast.

3.10. Summary

This unit delves into advanced methods in multiple linear regression analysis, focusing on enhancing model estimation, testing hypotheses, and improving predictive capabilities. Topics include restricted regression estimation, enabling the incorporation of constraints to align models with theoretical or practical considerations, and tests for linear

restrictions to evaluate specific coefficient relationships. The unit introduces the deviation form of regression models, simplifying computations and improving interpretability. Measures of model fit, including R-square and adjusted R-square, are discussed alongside the application of Analysis of Variance (ANOVA) in regression to dissect variability. Techniques for interval estimation of regression coefficients and both point and interval predictions are also covered, equipping learners to construct robust and insightful regression models. This unit provides the analytical tools necessary for advanced regression modeling, emphasizing precision, reliability, and applicability in diverse contexts.

3.11. References

- Gujarati, D. and Guneshker, S. (2007): Basic Econometrics, 4th Ed., McGraw Hill Companies.

- Johnston, J. and Dinardo, J. (1997): Econometric Methods, 2nd Ed., McGraw Hill International.

3.12. Further Readings

- Judge, G.G., Griffiths, W.E., Hill Lütkepohl, R.C. and Lee, T. C. (1985): The theory and practice of econometrics, Wiley.

- Koutsoyiannis, A. (2004): Theory of Econometrics, 2 Ed., Palgrave Macmillan Limited.

- Maddala, G.S. and Lahiri, K. (2009): Introduction to Econometrics, 4 Ed., John Wiley & Sons.

Unit 4 Model with qualitative independent variables

4.1. Introduction

In econometrics and statistical modeling, data often include categorical variables that need special treatment when used in regression models. These variables, referred to as **dummy variables**, are essential tools to represent qualitative characteristics, enabling the inclusion of such data in quantitative models. Additionally, models may involve discrete or limited dependent variables, which require specific techniques to ensure accurate estimation and interpretation.

A dummy variable is a numerical variable used in regression analysis to represent subgroups or categories within the data. Typically coded as 0 or 1, dummy variables capture the presence or absence of a particular attribute.

For example, a variable for gender might be coded as:

- 1 for "Male"

- 0 for "Female"

Dummy variables allow models to include qualitative factors, enabling analysis of their effects on the dependent variable.

Dummy variables serve various purposes in regression modeling:

(i) Capturing group effects: They differentiate between categories (e.g., urban vs. rural).

(ii) Interaction effects: Allow investigation of whether the impact of one variable differs across groups.

(iii) Structural changes: Model changes over time or across different periods.

A dependent variable may not always be continuous. Instead, it might be:

(i) Discrete: Taking on a limited number of distinct values (e.g., binary outcomes like "yes" or "no").

(ii) Limited: Subject to boundary constraints (e.g., censored or truncated data).

Examples:

(i) Binary outcomes: Whether a student passes or fails (1 = pass, 0 = fail).

(ii) Ordered outcomes: Satisfaction ratings (e.g., 1 = dissatisfied, 2 = neutral, 3 = satisfied).

(iii) Count data: Number of accidents on a road in a year.

To appropriately handle Dummy and Limited Variables, specialized models are employed. The Logit and Probit Models address the limitations of LPM by modeling probabilities in a nonlinear fashion. The Tobit Models are used for censored dependent variables, combining regression with a limit-dependent component. The Multinomial and Ordered Models extend the analysis to categorical outcomes with more than two categories.

In brief, the integration of dummy variables and the modeling of discrete or limited dependent variables expands the applicability of regression analysis to diverse datasets. By transforming qualitative attributes into quantitative forms and tailoring methods to specific data characteristics, these tools ensure robust, interpretable, and actionable results**.**

4.2 Objectives

The chapter aims to equip readers with a comprehensive understanding of the conceptual foundations, practical applications, and analytical techniques involved in modeling with dummy variables and discrete or limited dependent variables. The specific objectives are:

- To explain the purpose and importance of dummy variables in regression models.
- To learn how dummy variables, represent categorical or qualitative data in quantitative analysis.
- To understand the creation, interpretation, and use of dummy variables in different scenarios, such as group comparisons, seasonal effects, and policy evaluation.

- To introduce the characteristics and challenges of discrete and limited dependent variables, such as binary, ordered, or count data.
- To explain the limitations of standard regression models when applied to such dependent variables.
- To familiarize readers with specific models designed for discrete and limited dependent variables, including logit, probit, and Tobit models.
- To demonstrate the formulation and estimation of models incorporating dummy variables and non-continuous dependent variables.
- To illustrate how to interpret coefficients and marginal effects in these models.
- To emphasize the importance of diagnostic checks and goodness-of-fit measures in evaluating model performance.
- To provide examples of how dummy variables can be used to analyze qualitative factors like gender, location, or policy interventions.
- To showcase the use of discrete and limited dependent variable models in various domains, including economics, health, marketing, and social sciences.
- To highlight the role of such models in decision-making and prediction.
- To identify common pitfalls in the use of dummy variables, such as multicollinearity and omitted variable bias.
- To discuss challenges in modeling discrete outcomes, such as sample size requirements and model misspecification.
- To offer strategies for addressing these challenges and improving model robustness.

4.3. Models with Dummy Explanatory Variables

Dummy Variables are defined for some unusual variables like seasonal variables, qualitative variables etc. These variables usually take values 1 or 0 to indicate the absence or presence of some categorical effect/qualitative characteristic.

Seasonal Dummies:

Example 4.3.1: Suppose the dependent variable y is rainfall. We have quarterly data on rainfall. Apart from other variables, rainfall depends upon different quarters also. How to accommodate different quarters in the model?

For $i = 1,2,3,4,$ we define a quarterly dummy variable as follows:

$$Q_{it} = \begin{cases} 1, & if\ t^{th}\ observation\ is\ from\ i^{th}\ quarter \\ 0, & otherwise \end{cases}$$

For four quarters of each year these dummies are

$Q_1\ Q_2\ Q_3\ Q_4$ 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1

Let x_t be the vector of other explanatory variables at time t. Then, the model may be written as

$$y_t = \alpha_1 Q_{1t} + \alpha_2 Q_{2t} + \alpha_3 Q_{3t} + \alpha_4 Q_{4t} + x_t'\beta + u_t \qquad (4.1)$$

x_t should not contain a column of ones, otherwise $Q_{1t} + Q_{2t} + Q_{3t} + Q_{4t} = 1$ would imply that the data matrix is not of full column rank. The model has four intercept terms corresponding to four seasons, i.e., $\alpha_1, \alpha_2, \alpha_3, \alpha_4$.

An alternative formulation is

$$y_t = \alpha_1 + \gamma_2 Q_{2t} + \gamma_3 Q_{3t} + \gamma_4 Q_{4t} + x_t'\beta + u_t \qquad (4.2)$$

Here $\gamma_2 = \alpha_2 - \alpha_1, \gamma_3 = \alpha_3 - \alpha_1, \gamma_4 = \alpha_4 - \alpha_1$.

One may be interested in the hypothesis of the form $H: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ or $H: \gamma_2 = \gamma_3 = \gamma_4 = 0$.

Qualitative Variables: Qualitative variables like education level, cast, sex etc. may be represented by dummy variables.

Example 4.3.2.: $Income = f(Cast, Sex, education\ level, age)$

Consider three cast groups

(i) General (GC) denoted by dummy variable C_1,

(ii) OBC denoted by C_2,

(iii) SC/ST denoted by C_3.

Dummy variables C_1, C_2, $C_3\ are$ defined as

$$C_1 = \begin{cases} 1, if\ person\ is\ from\ GC \\ 0,\ otherwise \end{cases}$$

$$C_2 = \begin{cases} 1, if\ person\ is\ from\ OBC \\ 0,\ otherwise \end{cases}$$

$$C_3 = \begin{cases} 1, if\ person\ is\ from\ SC/ST \\ 0,\ otherwise \end{cases}$$

Similarly, corresponding to two categories of sex (male, female), we define two dummy variables, say S_1, S_2.

literacy level has four categories (i) illiterate, (ii) below graduation, (iii) graduate, (iv) above graduation. We define four dummy variables, say E_1, E_2, E_3, E_4 for each category. In case of one dummy variable either we do not include the constant term or if we include the constant term, we drop one of the dummy variables. For two dummy variables, sum of first set minus the sum of second set is $0, [(S_1 + S_2) - (C_1 + C_2 + C_3) = 0]$.

Thus

(i) Either drop the constant term and one of the dummy variables from one set.

(ii) Alternatively, retain the constant term and drop one of the dummy variables from each set.

This rule also appears for three or more dummy variables.

Example 4.3.3.: Suppose the two dummy variables are Sex (S) and Category (C). Then we formulate the model as

$$y_t = \alpha_1 S_1 + \alpha_2 S_2 + \gamma_1 C_1 + \gamma_2 C_2 + x_t'\beta + u_t$$

or alternatively

$$y_t = \alpha_1 + S_1 + \gamma_1 C_1 + \gamma_2 C_2 + x_t'\beta + u_t$$

4.4. Model with Limited Dependent Variable

Discrete and limited dependent variables are types of outcomes in statistical modelling that differ from continuous variables in how they are measured and modelled.

Discrete Dependent Variable:

A discrete dependent variable is one that can only take on a finite number of values or a countable number of values.

Examples: include binary outcomes (yes/no, 0/1), counts of events (number of accidents in a month), or categorical outcomes with a limited number of categories (like low/medium/high).

These variables are usually modelled using logistic regression (for binary outcomes) or Poisson regression (for count data).

Limited Dependent Variable:

Limited dependent variables are those that are restricted in some way, often in terms of their range or distribution. This could mean variables that are bounded (like proportions that range from 0 to 1), censored (where observations are not fully observed beyond a certain point), or truncated (where data points beyond a certain threshold are missing).

Examples include wages (which cannot be negative), percentages (which are bounded between 0 and 100), or survival times (where the observation is censored if the event of interest has not occurred by the end of the study).

Techniques used for modelling limited dependent variables include Tobit models (for censored or truncated data), logit and probit models (for bounded outcomes).

4.4.1. Differences between Discrete dependent variables and limited dependent variables:

(i) Nature of Values: Discrete dependent variables take on distinct, separate values, while limited dependent variables are restricted in terms of their range or distribution.

(ii) Modelling Approaches: Both types often require specialized modelling techniques beyond ordinary least squares regression used for continuous variables. Logistic regression, Poisson regression, Tobit models, and others are tailored to handle the characteristics of these variables

(iii) Application: Discrete dependent variables are common in binary and categorical outcomes, whereas limited dependent variables are encountered when dealing with bounded, censored, or truncated data.

4.5 Uses of dummy variables

First, we define the Dichotomous, and Polychotomous Variables.

Dichotomous, binary, or dummy variables usually take on a value 1 or 0 depending upon which of two possible results occur.

Example: Suppose y_i^* is the tolerance of an insect to a particular insecticide. Since y_i^* is unobservable, it may be replaced by a dummy variable defined as

$$y_i = \begin{cases} 1, & if \ \ i^{th} insect \ dies \\ 0, & otherwise \end{cases}$$

Another example is

$$y_i = \begin{cases} 1, & if \ \ i^{th} person \ is \ employed \ in \ a \ week \\ 0, & otherwise \end{cases}$$

Polychotomous variables, also known as categorical variables or multinomial variables, are variables that can take on more than two distinct categories or levels. Unlike

binary (dichotomous) variables that have only two categories (e.g., yes/no, true/false), polychotomous variables have multiple categories.

Example:

$$y_i = \begin{cases} 1, & if\ \ i^{th} person\ is\ in\ poor\ health \\ 2, & if\ i^{th}\ person\ is\ in\ fair\ health \\ 3, & if\ i^{th}\ person\ is\ in\ excellent\ health \end{cases}$$

Some other examples of polychotomous variables include:

Colour of a Car: Red, Blue, Green, etc.

Education Level: High School, Bachelor's Degree, Master's Degree, PhD, etc.

Type of Employment: Full-time, Part-time, Self-employed, Unemployed, etc.

First, we consider models for Dichotomous variables. Approaches for such models are

1. Linear probability Model (L P M)

2. Logit Model

3. Probit Model

Models with single explanatory variable:

First, we consider different models for the case when we have just one explanatory variable.

Linear Probability Model

Consider the model

$$y_i = \beta_1 + \beta_2 x_i + u_i;\ i = 1,2,\ldots,n \qquad (4.3)$$

Here

$$y_i = \begin{cases} 1, & if\ \ i^{th} insect\ dies \\ 0, & otherwise \end{cases}$$

x_i: Amount of insecticide used

Model (4.3) expresses a dichotomous dependent variable y as a linear function of the explanatory variable x_i. Such models are called LPM.

Let $E(u_i) = 0, \forall i, then$

$$E(x_i) = \beta_1 + \beta_2 x_i \qquad (4.4)$$

Let $p_i = P(y_i = 1), 1 - p_i = P(y_i = 0)$.

Then

$$E(y_i) = p_i = \beta_1 + \beta_2 x_i$$

Since $0 \leq p_i \leq 1,\ it\ implied\ that\ 0 \leq E(y_i) \leq 1$.

Estimation of LPM: Difficulties in applying OLS

1. Normality: We observe that $u_i = y_i - \beta_1 - \beta_2 x_i$

Therefore

$$u_i = \begin{cases} 1 - \beta_1 - \beta_2 x_i & if\ y_i = 1\ with\ probability\ p_i \\ -\beta_1 - \beta_2 x_i & if\ y_i = 0\ with\ probability\ 1 - p_i \end{cases}$$

Thus, u_i does not follow a normal distribution. However, in applying OLS normality of u_i is not required.

2. Heteroscedasticity: Let

$$E(u_i u_j) = 0, \forall i \neq j. Var(u_i) = E(u_i^2)$$

$$= (1 - \beta_1 - \beta_2 x_i)^2 p_i + (-\beta_1 - \beta_2 x_i)^2 (1 - p_i)$$

$$= (1 - p_i)^2 p_i + p_i^2 (1 - p_i)$$

$$= E(x_i)[1 - E(x_i)]$$

$$= w_i \tag{4.5}$$

$Thus, u_i's\ are$ heteroscedastic. Dividing both sides of (4.3) by $\sqrt{w_i}$ gives

$$\frac{y_i}{\sqrt{w_i}} = \frac{1}{\sqrt{w_i}}\beta_1 + \beta_2\frac{x_i}{\sqrt{w_i}} + \frac{u_i}{\sqrt{w_i}} \tag{4.6}$$

Since $E\left(\frac{u_i}{\sqrt{w_i}}\right)^2 = 1$, we can apply OLS to model (4.6). Since $w_i's$ are unknown we adopt the following two step procedure:

Step 1: Using OLS to model (4.3), obtain $\hat{y}_i$.

Then estimate w_i by $\hat{w}_i = \hat{y}_i(1 - \hat{y}_i)$.

Step 2: Transform the data by dividing by $\sqrt{\hat{w}}_i$ and run OLS to the transformed data.

Step 3: However, there is no guarantee that $\hat{y}_i$, the estimate of $E(x_i)$, lies between 0 and 1. We may take $\hat{y}_i = 0$ if OLS estimate is less than 0 and $\hat{y}_i = 1$ if OLS estimate is greater than 1.

Logit Model

Suppose x_i is the income of i^{th} person

$$y_i = \begin{cases} 1, & if\ \ i^{th}\ person\ owns\ a\ house \\ 0, & otherwise \end{cases}$$

Logit model for the above house ownership example is

$$p_i = E[x_i] = \frac{1}{1 + exp\ exp\ [-(\beta_1 + \beta_2 x_i)]} \tag{4.7}$$

or

$$p_i = \frac{1}{1+e^{-z_i}}$$

$$= \frac{e^{z_i}}{1+e^{z_i}} \qquad (4.8)$$

where $z_i = \beta_1 + \beta_2 x_i$. As $-\infty < z_i < \infty$, p_i ranges from 0 to 1. Since p_i has a non-linear relationship with x_i and β_1, β_2, we cannot directly use OLS.

Now

$$1 - p_i = 1 - \frac{1}{1+e^{-z_i}} = \frac{e^{-z_i}}{1+e^{-z_i}}$$

or

$$\frac{p_i}{1-p_i} = e^{z_i}$$

or

$$L_i = log\ log\left(\frac{p_i}{1-p_i}\right)$$

$$= \beta_1 + \beta_2 X_i \ = z_i$$

$$\frac{p_i}{1-p_i} : Odds\ ratio\ in\ favour\ of\ event\ that\ insect\ dies L_i = log\ log\left(\frac{p_i}{1-p_i}\right)$$

L_i is known as the LOGIT.

Probit Model

We consider the family house ownership example. Let the decision of the i^{th} family to own a house depends upon a utility I_i and

$$I_i = \beta_1 + \beta_2 x_i$$

Let I_i^* be the critical or threshold level of the utility index for the i^{th} family and if $I_i > I_i^*$, the family own a house. I_i^* is also unobservable.

Assume that I_i^* follows a normal distribution with the same mean and variance $\forall\, i$. Then $p_i = p(y_i = 1)$

$$= p[I_i^* \leq I_i]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-\frac{t^2}{2}}\, dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 X_i} e^{-\frac{t^2}{2}}\, dt$$

$$= F(\beta_1 + \beta_2 X_i)$$

So that

$$I_i = F^{-1}(p_i)$$

$$= \beta_1 + \beta_2 x_i$$

F^{-1} is the inverse of normal cdf. I_i is known as the normal equivalent deviate (ned). Since $I_i < 0$ whenever $p_i < 0.5$, we add 5 to the ned and get the PROBIT.

$$Probit = I_i + 5$$

We write $S = \left(b - (X'X)^{-1}X'y\right)' X'X \left(b - (X'X)^{-1}X'y\right) + v$ as

$$I_i = \beta_1 + \beta_2 x_i + u_i$$

For estimating β_1, β_2, we proceed as follows:

Step 1: Estimate p_i by $\hat{p}_i = \frac{n_i}{N_i}$.

Step 2: Obtain $F^{-1}(\hat{p}_i) = \hat{I}_i$ by using the normal table.

Step 3: Use $\hat{I}_i$ or if required convert them *need* into Probit by adding 5 and use Probit as the dependent variable in $\left(b - \left(X'X\right)^{-1}X'y\right)' X'X\left(b - \left(X'X\right)^{-1}X'y\right)$.

Step 4: Apply GLS for estimating β_1 , β_2 .

Notice that the slope coefficient β_2 and R^2 will remain the same whether we use *ned* or PROBIT as the dependent variable and only the intercept term β_1 will change.

General case of k Explanatory Variables:

Now we consider the general case, when there are k explanatory variables.

Formulating a probability model

Let x be a $k \times 1$ vector of explanatory variables, and

$$y = \begin{cases} 1, & if\ unit\ own\ a\ character\ C \\ 0, & if\ unit\ does\ not\ own\ C \end{cases}$$

$$P(x) = F(x'\beta), P(y = 0) = 1 - F(x'\beta)$$

$F(.)$ is the link function lying between 0 and 1. Then

$$E(y|x) = F(x'\beta)$$

Linear Probability Model:

One possibility is to retain linear regression

$$F(x'\beta) = x'\beta$$

Thus

$$y = x'\beta + \left(y - E(y|x)\right) = x'\beta + u$$

Then

$$Var(u|x) = (1 - x'\beta)^2 P(x) + (x'\beta)^2 \big(1 - P(x)\big) l = (x'\beta)(1 - x'\beta)$$

Then, the disturbances $u's$ are heteroscedastic and we can apply GLS. However, there is no guarantee that the predicted value $P(y = 1|x) = x'\hat{\beta}$ lies between 0 and 1 or $Var(u|x)$ is non-negative. Thus, linear probability models are not frequently used.

Some requirements for a probability model are as follows:

(i) $0 \leq F(x'\beta) \leq 1$

(ii) $F(x'\beta) = 1$

(iii) $F(x'\beta) = 0$

Any continuous probability distribution satisfies these conditions and can be used as a link function.

Most frequently used models are

LOGIT Model:

$$F(x'\beta) = \frac{exp\ exp\ (x'\beta)}{1 + exp\ exp\ (x'\beta)} = \Lambda(x'\beta)$$

PROBIT Model:

$$F(x'\beta) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right) dz$$

Some other models are

Weibull model:

$$F(x'\beta) = exp\ [-\ exp\ (-x'\beta)\ ]$$

Log-log Model:

$$F(x'\beta) = 1 - exp\,[-\,exp\,(x'\beta)\,]$$

The LOGIT model is based on Logistic distribution whereas the PROBIT model is based on normal distribution.

Two distributions are similar in nature except in tail region. The Logistic distribution has much heavier tails. For intermediate values of $x'\beta$ (between -1.2 to 1.2) two distributions give almost similar probabilities. When $x'\beta$ is extremely small, logistic distribution gives larger probabilities to $y = 0$. It gives smaller probabilities to $y = 0$ when $x'\beta$ is extremely large. If the observations have very few responses $(y = 1)$ or very few non-responses $(y = 0)$, two models give different predictions.

Regression for probability model:

Let

$$E(x) = F(x'\beta)$$

The marginal effect is

$$\frac{\partial}{\partial x}E(x) = f(x'\beta)\beta$$

The density function corresponding to $F(t)$ is

$$f(t) = \frac{\partial}{\partial t}F(t)$$

For LOGIT Model, the marginal effect is given by

$$\frac{\partial}{\partial x}E(x) = \Lambda(x'\beta)[1 - \Lambda(x'\beta)]\beta$$

For PROBIT Model, the marginal effect is

$$\frac{\partial}{\partial x}E(x) = \phi(x'\beta)\beta$$

Here $\phi(.)$ denotes the pdf of a standard normal distribution. Further, for both the models, the m$arginal\ Effect$s depends on x.

4.5.1 LOGIT model

Let $y_1, \cdots, y_n$ be observations on binary response variable, and $x_i\ (i = 1, \ldots, n)\ are\ k \times 1$ vectors of observations on input variables. We take

$$F(x_i'\beta) = P(y_i = 1)$$

$$= p_i = \frac{exp\ exp\ (x_i'\beta)}{1 + exp\ exp\ (x_i'\beta)}$$

$$= \Lambda(x_i'\beta) \equiv \Lambda_i\ (say)$$

It ensures that $0 \leq P(y_i = 1) \leq 1$. Then

$$L_i = ln\ ln\ \left(\frac{p_i}{1 - p_i}\right) = x_i'\beta$$

<u>Estimation for grouped data:</u>

If out of N_i observations corresponding to x_i, n_i times $y_i\ is\ 1.$ Then estimate p_i by

$$\widehat{p}_i = \frac{n_i}{N_i}$$

so that

$$\widehat{L}_i = ln\ ln\ \left(\frac{\widehat{p}_i}{1 - \widehat{p}_i}\right) = x_i'\beta + u_i$$

For large N_i,

$$Var(u_i) \approx \frac{1}{N_i p_i (1 - p_i)}$$

Take

$$\widehat{w}_i = N_i\widehat{p}_i(1-\widehat{p}_i)L_i^* = \sqrt{\widehat{w}_i}\widehat{L}_i = \sqrt{\widehat{w}_i}x_i'\beta + \sqrt{\widehat{w}_i}u_i = x_i^{*'}\beta + u_i^*$$

or

$$L^* = X^*\beta + u^* \tag{4.9}$$

Applying OLS to (4.9), we can estimate β.

Estimation in Binary Choice Models

Let $y = (y_1, \dots, y_n)'$ be the observation vector on binary response variable and $X = (x_1 \dots x_k)$ is $n \times k$ matrix of observations on input variables. We can write the likelihood function as

$$L(y, X) \equiv L$$

$$= \prod_{i=1}^{n} [F(x_i'\beta)]^{y_i}[1 - F(x_i'\beta)]^{1-y_i}$$

Then, log likelihood function is

$$lnL = \sum_{i=1}^{n} \{y_i lnF(x_i'\beta) + (1 - y_i)\, ln\, ln\, [1 - F(x_i'\beta)]\}$$

Write

$$F_i = F(x_i'\beta), f_i = \frac{dF_i}{d(x_i'\beta)}$$

Then

$$\frac{\partial}{\partial\beta} lnL = \sum_{i=1}^{n} \left\{\frac{y_i f_i}{F_i} - (1 - y_i)\frac{f_i}{1 - F_i}\right\} x_i = 0$$

or

$$\sum_{i=1}^{n}\left\{\frac{y_i f_i}{F_i(1-F_i)}-\frac{f_i}{1-F_i}\right\}x_i=0$$

For LOGIT model $F_i=\Lambda_i, f_i=\Lambda_i(1-\Lambda_i)$,

$$\frac{f_i}{F_i}=(1-\Lambda_i), \frac{f_i}{F_i(1-F_i)}=1, \qquad \Lambda_i\equiv\Lambda(x_i'\beta)$$

Hence

$$\frac{\partial}{\partial\beta}lnL=\sum_{i=1}^{n}(y_i-\Lambda_i)x_i=0$$

If first element of x_i is 1, it implies that

$$\sum_{i=1}^{n}(y_i-\Lambda_i)=0\Rightarrow\sum_{i=1}^{n}y_i=\sum_{i=1}^{n}\Lambda_i$$

Thus, average of predicted probabilities is equal to proportion of ones in the sample. If we view $(y_i-\Lambda_i)$ as residuals, then sum of residuals is zero.

The matrix of second derivatives

$$H=\frac{\partial^2}{\partial\beta\partial\beta'}lnL$$

$$=\frac{\partial}{\partial\beta'}\sum_{i=1}^{n}(y_i-\Lambda_i)x_i$$

$$=-\sum_{i=1}^{n}\Lambda_i(1-\Lambda_i)x_i x_i'$$

The Hessian matrix H does not involve y_i and is negative definite. Thus, log likelihood is globally concave and Newton method converges to maximum and provides the maximum likelihood estimator of β.

Measures of Goodness of fit:

Some of the measures of goodness of fit are

Pearson Goodness of fit Statistic

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - n_i\hat{p}_i)^2}{n_i\hat{p}_i(1-\hat{p}_i)}$$

Deviance Statistic

$$G^2 = 2\sum_{i=1}^{n}\left[y_i \, log \, log\left(\frac{y_i}{n_i\hat{p}_i}\right) + (n_i - y_i) \, log \, log\left(\frac{n_i - y_i}{n_i(1-\hat{p}_i)}\right)\right]$$

Under the null hypothesis that the LOGIT model fits the data, both of the statistic follows $\chi^2(n-k-1)$.

4.5.2. PROBIT Model

Let y_i^* be a latent variable such that

$$y_i^* = x_i'\beta + u_i$$

We observe dummy variable y_i in place of y_i^* such that

$$y_i = \begin{cases} 1, & if \; y_i^* > 0 \\ 0, & otherwise \end{cases}$$

Assume that $u_i \sim N(0, \sigma_u^2 I_n)$. Further $z_i = \sigma_u^{-1} ui \sim N(0,1)$. Then

$$P(y_i = 1) = P(y_i^* > 0)$$

$$= P(u_i > -x_i'\beta)$$

$$= P(z_i > -\sigma_u^{-1} x_i'\beta),$$

$$= 1 - \Phi(-\sigma_u^{-1} x_i'\beta)$$

$$= (-\sigma_u^{-1} x_i' \beta) P(y_i = 1)$$

$$= 1 - (-\sigma_u^{-1} x_i' \beta)$$

Let $y_1, \cdots, y_n$ be n observations. Then the likelihood function is

$$L(\beta, \sigma_u) = \prod_{i=1}^{n} \left[\Phi\left(\frac{1}{\sigma_u} x_i' \beta\right)\right]^{y_i} \left[1 - \Phi\left(\frac{1}{\sigma_u} x_i' \beta\right)\right]^{1-y_i}$$

$L(\beta, \sigma_u)$ is a function of $\left(\frac{1}{\sigma_u}\beta\right)$. Thus, we cannot estimate β, σ_u separately. However, we can estimate $\left(\frac{1}{\sigma_u}\beta\right)$.

Without loss of generality, we take $\sigma_u = 1$. The log likelihood function is

$$lnL = \sum_{i:y_i=0}^{n} ln\ [1 - \Phi(x_i'\beta)] + \sum_{i:y_i=1}^{n} ln\ \Phi(x_i'\beta)$$

Then likelihood equation is

$$\frac{\partial}{\partial \beta} lnL = \sum_{i:y_i=0}^{n} \frac{-\phi_i}{1 - \Phi_i} x_i + \sum_{i:y_i=1}^{n} \frac{\phi_i}{\Phi_i} x_i$$

Let us write $q_i = 2y_i - 1$. We utilize the result that

$$1 - \Phi(x_i'\beta) = \Phi(-x_i'\beta)$$

Then

$$lnL = \sum_{i=1}^{n} q_i \frac{\phi(q_i x_i'\beta)}{\Phi(q_i x_i'\beta)} x_i = \sum_{i=1}^{n} \lambda_i x_i = 0$$

where

$$\lambda_i = q_i \frac{\phi(q_i x_i' \beta)}{\Phi(q_i x_i' \beta)}$$

Further

$$H = \frac{\partial^2}{\partial \beta \partial \beta'}$$

$$lnL = \sum_{i=1}^{n} -\lambda_i (\lambda_i + x_i' \beta) x_i x_i' \qquad (4.10)$$

We observe that when $y_i = 1$

$$E[u \leq x_i' \beta] = \int_{-\infty}^{x_i' \beta} z^2 \frac{\phi(z)}{\Phi(x_i' \beta)} dz$$

$$= -\frac{(x_i' \beta)\phi(x_i' \beta)}{\Phi(x_i' \beta)} + 1$$

$$= -\lambda_i x_i' \beta + 1E[u \leq x_i' \beta]$$

$$= -\frac{\phi(x_i' \beta)}{\Phi(x_i' \beta)}$$

$$= -\lambda_i$$

Hence

$$Var(u \leq \beta' x_i) - 1 = -\lambda_i(\lambda_i + x_i' \beta) \leq 0$$

as

$$Var(u \leq \beta' x) \leq Var(u) = 1.$$

Similarly, for $y_i = 0$,

$$Var(u \geq \beta' x_i) - 1 = -\lambda_i(\lambda_i + x_i' \beta) \leq 0$$

Hence, H given in (4.10), is negative definite. Here also Newton's method leads to maximum of the log likelihood.

4.5.3. Censored or Truncated Data

Suppose we do not observe values above or below a certain magnitude, due to a censoring or truncation mechanism.

Example:

- SEBI intervenes to stop trading in stock market if it falls or goes above certain levels.
- Company paying no Dividend until its earnings reach some threshold value.
- A pesticide affects the insect once its dose reaches a threshold level.
- Estimate the relationship between hours worked by a labor and characteristics such as age, education, sex, family status. For unemployed, data on the number of hours they would have worked, in case employed, are not available but their age, education, sex and family status are available.
- Suppose 100 researchers have applied for a grant out of which only 30 have received the grant. To model amount of grant received using number of publications, amount received in past grants, quality of proposal measured on a scale, amount received cannot be negative and for those not received grant, observations on other variables are available.

4.5.4. TOBIT Model

TOBIT model was developed by James Tobin in 1958 to overcome the problem of zero-inflated data. By zero-inflated, we mean, many zero cells in data matrix (here $y_i's$).

Example: Let y^* denotes a person's desire to own a car, which is unobservable. In LOGIT and PROBIT models, we define a dummy variable

$$y_i = \begin{cases} 1, & if\ i^{th}\ person\ purchases\ a\ car \\ 0, & if\ i^{th}\ person\ does\ not\ purchase\ a\ car \end{cases}$$

In TOBIT model

$$y_i = \begin{cases} price\ of\ car\ y_i^*, & if\ \ i^{th}\ person\ purchases\ a\ car \\ 0, & if\ i^{th}\ person\ \ does\ not\ purchase\ a\ car \end{cases}$$

Hence, y_i is a censored variable.

Let x_i be the income of i^{th} person in the sample. Then the model is

$$y_i^* = \beta_1 + \beta_2 x_i + u_i; i = 1,2, \dots, n$$

$$y_i = \{y_i^* = \beta_1 + \beta_2 x_i + u_i, if\ y_i^* > 0\ 0, \qquad if\ y_i^* \leq 0$$

The model is called a censored regression model. Let y^* be a dependent variable with $y_i^* = x_i'\beta + u_i\ \ (i = 1,2, \dots, n);\ u_i \sim N(0, \sigma_u^2) \forall i.$

Then y_i^* is observed using censoring

$$y_i = \{y_i^*, if\ y_i^* > 0\ 0, if\ y_i^* \leq 0$$

We can write the model in terms of y_i as

$$y_i = (\ 0, x_i'\beta + u_i)$$

Obviously

$$P(y_i > 0) = P(u_i > -x_i'\beta) = P(u_i < x_i'\beta)$$

<u>Maximum Likelihood Estimation:</u>

For an observation $y_i^* > 0$, the contribution to the likelihood function is

$$P(y_i^* > 0)\phi(y_i^* | y_i^* > 0) = \Phi\left(\frac{x_i'\beta}{\sigma_u}\right)\frac{\frac{1}{\sqrt{2\pi}\sigma_u} exp\ exp\left(-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma_u^2}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma_u}\right)}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_u} exp\ exp\left(-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma_u^2}\right)$$

$$P(y_i^* \leq 0) = P\left(\frac{u_i}{\sigma_u} \leq -\frac{x_i'\beta}{\sigma_u}\right) = 1 - \Phi\left(\frac{x_i'\beta}{\sigma_u}\right)$$

$$\frac{\frac{1}{\sqrt{2\pi}\sigma_u} exp\ exp\left(-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma_u^2}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma_u}\right)}\ is\ the\ pdf\ of\ a\ truncated\ normal\ distribution.$$

The likelihood function is

$$L = \prod_{i:y_i=0}\left[1 - \Phi\left(\frac{x_i'\beta}{\sigma_u}\right)\right]\prod_{i:y_i>0}\frac{1}{\sqrt{2\pi}\sigma_u} exp\left(-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma_u^2}\right)$$

Log likelihood is

$$ln\ L = \sum_{i:y_i=0} ln\left[1 - \Phi\left(\frac{x_i'\beta}{\sigma_u}\right)\right] + \sum_{i:y_i>0}\left\{\left(-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma_u^2}\right) + ln\ ln\ \frac{1}{\sqrt{2\pi}\sigma_u}\right\}$$

The MLE can be obtained by maximizing lnL w.r.t. β, σ_u.

Some other estimation procedures are:

1) Symmetrically trimmed least squares method
2) Censored least absolute deviation (CLAD) method.

(1) Symmetrically trimmed least squares method

$$y^* = X\beta + uy_i = \{y_i^*, if\ y_i^* > 0\ or\ if\ u_i > -x_i'\beta\ 0, if\ y_i^* \leq 0\ or\ if\ u_i \leq -x_i'\beta$$

The OLS estimator is inconsistent because of asymmetry in the distribution of the error term around 0, as the observations corresponding to $u_i \leq -x_i'\beta$ are omitted. Now observations corresponding to $u_i \geq x_i'\beta$ (or $y_i \geq 2x_i'\beta$) are also truncated (trimmed) and we take

$$y_i = \{y_i^*, 2x_i'\beta_0\}, \beta_0 \text{ is true value of } \beta$$

The distribution of u_i becomes symmetric and OLS estimator becomes consistent.

Equivalently we define

$$u_i^* = \{u_i, -x_i'\beta_0\}$$

and replace u_i with $u_i^* = \{u_i^*, x_i'\beta_0\}$ if $x_i'\beta_0 > 0$, Delete the observation otherwise.

The true value of the coefficient β_0 would satisfy the normal equation

$$\sum_{i=1}^{n} I(x_i'\beta_0 > 0)[\{y_i, 2x_i'\beta_0\}]x_i = 0.$$

The normal equation is obtained by minimizing

$$M(\beta) = \sum_{i=1}^{n}\left[y_i - \left(\frac{1}{2}y_i, x_i'\beta\right)\right]^2 + \sum_{i=1}^{n} I(y_i > 2x_i'\beta)\left\{\left(\frac{1}{2}y_i\right)^2 - [(0, x_i'\beta)]^2\right\}$$

Algorithm to obtain a consistent estimator of β

(i) Compute the initial estimate b, say OLS, on the original data.

(ii) Compute predicted value $\hat{y}_i = x_i'b$. If $\hat{y}_i < 0$, set observation to missing. If $y_i > 2\hat{y}_i$, set $y_i = 2x_i'b$.

(iii) Run OLS to these altered data.

(iv) Use this β in the original data and repeat until β stabilizes.

(2) Censored Least Absolute Deviation Method

In this method the estimator of β is obtained by minimizing the sum of absolute deviations

$$\Sigma_{i=1}^{n}|y_i^* - x_i'\beta|.$$

The procedure is the same as that of symmetrically trimmed estimator.

Multinomial Choice Models

Let Y be the result of a single decision among more than 2 alternatives. For Unordered choice set such as categories or qualitative choices, we use multinomial LOGIT, or conditional LOGIT models. For ordered choice set (rankings), the models for ordered data include ordered PROBIT model.

Example (Unordered): In Occupational field, let y be 0 for labor, 1 Professional, 2 White collar, 3 Blue collar (workers in a division of manufacturers) and x include the education, parent's income, etc.

Example (Ordered): Opinions of a survey are coded (y) as, say, 1 for "strongly disagree", 2 "disagree", 3 "neutral", 4 "agree", 5 "strongly agree" (rankings). Further, x include the monthly income, education level, caste group, etc.

Random Utility Models

Example: A car consumer decided to choose one of two cars A and B which are nearly identical except the car B has enhanced safety features and costs Rs 20,000 more.

Marginal utility derived from car B is Rs 20,000.

If 10,000 consumers preferred car B to car A, consumers overall received $10{,}000 \times Rs\ 20{,}000 = Rs\ 20$ crore worth of incremental utility from the safety features of car B.

Utility is derived from the consumer's belief that they are likely to have fewer accidents due to the added safety features.

U_{ij}: Utility for i^{th} customer if he makes choice j among m possible utilities $(j = 1,2, \dots, m)$

Customer makes the choice j if U_{ij} is maximum. Statistical model for utility is driven by $P(U_{ij} > U_{ij'}) \forall j' \neq j$.

Linear utility model:

$U_{ij} = \eta_{ij} + u_{ij}$ η_{ij} links the agent utility to factors that can be observed.

4.5.5. Multinomial Logit Model

Explanatory variables contain only individual characteristics

$$\eta_{ij} = x_i' \beta_j$$

Here x is individual characteristics constant across all the alternatives j.

Estimated equations assign a set of probabilities to m classes with observed characteristics x_i.

Model to define the probabilities for different classes is

$$P(y_i = j|x_i) = \frac{e^{x_i' \beta_j}}{\sum_{k=1}^{m} e^{x_i' \beta_k}}, (j = 1, \dots, m, i = 1, \dots, n) \quad (4.11)$$

For a vector c, if $\beta_k^* = \beta_k + c$, then the probabilities computed in (4.11) remain same as all the terms involving c will drop out. Sum of all the probabilities is one. Only

$(m - 1)$ parameters are needed to define m probabilities. Thus, we can write

$$P_{ij} = P(y_i = j|x_i) = \frac{e^{x_i' \beta_j}}{1 + \sum_{l=2}^{m} e^{x_i' \beta_l}} \quad (j = 2, \dots, m)$$

$$P_{i1} = \frac{1}{1 + \sum_{l=2}^{m} e^{x_i' \beta_l}}, (set\ \beta_1 = 0)$$

Odds ratio of alternatives *j* and *l*:

$$\frac{P_{ij}}{P_{il}} = \frac{\frac{e^{x_i' \beta_j}}{\sum_{k=1}^{m} e^{x_i' \beta_k}}}{\frac{e^{x_i' \beta_l}}{\sum_{k=1}^{m} e^{x_i' \beta_k}}}$$

$$= e^{x_i'(\beta_j - \beta_l)}\ ln\left(\frac{P_{ij}}{P_{il}}\right)$$

$$= x_i'(\beta_j - \beta_l)$$

The odds ratio of alternatives j and l does not depend on any other alternative.

Let;

$$d_{ij} = \begin{cases} 1, & if\ alternative\ j\ is\ chosen\ by\ individual\ i \\ 0, & otherwise \end{cases}$$

Log likelihood is

$$lnL = \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij}\, lnP(y_i = j)$$

Derivative of log likelihood is

$$\frac{\partial}{\partial \beta_j} lnL = \sum_{i}^{n} (d_{ij} - P_{ij})\, x_i \qquad \forall j = 1, \ldots, m$$

Second derivative matrix has $m^2 k \times k$ blocks

$$\frac{\partial^2}{\partial \beta_j \partial \beta_l'} lnL = -\sum_{i}^{n} P_{ij}\, (I(j = l) - P_{il}) x_i x_i' = \begin{cases} 1, & if\ j = l \\ 0, & otherwise \end{cases}$$

The Hessian matrix does not involve d_{ij} and Newton's method can be applied to obtain MLEs.

The main weaknesses of Newton's method is that it has too many parameters.

Multinomial LOGIT model:

$\eta_{ij} = Z_{ij}^{'}\gamma$.

Z_{ij}: Characteristics of the choice j and individual i.

$Z_{ij} = (X_{ij}\ w_i)X_{ij}$ *include variables specific to individuals as well as choices*

w_i *include characteristics of individuals* same for all choices

Model:

$$P(y_i = j) = P_{ij} = P[Z_{ij}'\gamma \geq Z_{il}'\gamma] = \frac{e^{Z_{ij}'\gamma}}{\sum_{k=1}^{m} e^{Z_{ik}'\gamma}}$$

$$or\ P_{ij} = \frac{e^{X_{ij}'\beta + w_i'\alpha}}{\sum_{k=1}^{m} e^{X_{ik}'\beta + w_i'\alpha}} = \frac{e^{X_{ij}'\beta}}{\sum_{k=1}^{m} e^{X_{ik}'\beta}}$$

P_{ij} is independent of individual specific effects.

Probability ratio is

$$\frac{P_{ij}}{P_{il}} = e^{(X_{ij} - X_{il})'\beta}$$

Probability ratio is independence from irrelevant alternatives *(IIA)*, i.e., it does not depend on alternatives other than j and l.

We can write the log likelihood just like the multinomial Logit model.

4.5.6. Poisson Regression Model

Poisson regression is indeed a type of generalized linear model (GLM) specifically designed for modeling count data. It assumes that the response variable Y follows a Poisson distribution, which is suitable for modeling non-negative integer outcomes such as counts of events. The objective of the model is to develop a relationship between observed counts and potentially useful regressors.

Example: 1. Defects in a unit of manufacturing products

2. errors in software

3. counts of pollutants in environment

Assume that response variable y_i is a count such that observations are $y_i = 0, 1, 2, \ldots$.. Probability model for count data is Poisson distribution:

$$f(y) = \frac{e^{-\mu}\mu^{y}}{y!}; y = 0,1,2, \ldots; \ \mu > 0.$$

For Poisson distribution

$$E(y) = \mu \ and \ V(y) = \mu$$

Both mean and variance are equal to parameter μ.

Poisson Regression Model

Let

$$y_i = E(y_i) + \epsilon_i; \qquad i = 1,2, \ldots, n.$$

We assume, expected value of observed response can be written as

$$E(y_i) = \mu_i$$

and there is a function g called link function that relates or links mean response with linear predictor.

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

where, $x_i^{'} = (1, x_{i1}, x_{i2}, \dots, x_{ik})$

so the relationship between mean response and linear predictor is

$$\mu_i = g^{-1}(x_i^{'}\beta)$$

Several link functions are used with Poisson distribution one of them is identity link

$$\mu_i = g(\mu_i) = x_i^{'}\beta$$

When this link function is used

$$E(y_i) = \mu_i = g^{-1}(x_i'\beta) = x_i'\beta$$

Another popular link with Poisson distribution is *log link*.

$$g(\mu_i) = ln\ (\mu_i) = x_i'\beta$$

$$\Rightarrow \qquad E(y_i) = \mu_i = g^{-1}(x_i'\beta) = e^{x_i'\beta}$$

Log-link is particularly attractive for Poisson distribution because it ensures that all predicted values for response variable will be non-negative. For estimation of parameters method of maximum likelihood is used (approach like logistic regression).

Let $y_i; i = 1, 2, \dots, n$ is a random sample of n observation.

Likelihood function is

$$l(y, \beta) = \prod_{i=1}^{n} f_i(y_i)$$

$$= \prod_{i=1}^{n} \frac{e^{-\mu_i}{\mu_i}^{y_i}}{y_i!}$$

$$= \frac{(\prod_{i=1}^{n} {\mu_i}^{y_i}) e^{-\sum_{i=1}^{n}\mu_i}}{\prod_{i=1}^{n}(y_i!)}$$

Log likelihood function

$$L(y,\beta) = log\ \ l(y,\beta)\ = \sum_{i=1}^{n} y_i\ ln\ (\mu_i) - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} ln\ (y_i!)$$

where, $\mu_i = g^{-1}(x_i'\beta)$

Once the link function is specified, we maximize log-likelihood to find MLE's. Iteratively reweighted least squares (IRLS) can be used following an approach similar to logistics regression.

For Poisson distribution if $\mu_i = e^{x_i'\beta}$ *i.e.* if the link is log-link then

$$L(y,\beta) = \sum_{i=1}^{n} y_i\ (x_i'\beta) - \sum_{i=1}^{n} e^{x_i'\beta} - \sum_{i=1}^{n} ln\ (y_i!)$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} y_i\, x_i - \sum_{i=1}^{n} e^{x_i'\beta}\, x_i$$

$$= \sum_{i=1}^{n} \left(y_i - e^{x_i'\beta}\right) x_i$$

$$= \sum_{i=1}^{n} (y_i - \mu_i)\, x_i$$

Equating to zero vector we get maximum likelihood score equations

$$\sum_{i=1}^{n} (y_i - \mu_i)\, x_i = 0$$

$$\Rightarrow \qquad X'(y - \mu) = 0$$

where, $\mu' = (\mu_1, \mu_2, \ldots, \mu_n)$.

If intercept is included in the model, then $\sum_{i=1}^{n} (y_i - \mu_i) = 0$.

4.6. Self-Assessment Exercise

1. Discuss the limitations of the linear probability model. Can it be applied to all binary dependent variables? Why or why not?
2. How do the Tobit and Probit models differ in their handling of limited dependent variables?
3. Evaluate the implications of selecting an inappropriate model for discrete or limited dependent variables in terms of prediction accuracy and interpretation.
4. What are dummy variables, and how are they created from categorical data?
5. Define limited dependent variables and provide examples.
6. Why is a standard OLS regression inappropriate for limited dependent variables?
7. Discuss the key challenges in modeling limited dependent variables and suggest solutions.
8. Describe the differences between binary, ordered, and count data as limited dependent variables.
9. What is the primary purpose of the logit and probit models?
10. Describe Logit and Probit models. Compare the logit and probit models in terms of assumptions, computation, and use cases.
11. How do we estimate the parameters of logit and probit models?
12. Explain the difference between censoring and truncation in data with examples.
13. Describe the practical implications of ignoring censoring or truncation in data analysis.
14. What is the Tobit model, and when should it be used? Explain how the Tobit model differs from standard linear regression.

15. What are the assumptions underlying the Tobit model? How do we estimate the parameters of the model?

16. What is the multinomial logit model, and how does it differ from the standard logit model? When is it appropriate to use a multinomial logit model?

17. What type of dependent variable is suitable for analysis using a Poisson regression model?

18. Discuss the assumptions of the Poisson regression model. How do we estimate its parameters?

4.7. Summary

This unit offers a comprehensive overview model with dummy explanatory variable and limited dependent variables. The discussion then extends to specialized models for non-normal response variables. These include logistic (LOGIT) and probit (PROBIT) models for binary outcomes, TOBIT models for censored data, and Poisson regression for count data. Each model is analyzed in terms of its objectives, underlying assumptions, and parameter estimation techniques.

4.8. References

- Gujarati, D. and Guneshker, S. (2007): Basic Econometrics, 4th Ed., McGraw Hill Companies.

- Johnston, J. and Dinardo, J. (1997): Econometric Methods, 2nd Ed., McGraw Hill International.

4.9. Further Readings

- Judge, G.G., Griffiths, W.E., Hill Lütkepohl, R.C. and Lee, T. C. (1985): The theory and practice of econometrics, Wiley.

- Koutsoyiannis, A. (2004): Theory of Econometrics, 2 Ed., Palgrave Macmillan Limited.

- Maddala, G.S. and Lahiri, K. (2009): Introduction to Econometrics, 4 Ed., John Wiley & Sons.
- Ullah, A. and Vinod, H.D. (1981): Recent advances in Regression Methods, Marcel Dekker.

Introduction of unit on model with nonspherical disturbances, seemingly unrelated regression model, panel data model

Unit 5. Non-spherical disturbances

5.1. Introduction

5.2. Objectives

5.3. Model with non-spherical disturbances and estimation of parametric by generalized equation

5.4. Seemingly unrelated regression equations (SURE) model and its estimation

5.4.1. SUR Model and Simultaneous Equation Model

5.4.2. Two Equations Seemingly Unrelated Regression (SUR) Model

5.4.3. Maximum Likelihood Estimation

5.5. Panel data models

5.5.1. Benefits of Panel Data Model

5.5.2. Limitations of Panel Data Model

5.5.3. Heterogeneity across individuals and time

5.6. Estimation in random effect and fixed effect models

5.6.1. Fixed effects model with more than two time periods

5.6.2. Steps to Implement fixed effects estimator

5.6.3. Random Effects Model

5.6.4. Random effects as a combination of within and between estimators

5.6.5. Estimation of σ_η^2 σ_B^2 and σ_μ^2

5.1. Introduction

1. Model with Nonspherical Disturbances

This unit covers the models with nonspherical disturbances, seemingly unrelated regression model, panel data model.

In econometrics, **nonspherical disturbances** refer to a situation where the error terms (disturbances) in a regression model are not independently and identically distributed (i.i.d.) and may exhibit heteroskedasticity or autocorrelation. Such violations of the classical assumptions can lead to inefficient and biased estimators, if standard ordinary least squares (OLS) is used.

To address nonspherical disturbances:

- Generalized Least Squares (GLS) or Feasible GLS (FGLS) methods are often employed.
- GLS transforms the model to ensure the error terms become i.i.d., improving the efficiency of the estimates.

The **seemingly unrelated regression (SUR)** model is used when multiple regression equations are estimated simultaneously, and the error terms across equations are correlated. While the equations may appear unrelated in terms of their regressors, the correlation of their disturbances creates interdependence.

Key Features of the SUR model are

- SUR is more efficient than estimating the equations separately, especially if the error terms are highly correlated.
- It is particularly useful in applications like demand systems or sectoral economic models.
- Estimation is typically done using GLS or Maximum Likelihood Estimation (MLE).

A **panel data model** combines cross-sectional and time-series data, observing the same units (individuals, firms, countries) over multiple time periods. Panel data models help capture dynamics across time while accounting for individual heterogeneity.

Key Variants: of panel data models are

1. **Pooled OLS Model**: Ignores individual or time-specific effects and pools all observations, assuming homogeneity.
2. **Fixed Effects (FE) Model**: Accounts for unobserved individual-specific characteristics that are constant over time.
3. **Random Effects (RE) Model**: Assumes individual-specific effects are randomly distributed and uncorrelated with regressors.

Advantages:

- Controls for unobserved heterogeneity.
- Enables analysis of time-invariant and time-variant factors.
- Improves estimation efficiency by utilizing both cross-sectional and temporal dimensions.

5.2. Objectives

After completing this Block, students should have developed a clear understanding of:

- Model with non-spherical disturbances and estimation of parametric by generalized equation.
- Seemingly unrelated regression equations (SURE) model and its estimation.
- Panel data models.
- Estimation in random effect and fixed effect models

5.3. <u>Model with non-spherical disturbances</u>

The Model is

$y = X\beta + u$

Assumption of Spherical Disturbances is $E(uu') = \sigma_u^2 I_n$.

This assumption may be violated in many practical applications

- Heteroskedastic disturbances $\Rightarrow$ $Var\ (u_i)$ is different for different $i's$
- Cross-observation correlations $\Rightarrow$ $E(u_i u_j) \neq 0\ for\ i \neq j$
- Cluster effects $\Rightarrow$ Observations come in groups having correlations within group, but not across groups.

Suppose the Spherical disturbances are non-spherical then assumption is $E(uu') = V = \sigma_u^2 \Omega$ (say).

where

V and Ω are Positive definite, symmetric matrices.

For non-spherical disturbances, we refer to model as generalized linear model.

<u>Note:</u> We write covariance matrix in the form $\sigma_u^2 \Omega$ so that if we set $\Omega = I_n$, we get classical model with $E(uu') = \sigma_u^2 I_n$.

Special cases

(i) **Heteroscedasticity:** Disturbances are said to be heteroscedastic when they have different variances.

Suppose n observations are in g groups with n_i observations in i^{th} group, so that $n = \sum_{i=1}^{g} n_i$. The disturbance variance varies in difference groups. The model is

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_g \end{pmatrix}_{n\times 1} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_g \end{pmatrix}_{n\times k} \beta + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_g \end{pmatrix}_{n\times 1}$$

where $y_i: n_i \times 1; X_i: n_i \times k; u_i: n_i \times 1\ (i = 1,2, \dots, g)$.

Further,

$$E(u_i u_i') = \sigma_{ui}^2 I_{n_i}, E(u_i u_{i'}') = 0\ \forall i \neq i'$$

We can write the model as

$$y = X\beta + u$$

with

$$E(uu') = E\begin{bmatrix} u_1u_1' & u_1u_2' & \dots & u_1u_g' \\ u_2u_1' & u_2u_2' & \dots & u_2u_g' \\ \vdots & \vdots & \ddots & \vdots \\ u_gu_1' & u_gu_2' & \dots & u_gu_g' \end{bmatrix} = V = \begin{bmatrix} \sigma_1^2 I_{n_1} & 0 & \dots & 0 \\ 0 & \sigma_2^2 I_{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_g^2 I_{n_g} \end{bmatrix}$$

(ii) **Autocorrelated Disturbances:** Autocorrelation is usually observed for time series data. The disturbances usually have common variance but autocorrelated.

Suppose $E(u_t^2) = \sigma_u^2\ \forall\ t = 1, \dots, n\ and\ E(u_t u_{t+i}) = \sigma_u^2 \rho_i, \forall\ i > 0.$ Then

$$E(uu') = \sigma_u^2\Omega = \sigma_u^2\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}$$

Suppose u_t follows an AR(1) process

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where ρ is the autoregressive coefficient and $\varepsilon_t's$ are *iid* random errors with

$E(\varepsilon_t^2) = \sigma_\varepsilon^2;\ E(\varepsilon_t\varepsilon_{t'}) = 0\ \forall t \neq t'$. Then

$$E(u_t u_{t+i}) = \sigma_u^2\rho^i\ \forall\ t = 1, \ldots, n; i \geq 1$$

$$\Omega = \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix}.$$

Some other examples of models with non-spherical disturbances

(i) Model with Autoregressive Conditional Heteroscedastic (ARCH) or Generalized ARCH (GARCH) disturbances.

(ii) Panel data models.

(iii) Models having spatial autocorrelation

<u>Finite Sample Properties of OLS Estimator for Generalized Linear Model:</u>

<u>Result 5.1:</u> For the model with non-spherical disturbances

(i) The OLS estimator b is unbiased

(ii) The variance-covariance matrix of b is

$$E(b-\beta)(b-\beta)' = \sigma_u^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

(iii) If $u \sim N(0, \sigma_u^2 \Omega)$, then

$$b \sim N(\beta, \sigma_u^2 (X'X)^{-1} X' \Omega X (X'X)^{-1})$$

Proof: We can write the OLS estimator as

$$b = (X'X)^{-1} X'y = \beta + (X'X)^{-1} X'u$$

(i) As long as $E(u|X) = 0$,

$$E(b) = \beta \qquad \forall \beta$$

Thus, b is still unbiased for β.

(ii) Variance-Covariance matrix of b is

$$E(b - \beta)(b - \beta)' = E[(X'X)^{-1} X' u u' X (X'X)^{-1}]$$

$$= (X'X)^{-1} X' E(uu') X (X'X)^{-1}$$

$$= \sigma_u^2 (X'X)^{-1} X' \Omega X (X'X)^{-1}.$$

(iii) Since

$$b = \beta + (X'X)^{-1} X'u,$$

is a linear function of u and $u \sim N(0, \sigma_u^2 \Omega)$, we have

$b \sim N(\beta, \sigma_u^2 (X'X)^{-1} X' \Omega X (X'X)^{-1})$■

b is the linear unbiased estimator of β.

Consistency of OLS estimator for General Linear Model

Result 5.2: If $plim\left(\frac{1}{n}X'X\right) = Q$ and $plim\left(\frac{1}{n}X'\Omega X\right)$ are finite positive definite matrices, then the OLS estimator b is a consistent estimator of β.

Proof: We have

$$plim(b) = \beta + plim\left[\left(\frac{1}{n}X'X\right)^{-1}\right] plim\left(\frac{1}{n}X'u\right) = \beta + Q^{-1} plim\left(\frac{1}{n}X'u\right)$$

Further

$$plim\left(\frac{1}{n}X'u\right) = 0$$

Hence

$$plim(b) = \beta.$$

Further

$$p\lim E(b-\beta)(b-\beta)' = \sigma_u^2\, p\, lim[(X'X)^{-1}(X'\Omega X)(X'X)^{-1}]$$

$$= \frac{\sigma_u^2}{n} p\, lim\left[\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\Omega X\right)\left(\frac{1}{n}X'X\right)^{-1}\right] = 0.$$

Hence b is a consistent estimator of β■

Generalized Least Squares (GLS) Estimator:

Since Ω is positive definite, $\exists$ a nonsingular matrix T such that $T\Omega\, T' = I_n$.

We write

$$Ty = TX\beta + Tu$$

$$\text{or } y^* = X^*\beta + u^* \qquad (5.1)$$

where $y^* = Ty, X^* = TX \text{ and } u^* = Tu$.

Then

$$E(u^*u^{*\prime}) = E(T\,u\,u'\,T') = T(\sigma_u^2\,\Omega)T' = \sigma_u^2 T\Omega\,T' = \sigma_u^2 I_n.$$

Further $T'T = \Omega^{-1}$.

The GLS estimator of β is obtained by minimizing

$$S^* = (y^* - X^*\beta)'(y^* - X^*\beta)$$

$$= (y - X\beta)'T'T(y - X\beta)$$

$$= (y - X\beta)'\Omega^{-1}(y - X\beta)$$

Notice that the OLS estimator is obtained by minimizing $S = (y - X\beta)'(y - X\beta)$, whereas GLS estimator is obtained by minimizing S^*.

Result 5.3: The GLS estimator of β is

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \qquad (5.2)$$

Further

(i) GLS estimator $\hat{\beta}$ is an unbiased estimator of β.

(ii) Variance-covariance matrix of $\hat{\beta}$ is

$$E\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)' = \sigma_u^2(X'\Omega^{-1}X)^{-1}$$

Proof: The transformed model (5.1) satisfies the assumption of spherical disturbances $E(u^*u^{*\prime}) = \sigma_u^2 I_n$. For obtaining the GLS estimator, we have to minimize $S^* = (y^* - X^*\beta)'(y^* - X^*\beta)$, *i.e.*, we have to apply OLS to (5.1).

Applying OLS to (5.1), we have the following estimator for β:

$$\hat{\beta} = (X^{*\prime}X^{*})^{-1}X^{*}y^{*}$$

$$= (X'T'TX)^{-1}X'^{T'}Ty$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,$$

which leads to (5.2).

(i) For proving unbiasedness of $\hat{\beta}$, we write

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}[X\beta + u]$$

$$= \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u$$

Hence

$$E(\hat{\beta}) = \beta, \ \ \forall\beta$$

(ii) Again

$$E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(uu')\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$

$$= \sigma_u^2(X'\Omega^{-1}X)^{-1}\blacksquare$$

Generalized Gauss Markov Theorem:

Result 5.4:

Let $\lambda = (\lambda_1, \dots, \lambda_k)'$ and $\lambda'\beta = \lambda_1\beta_1 + \cdots + \lambda_k\beta_k$. Then $\lambda'\hat{\beta}$ is a Best Linear Unbiased Estimator (BLUE) of β in the sense that (i) it is an unbiased estimator of $\lambda'\beta$, (ii) for any linear unbiased estimator $a'y$ of $\lambda'\beta$, $\mathrm{Var}(\lambda'\hat{\beta}) \leq \mathrm{Var}(a'y)$ for all β.

Proof: We observe that

(i) $\mathrm{E}(\lambda'\hat{\beta}) = \lambda' E(\hat{\beta}) = \lambda'\beta \ \ \forall\beta$

(ii) We write $a'y = a^{*'}y^{*}, \text{with } a^{*} = T'^{-1}a.$ Further

$$\mathrm{E}(a^{*'}y^{*}) = a^{*'}X^{*}\beta = \lambda'\beta \;\forall\beta \Rightarrow a^{*'}X^{*} = \lambda'$$

$$Var(a^{*'}y^{*}) = \sigma_u^2 a^{*'}a^{*}Var\left(\lambda'\hat{\beta}\right) = \sigma_u^2\lambda'(X^{*'}X^{*})^{-1}\lambda = \sigma_u^2 a^{*'}X^{*}(X^{*'}X^{*})^{-1}X^{*'}a^{*}$$

Hence

$$Var(a^{*'}y^{*}) - Var\left(\lambda'\hat{\beta}\right) = \sigma_u^2 a^{*'}(I_n - X^{*}(X^{*'}X^{*})^{-1}X^{*'})a^{*} \geq 0\blacksquare$$

Result 5.5: If $plim\left(\frac{1}{n}X^{*'}X^{*}\right) = plim\left(\frac{1}{n}X'\Omega^{-1}X\right) = Q^{*}$ is a finite positive definite matrix then $\hat{\beta}$ is a consistent estimator of β.

Proof: The result can be easily verified as $\hat{\beta}$ is obtained by applying OLS to model (5.1).

Maximum Likelihood Estimation

Result 5.6: If $u \sim N(0, \sigma_u^2\Omega)$, the maximum likelihood estimators of β and σ_u^2 are, respectively, given by

$$\hat{\beta}_{ML} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \hat{\sigma}_{ML}^2 = \frac{1}{n}\left(y - X\hat{\beta}\right)\Omega^{-1}\left(y - X\hat{\beta}\right).$$

Proof: $u \sim N(0, \sigma_u^2\Omega).$ The log likelihood function is given by

$$lnL = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}ln\sigma_u^2 - \frac{1}{2\sigma_u^2}(y - X\beta)'\Omega^{-1}(\mathrm{y} - \mathrm{X}\beta) - \frac{1}{2}\ln|\Omega|$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}ln\sigma_u^2 - \frac{1}{2\sigma_u^2}(y^{*} - X^{*}\beta)'(\mathrm{y}^{*} - \mathrm{X}^{*}\beta) - \frac{1}{2}\ln|\Omega|$$

Then $\frac{\partial}{\partial\beta}lnL = \frac{1}{\sigma_u^2}(X^{*'}\mathrm{y}^{*} - \mathrm{X}^{*'}\mathrm{X}^{*}\beta) = 0$

$$\frac{\partial}{\partial\sigma_u^2}lnL = -\frac{n}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}(y^{*} - X^{*}\beta)'(\mathrm{y}^{*} - \mathrm{X}^{*}\beta) = 0$$

Substituting these derivatives equal to zero, we obtain the following ML estimators of β and σ_u^2:

$$\hat{\beta}_{ML} = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,$$

$$\hat{\sigma}_{ML}^2 = \frac{1}{n}(y^* - X^*\hat{\beta})'(y^* - X^*\hat{\beta})$$

$$= \frac{1}{n}(y - X\hat{\beta})\Omega^{-1}(y - X\hat{\beta})\blacksquare$$

ML estimator of β is the same as the GLS estimator $\hat{\beta}$. Further

$$E(\hat{\sigma}_{ML}^2) = \frac{n-k}{n}\sigma_u^2 Bias(\hat{\sigma}_{ML}^2) = -\frac{k}{n}\sigma_u^2$$

Feasible Generalized Least Squares Estimator

In practice Ω is usually unknown. Suppose it is a function of a $q\times1$ unknown parameter vector θ, so that we can write $\Omega\equiv\Omega(\theta)$. Suppose a consistent estimator $\hat{\theta}$ of θ is available. Then we estimate Ω by $\hat{\Omega} \equiv \Omega(\hat{\theta})$.

A feasible generalized least squares (FGLS) estimator of β is given by

$$\tilde{\beta} = \left(X'\hat{\Omega}^{-1}X\right)^{-1}X'\hat{\Omega}^{-1}y$$

5.4. Seemingly unrelated regression equations (SURE) model and its estimation

- Multiple regression model describes the behavior of a particular study variable using a set of explanatory variables.
- For explaining the whole economic system, the model may involve a set of multiple regression equations with each equation explaining a particular economic phenomenon.

- Should different equations be treated separately for estimation purpose?
- Equations may be seemingly unrelated with each other structurally but linked through some statistical interaction among them.
- The interaction is through the correlation between random error terms of different equations.
- Is it possible to improve the estimators based on individual equations by developing estimators based on the entire system of equations?
- How to estimate the entire system of equations?

SUR Models

$\Rightarrow$ Jointness of equations is through the covariance matrix of the associated disturbances

$\Rightarrow$ Jointness provide additional information over and above the information available with the individual equations.

Example : Objective is to estimate demand relationships for a particular commodity for several households.

Price and income data are the exogeneous variables.

One may expect jointness of demand equations for different households through their error covariances.

Example: While studying consumption pattern of a country with 20 states, each state has a consumption equation. Different equations may involve different variables and apparently look unrelated. Consumption pattern of different states may have some kind of statistical interdependence. Correlation may exist between the error terms associated with the equations. The equations are apparently or "seemingly" unrelated regressions but not independent relationships.

Example: Capital Asset Pricing Model (CAPM)

CAPM of finance evolves as a way to measure the systematic risk. For the i^{th} security

$R_{it} - R_{ft} = \alpha + \beta_i (R_{mt} - R_{ft}) + u_{it}$

$R_{it} =$ Expected return on i^{th} security

$R_{ft} =$ Risk free rate

$R_{mt} =$Expected return of the market

$\beta_i =$ Beta of the security (a measure of systematic risk of a security or portfolio compared to the market as a whole).

$R_{mt} - R_{ft} =$ Equity market premium

u_{it}: Disturbances

CAPM describes the relationship between systematic risk and expected return for security.

Risk-free rate accounts for the time value of money.

Other components of CAPM formula account for the investor taking on additional risk.

Beta is a measure of how much risk the investment will add to a portfolio that looks like the market.

If a stock has beta greater than one, stock is riskier than the market.

If a stock has a beta of less than one, it will reduce the risk of a portfolio.

The information about the return of a security to exceed risk free rate by a given amount will provide the information about the excess return of some other securities.

Disturbances are then obviously correlated across securities.

Estimating the securities jointly may provide better estimates than the estimates based on individual equations.

Example: Investment Model

Consider two IT companies; say Tata consultancy Services (TCS) and Infosys. Let

$N1: 25\times1$ Vector of observations on investment by TCS

$V1: 25\times1$ Vector of observations on stock market value of TCS

$K1: 25\times1$ Vector of observations on year capital stock of TCS

$N2, V2, K2:$ Corresponding 25×1 vectors of observations for Infosys

$l: 25\times1$

$$N_1 = \beta_{11} l + \beta_{12} V_1 + \beta_{13} K_1 + u_1: \text{ Model for TCS} N_2$$

$$= \beta_{21} l + \beta_{22} V_2 + \beta_{23} K_2 + u_2: \text{ Model for Infosys}$$

We assume that

$$E(u_1 u_1') = \sigma_{11} I_{25};\ E(u_2 u_2') = \sigma_{22} I_{25}$$

$l = (1\ \ 1\ \ \ldots\ 1)': 25\times1$ Vector

Let us write

$y_1 = N_1; X_1 = [l\ \ V_1\ \ K_1]: 25\times3$ matrix

$y_2 = N_2; X_2 = [l\ \ V_2\ \ K_2]: 25\times3$ matrix

$$\beta_1 = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \end{pmatrix}; \beta_2 = \begin{pmatrix} \beta_{21} \\ \beta_{22} \\ \beta_{23} \end{pmatrix}$$

Then

$$\left.\begin{aligned} y_1 &= X_1\beta_1 + u_1: \quad \text{Model for TCS} \\ y_2 &= X_2\beta_2 + u_2: \quad \text{Model for Infosys} \end{aligned}\right\}$$

Applying least squares to two equations separately, we get the estimators of β_1 and β_2 as

$b_1 = (X_1'X_1)^{-1}X_1'y_1 b_2 = (X_2'X_2)^{-1}X_2'y_2.$

Two firms are working in the same branch of industry and investments (or errors) of two models may be correlated, *i.e.*, $E(u_{1i}u_{2i}) = \sigma_{12}\ \forall\ i \text{ or } E(u_1u_2') = \sigma_{12}I_{25}$.

Reason for such a correlation

The state of economy whose effect is felt through u_1 andu_2 is likely to have similar effects on each of the firm.

"Is it possible to pool the two equations and estimate β_1 and β_2 more efficiently taking into account the correlation between them?"

5.4.2. Two Equations Seemingly Unrelated Regression (SUR) Model

Consider the following two equations SUR model:

$$\left.\begin{aligned} y_1 &= X_1\beta_1 + u_1 \\ y_2 &= X_2\beta_2 + u_2 \end{aligned}\right\} \qquad (5.3)$$

y_1, y_2: $n\times1$ vectors of observations on dependent variable

$X_1(n\times k_1), X_2\ (n\times k_2)$:Matrices of observations on explanatory variables

$$\left.\begin{aligned} u_1 &= (u_{11}, u_{12}, \dots, u_{1n})' \\ u_2 &= (u_{21}, u_{22}, \dots, u_{2n})' \end{aligned}\right\}: \text{Disturbances Vector}$$

$E(u_{1t}) = 0 = E(u_{2t}),\ \forall\ t = 1,2, \dots, n$

$E(u_{1t}^2) = \sigma_{11},\ E(u_{2t}^2) = \sigma_{22}\ \forall\ t = 1, \dots, n$

$E(u_{1t}u_{2t}) = \sigma_{12} \forall\ t\ = 1, \dots, n$

In matrix notations, we can write the conditions as

$E(u_1) = E(u_2) = 0$

$$E(u_1u_1') = \begin{bmatrix} E(u_{11}^2) & E(u_{11}u_{12}) & \cdots & E(u_{11}u_{1n}) \\ \vdots & \vdots & \cdots & \vdots \\ E(u_{1n}u_{11}) & E(u_{1n}u_{12}) & \cdots & E(u_{nn}^2) \end{bmatrix} = \sigma_{11}I_n$$

;

$$E(u_2u_2') = \sigma_{22}I_n$$

$$E(u_1u_2') = \sigma_{12}I_n$$

We can combine the two equations (5.3) as

$$y = X\beta + u$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} : 2n \times 1;$$

$$X = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} : 2n \times (k_1 + k_2)$$

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} : 2n \times 1; \text{ Disturbances vector}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} : (k_1 + k_2) \times 1 \text{ Vector of regression coefficients}$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

$$y_1 = X_1\beta_1 + u_1 \ , \quad y_2 = X_2\beta_2 + u_2$$

Then

$$E(uu') = E\begin{pmatrix} u_1u_1' & u_1u_2' \\ u_2u_1' & u_2u_2' \end{pmatrix} = \begin{pmatrix} \sigma_{11}I_n & \sigma_{12}I_n \\ \sigma_{21}I_n & \sigma_{22}I_n \end{pmatrix} = \Sigma \otimes I_n = V \text{ (say)}$$

$\otimes$ denotes the Kronecker product operator

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

The GLS estimator of β is

$$\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$$
$$= \{X'(\Sigma^{-1} \otimes I_n)X\}^{-1}X'(\Sigma^{-1} \otimes I_n)y$$

Covariance matrix of $\tilde{\beta}$ is

$$E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' = (X'V^{-1}X)^{-1} = \{X'(\Sigma^{-1} \otimes I_n)X\}^{-1}$$

Σ is usually unknown. For estimating β, we proceed as follows:

i. Apply OLS to each equation in (5.3). Obtain the OLS residuals

$$e_1 = [I_n - X_1(X_1'X_1)^{-1}X_1']y_1$$

$$e_2 = [I_n - X_2(X_2'X_2)^{-1}X_2']y_2$$

ii. Estimate $\sigma_{ii}\ (i = 1,2)$ by

$$s_{ii} = \frac{e_i'e_i}{n - k_i}; i = 1,2$$

iii. Estimate $\sigma_{12} = \sigma_{21}$ by

$$s_{12} = \frac{e_1'e_2}{(n - k_1)^{\frac{1}{2}}(n - k_2)^{\frac{1}{2}}} = s_{21}$$

$$\text{or} \quad \tilde{s}_{12} = \frac{e_1'e_2}{n - max(k_1, k_2)} = \tilde{s}_{21}$$

iv. Obtain

$$\hat{\Sigma}^{-1} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}^{-1} = \begin{pmatrix} s^{11} & s^{12} \\ s^{21} & s^{22} \end{pmatrix}$$

v. The feasible GLS estimator of β is

$$\hat{\beta} = \left\{X'(\hat{\Sigma}^{-1} \otimes I_n)X\right\}^{-1}X'(\hat{\Sigma}^{-1} \otimes I_n)y.$$

Kronecker Product:

$A = \left((a_{ij})\right): m \times n$ matrix

$B = \left((b_{kl})\right): p \times q$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}: mp \times nq$$ matrix

Algebra of Kronecker Products

Result 5.8: We have

$(i)\ (A \otimes B)(C \otimes D) = AC \otimes BD\, (ii)\ (A \otimes B)' = (A' \otimes B')$

From (i), we observe that for square matrices $A\ (m \times m)$ and $B\,(p \times p)$

$(A^{-1} \otimes B^{-1})(A \otimes B) = (A^{-1}A \otimes B^{-1}B) = I_m \otimes I_p = I_{mp}$

Hence $(A \otimes B)^{-1} = (A \otimes B)$.

SUR Model: General Case of M Equations

$y_1 = X_1\beta_1 + u_1$

$y_M = X_M\beta_M + u_M$

$E(u_m) = 0; E(u_m u_m') = \Sigma, \forall m = 1, \ldots, M$

$y_m: n \times 1; X_m: n \times k_m;\ \beta_m: k_m \times 1;\ u_m: n \times 1$

Let

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix},$$

$$X = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_M \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_M \end{pmatrix}; \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_M \end{pmatrix}$$

Then the model can be expressed as

$$y = X\beta + u$$

$$E(u) = 0;$$

$$E(uu') = \Sigma \otimes I_n$$

The GLS estimator of β is

$$\hat{\beta} = \{X'(\Sigma^{-1} \otimes I_n)X\}^{-1} X'(\Sigma^{-1} \otimes I_n)y$$

Let $\sigma^{ij}: (i,j)^{th}\ element\ of\ \Sigma^{-1}$

Expending the Kronecker product, we obtain

$$\hat{\beta} = \begin{pmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2 & \cdots \sigma^{1M}X_1'X_M \\ \sigma^{21}X_2'X_1 & \sigma^{22}X_2'X_2 & \cdots \sigma^{2M}X_2'X_M \\ \vdots & \vdots & \ddots \quad \vdots \\ \sigma^{M1}X_M'X_1 & \sigma^{M2}X_M'X_2 & \cdots \sigma^{MM}X_M'X_M \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{M} \sigma^{1j}X_1'y_j \\ \sum_{j=1}^{M} \sigma^{2j}X_2'y_j \\ \vdots \\ \sum_{j=1}^{M} \sigma^{Mj}X_M'y_j \end{pmatrix} \qquad (5.4)$$

Result 5.9: *When* $\sigma_{ij} = 0\ \forall\ i \neq j$, *the GLS estimator of* β *reduces to the OLS estimator* $b = \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix}$, *where* $b_m = (X_m'X_m)^{-1}X_m'y_m$.

Proof: If $\sigma_{ij} = 0\ \forall\ i \neq j,\ \text{then}\ \sigma^{ii} = \sigma_{ii}^{-1}\ \text{and}\ \sigma^{ij} = 0\ \forall i \neq j.$ Hence (5.4) becomes

$$\hat{\beta} = \begin{pmatrix} \sigma^{11}X_1'X_1 & 0 & \cdots & 0 \\ 0 & \sigma^{22}X_2'X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots\sigma^{MM}X_M'X_M \end{pmatrix}^{-1} \begin{pmatrix} \sigma^{11}X_1'y_1 \\ \sigma^{22}X_2'y_2 \\ \vdots \\ \sigma^{MM}X_M'y_M \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix}, b_m$$

$$= (X_m'X_m)^{-1}X_m'y_m \blacksquare$$

Result 5.10: *When* $X_1 = \cdots = X_M$*, the GLS estimator becomes the OLS estimator.*

Proof: Let $X_1 = \cdots = X_M = Z$, so that

$$X = \begin{pmatrix} Z & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z \end{pmatrix} = I_M \otimes Z \text{Then}\ X'(\Sigma^{-1}\otimes I_n)X$$

$$= (I_M\otimes Z')(\Sigma^{-1}\otimes I_n)(I_M\otimes Z) \qquad = \Sigma^{-1}\otimes Z'ZX'(\Sigma^{-1}\otimes I_n)y$$

$$= (\Sigma^{-1}\otimes Z')y$$

$$\text{Thus}\hat{\beta} = \{\Sigma^{-1}\otimes Z'Z\}^{-1}(\Sigma^{-1}\otimes Z')y = \{I_M\otimes(Z'Z)^{-1}Z'\}\begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix},\ b_m$$

$$= (Z'Z)^{-1}Z'y_m$$

Hence the result follows $\blacksquare$

$\boldsymbol{\Sigma}$ unknown: Feasible Generalized Least Squares

(i) Use of unrestricted residuals for Estimating $\boldsymbol{\Sigma}$

Let K^* be the total number of distinct explanatory variables out of $k_1, \ldots, k_M$ variables in the full model Z be $n \times K^*$ observation matrix of these variables Regress each of the *M* study variables on the column of *Z* and obtain $(n \times 1)$ residual vectors

$$\hat{u}_i = y_i - Z(Z'Z)^{-1}Z'y_i = \bar{H}_Z y_i; i = 1,2,\ldots,M$$

where $\bar{H}_Z = I_n - Z(Z'Z)^{-1}Z'.$

Obtain

$$s_{ij} = \frac{1}{n}\hat{u}_i'\hat{u}_j = \frac{1}{n}y_i'\bar{H}_Z y_j$$

and construct the matrix $S = \left((s_{ij})\right)$. Since X_i is a submatrix of Z, we can write $X_i = ZJ_i$, where J_i is a $(K \times k_i)$ selection matrix. Then

$$\bar{H}_Z X_i = X_i - Z(Z'Z)^{-1}Z'X_i = X_i - ZJ_i = 0$$

Thus

$$y_i'\bar{H}_Z y_j = (\beta_i'X_i' + u_i')\bar{H}_Z\left(X_j\beta_j + u_j\right) = u_i'\bar{H}_Z u_j.$$

Hence

$$E\left(s_{ij}\right) = \frac{1}{n}E\left(u_i'\bar{H}_Z u_j\right) = \frac{1}{n}\sigma_{ij}tr(\bar{H}_Z) = \left(\frac{n-K^*}{n}\right)\sigma_{ij} \text{ or } E\left(\frac{n}{n-K^*}s_{ij}\right) = \sigma_{ij}.$$

Thus, an unbiased estimator of σ_{ij} is

$$\hat{\sigma}_{ij} = \frac{n}{n-K^*}s_{ij}$$
$$= \frac{1}{n-K^*}\hat{u}_i'\hat{u}_i.$$

The FGLS estimator of β is

$$\hat{\beta} = \left\{X'\left(\hat{\Sigma}^{-1}\otimes I_n\right)X\right\}^{-1}X'\left(\hat{\Sigma}^{-1}\otimes I_n\right)y$$

(ii) <u>Use of restricted residuals for Estimating $\boldsymbol{\Sigma}$</u>

Regress y_i on X_i for each equation, $i = 1,2,\ldots,M$, and obtain the OLS residual vector

$$\tilde{u}_i = [I - X_i(X_i'X_i)^{-1}X_i']y_i = \bar{H}_{X_i}y_i.$$

A consistent estimator of σ_{ij} is obtained as

$$s_{ij}^* = \frac{1}{n}\tilde{u}_i'\tilde{u}_j = \frac{1}{n}y_i'\bar{H}_{X_i}\bar{H}_{X_j}y_j$$

where

$$\bar{H}_{X_i} = I - X_i(X_i'X_i)^{-1}X_i\bar{H}_{X_j}$$
$$= I - X_j\left(X_j'X_j\right)^{-1}X_j.$$

Using s_{ij}^{*}, a consistent estimator S of Σ can be constructed. Further

$$\mathrm{E}\left(y_i'\bar{H}_{X_i}\bar{H}_{X_j}y_j\right) = \sigma_{ij}\,tr\left(\bar{H}_{X_i}\bar{H}_{X_j}\right)$$
$$= \sigma_{ij}\left[n - k_i - k_j + tr(X_i'X_i)^{-1}X_i'X_j\left(X_j'X_j\right)^{-1}X_j'X_i\right]$$

Hence

$$\tilde{s}_{ij} = \frac{\tilde{u}_i'\tilde{u}_i}{\left[n - k_i - k_j + tr(X_i'X_i)^{-1}X_i'X_j\left(X_j'X_j\right)^{-1}X_j'X_i\right]}$$

is an unbiased estimator of σ_{ij}.

5.4.3. Maximum Likelihood Estimation

For obtaining MLE, consider t^{th} observation $(t = 1, \dots, n)$ on each of M dependent variables and corresponding regressors. Arranging these observations horizontally, we can write the model as

$$(y_1\ y_2\ \dots y_M)_t = [x_t^*]'(\pi_1\ \pi_2\ \dots \pi_M) + (u_1\ u_2\ \dots u_M)_t$$

or

$$Y = X^*\Pi' + U$$

Here x_t^* is the set of K^* different explanatory variables.

$x_t^{*\prime}$ is t^{th} row of X^* $(n \times K^*)$,

$Y: n \times M, U: n \times M$

Π' has one column for each equation and i^{th} column π_i has number of zeros in it each one imposing exclusion restriction.

For example, let

$$y_1 = \alpha_1 + \beta_{11}x_1 + \beta_{12}x_2 + u_1y_2$$
$$= \alpha_2 + \beta_{22}x_2 + \beta_{23}\,x_3 + u_2$$

Then, corresponding to the t^{th} observation

$$[y_1\ \ y_2]_t = [1\ \ x_1\ \ x_2\ \ x_3]_t \begin{pmatrix} \alpha_1 & \alpha_2 \\ \beta_{11} & 0 \\ \beta_{12} & \beta_{22} \\ 0 & \beta_{23} \end{pmatrix} + [u_1\ \ u_2]_t$$

Log of joint normal density of M disturbances is

$$logL_t = -\frac{M}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}u_t'\Sigma^{-1}u_t$$

Log likelihood is

$$logL = -\frac{Mn}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{t=1}^{n} u_t'\Sigma^{-1}u_t$$

$$= -\frac{n}{2}[M\log(2\pi) + \log|\Sigma| + tr(\Sigma^{-1}W)] \qquad \{W = ((W_{ij}));\ W_{ij}\} \qquad (5.5)$$

$$= \frac{1}{n}\sum_{t=1}^{n} u_{ti}u_{tj}$$

Now

$$\frac{\partial}{\partial\Pi'}logL = 0 \Rightarrow X^{*\prime}\Sigma^{-1}U = 0$$

$$\frac{\partial}{\partial\Sigma}logL = 0 \Rightarrow \Sigma^{-1}(\Sigma - W)\Sigma^{-1} = 0 \qquad (5.6)$$

We have utilized the following results:

$$\frac{\partial tr(\Sigma^{-1}W)}{\partial \Sigma} = -\Sigma^{-1}\frac{\partial tr(\Sigma^{-1}W)}{\partial \Sigma^{-1}}\Sigma^{-1}$$

$$= -\Sigma^{-1}W\Sigma^{-1}\frac{\partial \log|\Sigma|}{\partial \Sigma}$$

$$= \frac{1}{|\Sigma|}\frac{\partial|\Sigma|}{\partial \Sigma}$$

$$= \frac{1}{|\Sigma|}|\Sigma|\Sigma^{-1}$$

$$= \Sigma^{-1}$$

$$\frac{\partial}{\partial \Pi'}tr(\Sigma^{-1}W) = \frac{1}{n}\frac{\partial}{\partial \Pi'}tr\big((Y - X^{*}\Pi')'\Sigma^{-1}(Y - X^{*}\Pi')\big)$$

$$= -\frac{2}{n}X^{*\prime}\Sigma^{-1}U$$

From (5.6), given slope parameters, the MLE of Σ is W. Replacing Σ by W in (5.5), the concentrated-log likelihood is

$$logL_c = -\frac{n}{2}[M\log(2\pi) + \log|\mathrm{W}| + M]$$

Thus

$\hat{\beta}_{ML} = Min_{\beta}\frac{1}{2}\log|W|$ subject to exclusion restrictions.

Goodness of fit measure R^2 is defined as

$$R^2 = 1 - \frac{\hat{u}'\hat{\Omega}^{-1}\hat{u}}{\sum_{i=1}^{M}\sum_{j=1}^{M}\hat{\sigma}^{ij}\left[\sum_{t=1}^{n}(y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j)\right]}$$

$$= 1 - \frac{M}{tr\left(\hat{\Sigma}^{-1}S_{yy}\right)}$$

$\hat{\beta}$ is FGLS estimator of β

$\hat{e} = y - X\hat{\beta}$,

$\hat{\Omega} = \hat{\Sigma} \otimes I_n$,

$\bar{y}_i = \frac{1}{n}\sum_{l=1}^{n} y_{il}$

$\hat{\sigma}^{ij}$: $(i,j)^{th}$ element of $\hat{\Sigma}^{-1}$

$S_{yy} = \left((s_{ij})\right)$,

$s_{ij} = \frac{1}{n}\sum_{t=1}^{n} (y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j)$

5.5. Panel data models

Three types of data sets:

(i) Time series data,

(ii) Cross section data,

(iii) Longitudinal or Panel data: Data that contains observations on different cross sections across time.

If the same people or states or countries, sampled in the cross section, are then re-sampled at different time points, we get longitudinal or panel data set.

- Panel data contains observations collected at a regular frequency, chronologically and observations across a collection of individuals.
- Longitudinal/ Panel data sets are very common in Economics, Medical and Biostatistical studies.
- Becoming popular due to the widespread use of the computer making it easy to organize, produce and analyze such data.

Examples:

- Annual data on unemployment rates, GDP, % of people living below the poverty line for 28 states of India over 2011-2020.
- Quarterly sales of Cars (small hatchbacks) of different brands over several quarters.

- Currency values of developing countries at regular intervals during last ten years.
- Daily closing prices of different stocks of IT sector for the past one year.

5.5.1. Benefits of Panel Data

- ***More accurate inference of model parameters***

Panel data provide more informative data, more variability, less collinearity, more degrees of freedom and more efficiency.

Provide a large number of data points, increasing the degrees of freedom and reducing the multicollinearity in explanatory variables. Thus lead to more efficient estimates.

Example:

Objective is to model yearly demand of car of Kia Motors in a city using explanatory variables such as peoples income, city size etc.

Time series data on a particular city may not have enough data points

If we fit model for panel data collected from different cities, the sample size increases. Data involves more heterogeneity, leading to more efficient estimates.

One has to assume that the same relationship holds for different cities.

If predictions for an individual are based on short time series, using panel data increases the sample information and the accuracy of predictions if behavior of individuals are similar conditional on certain variables.

- ***Greater capacity for capturing the complexity of human behavior than a single cross-section or time series data***

Able to identify and measure effects that are not detectable using only cross section or time series data.

Able to address important questions which can not be answered using only time series or cross section data.

Example:

While analyzing labor force participation cross-section data of women, an observed participation rate of, say, 50% implies that each women in a homogeneous population has 50% chance of being (spends 50% of her life) in the labor force.

It does not address the issue that whether the women is working or not in the past, which may be a good predictor of labor force participation. This kind of issues can be addressed if we analyze panel data.

Example:

Evaluating the effects of legalizing in a state A on marijuana smoking behavior by comparing the differences between A and other states that were still non-legalized.

The panel data would allow the possibility of observing the before- and affect effects on individuals of legalization as well as providing the possibility of isolating the effects of treatment from other factors affecting the outcome.

- ***Better understanding of dynamics of adjustment***

A single time series model cannot provide good estimates of dynamic coefficients.

For instance, in distributed lag model

$$y_t = \sum_{j=0}^{k} \beta_j x_{t-j} + u_t; t = 1, \dots, N$$

x_t is close to x_{t-1} or $2x_{t-1} - x_{t-2}$. Thus data are nearly multicollinear.

The panel data reduces multicollinearity using inter individual differences in x values.

- ***Able to overcome the bias arising from Omitted Variables***

If omitted variables are correlated with included explanatory variables, it may lead to correlation between these variables and disturbances. Panel data controls the impact of omitted variables leading to individual or time heterogeneity.

Consider

$$y_{it} = \alpha + \beta x_{it} + \gamma z_{it} + u_{it}; i = 1, \dots, N; t = 1, \dots, T$$

If z_{it} is omitted, it makes the OLS estimators of α and β biased. However, in panel data, we may get rid of this problem.

For instance, if $z_{it} = z_i \, \forall t$, we may fit the model for first differences

$$y_{it} - y_{i,t-1} = \beta(x_{it} - x_{i,t-1}) + (u_{it} - u_{i,t-1})$$

If $z_{it} = z_t \, \forall i$ then we may transform the model as

$$y_{it} - \bar{y}_{.t} = \beta(x_{it} - \bar{x}_{.t}) + (u_{it} - \bar{u}_{.t})$$.

- ***Overcome the problem of measurement error***

Measurement error leads to under identification of the model.

The availability of multiple observations in panel data allows to make transformation in the model to make it identifiable.

- ***Providing micro foundations for aggregate data analysis and overcome the problem of biases arising from aggregation***

If micro units are heterogeneous in nature, the time series properties of macro data based on aggregate information may be misleading.

With panel data having time series observations on individuals, investigation of homogeneity versus heterogeneity is possible.

- ***More accurate predictions for individual outcomes***

Pooling leads to improved predictions in comparison to generating predictions of individual outcomes using the data on the individuals.

If individual behaviors are similar conditional on explanatory variables, panel data provides the possibility of learning an individual's behavior by observing the behavior of others.

- ***Availability of panel data simplifies computation and inference. Also Allows to construct and test more complicated behavioral models than purely cross-section or time series data.***

In nonstationary time series, the large sample approximations for the distribution of least squares or several other statistic is not normal.

For panel data, one can use central limit theorem across the cross-section to derive the limiting distribution of several statistic and to prove asymptotic normality.

5.5.2. Limitations of Panel Data Model:

- ***Design and Data collection***

Include problems of coverage (incomplete account of the population of interest), nonresponse, recall (respondent not remembering correctly), frequency of interviewing, time-in-sample bias etc.

- ***Distortions of measurement errors***

Faulty responses due to unclear questions, memory errors, deliberate distortion of responses (e.g. prestige bias), inappropriate informants, mis recording of responses and interviewer effects.

- ***Selectivity problem***

Self-Selectivity: Labor refuse to work as offered wage is lower than reservation wage (lowest wage rate at which a worker would be willing to accept a job). We can observe other variables but not wages of such people. If we don't take data on such people, we get truncated sample.

Nonresponse: Usually occurs at the initial waves of panel. In surveys some of the questions may remain unanswered or complete nonresponse may occur.

Attrition: Nonresponse may occur in cross-sectional studies in the subsequent waves of the panel.

- ***Short time-series dimension***

Micro panels involve annual data covering a short time span for each individual. Usually the asymptotic arguments rely crucially on the number of individuals tending to infinity.

- ***Cross-section dependence***

Macro panels on countries or regions with long time series that do not account for cross-country dependence may lead to misleading inference.

5.5.3. Heterogeneity across individuals and time:

Different individuals may be influenced by different factors leading to heterogeneity across individuals.

Assumption that y is generated from probability distribution $f(y|\theta)$, with θ constant for all individuals is not valid if some significant factors are left out from the model specification. We characterize the distribution of y_{it} as $f(y_{it}|x_{it}, \theta_{it})$.

If θ_{it} is decomposed into β, γ_{it}, then β are called structural parameters.

γ_{it} vary across individuals and time and called *incidental parameters*.

If γ_{it} are random variables, the model is called *random effects model*. If γ_{it} are fixed unknown constants, it is called fixed effects model.

Consider following models with single explanatory variable

Model with no heterogeneity $(\alpha_i = \alpha,\ \beta_i = \beta\ \forall i)$ is

$$y_{it} = \alpha + \beta x_{it} + u_{it}; i = 1, \dots, N; t = 1, \dots, T$$

Model with heterogeneity in intercept $(\beta_i = \beta\ \forall i)$:

$$y_{it} = \alpha_i + \beta x_{it} + u_{it}; i = 1, \dots, N; t = 1, \dots, T$$

Model with heterogeneity in intercept and slope coefficient

$$y_{it} = \alpha_i + \beta_i x_{it} + u_{it}; i = 1, \dots, N; t = 1, \dots, T$$

Balanced panels: Same number of observations on each cross-section unit. $(i = 1,2, \dots, N; t = 1,2, \dots, T)$

Unbalanced panels: Unequal number of observations on each cross-section unit. $(i = 1,2, \dots, N; t = 1,2, \dots, T_i)$

We consider the case of balanced panels only.

For $T = 1$, we get cross-section observations. For $N = 1$, we get time series observations.

In panel data estimation methods, we consider the case when $N > 1$ and $T > 1$.

5.6. Estimation in random effect and fixed effect models

5.6.1. Fixed effects model with more than two time periods:

Consider the model

$y_{it} = \alpha + X_{it}\beta + \mu_i + \eta_{it}; (t = 1,2, \dots, T)$

and $\mathrm{E}(X_{it}\mu_i) \neq 0$.

As N increases, number of parameters $(\mu_i's)$ also increases.

We cannot estimate $\mu_i's$ consistently but we can estimate other parameters consistently.

We write the model as

$y = \alpha l_{NT} + X\beta + D\mu + \eta$

$D = I_N \otimes l_T$ is a set of N dummy variables.

We regress y on D and obtain OLS residuals $Q_D y$.

Then regress X on D to get OLS residuals $Q_D X$.

Running regression between OLS residuals $Q_D y$ and $Q_D X$, the estimator of β is obtained as

$$\hat{\beta}_W = (X' Q_D X)^{-1} X' Q_D y$$

This is the *within estimator*. The estimator is also called least squares dummy variable (LSDV) estimator.

Any transformation that deletes the fixed effect produces a fixed effects estimator.

For instance, pre multiplying a $T \times 1$ vector by a $T \times (T-1)$ matrix

$$F = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

produces a $(T-1) \times 1$ vector of first differences.

If we pre multiply the model by F, we get rid of fixed effect and then we can consistently estimate β by applying OLS.

Again

$$\bar{y}_{i.} = \bar{X}_{i.}\beta + \mu_i + \bar{\eta}_{i.}$$

The model in deviation form is

$$y_{it} - \bar{y}_{i.} = (X_{it} - \bar{X}_{i.})\beta + (\eta_{it} - \bar{\eta}_{i.})$$

which satisfies the orthogonality condition. OLS can be applied to produce consistent estimators.

First differences or differencing from person-specific means produce consistent (and unbiased) estimators of β.

5.6.2. Steps to Implement fixed effects estimator are:

⇒Transform all the variable by subtracting person-specific means

$\Rightarrow$Run OLS on transformed variables

The matrix of standard errors of fixed effects estimators is

$\sigma_\eta^2 (X' Q_D X)^{-1}$

The estimator of σ_η^2 is $\hat{\sigma}_\eta^2 = \frac{\hat{u}_W' \hat{u}_W}{NT-N-K}$, where $\hat{u}_W$ is within OLS residual.

Testing for Fixed Effects

We assume $\sum_{i=1}^{N} \mu_i = 0$

For testing $H_0: \mu_1 = \mu_2 = \cdots = \mu_{N-1} = 0$, we apply Chow test and obtain

Restricted regression SS (RRSS) by using pooled model.

Unrestricted regression SS (URSS) by using within estimator.

The test statistic is

$$F = \frac{(RRSS - URSS)/(N-1)}{URSS/(NT-N-K)}$$

$\sim F(N-1, N(T-1)-K)$ under H_0.

5.6.3. Random Effects Model:

Random effect model is given by

$y_{it} = \alpha + X_{it}\beta + u_{it}$

where

$u_{it} = \mu_i + \eta_{it}$

Here μ_i is uncorrelated with X_{it}.

We assume that

$$E(\eta) = 0,\ E(\eta\eta') = \sigma_\eta^2 I_N,\ E(\mu_i) = 0,\ E(\mu_i \mu_j) = 0\ \forall\ i \neq j,\ E(\mu_i^2) = \sigma_\mu^2 I_N,\ E(\mu_i \eta_{jt}) = 0$$

Hence

$$E(u_i u_i') = \sigma_\eta^2 I_T + \sigma_\mu^2 l_T l_T' = \begin{pmatrix} \sigma_\eta^2 + \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\ \vdots & \ddots & \vdots \\ \sigma_\mu^2 & \cdots & \sigma_\eta^2 + \sigma_\mu^2 \end{pmatrix} \qquad = \Sigma \text{ (say)}$$

where l_T is a $T \times 1$ vector with all elements 1.

Hence $\quad E(uu') = I_N \otimes \Sigma = \Omega$ (say)

Using the result

$$(A + aa')^{-1} = A^{-1} - \frac{1}{1 + a' A^{-1} a} A^{-1} a a' A^{-1}$$

We have

$$\Sigma^{-1} = \left(\sigma_\eta^2 I_T + \sigma_\mu^2 l_T l_T'\right)^{-1} = \frac{1}{\sigma_\eta^2}\left[I_T - \frac{1-\theta^2}{T} l_T l_T'\right]$$

Thus

$$\Sigma^{-\frac{1}{2}} = \frac{1}{\sigma_\eta}\left[I_T - \left(\frac{1-\theta}{T} l_T l_T'\right)\right]; \theta = \left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + T\sigma_\mu^2}\right)^{\frac{1}{2}} = \left(\frac{\sigma_\eta^2}{T\sigma_B^2}\right)^{\frac{1}{2}}, \sigma_B^2 = \frac{1}{T}\sigma_\eta^2 + \sigma_\mu^2$$

FGLS estimator of $\beta \Rightarrow$ We need to estimate σ_η^2 and σ_μ^2.

5.6.4. Random effects as a combination of within and between estimators:

Consider the i^{th} equation

$$y_i = X_i \beta + u_i$$

Then

$$\frac{1}{T} l_T' y_i = \frac{1}{T} \sum_{i=1}^{T} y_{it} = \bar{y}_{i.}$$

Similarly, we define $\bar{X}_{i.}$

Define a $NT \times N$ matrix D of N dummy variables

$$D = \begin{pmatrix} l_T & 0 & \dots & 0 \\ 0 & l_T & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & l_T \end{pmatrix} = I_N \otimes l_T$$

$P_D = D(D'D)^{-1}D'$: $NT \times NT$ symmetric, idempotent matrix

Pre multiplying the model by P_D transforms the data into the means and the model becomes

$$P_D y = P_D X\beta + error$$

Applying OLS to above model, we get the estimator

$$\hat{\beta}_B = (X' P_D X)^{-1} X' P_D y$$

This estimator is called the between estimator or Wald estimator.

The between estimator is consistent but not efficient.

If T is large, this estimator is robust to measurement errors in X variables, provided orthogonality condition μ_i uncorrelated with X_{it} is satisfies for the correct data.

The information thrown away by the between estimator can be used to construct following within estimator

Let $Q_D = I_{NT} - P_D = I_{NT} - D(D'D)^{-1}D'$,

Q_D: Symmetric idempotent matrix

Pre multiplying by Q_D, we obtain

$$Q_D y = Q_D X\beta + Q_D u$$

$Q_D y$: Residual when we run regression between y and dummy variables D

$Q_D X$: Residual when we run regression between X and D

Thus, pre multiplying by Q_D transforms the data into residuals from auxiliary regressions of all the variables on a complete set of individual specific constants.

Predicted value from such a regression is the individual specific means. Thus, residuals are deviations from personal specific means.

$$Q_D y = y - \frac{1}{T}(I_N \otimes J_T)y = y - \begin{pmatrix} \bar{y}_1 l_T \\ \vdots \\ \bar{y}_N l_T \end{pmatrix}$$

Pre multiplying the model by Q_D and applying OLS leads to the within estimator

$$\hat{\beta}_W = (X' Q_D X)^{-1} X' Q_D y$$

Estimator $\hat{\beta}_W$ is obtained by running OLS on the following equation:

$$y_{it} - \bar{y}_{i.} = (X_{it} - \bar{X}_{i.})\beta + u_{W,it}$$

The estimator is called the within estimator because it uses only variation within each cross-section unit.

The estimator is consistent but not efficient as it uses N unnecessary extra variables. We can write pooled OLS estimator as weighted sum of between and within estimators:

$$\hat{\beta} = (X'X)^{-1} X' y = (X'X)^{-1}[X' P_D y + X' Q_D y]$$

$$= (X'X)^{-1}(X' P_D X)\hat{\beta}_B + (X'X)^{-1}(X' Q_D X)\hat{\beta}_W$$

Where $\hat{\beta}_B = (X' P_D X)^{-1}(X' P_D y), \hat{\beta}_W = (X' Q_D X)^{-1}(X' Q_D y)$

The pooled OLS estimator is consistent but inefficient as it does not incorporate information about heteroscedasticity resulting from repeated observations from same cross section units.

5.6.5. Estimation of σ_η^2 σ_B^2 and σ_μ^2:

Estimators of σ_η^2 σ_B^2 and σ_μ^2 based on standard ANOVA are

$$\hat{\sigma}_\eta^2 = \frac{1}{NT - NK - N} \hat{u}_W' \hat{u}_W;$$

$$\hat{\sigma}_B^2 = \frac{1}{N-K}\hat{u}_B'\hat{u}_B;$$

$$\hat{\sigma}_\mu^2 = \hat{\sigma}_B^2 - \frac{\hat{\sigma}_\eta^2}{T}.$$

$\hat{u}_W$: Residuals from the within regression

$\hat{u}_B$: Residuals from the between regression

Degrees of freedom for $\hat{u}_W'\hat{u}_W$ is

$$NT - (NK - N) = N(T - K - 1)$$

as the i^{th} equation has K explanatory variables, but $\forall i = 1, \dots, N \ \sum_{t=1}^{T} \hat{u}_{W,it} = 0$. These estimators are consistent estimators of corresponding variances. Then

$$\hat{\theta} = \left(\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\eta^2 + T\hat{\sigma}_\mu^2}\right)^{\frac{1}{2}}.$$

5.6.6. Steps for estimating random effects panel data model:

i. Compute within and between estimators

ii. Compute corresponding residuals and use residuals to estimate variance terms $\hat{\sigma}_\eta^2$ and $\hat{\sigma}_\mu^2$.

iii. Obtain

$$\hat{\theta} = \left(\frac{\hat{\sigma}_\eta^2}{\hat{\sigma}_\eta^2 + T\hat{\sigma}_\mu^2}\right)^{1/2}$$

iv. Run OLS between transformed variables $\tilde{y}$ and $\tilde{X}$ where

$$\tilde{y}_{it} = y_{it} - \bar{y}_{i.} + \hat{\theta}\bar{y}_{i.} \ ; \ \tilde{X}_{it} = X_{it} - \bar{X}_{i.} + \hat{\theta}\bar{X}_{i.}$$

5.6.7. GLS Estimation:

Consider the model

$$y = \alpha l_{NT} + X\beta + u, \;\; u = D\mu + \eta$$

Where

$$D = I_N \otimes l_T \, , Q_D = I_{NT} - P_D \, , P_D = D(D'D)^{-1}D' = I_N \otimes \bar{J}_T$$

$$\bar{J}_T = \frac{1}{T} J_T = \frac{1}{T} l_T l_T' \, ,$$

$J_T : T \times T$ with all elements 1

$$\bar{J}_T = \frac{1}{NT} J_{NT} = \frac{1}{NT} l_{NT} l_{NT}'$$

We consider

Within Model:

$$Q_D y = Q_D X\beta + Q_D \eta \qquad (5.7)$$

Between model:

$$\sqrt{T}(\bar{y}_{i.} - \bar{y}_{..}) = \sqrt{T}(\bar{X}_{i.} - \bar{X}_{..})\beta + \sqrt{T}(\bar{u}_{i.} - \bar{u}_{..}), \forall i \qquad (5.8)$$

For each i, equation is repeated T times. So, we multiply by $\sqrt{T}$. We can write (5.8) as

$$(P_D - \bar{J}_{NT})y = (P_D - \bar{J}_{NT})X\beta + (P_D - \bar{J}_{NT})u \qquad (5.9)$$

$$Var(Q_D u) = Var(Q_D \eta) = \sigma_\eta^2 Q_D$$

$$Var[(P_D - \bar{J}_{NT})u] = \sigma_1^2 (P_D - \bar{J}_{NT})$$

$$\sigma_1^2 = T\sigma_B^2 = T\sigma_\mu^2 + \sigma_\eta^2$$

Combining (5.7) and (5.9), we get

$$\begin{pmatrix} Q_D y \\ (P_D - \bar{J}_{NT})y \end{pmatrix} = \begin{pmatrix} Q_D X \\ (P_D - \bar{J}_{NT})X \end{pmatrix} \beta + \begin{pmatrix} Q_D u \\ (P_D - \bar{J}_{NT})u \end{pmatrix}$$

Covariance matrix of $\begin{pmatrix} Q_D u \\ (P_D - \bar{J}_{NT})u \end{pmatrix}$ is

$$\begin{pmatrix} \sigma_\eta^2 Q_D & 0 \\ 0 & \sigma_1^2 (P_D - \bar{J}_{NT}) \end{pmatrix}$$

Applying GLS to (5.9), we obtain

$$\hat{\beta}_{GLS} = \left[\frac{1}{\sigma_\eta^2} X' Q_D X + \frac{1}{\sigma_1^2} X' (P_D - \bar{J}_{NT}) X\right]^{-1} \times \left[\frac{1}{\sigma_\eta^2} X' Q_D y + \frac{1}{\sigma_1^2} X' (P_D - \bar{J}_{NT}) y\right]$$

$$= [W_{XX} + \theta^2 B_{XX}]^{-1} [W_{Xy} + \theta^2 B_{Xy}]$$

$$= W_1 \hat{\beta}_{WI} + W_2 \hat{\beta}_{BW}$$

where

$W_{XX} = X' Q_D X,$

$W_{Xy} = X' Q_D y$

$B_{XX} = X' (P_D - \bar{J}_{NT}) X,$

$B_{Xy} = X' (P_D - \bar{J}_{NT}) y$

$W_1 = [W_{XX} + \theta^2 B_{XX}]^{-1} W_{XX}$

$W_2 = [W_{XX} + \theta^2 B_{XX}]^{-1} \theta^2 B_{XX} = I - W_1 \qquad (\because W_1 + W_2 = I)$

$\hat{\beta}_{WI} = W_{XX}^{-1} W_{Xy}$

$\hat{\beta}_{BW} = B_{XX}^{-1} B_{Xy}$

For $\sigma_\mu^2 = 0, \theta = 0$ and $\hat{\beta}_{GLS}$ reduces to the OLS estimator

$$b_{OLS} = W_{XX}^{-1} W_{Xy} = (X' Q_D X)^{-1} X' Q_D y$$

5.7. Self-Assessment Exercise

19. Explain the concept of nonspherical disturbances and why they pose a challenge in econometric modelling.

20. What are the key differences between heteroskedasticity and autocorrelation in the context of regression analysis?

21. Explain the purpose of Generalized Least Squares (GLS) and how it improves upon Ordinary Least Squares (OLS) in the presence of nonspherical disturbances.

22. How does the Seemingly Unrelated Regression (SUR) model improve efficiency in estimation? Provide a real-world example.

23. Compare and contrast the Fixed Effects (FE) and Random Effects (RE) models in panel data analysis.

24. Describe a scenario where panel data models are more suitable than cross-sectional or time-series models.

25. A researcher is analysing two interrelated equations for household spending on food (Y_1) and clothing (Y_2), where the error terms of the two equations are correlated. Explain how a SUR model can be used in this situation, and outline the steps for estimation.

26. Given panel data on the annual income of individuals across 10 years, propose a suitable econometric model to examine the impact of education and experience on income. Justify your choice and discuss how to handle unobserved heterogeneity.

27. Suppose a dataset exhibits heteroskedasticity in the error terms. Show how GLS can be applied to transform the model and estimate the parameters efficiently.

5.8. Summary

This unit introduces three advanced econometric models commonly used in analyzing complex economic relationships. The focus is on addressing limitations of basic regression models, particularly in cases involving nonspherical disturbances, interrelated systems, or combined cross-sectional and time-series data.

1. Models with Nonspherical Disturbances

- Explores situations where the assumptions of homoscedasticity and no autocorrelation in error terms are violated.
- Discusses techniques such as Generalized Least Squares (GLS) and Feasible GLS (FGLS) for addressing these issues.
- Applications: Time-series models with serial correlation, cross-sectional models with heteroskedasticity.

2. Seemingly Unrelated Regression (SUR) Models

- Focuses on systems of multiple regression equations with correlated error terms across equations.
- Highlights efficiency gains from jointly estimating the equations using GLS or MLE.
- Applications: Sectoral studies, consumption patterns, or interdependent economic models.

3. Panel Data Models

- Combines cross-sectional and time-series data to analyze entities over time.
- Introduces key estimation techniques including:
    i. Pooled OLS for simplicity.
    ii. Fixed Effects (FE) for controlling unobserved, entity-specific characteristics.
    iii. Random Effects (RE) for cases where individual effects are random and uncorrelated with regressors.

5.9. References

- Gujarati, D. and Guneshker, S. (2007): Basic Econometrics, 4th Ed., McGraw Hill Companies.
- Johnston, J. and Dinardo, J. (1997): Econometric Methods, 2nd Ed., McGraw Hill International.
- Baltagi, B. H. (2021). Econometrics of Panel Data: Methods and Applications. 6th Edition. Springer.
- Srivastava, V. K., & Giles, D. E. (1987). Seemingly Unrelated Regression Equations Models: Estimation and Inference. New York: Marcel Dekker.

5.10. Further Readings

- Hsiao, C. (2022). *Analysis of Panel Data*. 4th Edition. Cambridge University Press.
- Judge, G.G., Griffiths, W.E., Hill Lütkepohl, R.C. and Lee, T. C. (1985): The theory and practice of econometrics, Wiley.
- Koutsoyiannis, A. (2004): Theory of Econometrics, 2 Ed., Palgrave Macmillan Limited.
- Maddala, G.S. and Lahiri, K. (2009): Introduction to Econometrics, 4 Ed., John Wiley & Sons.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 303N /MASTAT – 303N

Econometrics

Block: 2 Simultaneous Equation Models and Forecasting

Unit – 6 : Structural and Reduced from of the Model and Identification Problem

Unit – 7 : Estimators in Simultaneous Equation Models - I

Unit – 8 : Estimators in Simultaneous Equation Models - I

Unit – 9 : Forecasting

Unit – 10 : Instrumental Variable Estimation

Course Design Committee

Dr. Ashutosh Gupta — **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi — **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha — **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey — **Member**
Department of Statistics, D. D. U. Gorakhpur University, Gorakhpur.

Prof. Shruti — **Member-Secretary**
Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Sandeep Mishra — **Writer**
Department of Statistics (***Unit 6, 7, 8, 9 & 10***)
Shahid Rajguru College of Applied Sciences for Women
University of Delhi, New Dehi

Prof. Anoop Chaturvedi — **Editor**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Shruti — **Course Coordinator**
School of Sciences,
U. P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 303N/ MASTAT – 303N **ECONOMETRICS**
©UPRTOU
First Edition: ***July 2024***
ISBN :

Printed By:

Block & Units Introduction

The present SLM on ***Econometrics*** consists of fourteen units with three blocks.

The ***Block – 2 - Simultaneous Equations Models and Forecasting,*** is the second block, which is divided into five units.

The ***Unit – 6 - Structural and reduced form of the model and identification problem,*** deals with the Simultaneous equations model, concept of structural and reduced forms, problem of identification, rank and order conditions of identifiability.

The ***Unit – 7 - Estimators in Simultaneous Equation Models – I,*** deals with the Limited and full information estimators, indirect least squares estimators, two stage least squares estimators, three stage least squares estimators and k class estimator.

The ***Unit – 8 - Estimators in Simultaneous Equation Models – I,*** deals with the Limited information maximum likelihood estimation, full information maximum likelihood estimation, prediction, and simultaneous confidence interval.

The ***Unit – 9 - Forecasting,*** deals with the Forecasting, exponential and adaptive smoothing methods, periodogram and correlogram analysis.

The ***Unit – 10 - Instrumental Variable Estimation,*** deals with the Review of GLM, analysis of GLM and generalized leased square estimation, Instrumental variables, estimation, consistency properties, asymptotic variance of instrumental variable estimators.

At the end of every unit the summary, self-assessment questions and further readings are given.

UNIT 6 STRUCTURAL AND REDUCED FORM OF THE MODEL AND IDENTIFICATION PROBLEM

Structure

6.1 Introduction

In econometrics, models are typically categorized into structural and reduced-form models based on their complexity and the relationships they describe:

1. *Structural Models:* These models explicitly specify the relationships between variables based on economic theory. They are grounded in economic principles and attempt to capture the underlying mechanisms driving economic phenomena. Structural models often involve a system of equations that describe how different variables interact. They aim to uncover causal

relationships and are typically used for policy analysis and theoretical exploration. For examples include models of consumer behavior, investment decisions, production functions, etc.

2. *Reduced-Form Models*: These models focus on describing the statistical relationships between variables without explicit reference to underlying economic theory. Reduced-form models are derived from structural models and emphasize empirical relationships rather than theoretical foundations. They are commonly used for empirical analysis, forecasting, and to test specific hypotheses. For example, regression models where variables are regressed against each other without a clear theoretical model.

In econometrics, the identification problem refers to the ability to uniquely determine the parameters of a model based on the available data. It arises when the model does not have enough information to estimate all the parameters separately and unambiguously. Identifiability is crucial because, without it, the estimates of model parameters can be biased, inconsistent, or even meaningless. The identification problem in econometrics underscores the importance of careful model specification and appropriate estimation techniques to ensure that the estimated parameters are reliable and meaningful for analysis and policy-making.

6.2 Objectives

After completing this course, there should be a clear understanding of:

- Simultaneous equations model
- Concept of structural and reduced forms
- Problem of identification
- Rank and order conditions of identifiability

6.3 Simultaneous Equations Model: Introduction

A simultaneous equations model (SEM) is a type of econometric model that consists of multiple equations where each equation depends on the endogenous variables (variables that are determined within the model) of the other equations. This interdependence distinguishes

simultaneous equations model from single-equation model, such as simple regression models. The main features of the simultaneous equations model are:

1. Endogeneity: Variables in a simultaneous equations model are endogenous, meaning they are jointly determined within the model rather than being exogenously given.

2. System of Equations: In a system of equations (SEM), each equation represents a relationship between variables.

Example: In a basic economic model, you might have equations for demand and supply that interact to determine equilibrium price and quantity.

3. Simultaneity Bias: Simultaneous equations models can suffer from simultaneity bias, where the estimation of coefficients in one equation is biased due to endogeneity with respect to variables in other equations. Special techniques like instrumental variables or two-stage least squares are often used to address simultaneity bias.

4. Identification Issues: Identification refers to the ability to separately estimate the parameters of the model equations. Simultaneous equations models require careful consideration of identification to ensure reliable estimation.

Linear regression model involves a single equation which explains a dependent variable (y) in terms of a set of independent or explanatory variables $(x's)$ and then the relationship is unidirectional, *i.e.*, $x's$ explain y but y do not explain $x's$. This is called one way causality. In many economic theories, we usually built upon a system of relationships and in this kind of relationship we normally find variables which determine each other.

Example: In microeconomic theory for a particular commodity, there is a demand equation and a supply equation both involving price and the phenomenon is explained by the system of demand-supply equations. The price and quantity are interdependent and determined by the interaction of demand-supply equations.

6.3.1 Simultaneous equation models

Model is in the form of a *set of linear simultaneous equations*. The system is jointly determined by the set of equations in the system. A particular equation explaining a

dependent variable may involve the dependent variables of other equations among the explanatory variables. When a relationship is a part of a system, some of the explanatory variables are stochastic and correlated with the disturbances. The assumption that the explanatory variable and disturbance are uncorrelated is not satisfied leading to inconsistent OLS.

How to overcome this problem and estimate the model?

1) Variables Classification in Single Equation Linear Model
    i. Dependent Variable
    ii. Independent Variables
2) Variables Classification in Simultaneous Equations Model
    i. Endogenous variables
    ii. Exogenous variables

6.3.1.1 Endogenous variables or jointly determined variables

The variables, which are explained by the functioning of system, are known as endogenous variables. The values are determined by the simultaneous interaction of the relations in the model. Endogenous variables are those variables that are determined within the model itself, rather than being exogenously given. These variables are typically the main variables of interest whose values are simultaneously determined by the equations in the model.

Example: In an economic system, price and quantity in a market might be endogenous variables since the equilibrium price and quantity are determined by the interaction of supply and demand equations.

Thus, endogenous variables in a simultaneous equations model are the variables whose values are determined by the interactions and relationships specified within the model itself, making them central to the analysis of the system's behaviour and outcomes.

6.3.1.2 Exogenous variables

Exogenous variables provide explanations for the endogenous variables. The values are determined from outside the model. These variables help is explaining the variations in endogenous variables and influence the endogenous variables but are not influenced by them. Exogenous variables are independent variables in the model whose values are not determined by the relationships specified within the model. They are external to the system being modeled and are often considered as constants or variables whose values are given from outside the model.

Example: In economic models, exogenous variables could include factors such as government policies, external shocks to the economy (like changes in international prices), or other variables that are assumed to be outside the control of the model and whose values are taken as given.

6.3.1.3 Predetermined variables

Exogenous variables and lagged endogenous variables form predetermined variables. Since exogenous variables are predetermined, so they are independent of disturbance term in the model. These variables satisfy the assumption that explanatory variables satisfy in the usual regression model.

6.3.2 Reduced Form

- Economic model involving several endogenous variables in each equation is called the *structural form* of the model.
- If we transform the structural form so that each equation has one endogenous variable as a function of only exogenous and lagged endogenous variables, the new form is called the *reduced form*.
- The structural form cannot be estimated using least squares as it includes endogenous variables on its right-hand side.

In econometrics, the reduced form of a model refers to an equation that expresses the endogenous variables (variables influenced by the model's parameters) solely in terms of exogenous variables (variables not influenced by the model's parameters) and error terms. In econometrics, reducing the model to its reduced form is useful for understanding the

relationships between variables without directly dealing with the complexities of the underlying structural model. It simplifies analysis and interpretation, especially in cases where the structural form involves multiple equations or complex interactions. The reduced form of a model typically refers to a simplified version that retains essential elements while omitting less significant details. In economics, for instance, it often involves eliminating endogenous variables by expressing them in terms of exogenous variables. This simplification aids in analysis and understanding by focusing on the core relationships within the model. The reduced form can be estimated by least squares.

Example: Let us consider the following wage and price function

$$\left.\begin{aligned}\text{Wage Equation: } W_t &= \alpha_0 + \alpha_1 U_t + \alpha_2 P_t + u_{1t},\\ \text{Price Equation: } P_t &= \beta_0 + \beta_1 W_t + \beta_2 R_t + \beta_3 M_t + u_{2t}\end{aligned}\right\} \quad (1)$$

where,

$\left.\begin{aligned}W_t &= \text{Rate of change in money wage}\\ P_t &= \text{Rate of change in prices}\end{aligned}\right\}$; both are Endogenous Variables.

$U_t = \%$ Unemployment rate, $M_t =$ Supply of money,

$R_t =$ Rate of change in cost of capital

$u_{1t}, u_{2t} =$ Stochastic disturbances

W and P are jointly dependent and may be correlated with the stochastic disturbance terms.

Reduces Form: We can write

$$W_t = \alpha_0 + \alpha_1 U_t + \alpha_2(\beta_0 + \beta_1 W_t + \beta_2 R_t + \beta_3 M_t + u_{2t}) + u_{1t}$$

or

$$W_t = \pi_{10} + \pi_{11} U_t + \pi_{12} R_t + \pi_{13} M_t + v_{1t}$$

$$\pi_{10} = \frac{\alpha_0 + \alpha_2\beta_0}{1 - \alpha_2\beta_1}, \pi_{11} = \frac{\alpha_1}{1 - \alpha_2\beta_1},$$

$$\pi_{12} = \frac{\alpha_2\beta_2}{1 - \alpha_2\beta_1}, \pi_{13} = \frac{\alpha_2\beta_3}{1 - \alpha_2\beta_1},$$

$$v_{1t} = \frac{u_{1t} + \alpha_2 u_{2t}}{1 - \alpha_2\beta_1}$$

Similarly

$$P_t = \pi_{20} + \pi_{21}U_t + \pi_{22}R_t + \pi_{23}M_t + v_{2t}$$

$$\pi_{20} = \frac{\beta_0 + \alpha_0\beta_1}{1 - \alpha_2\beta_1}, \pi_{21} = \frac{\alpha_1\beta_1}{1 - \alpha_2\beta_1},$$

$$\pi_{22} = \frac{\beta_2}{1 - \alpha_2\beta_1}, \pi_{23} = \frac{\beta_3}{1 - \alpha_2\beta_1},$$

$$v_{2t} = \frac{\beta_1 u_{1t} + u_{2t}}{1 - \alpha_2\beta_1}$$

Is it possible to estimate the reduced form coefficients and then with the help of those estimates, estimate the structural form coefficients?

Here, we discuss some examples to explain different problems that arise in the analysis of a simultaneous equation model.

Example: Consider the following consumption function with a national income identity:

$$\left.\begin{array}{ll} (i) & C_t = \alpha + \beta Y_t + u_t \\ (ii) & Y_t = C_t + Z_t \end{array}\right\} \qquad (2)$$

where, *C* = aggregate consumption expenditure; *Y*= national income

Z = non-consumption expenditure; *u* = disturbance term

The model determines the consumption expenditure and the national income, which are jointly dependent or *endogenous* variables. The model contains a (i) *behavioral equation* and (ii) *equilibrium condition.* The equilibrium conditions have no disturbance term and are exact. Non-consumption expenditure is determined outside the model, which makes it *exogeneous* variable. We assume

i. $u \sim N(0, \sigma_u^2 I_n)$ where $u = (u_1, \ldots, u_n)'$ and

ii. Z and u are independent.

Solving (2) for *C* and *Y*, we get the *reduced form*

$$\left.\begin{aligned} C_t &= \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} Z_t + v_t \\ Y_t &= \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} Z_t + v_t \end{aligned}\right\} \qquad (3)$$

where

$$v_t = \frac{1}{1-\beta} u_t \sim N\left(0, \frac{\sigma_u^2}{(1-\beta)^2}\right).$$

Now

$$plim\left(\frac{1}{n}\sum Y_t v_t\right)$$
$$= plim\left[\frac{1}{n}\sum\left\{\frac{\alpha}{1-\beta} + \frac{1}{1-\beta} Z_t + v_t\right\} v_t\right]$$
$$= plim\left(\frac{1}{n}\sum v_t^2\right)$$
$$= \frac{\sigma_u^2}{(1-\beta)^2} \neq 0.$$

OLS estimates of α, β obtained from (2) would be inconsistent. So, we require some alternative estimation procedures.

6.4 Explaining estimation procedures using example

Selecting a model should be based on strong theoretical justification rather than just convenience of statistical estimation. Each model should be evaluated independently, considering the random fluctuations that may be present. It is not possible to apply OLS directly; instead, strategies for handling the estimation issue must be developed. There are

several other estimation techniques available, however some of them could need a lot of work. Some methods are:

6.4.1 Instrumental variable (I V) estimation

Let *Z* be uncorrelated with *u* and correlated with *Y*. Applying IV estimator using *Z* as the IV leads to

$$\left.\begin{aligned} a_{IV} &= \bar{C} - b_{IV}\bar{Y} \\ b_{IV} &= \frac{\sum(C_t - \bar{C})(Z_t - \bar{Z})}{\sum(Y_t - \bar{Y})(Z_t - \bar{Z})} = \frac{\sum c_t z_t}{\sum y_t z_t} \end{aligned}\right\} \qquad (4)$$

where $c_t = C_t - \bar{C}, y_t = Y_t - \bar{Y},\ z_t = Z_t - \bar{Z}$.

Using (4), we have

$$\sum c_t z_t = \sum \left(\frac{\beta}{1-\beta} z_t + v_t\right) z_t$$

$$= \frac{\beta}{1-\beta}\sum z_t^2 + \sum v_t z_t$$

$$\sum y_t z_t = \sum \left(\frac{1}{1-\beta} z_t + v_t\right) z_t$$

$$= \frac{1}{1-\beta}\sum z_t^2 + \sum v_t z_t$$

Hence

$$plim\left(\frac{1}{n}\sum c_t z_t\right) = \frac{\beta}{1-\beta} m_{zz},$$

$$m_{zz} = plim\left(\frac{1}{n}\sum z_t^2\right)$$

$$plim\left(\frac{1}{n}\sum y_t z_t\right) = \frac{1}{1-\beta} m_{zz}.$$

Thus

$$plim(b_{IV}) = \beta$$

$$plim(a_{IV}) = plim\left[\frac{1}{n}\sum(C_t - b_{IV}Y_t)\right]$$

$$= plim\left[\frac{1}{n}\sum\{\alpha - (b_{IV} - \beta)Y_t + u_t\}\right]$$

$$= \alpha$$

6.4.2 Indirect least squares (ILS)

Apply OLS to reduced form equations, which satisfy the conditions under which OLS estimators are consistent and BLUE.

$$\left.\begin{aligned} C_t &= \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta}Z_t + v_t \\ Y_t &= \frac{\alpha}{1-\beta} + \frac{1}{1-\beta}Z_t + v_t \end{aligned}\right\} \qquad (5)$$

$$\frac{\sum c_t z_t}{\sum z_t^2}: \text{consistent estimator of } \frac{\beta}{(1-\beta)}$$

$$\frac{\sum y_t z_t}{\sum z_t^2}: \text{consistent estimator of} \frac{1}{(1-\beta)}$$

Hence the estimator of β is the ratio

$$b_{ILS} = \frac{\sum c_t z_t}{\sum z_t^2} \div \frac{\sum y_t z_t}{\sum z_t^2}$$

$$= \frac{\sum c_t z_t}{\sum y_t z_t}.$$

In ILS, reduced form coefficients are estimated by OLS and then structural coefficients are estimated by an appropriate transformation of the estimates of reduced form coefficients.

6.4.3 Two-stage least squares estimation (2SLS)

We first regress y on the exogenous variable. Writing second eq. in (3) in deviation form we have

$$y_t = \delta z_t + v_t; \qquad \delta = \frac{1}{1-\beta} \tag{6}$$

Apply OLS to (6)

$$\hat{\delta} = \frac{\sum y_t z_t}{\sum z_t^2}$$

Hence, an estimated value of y_t is

$$\begin{aligned} \hat{y}_t &= \hat{\delta} z_t \\ &= \left(\frac{\sum y_t z_t}{\sum z_t^2}\right) z_t \\ &= \left(\delta + \frac{\sum v_t z_t}{\sum z_t^2}\right) z_t \end{aligned} \tag{7}$$

Then

$$\sum \hat{y}_t u_t = \delta \sum z_t u_t + \sum u_t z_t \frac{\sum v_t z_t}{\sum z_t^2}$$

Since,

$$plim\left(\frac{1}{n}\sum z_t u_t\right) = plim\left(\frac{1}{n}\sum z_t v_t\right) = 0,$$

we have

$$plim\left(\frac{1}{n}\sum \hat{y}_t u_t\right) = 0.$$

We write the first eq. of (2) as

$$C_t = \alpha + \beta \hat{Y}_t + \left[u_t + \beta\left(Y_t - \hat{Y}_t\right)\right] \tag{8}$$

Since $\hat{Y}_t$ has zero correlation (in limit) with u_t and $\left(Y_t - \hat{Y}_t\right), \hat{Y}_t$ has zero correlation with the combined disturbance term $\left[u_t + \beta\left(Y_t - \hat{Y}_t\right)\right]$ of (8).

Applying OLS to (8) gives the following 2SLS estimate of β, which is a consistent estimator:

$$b_{2SLS} = \frac{\sum c_t \hat{y}_t}{\sum \hat{y}_t^2}$$
$$= \frac{\hat{\delta} \sum c_t z_t}{\hat{\delta}^2 \sum z_t^2}$$
$$= \frac{\sum c_t z_t}{\sum y_t z_t}.$$

In this example

$$b_{IV} = b_{ILS} = b_{2sls}.$$

The reason is that both the equations are just identified.

6.5 General form of the Simultaneous Equation model

Let us consider a model containing M structural equations.

$$\left.\begin{array}{l} \beta_{11}y_{1t} + \cdots + \beta_{1M}y_{Mt} + \gamma_{11}x_{1t} + \cdots + \gamma_{1K}x_{Kt} = u_{1t} \\ \beta_{21}y_{1t} + \cdots + \beta_{2M}y_{Mt} + \gamma_{21}x_{1t} + \cdots + \gamma_{2K}x_{Kt} = u_{2t} \\ \vdots \\ \beta_{M1}y_{1t} + \cdots + \beta_{MM}y_{Mt} + \gamma_{M1}x_{1t} + \cdots + \gamma_{MK}x_{Kt} = u_{Mt} \end{array}\right\} (t = 1, \ldots\ldots, n) \qquad (9)$$

where,

$y_{1t}, \cdots, y_{Mt}$: Observations on endogenous variables

$x_{1t}, \ldots, x_{Kt}$: Observations on predetermined (exogenous and lagged endogenous) variables

$\beta_{i1}, \ldots, \beta_{iM}, \gamma_{i1}, \ldots, \gamma_{iK}$: Regression coefficients

u_{it}: disturbance term $(i = 1, \ldots\ldots, M), (t = 1, \ldots\ldots, n)$

In equation (9) some of the $\beta's$ and $\gamma's$ are zero otherwise all the equation in the model look alike and no equation could be identified.

We can write the model in matrix form as

$$By_t + \Gamma x_t = u_t;\ t = 1,..,n \qquad (10)$$

$$B = \begin{pmatrix} \beta_{1t} & \cdots & \beta_{1M} \\ \vdots & \ddots & \vdots \\ \beta_{M1} & \cdots & \beta_{MM} \end{pmatrix}, \Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1K} \\ \vdots & \ddots & \vdots \\ \gamma_{M1} & \cdots & \gamma_{MK} \end{pmatrix}$$

$$y_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{Mt} \end{pmatrix}, \qquad x_t = \begin{pmatrix} x_{1t} \\ \vdots \\ x_{Kt} \end{pmatrix}, \qquad u_t = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{Mt} \end{pmatrix}$$

Suppose B is nonsingular. The reduced form of the model is

$$y_t = \Pi x_t + v_t, t = 1,..,n \qquad (11)$$

where $\Pi = -B^{-1}\Gamma, v_t = B^{-1}u_t$.

Note:

Structural form relations: The interaction between endogenous and predetermined variables taking place inside the model.

Reduced form relation: It express as jointly dependent (endogenous) variables as linear combination of predetermined variables.

6.6 Identification Problem

In econometrics, the identification problem refers to the ability to uniquely determine the parameters of a model based on the available data. It arises when the model does not have enough information to estimate all the parameters separately and unambiguously. Identifiability is crucial because without it, the estimates of model parameters can be biased, inconsistent, or even meaningless.

The identification problem is whether the estimates of the structural parameters of the model can be obtained from the estimates of the reduced form coefficients?

It has three possibilities:

(i) Unique estimates of the structural parameters of a particular equation can be obtained. The equation is said to be *exactly (or fully or just) identified*.

(ii) More than one estimates of the structural parameters can be obtained. The equation is said to be *over identified*.

(iii) It is not possible to estimate of the structural parameters by using the estimates of the reduced form coefficients. The equation is said to be *un-identified* or *under identified*.

Example: Under identified

Consider the demand supply model,

Demand Function (DF): $Q_t^d = \alpha_0 + \alpha_1 P_t + u_{1t};\ \alpha_1 < 0 \quad (12)$

Supply Function (SF): $Q_t^s = \beta_0 + \beta_1 P_t + u_{2t};\ \beta_1 > 0 \quad (13)$

Equilibrium Condition (EC): $Q_t^d = Q_t^s \quad (14)$

Q_t^d = quantity demanded, Q_t^s = quantity supplied, t = time

By EC

$$\alpha_0 + \alpha_1 P_t + u_{1t} = \beta_0 + \beta_1 P_t + u_{2t}$$

This gives the equilibrium price,

$$P_t = \pi_0 + v_t \quad (15)$$

where

$$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1},$$

$$v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$$

Substituting P_t in DF or SF gives the equilibrium quantity

$$Q_t = \pi_1 + w_t \quad (16)$$

where

$$\pi_1 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1},$$

$$w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

Equations (15) & (16) are reduced form equations.

Using (15) and (16) we can estimate π_0 and π_1.

We cannot estimate four structural parameters ($\alpha_0, \beta_0, \alpha_1, \beta_1$) from the estimates of two reduced form coefficients.

Multiplying DF (12) by λ and SF (13) by (1-λ), (0<λ<1) and adding we get,

$$Q_t = \gamma_0 + \gamma_1 p_t + w_t \qquad (17)$$

where

$$\gamma_0 = \alpha_0\lambda + \beta_0(1-\lambda) \gamma_1 = \alpha_1\lambda + \beta_1(1-\lambda)$$
$$w_t = \lambda u_{1t} + (1-\lambda)u_{2t}$$

Equation (17) involves the regression of Q_t and P_t.

It is observationally equivalent or indistinguishable from (12) and (13).

For a given set of data on (Q_t, P_t) we are unable to say which one of these models we are fitting.

Example: Just or exact identification

$$DF: Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t}, (\alpha_1 < 0, \alpha_2 > 0) \qquad (18)$$

$$\text{SF: } Q_t^s = \beta_0 + \beta_1 P_t + u_{2t}, (\beta_1 > 0) \qquad (19)$$

I_t: income of a consumer

Using the equilibrium condition $Q_t^d = Q_t^s$, we get equilibrium price and equilibrium quantity

$$P_t = \pi_0 + \pi_1 I_t + v_t \quad (20)$$

$$Q_t = \pi_2 + \pi_3 I_t + w_t \quad (21)$$

where

$$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}, \quad \pi_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1},$$

$$v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}, \pi_2 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1},$$

$$\pi_3 = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1},$$

$$w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

We can obtain OLS estimates of $\pi_0, \pi_1, \pi_2, \pi_3$.

We cannot get the unique solutions for five structural parameters $(\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1)$. However

$$\beta_0 = \pi_2 - \frac{\pi_3}{\pi_1}\pi_0, \ \beta_1 = \frac{\pi_3}{\pi_1}.$$

SF can be identified but DF remains unidentified.

Multiplying (20) by λ and (21) by $(1 - \lambda)$ and adding gives

$$Q_t = \gamma_0 + \gamma_1 P_t + \gamma_2 I_t + \varepsilon_t \quad (22)$$

where

$$\gamma_0 = \alpha_0\lambda + \beta_0(1 - \lambda),$$

$$\gamma_1 = \alpha_1\lambda + \beta_1(1 - \lambda),$$

$$\gamma_2 = \lambda\alpha_2, \varepsilon_t = \lambda u_{1t} + (1 - \lambda)u_{2t}$$

(22) is distinguishable from SF but indistinguishable from DF.

Consider

$$DF: Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} \ (\alpha_1 < 0, \alpha_2 > 0) \quad (23)$$

$$SF: Q_t^S = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} \quad (\beta_1 > 0, \beta_2 < 0) \quad (24)$$

P_{t-1}: Price lagged one period which is a predetermined variable at time t.

Equilibrium price and demand are

$$P_t = \pi_0 + \pi_1 I_t + \pi_2 P_{t-1} + v_t \quad (25)$$

$$Q_t = \pi_3 + \pi_4 I_t + \pi_5 P_{t-1} + w_t \quad (26)$$

where

$$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}, \ \pi_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1},$$

$$\pi_2 = \frac{\beta_2}{\alpha_1 - \beta_1} \pi_3 = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1},$$

$$\pi_4 = -\frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1}, \pi_5 = \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1}$$

$$v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}, w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

$$\lambda Q_t^d + (1 - \lambda) Q_t^S$$
$$= \lambda(\alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t}) + (1 - \lambda)(\beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t})$$
$$= \gamma_0 + \gamma_1 P_t + \gamma_2 I_t + \gamma_3 P_{t-1} + \varepsilon_t \quad (27)$$

where

$$\gamma_0 = \alpha_0 \lambda + \beta_0 (1 - \lambda),$$
$$\gamma_1 = \alpha_1 \lambda + \beta_1 (1 - \lambda),$$
$$\gamma_2 = \lambda \alpha_2,$$
$$\gamma_3 = (1 - \lambda) \beta_2,$$
$$\varepsilon_t = \lambda u_t + (1 - \lambda) u_t.$$

Then equation (27) is distinguishable from both the DF (23) and SF (24). Both the DF and SF are just identifiable.

Example: Over identification

$$DF: Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t}; (\alpha_1 < 0, \alpha_2, \alpha_3 > 0) \quad (28)$$

$$SF: Q_t^S = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}; (\beta_1 > 0, \beta_2 < 0) \quad (29)$$

R_t: Wealth of the consumer

Equilibrium price and quantity are obtained as

$$P_t = \pi_0 + \pi_1 I_t + \pi_2 R_t + \pi_3 P_{t-1} + v_t \quad (30)$$

$$Q_t = \pi_4 + \pi_5 I_t + \pi_6 R_t + \pi_7 P_{t-1} + w_t \quad (31)$$

where

$$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}, \pi_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1},$$

$$\pi_2 = -\frac{\alpha_3}{\alpha_1 - \beta_1} \; \pi_3 = \frac{\beta_2}{\alpha_1 - \beta_1},$$

$$\pi_4 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}, \pi_5 = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1},$$

$$\pi_6 = -\frac{\alpha_3\beta_1}{\alpha_1 - \beta_1}, \pi_7 = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}$$

$$v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}, w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

DF and SF have seven structural parameters. There are eight reduced form coefficients to estimate them. Hence unique solutions for all the structural parameters are not possible. For instance,

$$\beta_1 = \frac{\pi_6}{\pi_2} \quad \text{and} \quad \beta_1 = \frac{\pi_5}{\pi_1}$$

This is the case of *Over Identification*. DF & SF both are distinguishable.

Let us summarize the above possibilities:

(i) Under identifiable or unidentified:

The estimation of parameters is not at all possible in this case. No enough estimates are available for structural parameters. *If more than one theory is consistent with the same data, the theories are observationally equivalent in the sense that we cannot distinguish them.*

(ii) *Exactly identifiable*:

The estimation of parameters is possible in this case. The relationship between the reduced form coefficients and structural parameters is one to one. Thus, the OLSE of reduced form coefficients lead to unique estimates of structural coefficients.

(iii) *Over identifiable*:

The estimation of parameters in this case is possible. The OLS estimator of reduced form coefficients leads to multiple estimates of structural coefficients.

Notice that the identification problem is not the sampling problem. We can not overcome this problem by increasing the sample size.

In the reduced form

$$y_t = \Pi x_t + v_t, t = 1, .., n;$$
$$\Pi = -B^{-1}\Gamma$$

we can consistently estimate Π using OLS. If B is known, we can simply estimate Γ using the relation $\Gamma = -\mathrm{B}\Pi$.

Our problem come to the estimation of B. This makes sense because if B is known, we can write $By_t = z_t$ (say) and apply OLS to

$$z_t = -\Gamma x_t + u_t$$

and estimate Γ.

We can put the identification problem as:

We can observe the reduced form and must be able to deduce the structural form from the reduced form. If more than one structural form leads to the same reduced form and we are not sure which structure we are estimating. Thus, we cannot estimate the structure.

6.6.1 Structural form of the model

Consider the structural form of the model

$$\text{Model } S: By_t + \Gamma x_t = u_t;\ t = 1,..,n \qquad (32)$$

where,

$B: M \times M, \qquad \Gamma: M \times K,$

$u_t: M \times 1, \qquad y_t: M \times 1, \qquad x_t: K \times 1$

and $u_t's:$ *iid* following $N(0,\Sigma)$. The reduced form of the model is

$$y_t = \Pi x_t + v_t \qquad (33)$$

Here

$\Pi = -B^{-1}\Gamma, \qquad v_t = B^{-1}u_t$

and

$v_t's$ are *iid* following $N(0,\Omega)$ with $\Omega = B^{-1}\Sigma B'^{-1}$.

Further

$E(u_t) = 0,$
$E(u_t u_t') = \Sigma, E(u_t u_l') = 0\ \forall\ t \neq l.$

$E(v_t) = 0,$
$E(v_t v_t') = \Omega, E(v_t v_l') = 0\ \forall t \neq l.$

Now, we can conclude the following points:

(1) We can consistently estimate Π using OLS.

(2) If B is known, we can estimate Γ using the relation $\Gamma = -B\Pi$. This makes sense because if B is known, we can write $By_t = z_t$ (say) and apply OLS to $z_t = -\Gamma x_t + u_t$ and estimate Γ.

(3) The problem come to the estimation of B. For this, we must be able to deduce the structural form from the reduced form. If more than one structural form led to the same reduced form, then we are not sure which structure we are estimating.

6.6.2 Identification Problem and Likelihood Function

The identification problem is directly related to the likelihood function. If a model is not identified, the likelihood function does not have a unique maximum because multiple parameter values produce the same likelihood. Therefore, for a model to be estimable (using MLE or other methods), it must be identified, meaning the likelihood function should have a unique maximum at the true parameter values.

Consider the joint *pdf* of y_t given x_t is

$$\begin{aligned} p(y_t|x_t) &= p(v_t) \\ &= p(u_t)\left|\frac{\partial u_t}{\partial v_t}\right| \\ &= p(u_t)|\det(B)| \end{aligned}$$

where, $|\det(B)|$ denotes the absolute value of determinant of B.

Notice that,

$$u_t = Bv_t$$

$$\Rightarrow \frac{\partial u_t}{\partial v_t} = B$$

where,

$\left|\frac{\partial u_t}{\partial v_t}\right| = |\det(B)|$ is the Jacobean of the transformation.

The likelihood function corresponding to S is given by

$$L = p(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n)$$

$$= \prod_{t=1}^{n} p(y_t | x_t)$$

$$= |det(B)|^n \prod_{t=1}^{n} p(u_t). \qquad (34)$$

Applying a nonsingular linear transformation on structural equations (32) with a nonsingular matrix D we get

$$DBy_t + D\Gamma x_t = Du_t$$

Writing

$$B^* = DB, \Gamma^* = D\Gamma, u_t^* = Du_t,$$

we obtain

$$\text{Model } S^*: B^* y_t + \Gamma^* x_t = u_t^*, t = 1, \dots, n \qquad (35)$$

and

$$E(u_t^* u_t^{*\prime}) = D\Sigma D' = \Sigma^* \text{ (say)}$$

The reduced form for S^* is also $(33), i.e.,$

$$y_t = \Pi x_t + v_t.$$

Thus

S: (B, Γ, Σ) is the true structure and

S^*: $(B^*, \Gamma^*, \Sigma^*)$ is the structure after nonsingular transformation.

Both the structures have the same reduced form. Statistically we cannot distinguish them and both are observationally equivalent.

We find $p(y_t|x_t)$ with S^* as follows:

$$p(y_t|x_t)$$
$$= p(u_t^*)\left|\frac{\partial u_t^*}{\partial v_t}\right|$$
$$= p(u_t^*)|det(B^*)|.$$

Also

$$u_t^* = Du_t$$
$$\Rightarrow \frac{\partial u_t^*}{\partial u_t} = D.$$

Thus

$$p(y_t|x_t)$$
$$= p(u_t)\left|\frac{\partial u_t}{\partial u_t^*}\right||det(B^*)|$$
$$= p(u_t)|det(D^{-1})||det(DB)|$$
$$= p(u_t)|det(B)|.$$

The LF corresponding to S^* is

$$L^* = |det(B^*)|^n \prod_{t=1}^{n} p(u_t^*)$$
$$= |det(D^*)|^n |det(B)|^n \prod_{t=1}^{n} p(u_t)\left|\frac{\partial u_t}{\partial u_t^*}\right|$$
$$= |det(D^*)|^n |det(B)|^n \prod_{t=1}^{n} p(u_t)\,|det(D^{-1})|$$
$$= L.$$

Both the structural forms S and S^* have the same LF. LF forms the basis of statistical analysis, both S and S^* have same implications. One cannot identify whether the LF corresponds to S and S^*. So, we are unable to distinguish whether we are estimating true structure S or transformed structure S^*. In this sense, both S and S^* are observationally equivalent.

Remark:

1. *A parameter is said to be identifiable within the model if the parameter has the same value for all equivalent structures contained in the model.*

2. *A structural equation is identifiable if all the parameters in structural equation are identifiable.*

For a given structure, we can find many observationally equivalent structures by non-singular linear transformation. Then the presence and/or absence of certain variables help in the identifiability. So, a priori restrictions on B and Γ may help in the identification of parameters. These a priori restrictions may arise from various sources like economic theory. For the identification of an equation, some of the variables must be excluded from the equation which are present elsewhere in the model. This is known as *exclusion (of variables) criterion or zero restriction criterion*.

6.6.3 Condition for Identification

1. *Order condition*:

"A necessary condition for an equation to be identified is that it must exclude at least M-1 endogenous and predetermined variables."

- If it excludes exactly M-1 variables, the equation is just identified.

- If it excludes more than M-1 variables, it is over identified.

"Equivalently for an equation to be identified, the number of predetermined variables excluded from the equation must not be less than the number of endogenous variables included in that equation less one."

Let

k: Number of predetermined variables included in the equation

m: Number of endogenous variables included in that equation.

Then it must exclude at least M-1 endogenous and predetermined variables which results in $(K-k)+(M-m) \geq M-1$.

Equivalently, if $K-k \geq m-1$.

Thus

Number of predetermined variables excluded from the equation $\geq$ *Number of endogenous variables included in that equation less one.*

This gives

- If $K-k=m-1$, the equation is just identified.
- If $K-k>m-1$, then the equation is over identified.

Example 1:

Let

$$DF \quad Q_t = \alpha_0 + \alpha_1 P_t + u_{1t}$$
$$SF \quad Q_t = \beta_0 + \beta_1 P_t + u_{2t}$$

Here,

Number of endogenous variables $(M) = 2$

Number of predetermined variables $(K) = 1$

Each of these equations must exclude $M-1=1$ variables. Since this does not hold for any of two equations, neither equation is identified.

Example 2:

Let

$DF \quad Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t}$

$SF \quad Q_t = \beta_0 + \beta_1 P_t + u_{2t}$

Here $M = 2, K = 2$. SF is just identified as it excludes $M - 1 = 1$ variable but DF is unidentified.

Example 3:

Let

$DF \quad Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t}$

$SF \quad Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}$

Here $M = 2, K = 3$. Since each equation excludes exactly $M - 1 = 1$ variable, both DF and SF are just identified.

Example 4:

Let

$DF \quad Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t}$

$SF \quad Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}$

Here $M = 2, K = 4$.

1, I_t, R_t, P_{t-1}: predetermined variables

DF excludes $M - 1 = 1$ variables, so it is just identified and

SF excludes $2 (> M - 1 = 1)$ variables, thus it is over identified.

2. *Rank condition of identifiability:*

The Order condition is the necessary but not sufficient condition for the identification of an equation. An equation may be unidentified even if the order condition is satisfied. The Rank condition stated below is the sufficient condition for the identification of an equation.

"In a simultaneous equations model containing M equations in M endogenous variables, an equation is identified if and only if there exist at least one non-zero determinant of order (M-1) from the coefficients of endogenous and predetermined variables excluded from that particular equation but included in the other equations of the model".

Note: The rank condition states whether the equation is identified or not. From the order condition we know whether the equation is just or over identified.
Example 1:

$(i)\ y_{1t} - \beta_{10} - \beta_{12}y_{2t} - \beta_{13}y_{3t} - \gamma_{11}x_{1t} = u_{1t}$

$(ii)\ y_{2t} - \beta_{20} - \beta_{23}y_{3t} - \gamma_{21}x_{1t} - \gamma_{22}x_{2t} = u_{2t}$

$(iii)\ y_{3t} - \beta_{30} - \beta_{31}y_{1t} - \gamma_{31}x_{1t} - \gamma_{32}x_{2t} = u_{3t}$

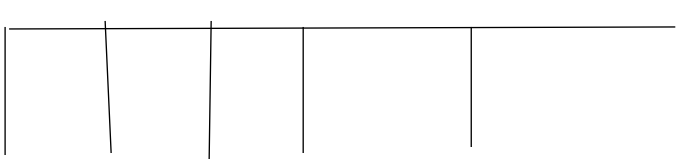
$(iv) y_{4t} - \beta_{40} - \beta_{41}y_{1t} - \beta_{42}y_{2t} - \gamma_{43}x_{3t} = u_{4t}$

Here, $M = 4, K = 4$ from each equation, $M - 1 = 3$ variables are excluded. Hence by order condition each of these equations is just identified.

To check the rank condition, write the system in the tabular form

$eq.\ no.$	1	y_1	y_2	y_3	y_4	x_1	x_2	x_3
i	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
ii	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
iii	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
iv	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

Then, strike out the columns corresponding to nonzero coefficients of the first equation. Also, strike out the coefficients of the row in which the first equation appears.



$eq.\,no.$	1	y_1	y_2	y_3	y_4	x_1	x_2	x_3
i	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
ii	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
iii	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
iv	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

The remaining entries will give the following matrix:

$$A = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix}$$

By rank condition, the first equation is identified if there exist a determinant of order M-1=3 formed from the entries of A. Since $|A| = 0$, no non-zero determinant of order 3 exists. Hence the equation is unidentified.

For the eq. (ii)

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -\beta_{31} & 0 & 0 \\ -\beta_{41} & 1 & -\gamma_{43} \end{bmatrix}$$

Again, no nonzero determinant of order 3 exists and eq. (ii) is under identified.

For the eq. (iii)

$$A = \begin{bmatrix} -\beta_{12} & 0 & 0 \\ 1 & 0 & 0 \\ -\beta_{42} & 1 & -\gamma_{43} \end{bmatrix}$$

$|A| = 0$ and eq. (iii) is unidentified.

For the Eq. (iv), $|A| \neq 0$, thus we have a 3×3 non zero matrix. Hence equation (iv) is identified.

Example 2: (y's are endogenous, $x's$ are exogenous)

$(i)\ y_{1t} + \beta_{12}y_{2t} + \gamma_{11}x_{1t} = u_{1t}$

(ii) $y_{2t} + \beta_{21}y_{1t} + \gamma_{22}x_{2t} + \gamma_{23}x_{3t} = u_{2t}$

Order condition: M=2, K=3

In equation (i) m=2, k=1. Hence K-k=2>m-1=1, so that the first equation is over identified.

In the equation (ii) m=2, k=2. Hence K-k=1=m-1=1, so the equation (ii) is just identified.

Rank condition:

Eq. (i) $A = [\gamma_{22} \quad \gamma_{23}]$

We can form a nonzero determinant from A of order M-1=1. So, the equation (i) is (over) identified.

For $\text{Eq. (ii)}\ A = [\gamma_{11}]$

So, equation (ii) is also (just) identified.

Example 3: Keynesian model of income determination

Let us consider the following Keynesian model

$Consumption\ function\ (CF): C_t = \beta_1 + \beta_2 Y_t - \beta_3 T_t + u_{1t}$

Investment function $(IF): I_t = \alpha_0 + \alpha_1 Y_{t-1} + u_{2t}$

Taxation function $(TF): T_t = \gamma_0 + \gamma_1 Y_t + u_{3t}$

Income identity $(II): Y_t = C_t + I_t + G_t$

where, C: consumption, I: investment, T: taxes, Y: Income, G: Government expenditure

and

C_t, I_t, T_t, Y_t: endogenous variables

G_t, Y_{t-1}: predetermined variables

This model includes an income identity. The identities do not raise any identification problem as the coefficients are known.

Here, M=4, K=3

Order condition:

CF: m=3, k=1, K-k=2, m-1=2, just identified

IF: m=1, k=2, K-k=1, m-1=0 over identified

TF: m=2, k=1, K-k=2, m-1=1 over identified

Rank Condition

$eq.no.$	C_t	I_t	T	Y_t	G_t	Y_{t-1}	1
CF	1	0	β_3	$-\beta_2$	0	0	$-\beta_1$
IF	0	1	0	0	0	$-\alpha_1$	$-\alpha_0$
TF	0	0	1	$-\gamma_1$	0	0	$-\gamma_0$
II	-1	-1	0	1	-1	0	0

For *CF*:

CF: $A = \begin{bmatrix} 1 & 0 & -\alpha_1 \\ 0 & 0 & 0 \\ -1 & -1 & 0 \end{bmatrix}$

We cannot form a nonzero determinant of order M-1=3. Thus, CF is unidentified.

For *IF*:

IF: $A = \begin{bmatrix} 1 & \beta_3 & -\beta_2 & 0 \\ 0 & 1 & -\gamma_1 & 0 \\ -1 & 0 & 1 & -1 \end{bmatrix}$

$\begin{vmatrix} 1 & \beta_3 & -\beta_2 \\ 0 & 1 & -\gamma_1 \\ -1 & 0 & 1 \end{vmatrix} = 1 + \gamma_1\beta_3 - \beta_2 \neq 0$. So, IF is identified.

For *TF*:

TF: $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -\alpha_1 \\ -1 & -1 & -1 & 0 \end{bmatrix}$

$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \end{vmatrix} = -1 \neq 0$. So, TF is identified.

So, according to order conditions, first equation is just identified and the second and third is over identified. But, from rank condition first equation is unidentified, so our result is the first equation is unidentified. From rank condition second and third equation is identified, if we combine it with the order condition the results, we get is second and third equation is over identified.

3. *Determination of Rank condition*:

Let us consider the following model,

Model: $By_t + \Gamma x_t = u_t;\ t = 1, .., n$ (36)

where,

$B: M \times M, \quad \Gamma: M \times K, \quad u_t: M \times 1, \qquad y_t: M \times 1, \quad x_t: K \times 1$

or $Az_t = u_t, A = (B \quad \Gamma), z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}$ (37)

We consider the identification of first equation.

Suppose the first structural equation is

$\alpha_1 z_t = u_{1t}$ (38)

α_1 is the first row of A. The element of α_1 have some linear restrictions. The most common restrictions are exclusion restrictions.

Suppose a priori restrictions are expressed as

$\alpha_1 \Phi = 0$ (39)

where, Φ is (M+K)×R matrix of R restrictions on elements of α_1.

In addition to restrictions (39), we have restrictions on α_1 arising from the relation between structural and reduced form coefficients

$$B\Pi + \Gamma = 0 \text{ or } \quad AW = 0 \qquad (40)$$

where, $W = \begin{pmatrix} \Pi \\ I_K \end{pmatrix}$

Thus, restrictions on α_1 are

$$\alpha_1 W = 0 \qquad (41)$$

Combining (39) and (41), we get

$$\alpha_1 [W \quad \Phi] = 0 \qquad (42)$$

α_1 has M+K unknowns. Further $[W \quad \Phi]$ is of order (M+K)×(K+R).

Hence (42) has (K+R) equations in (M+K) unknowns.

The first element of α_1 is 1 ($\beta_{11} = 1$). Hence to determine α_1 uniquely, the condition is

$\rho[W \quad \Phi] = M + K - 1$, so it is the case of Exact identification.

If $\rho[W \quad \Phi] > M + K - 1$, we get more than one solution for α_1, *i.e.*, the equation is over identified.

The rank condition cannot hold if $[W \quad \Phi]$ does not have at least $M + K - 1$ columns. Thus, a necessary condition for rank condition to satisfy is

$K + R \geq M + K - 1$

or $R \geq M - 1$

The necessary order condition for exclusion restrictions is

"The no. of variables excluded from the equation should be greater than or equal to no. of equations in the model less 1". i.e.

$$(K-k)+(M-m)\geq M-1$$

or $K-k\geq m-1$

Result:

Let $A=(B \quad \Gamma)$. Then

$$\rho(\Phi \quad W)=\rho(A\Phi)+K \quad (43)$$

Proof: The model is

$$By_t+\Gamma x_t=u_t$$

The relation between structural parameters and reduced form parameters is given by

$$B\Pi+\Gamma=0$$

or $AW=0$,

$$W=\begin{pmatrix}\Pi\\ I_K\end{pmatrix}$$

For the first equation the restriction due to relation between structural and reduced form parameters is

$$\alpha_1 W=0 \qquad (44)$$

We can write

$$A=(B \quad \Gamma)$$

$$=B(I_M \quad -\Pi)$$

$$[A' \quad W]=\begin{pmatrix}I_M & \Pi\\ -\Pi' & I_K\end{pmatrix}\begin{pmatrix}B' & 0\\ 0 & I_K\end{pmatrix}$$

For any (M+K)×1 non-null vector $x=\begin{pmatrix}x_1\\ x_2\end{pmatrix}$

$$x'\begin{pmatrix} I & \Pi \\ -\Pi' & I_K \end{pmatrix} x = x_1'x_1 + x_2'x_2 = x'x > 0$$

Thus $\begin{pmatrix} I & \Pi \\ -\Pi' & I_K \end{pmatrix}$ is positive definite and thus non-singular. Further $\begin{vmatrix} B' & 0 \\ 0 & I_K \end{vmatrix} = |B| \neq 0$.

$\begin{pmatrix} B' & 0 \\ 0 & I_K \end{pmatrix}$ is also non-singular.

It follows that $[A' \quad W]_{M+K\times M+K}$ is also non-singular.

Hence, we can write Φ as

$$\Phi = (A' \quad W)\begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$$

$$= A'S_1 + WS_2$$

where S_1 is M×R and S_2 is K×R.

Since AW=0, we have $A\Phi = AA'S_1$.

AA' is a M×M positive definite matrix, and hence non-singular so that $\rho(A\Phi) = \rho(S_1)$.

Thus $\rho(AA') = \rho(A') = M$

Also

$$(\Phi \quad W) = (A' \quad W)\begin{pmatrix} S_1 & 0 \\ S_2 & I_K \end{pmatrix}$$

So that

$$\rho(\Phi \quad W)$$

$$= \rho\begin{pmatrix} S_1 & 0 \\ S_2 & I_K \end{pmatrix}$$

$$= \rho(S_1) + K$$

$$= \rho(A\Phi) + K$$

Hence, we get the required result■

Remark: *The above result leads to the following equivalent rank condition for identification:*

An equation is identified if

$\rho(A\Phi) \geq M - 1.$

Example: let us consider the model of the form

$y_{1t} + \beta_{12}y_{2t} + \gamma_{11}x_{1t} = u_{1t}$

$y_{2t} + \beta_{21}y_{1t} + \gamma_{22}x_{2t} + \gamma_{23}x_{3t} = u_{2t}$

Now

$$A = \begin{pmatrix} 1 & \beta_{12} & \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \beta_{21} & 1 & \gamma_{21} & \gamma_{22} & \gamma_{23} \end{pmatrix}$$

For the first equation

$$\Phi = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$\alpha_1\Phi = 0$ gives the restrictions $\gamma_{12} = 0, \gamma_{13} = 0.$ Now

$$A\Phi = \begin{pmatrix} 0 & 0 \\ \gamma_{22} & \gamma_{23} \end{pmatrix}$$

Then $\rho(A\Phi) = 1(= M - 1)$ and the equation is identified.

For the second equation

$$\Phi = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$\alpha_2\Phi = 0$ gives the restrictions $\gamma_{21} = 0.$ Now

$$A\Phi = \begin{pmatrix}\gamma_{11}\\ \gamma_{21}\end{pmatrix}$$

Then $\rho(A\Phi) = 1(= M - 1)$ and the equation is identified.

1.6.4 Identification from Reduced form

Order condition is same as that for structural form:

$$(K - k) \geq m - 1$$

Rank Condition from the reduced form: We have

$$\rho(A\Phi) = \rho(B(I_M \ - \Pi)\Phi) = \rho\big((I_M \ - \Pi)\Phi\big)$$

Thus, rank condition is

$$\rho\big((I_M \ - \Pi)\Phi\big) \ \geq M - 1.$$

Rank Condition for Exclusive Restrictions:

"An equation containing m endogenous variables is identified if and only if it is possible to construct a non-zero determinant of order $m - 1$ from the reduced form coefficients of exogenous (predetermined) variables excluded from that particular equation."

Steps:

1. Strike out the rows corresponding to endogenous variables excluded from that particular equation being examined for identifiability.

2. Strike out columns referring to exogenous variables included in the structural form excluded from the equation.

3. We are left with the reduced form coefficients excluded from the structural equation.

4. If the order of largest non-zero determinant is m-1, the equation is identified. If it is less than m-1, equation is un identified.

Example:

Consider the given Model:

$$y_1 = 3y_2 - 2x_1 + x_2 + u_1$$

$$y_2 = y_3 + x_3 + u_2$$

$$y_3 = y_1 - y_2 - 2x_3 + u_3$$

The reduced form of the above model is

$$y_1 = 4x_1 - 2x_2 + 3x_3 + v_1$$

$$y_2 = 2x_1 - x_2 + x_3 + v_2$$

$$y_3 = 2x_1 - x_2 + v_3$$

Equation (i) (m=2): Table of reduced form coefficients

	x_1	x_2	x_3
y_1	4	-2	**3**
y_2	2	-1	**1**
y_3	2	-1	0

Excluded endogenous variables: y_3 (delete 3rd row)

Included exogenous variables: x_1, x_2 (delete 1st & 2nd columns).

Table of π's of excluded exogenous variables and included endogenous variables is $\binom{3}{1}$. We can form a nonzero determinant of order 1(=m-1), the equation is just identified.

For equation (ii) (m=2)

	x_1	x_2	x_3
y_1	4	-2	3
y_2	**2**	**-1**	1
y_3	**2**	**-1**	0

Excluded endogenous variables: y_1 (delete 1st row)

Included exogenous variables: x_3 (delete 3rd column)

Table of π's of excluded exogenous variables and included endogenous variables is $\begin{pmatrix} 2 & -1 \\ 2 & -1 \end{pmatrix}$. Highest order non-zero determinant is of order 1. The equation is identified. By order condition verify that equation is over identified.

For the third equation (iii): $m = 3, m - 1 = 2$.

Table of π's of excluded exogenous variables and included endogenous variables is $\begin{pmatrix} 4 & -2 \\ 2 & -1 \\ 2 & -1 \end{pmatrix}$. Highest order non-zero determinant is of order 1. The equation is unidentified.

1.7 Self-Assessment Exercise

1. Write general form of simultaneous equations model and, using the likelihood function of the model, explain the problem of identification. Deduce rank and order conditions for identification.

2. Examine the identifiability of each of the equations in the following simultaneous equations model.

$$y_{1t} = a_0 + a_1 y_{2t} + a_2 x_{1t} + u_{1t}$$
$$y_{2t} = b_0 + b_1 y_{1t-1} + u_{2t}$$

3. For the following simultaneous equation model check the identifiability of both the equations.

Demand Model: $Q_t^D = \alpha_0 + \alpha_1 P_t + \alpha_2 ln C_t + u_{1t}$

Supply Model: $Q_t^S = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{1t}$

Equilibrium: $Q_t^D = Q_t^S$.

Here Q_t and P_t are endogeneous variables.

4. Verify if the second equation of following simultaneous equations model is exactly identified, unidentified or over identified:

$$y_{1t} = a_0 + a_1 y_{2t} + a_2 x_{1t} + a_3 x_{2t} + u_{1t}$$
$$y_{2t} = b_0 + b_1 y_{1t-1} + b_2 x_{2t} + b_3 y_{3t} + u_{2t}$$
$$y_{3t} = c_0 + c_1 y_{1t} + c_2 x_{1t} + u_{3t}$$

5. For the simultaneous equations model involving M equations and K predetermined variables $By_t+\Gamma x_t=u_t$, if $\Pi=B^{-1}\Gamma$, $A=[B \quad \Gamma]$, $W = \begin{bmatrix} \Pi \\ I_K \end{bmatrix}$, and the first column of A, say α_1 follows R restrictions of the form $\alpha_1 \Phi = 0$. Then prove that

$$rank(\Phi \quad W) = rank(A\Phi) + K$$

6. With an example, elaborate the identification of reduced form equations.

6.8 Summary

This unit delves into simultaneous equations models, which are used to analyze systems of interdependent relationships among multiple variables. It begins by introducing the concept of simultaneity in econometric modelling, distinguishing these models from single-equation frameworks. The unit explores the identification problem, emphasizing the conditions under which structural parameters can be uniquely determined, with a focus on rank and order conditions.

After the completion of this unit, you will be able to understand the concept of Simultaneous equations model and structural and reduced forms of the model. Also, you have a clear understanding of problem of identification, rank, and order conditions of identifiability

6.9 References

- Baltagi, Badi H. (2021): Econometric Analysis of Panel Data, Springer.
- Gujarati, D. and Guneshker, S. (2007). *Basic Econometrics*, 4th Ed., McGraw Hill Companies.
- Johnston, J. and J. Dinardo (1997). *Econometric Methods*, 2nd Ed., McGraw Hill International.
- Srivastava, V.K. and Giles D.A.E. (1987): Seemingly unrelated regression equations models, Marcel Dekker.

6.10 Further Readings

- Judge, G.G., W.E. Griffiths, R.C. Hill Lütkepohl and T. C. Lee (1985). The theory and practice of econometrics, Wiley.
- Koutsoyiannis, A. (2004). *Theory of Econometrics*, 2 Ed., Palgrave Macmillan Limited.
- Maddala, G.S. and Lahiri, K. (2009). *Introduction to Econometrics*, 4 Ed., John Wiley & Sons.
- Ullah, A. and Vinod, H.D. (1981). Recent advances in Regression Methods, Marcel Dekker.

UNIT 7 ESTIMATORS IN SIMULTANEOUS EQUATION MODELS

7.1 Introduction

In simultaneous equations models (SEMs), multiple interdependent relationships are modeled, making ordinary least squares (OLS) estimation biased and inconsistent due to endogeneity issues. To address this, several specialized estimators are used:

1. Two-Stage Least Squares (2SLS)

2. Three-Stage Least Squares (3SLS)

3. Full Information Maximum Likelihood (FIML)

4. Limited Information Maximum Likelihood (LIML)

5. Instrumental Variables (IV) Estimator

6.Generalized Method of Moments (GMM)

Each of these estimators has its strengths and is chosen based on the specific characteristics of the SEM, the availability of instruments, and the degree of correlation among the errors in different equations.

Limited and full information estimators are two broad categories of estimation techniques used in the context of simultaneous equations models (SEMs). The choice between them depends on the level of information utilized in the estimation process.

1. Limited Information Estimators

Limited information estimators focus on estimating a single equation within the system without explicitly considering the entire system of equations. These estimators use information only from the specific equation being estimated, not from the full system.

a. Two-Stage Least Squares (2SLS): 2SLS is the most common limited information estimator. It is used when there is endogeneity in the equation being estimated, meaning some regressors are correlated with the error term.

b. Limited Information Maximum Likelihood (LIML): LIML is another limited information estimator that estimates parameters by maximizing the likelihood function, considering only the information available in a single equation.

2. Full Information Estimators

Full information estimators, on the other hand, consider the entire system of equations simultaneously. They make use of all available information in the system, leading to potentially more efficient estimates.

a. Three-Stage Least Squares (3SLS): 3SLS is a full information estimator that extends 2SLS by considering the contemporaneous correlation of errors across the different equations in the system.

b. Full Information Maximum Likelihood (FIML): FIML estimates all parameters of the SEM simultaneously by maximizing the likelihood function for the entire system.

The choice between limited and full information estimators depends on the specific context, the complexity of the system, the reliability of the model specification, and the goals of the analysis.

Indirect Least Squares (ILS) is an estimation technique used in the context of simultaneous equations models (SEMs), particularly in overidentified systems. The key idea behind ILS is to transform the simultaneous equations system into a reduced form, estimate the parameters of this reduced form, and then use these estimates to infer the structural parameters.

Indirect Least Squares is a method that first estimates the reduced form of a simultaneous equations model and then uses these estimates to infer the structural parameters. It is a useful technique under specific conditions, particularly when the model is overidentified and the reduced form is straightforward to estimate.

7.2 Objective

After going through this unit, you should be able to acquire the knowledge of

- Limited and full information estimators

- Indirect least squares estimators
- Two stage least squares estimators
- Three stage least squares estimators and k class estimator.

7.3 Recursive or Triangular Models

The OLS approach is inappropriate for the estimation of an equation in a system of simultaneous equations model due to the correlation between the endogenous explanatory variable(s) and the stochastic disturbance component. When incorrectly applied, the estimators exhibit not only bias (in small samples) but also inconsistent behaviour, meaning that the bias persists regardless of sample size. Nonetheless, there is one instance in which OLS—even in the context of simultaneous equations—can be used effectively. This is the situation with the causal, triangular, or recursive models. *Example:* Consider the three equations model,

$$y_{1t} = \gamma_{10} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + u_{1t}$$

$$y_{2t} = \gamma_{20} + \beta_{21}y_{1t} + \gamma_{21}x_{1t} + \gamma_{22}x_{2t} + u_{2t}$$

$$y_{3t} = \gamma_{30} + \beta_{31}y_{1t} + \beta_{32}y_{2t} + \gamma_{31}x_{1t} + \gamma_{32}x_{2t} + u_{3t}$$

where, $y's$ and $x's$ are endogenous and exogenous variables.

We assume that

$$E(u_{1t}) = E(u_{2t}) = E(u_{3t}) = 0;$$
$$cov(u_{1t}, u_{2t}) = cov(u_{1t}, u_{3t}) = cov(u_{2t}, u_{3t}) = 0.$$

that is, there is no correlation between the same-period disturbances in distinct equations. This is the assumption of zero contemporaneous correlation**.**

The matrix formed by the coefficients of the endogenous variables is the following triangular matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ -\beta_{21} & 1 & 0 \\ -\beta_{31} & -\beta_{32} & 1 \end{bmatrix}$$

Such models are called *recursive* or *triangular* models.

Now consider the first equation of the equation model, *Equation (i)* has only exogenous variables on RHS as explanatory variable, which are uncorrelated with u_{1t}. Thus OLS can be applied.

Next consider the second equation of the model, *Equation (ii)* contains endogenous variable y_{1t} as an explanatory variable. u_{1t}, which effects y_{1t}, is uncorrelated with u_{2t}, implying that $cov(y_{1t}, u_{2t}) = 0$. Thus, all the explanatory variables are uncorrelated with u_{2t} and OLS can be used.

Now consider *Equation (iii)* of the equation model,

$Cov(y_{1t}, u_{3t}) = Cov(y_{2t}, u_{3t}) = 0$. OLS can be applied.

Here y_1 effects y_2 but y_2 does not affect y_1 and y_1, y_2 influence y_3 but y_3 does not influence y_1 and y_2.

So, in the Recursive systems or model OLS can be applied to each equation separately.

An illustration of a recursive system would be the wage and price determination model presented below:

Price equation: $P_t + \gamma_{10} + \gamma_{11} w_{t-1} + \gamma_{12} R_t + \gamma_{13} L_t = u_{1t}$

Wage equation: $W_t + \gamma_{20} + \gamma_{24} U_t + \beta_{21} P_t = u_{2t}$

where P = rate of change of price per unit of employee

W = rate of change of wage per employee

R = rate of change of price of capital

L = rate of change of labor productivity

U = unemployment rate, %

We can write the model as

$$\begin{pmatrix} 1 & 0 \\ \beta_{21} & 1 \end{pmatrix}\begin{pmatrix} P_t \\ W_t \end{pmatrix} + \begin{pmatrix} \gamma_{10} & \gamma_{11} & \gamma_{12} & \gamma_{13} & 0 \\ \gamma_{20} & 0 & 0 & 0 & \gamma_{24} \end{pmatrix}\begin{pmatrix} 1 \\ w_{t-1} \\ R_t \\ L_t \\ U_t \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

where, $\begin{pmatrix} 1 & 0 \\ \beta_{21} & 1 \end{pmatrix}$ is Triangular matrix.

7.3.1 General Form of Recursive or Triangular Model

Let us consider the Model: $By_t + \Gamma x_t = u_t;\ t = 1, .., n$

If B is upper (or lower) triangular, then the system of equations is called *triangular* or *recursive.*

The model is of the form

$$y_{1t} = g_1(x_t) + u_{1t}$$

$$y_{2t} = g_2(y_{1t}, x_t) + u_{2t}$$

$$\vdots$$

$$y_{Mt} = g_M(y_{1t}, \cdots, y_{M-1t}, x_t) + u_{Mt}$$

$$g_j(y_{1t}, \cdots, y_{j-1t}, x_t) = \beta_{j1}y_{1t} + \cdots + \beta_{j,j-1}\, y_{j-1,t} + \Gamma x_t$$

Determination of variables is recursive in the sense that:

(i) y_1 affects y_2 but y_2 does not affect y_1,

(ii) y_1 and y_2 affect y_3 but y_3 does not affect y_1 and y_2

and so on.

and $B = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \beta_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{M1} & \beta_{M2} & \cdots & 1 \end{pmatrix}$: Matrix of coefficients of endogenous variables is a lower triangular matrix.

So, we can apply OLS to each equation separately.

7.4 Limited and Full Information Estimation

In limited information Estimation only the information specific to the equation under investigation is utilized in the estimation process. Limited information estimators are econometric tools used in situations where the available data or information is incomplete or limited. It is commonly used in empirical research, especially in fields like labor economics, industrial organization, and macroeconomics, where endogeneity is a concern, and the full system of equations is not always identifiable or estimable. When dealing with incomplete data or endogeneity issues, providing a way to obtain consistent and potentially more efficient estimates compared to traditional methods like OLS this method is used.

The Estimation Procedures are

(i) Indirect Least Squares (ILS),

(ii) Two Stage Least Squares (2SLS),

(iii)Limited Information Maximum Likelihood (LIML).

Whereas, in full Information Estimation information present in other equations and the fact that the structural disturbances of various equations may be correlated is utilized in the estimation process. Full information estimators are econometric techniques used to estimate parameters in a model where all the equations and the relationships between them are considered simultaneously. These estimators are particularly relevant in the context of systems of simultaneous equations, where multiple interdependent equations need to be estimated together rather than separately. They offer more efficient estimates compared to

limited information methods but require careful model specification and are computationally demanding.

This Estimation Procedures includes

(i) Three Stage Least Squares (3SLS),

(ii) Full Information Maximum Likelihood (FIML)

7.4.1 Estimation of a just Identified Equation: *Indirect Least Squares (ILS)*

The process of estimating the structural coefficients for a just or exactly identified structural equation using the OLS estimates of the reduced-form coefficients is called the indirect least squares (ILS) method, and the resulting estimates are called the indirect least squares estimates.

The following three steps are involved in ILS:

Step 1: Obtain reduced form equations. These reduced-form equations are derived from the structural equations in such a way that the only endogenous variable in each equation is the dependent variable, which depends only on the stochastic error term(s) and the predefined (exogenous or lagged endogenous) variables.

Step 2: Obtain OLS estimates of reduced form coefficients. We treat each of the reduced-form equations separately using OLS. Because the explanatory variables in these equations are predetermined and hence uncorrelated with the stochastic disturbances, this operation is allowed. Thus, consistent estimations are achieved.

Step 3: Since the equation is just identified, there is one to one correspondence between the structural and reduced form coefficients. Obtain the estimates of structural coefficients form the estimated reduced form coefficients.

Example: Let us consider the demand and supply model:

Demand Function: $Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_{1t}$ (i)

Supply Function: $Q_t = \beta_0 + \beta_1 P_t + u_{2t}$ (ii)

where,

Q: Quantity, (Endogenous)

P: Price, (Endogenous)

X: Income.

Assume that P is endogenous variable and X is exogeneous variable. As we know that, the supply function is exactly identified whereas the demand function is not identified.

Rearranging the term of the equation (i) and (ii), we obtain the reduced form equations:

$$P_t = \pi_0 + \pi_1 X_t + v_t$$

$$Q_t = \pi_2 + \pi_3 X_t + w_t$$

where, $\pi's$ are reduced form parameter and

$$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1},$$

$$\pi_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1},$$

$$\pi_2 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1},$$

$$\pi_3 = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}.$$

It is important to note that every reduced-form equation only has one endogenous variable, the dependent variable, and that it depends only on the exogenous variable X (income) and its stochastic disturbances. Therefore, OLS may be used to estimate the parameters of the previous reduced-form equations. These OLS estimates of $\pi_0, \pi_1, \pi_2, \pi_3$ are

$$\hat{\pi}_1 = \frac{\sum p_t x_t}{\sum x_t^2},$$

$$\hat{\pi}_0 = \bar{P} - \hat{\pi}_1 \bar{X}$$

$$\hat{\pi}_3 = \frac{\sum q_t x_t}{\sum x_t^2},$$

$$\hat{\pi}_2 = \bar{Q} - \hat{\pi}_3 \bar{X},$$

$$x_t = X_t - \bar{X} \ ,$$

$$p_t = P_t - \bar{P} \ ,$$

$$q_t = Q_t - \bar{Q}$$

ILS estimators of β_0 and β_1 are

$$\hat{\beta}_0 = \hat{\pi}_2 - \hat{\beta}_1 \hat{\pi}_0,$$

$$\hat{\beta}_1 = \frac{\hat{\pi}_3}{\hat{\pi}_1}$$

Note that DF is unidentified and cannot be estimated.

General Form:

Consider the general structural model at time t:

$$By_t + \Gamma x_t = u_t; t = 1,2, \dots, n. \qquad (1)$$

where, $Y = \begin{pmatrix} y_1' \\ \vdots \\ y_n' \end{pmatrix}$: $n \times M$ matrix of observations on endogenous variables.

$X = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}$: $n \times K$ matrix of observations on predetermined variables.

$U = \begin{pmatrix} u_1' \\ \vdots \\ u_n' \end{pmatrix}$: $n \times M$ matrix of disturbances.

We may write (1) as

$$YB' + X\Gamma' = U \qquad (2)$$

Suppose we are interested in estimating the first equation is

$$y_1 = Y_1\beta + X_1\,\gamma + u_1, (3)$$

and $E(u_1) = 0, E(u_1u_1') = \sigma_{11}I_n$

where, $y_1: n \times 1$ vector of observations on endogenous variable in the equation

$Y_1: n \times (m-1)$ matrix of observations on other $(m-1)$ endogenous variable in the equation

$X_1: n \times k$ matrix of observations on k predetermined variable in the equation

$u_1: n \times 1$ vector of disturbances

Reduced form of model (2) $(YB' + X\Gamma' = U)$ is

$$Y = X\Pi' + V,$$
$$\Pi = -B^{-1}\Gamma, V = UB'^{-1} \quad (4)$$

Applying OLS to (4) gives the following estimator of Π':

$$\widehat{\Pi}' = (X'X)^{-1}X'Y \, (5)$$

We can write equation (3) as

$$(y_1 \;\; Y_1 \;\; Y_2 \;\; X_1 \;\; X_2)\begin{pmatrix} 1 \\ -\beta \\ 0 \\ -\gamma \\ 0 \end{pmatrix} = u \qquad (6)$$

Y_2 and X_2 are matrices of observations on endogenous and predetermined variables excluded from the equation.

The relation between the structural and reduced form parameters is

$$\Pi' B' = -\Gamma' \quad (7)$$

The relation for the coefficients of the structural equation (3) is

$$\Pi' \begin{pmatrix} 1 \\ -\beta \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma \\ 0 \end{pmatrix}$$

Then ILS estimator can be obtained by solving

$$(X'X)^{-1}X'Y \begin{pmatrix} 1 \\ -b \\ 0 \end{pmatrix} = \begin{pmatrix} c \\ 0 \end{pmatrix}$$

or

$$(X'X)^{-1}X'(y_1 \quad Y_1 \quad Y_2) \begin{pmatrix} 1 \\ -b \\ 0 \end{pmatrix} = \begin{pmatrix} c \\ 0 \end{pmatrix}$$

or

$$(X'X)^{-1}X'y_1 - (X'X)^{-1}X'Y_1 b = \begin{pmatrix} c \\ 0 \end{pmatrix}$$

or

$$X'Y_1 b + X'X \begin{pmatrix} c \\ 0 \end{pmatrix} = X'y_1$$

Writing $X = (X_1 \quad X_2)$ we have

$$X_1'Y_1 b + X_1'X_1 c = X_1'y_1$$

$$X_2'Y_1 b + X_2'X_1 c = X_2'y_1$$

We get unique solutions for *b* and *c* if the condition for exact identification is satisfied.

Let

$$\delta = (\beta' \quad \gamma')', Z_1 = (Y_1 \quad X_1)$$

Then ILS estimator of δ is given by

$$\hat{\delta} = \begin{pmatrix} b \\ c \end{pmatrix} = (X'Z_1)^{-1}X'y_1 \quad (8)$$

7.4.2 Instrumental Variable Estimator

Applying the instrumental variable (IV) method to one system equation at a time is a single equation method. It works well with over-identified models. The instrumental variables (IV) method is a statistical technique used in econometrics and other social sciences to estimate causal relationships when a model has endogenous explanatory variables—variables that are correlated with the error term, leading to biased and inconsistent estimates. The IV method is particularly useful in situations where controlled experiments are not possible, and it helps in obtaining consistent estimators of the causal effects despite the presence of endogeneity.

For instance, we can write the equation

$$y_1 = Y_1\beta + X_1\gamma + u_1$$

as

$$y_1 = Z_1\delta + u_1, (9)$$

where

$$Z_1 = (Y_1 \quad X_1), \qquad \delta = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}$$

Let W be a $n \times (m + k)$ matrix such that

$$plim(n^{-1}W'Z_1) = \Sigma_{WZ}: \text{a finite nonsingular matrix},$$

$$plim(n^{-1}W'u) = 0$$

$$plim(n^{-1}W'W) = \Sigma_{WW}: \text{a positive definite matrix}$$

Then IV estimator of δ *is*

$$d_{IV} = (W'Z_1)^{-1}W'y_1.$$

Here, d_{IV} is consistent and its asymptotic covariance matrix is

$$Asy\ Var(d_{IV})$$
$$= \frac{\sigma_{11}}{n} plim(n^{-1}W'Z_1)^{-1}(n^{-1}W'W)(n^{-1}Z_1'W)^{-1}$$
$$= \frac{\sigma_{11}}{n}\Sigma_{WZ}^{-1}\Sigma_{WW}\Sigma_{ZW}^{-1}.$$

A consistent estimator of $\sigma_{11}$is

$$s_{11} = \frac{1}{n}(y_1 - Z_1 d_{IV})'(y_1 - Z_1 d_{IV}).$$

Generalized Least Squares Estimator:

Pre multiplying equation (9) by X' leads to

$$X'y_1 = X'Z_1\delta + X'u_1 \qquad (10)$$

Applying OLS to (10) gives the ILS estimator

$$\hat{\delta} = (X'Z_1)^{-1}X'y_1.$$

The covariance matrix of $X'u_1$ is $\sigma_{11}X'X$.

Applying GLS to (10), we obtain

$$\tilde{\delta} = (Z_1'X(X'X)^{-1}X'Z_1)^{-1}Z_1'X(X'X)^{-1}X'y_1 \qquad (11)$$

where,

$$Z_1: n \times (m - 1 + k), \qquad \text{and } X: n \times K$$

For $Z_1'X(X'X)^{-1}X'Z_1: (m - 1 + k) \times (m - 1 + k)$ to be nonsingular, necessary condition is

$$m - 1 + k \leq K$$

or order condition of identification is

$$m - 1 \leq K - k$$

For just identified case $Z_1'X$ is of order $(m-1+k) \times (m-1+k)$ and

$$\tilde{\delta} = (Z_1'X(X'X)^{-1}X'Z_1)^{-1}Z_1'X(X'X)^{-1}y_1$$

$$= (X'Z_1)^{-1}X'X(Z_1'X)^{-1}Z_1'X(X'X)^{-1}y_1$$

$$= (X'Z_1)^{-1}X'y_1 = \hat{\delta}.$$

Overidentified equation

An equation is over-identified when there are more instruments (independent variables) available than the number of endogenous variables that need to be estimated. In other words, the system has more information than necessary to identify the parameters of interest. when a model is over-identified, the extra information provided by the additional instruments should be leveraged carefully to ensure the robustness of the estimates.

If we consider the relations between structural and reduced form parameters and solve it then it leads to more than one solution for structural parameters. So, if we apply here Indirect least square procedure, we are getting more than one estimator and ILS is not unique. Hence, ILS cannot be used. For getting a unique estimator the alternative estimation procedure is Two Stage Last Squares.

7.4.3 Estimation of an over identified Equation: *Two Stage Least Squares* (2SLS)

It is a popular method used to estimate the parameters of an over-identified equation in econometrics. It is particularly useful when dealing with models where some of the explanatory variables (endogenous variables) are correlated with the error term, which violates the assumptions of ordinary least squares (OLS) regression. The idea behind 2SLS is to replace the endogenous variables with their instrumented versions, which are uncorrelated with the error term. This eliminates the bias that would occur if the endogenous variables were used directly. Let us take an example for better understanding of this method.

Example: consider the following model

Income function (IF): $y_{1t} = \beta_{10} + \beta_{11}y_{2t} + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u_{1t}$

Money-supply function (M-SF): $y_{2t} = \beta_{20} + \beta_{21}y_{1t} + u_{2t}$

where,

$$\left.\begin{matrix} y_1 = \text{income} \\ y_2 = \text{stock of money} \end{matrix}\right\} \text{endogenous variable}$$

, and

$$\left.\begin{matrix} X_1 = \text{investment expendeture} \\ X_2 = \text{government expendeture} \end{matrix}\right\} \text{exogenous variable}$$

The money supply equation is overidentified while the income equation is under-identified, as can be seen by applying the order condition of identification. Apart from altering the model specifications, there isn't much that can be done about the income equation. ILS might not be able to estimate the overidentified money supply function because there are two estimates of β_{21}. So, we can apply the method of 2SLS for estimating M-SF.

Method involves the following steps:

Step 1*:* Regress $y_1$on all the predetermined variables in the system, here X_1 and X_2, and obtain

$$\hat{y}_{1t} = \hat{\pi}_0 + \hat{\pi}_1 X_{1t} + \hat{\pi}_2 X_{2t}$$

Then

$$y_{1t} = \hat{y}_{1t} + e_t \quad ; e_t : \text{OLS residual}$$

Step 2: M-SF can be written as

$$y_{2t} = \beta_{20} + \beta_{21}(\hat{y}_{1t} + e_t) + u_{2t} = \beta_{20} + \beta_{21}\hat{y}_{1t} + (\beta_{21}e_t + u_{2t})$$

$$= \beta_{20} + \beta_{21}\hat{y}_{1t} + u_t^* \; ; (u_t^* = \beta_{21}e_t + u_{2t})$$

$\hat{y}_{1t}$ is uncorrelated with $u_t^* = \beta_{21}e_t + u_{2t}$.

Apply OLS to this equation for obtaining 2SLS estimates of β_{20} and β_{21}.

7.4.4 Two Stage Least Squares as generalized classical linear (GCL) method

The generalized classical linear method in econometrics refers to the Generalized Least Squares (GLS) technique, which is an extension of the ordinary least squares (OLS) method. While OLS assumes that the errors (or residuals) in the regression model are homoscedastic (i.e., they have constant variance) and uncorrelated, these assumptions are often violated in real-world data. When this happens, OLS estimates can become inefficient, leading to biased standard errors and misleading statistical inferences. The generalized least squares estimator $\tilde{\delta}$ may be interpreted as a 2SLS estimator.

For estimating the equation

$$y_1 = Y_1\beta + X_1\gamma + u_1 \qquad (12)$$

with K-k≥m-1, the 2SLS estimator of $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}$ can be obtained by solving the equations

$$\begin{pmatrix} Y_1'Y_1 - V_1'V_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{pmatrix}\begin{pmatrix} b \\ c \end{pmatrix} = \begin{pmatrix} (Y_1 - V_1)'y_1 \\ X_1'y_1 \end{pmatrix}$$

Here

$$V_1 = [I_n - X(X'X)^{-1}X']Y_1$$
$$= Y_1 - \hat{Y}_1$$

$$\hat{Y}_1 = X(X'X)^{-1}X'Y_1$$

$$X_1'\hat{Y}_1 = X_1'(Y_1 - V_1)$$
$$= X_1'Y_1.$$

Notice that $X'V_1 = 0 \Rightarrow X_1'V_1 = 0, X_2'V_1 = 0.$

7.4.4.1 Derivation of 2-SLS estimator

Let us write

$Y = [y_1 \quad Y_1 \quad Y_2];$

$X = [X_1 \quad X_2]$

$y_1 : n \times 1,$

$Y_1 : n \times m - 1,$

$Y_2 : n \times M - m,$

$X_1 : n \times k,$

$X_2 : n \times K - k$

The reduced form is

$$Y = X\Pi' + V$$

Applying OLS to reduced form, the predicted value of Y is

$$\hat{Y} = X(X'X)^{-1}X'Y = PY,$$

where, $P = X(X'X)^{-1}X'$ is an idempotent matrix.

Then $\hat{Y}_1 = PY_1$.

Replacing Y_1 by $\hat{Y}_1$ in equation (12), we obtain

$$y_1 = \hat{Y}_1\beta + X_1\gamma + (u_1 - V_1\beta)$$

Applying OLS to this equation, we obtain 2SLS estimator

$$\begin{pmatrix} b \\ c \end{pmatrix} = \begin{pmatrix} \hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1 \\ X_1'\hat{Y}_1 & X_1'X_1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{Y}_1'y_1 \\ X_1'y_1 \end{pmatrix}$$

$$\text{or } \begin{pmatrix} b \\ c \end{pmatrix} = \begin{pmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'X(X'X)^{-1}X'Y_1 & X_1'X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_1X(X'X)^{-1}X'y_1 \\ X_1'y_1 \end{pmatrix}$$

Notice that

$$X'X(X'X)^{-1}X' = X'$$

$$\Rightarrow X_1'X(X'X)^{-1}X' = X_1'$$

$$\Rightarrow X_1'X(X'X)^{-1}X'X_1 = X_1'X_1.$$

Then

$$\begin{pmatrix} b \\ c \end{pmatrix} = \begin{pmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_1X(X'X)^{-1}X'y_1 \\ X_1'y_1 \end{pmatrix}$$

We can write (12) as

$$y_1 = Z_1\delta + u_1,$$

where $Z_1 = [Y_1 \quad X_1],\ \delta = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}$.

Then

$$d_{2SLS} = \begin{pmatrix} b \\ c \end{pmatrix}$$

$$= [Z_1'X(X'X)^{-1}X'Z_1]^{-1}Z_1'X(X'X)^{-1}X'y_1.$$

7.4.4.2 Interpretation of 2-SLS as an IV Estimator

Let us consider,

W: Matrix of instrumental variables.

The IV estimator of δ is

$$d_{IV} = (W'Z_1)^{-1}W'y_1.$$

If we set $W = (\hat{Y}_1 \quad X_1)$, where $\hat{Y}_1 = X(X'X)^{-1}X'Y_1$, then

$$W'Z_1$$

$$= \begin{pmatrix} \hat{Y}_1' \\ X_1' \end{pmatrix} (Y_1 \quad X_1)$$

$$= Z_1'X(X'X)^{-1}X'Z_1W'y_1$$

$$= \begin{pmatrix} \hat{Y}_1' y_1 \\ X_1' y_1 \end{pmatrix}$$

$$= Z_1' X (X'X)^{-1} X' y_1.$$

Then, the IV estimator d_{IV} reduces to the 2SLS estimator.

7.4.4.3 Consistency of 2SLS Estimator

Result 1: *Suppose, as* $n \to \infty$

$$plim(n^{-1} Y_1' X) = \Sigma_{Y_1 X}: a\ finite\ matrix,$$

$$plim(n^{-1} X'X) = \Sigma_{XX}: a\ finite\ matrix$$

$$plim(n^{-1} X' u_1) = 0.$$

Then 2SLS estimator is a consistent estimator.

Proof: Let us write

$$\mathrm{W} = \hat{Z}_1$$

$$= [X(X'X)^{-1} X' Y_1\ X_1]$$

$$= PZ_1,$$

$$P = X(X'X)^{-1} X': \text{Idempotant, symmetric matrix}$$

Further

$$\hat{Z}_1' \hat{Z}_1$$

$$= Z_1' P Z_1$$

$$= \hat{Z}_1' Z_1.$$

Thus, we have

$$d_{2SLS} = \left(\hat{Z}_1' \hat{Z}_1\right)^{-1} \hat{Z}_1' y_1$$

$$= \delta + \left(\hat{Z}_1' \hat{Z}_1\right)^{-1} \hat{Z}_1' u_1$$

$$d_{2SLS} - \delta$$
$$= \left(\hat{Z}_1'\hat{Z}_1\right)^{-1}\hat{Z}_1'u_1$$

$$\text{plim}(d_{2SLS} - \delta)$$
$$= \left(\text{plim}\frac{1}{n}\hat{Z}_1'\hat{Z}_1\right)^{-1}\text{plim}\left(\frac{1}{n}\hat{Z}_1'u_1\right)$$

Now

$$plim\left(\frac{1}{n}Z_1'X\right)$$
$$= \begin{pmatrix}\Sigma_{Y_1X}\\ \Sigma_{X_1X}\end{pmatrix}$$
$$= \Sigma_{Z_1X}\ (say)$$

$$plim\frac{1}{n}\hat{Z}_1'\hat{Z}_1$$
$$= plim\left(\frac{1}{n}Z_1'X\right)\left(plim\frac{1}{n}X'X\right)^{-1}plim\ \left(\frac{1}{n}X'Z_1\right)$$
$$= \Sigma_{Z_1X}\Sigma_{XX}^{-1}\Sigma_{XZ_1}$$

$$plim\left(\frac{1}{n}\hat{Z}_1'u_1\right) = 0$$

Hence

$$plim(d_{2SLS} - \delta) = 0$$

The asymptotic variance-covariance matrix of $(d_{2SLS} - \delta)$ is $\frac{\sigma_{11}}{n}\left(\Sigma_{Z_1X}\Sigma_{XX}^{-1}\Sigma_{XZ_1}\right)^{-1}$.

Therefore, as $n \to \infty$, the asymptotic variance-covariance matrix of $(d_{2SLS} - \delta)$ tends to 0. Hence the result follows■

A consistent estimator of the asymptotic variance covariance matrix is given by

$$s_{11}(Z_1'X(X'X)^{-1}X'Z_1)^{-1}$$

where

$s_{11} = \frac{(y_1 - Y_1 b - X_1 c)'(y_1 - Y_1 b - X_1 c)}{n}$: Consistent estimator of σ_{11}

For exactly identified equations (K-k=m-1),

$$[Z_1' X (X'X)^{-1} X' Z_1]^{-1} Z_1' X (X'X)^{-1} X'$$
$$= (X' Z_1)^{-1} X' X (Z_1' X)^{-1} Z_1' X (X'X)^{-1} X'$$
$$= (X' Z_1)^{-1} X'$$

So that

d_{2SLS}

$= [Z_1' X (X'X)^{-1} X' Z_1]^{-1} Z_1' X (X'X)^{-1} X' y_1$

$= (X' Z_1)^{-1} X' y_1$

Hence and 2SLS estimator coincides with ILS estimator when equation is just identified.

Example: Let us take again the example of demand and supply function

$DF \quad Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t}$

$SF \quad Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}$

Q_t, P_t: Endogeneous variables, $M = 2$

1, I_t, R_t, P_{t-1}: Predetermined variables $K = 4$

DF is just identified, SF is over identified

2SLS Estimation:

(i) Run OLS for

$P_t = \pi_0 + \pi_1 I_t + \pi_2 R_t + \pi_3 P_{t-1} + v_t$

and obtain OLS estimators $\hat{\pi}_0, \hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3$.

(ii) The predicted value of P_t is

$$\hat{P}_t = \hat{\pi}_0 + \hat{\pi}_1 I_t + \hat{\pi}_2 R_t + \hat{\pi}_3 P_{t-1}$$

(iii) For estimating the demand function run OLS for

$$Q_t = \alpha_0 + \alpha_1 \hat{P}_t + \alpha_2 I_t + \alpha_3 R_t + (u_{1t} + \alpha_1 \hat{v}_t),$$

$(\hat{v}_t = P_t - \hat{P}_t)$ and obtain 2SLS estimators of $\alpha_0, \alpha_1, \alpha_2, \alpha_3$.

(iv) For estimating supply function run OLS for

$$Q_t = \beta_0 + \beta_1 \hat{P}_t + \beta_2 P_{t-1} + (u_{2t} + \beta_1 \hat{v}_t).$$

Interpretation

(1) In over identified equations, ILS provides multiple estimates of parameters whereas, 2SLS provides only one estimate per parameter. So, it is unique.

(2) 2SLS requires total number of predetermined variables in the system.

(3) This method can also be applied to exactly identified equations.

(4) 2SLS provides a consistent estimator.

7.5 Family of k-class Estimators

In econometrics, K-class estimators are a family of estimators used for estimating the parameters of a structural equation in the presence of endogeneity, which occurs when an explanatory variable is correlated with the error term. The K-class estimators were introduced by Henri Theil in 1958 and are a generalization of the instrumental variable (IV) estimator. K-class estimators are primarily used to address the issue of endogeneity by generalizing the way instruments are used in estimation.

The family of k-class estimators is defined as

$$\hat{\delta}_i(k) = \begin{pmatrix} Y_i'Y_i - k\hat{V}_i'\hat{V}_i & Y_i'X_i \\ X_i'Y_i & X_i'X_i \end{pmatrix}^{-1} \begin{pmatrix} Y_i' - k\hat{V}_i' \\ X_i' \end{pmatrix} y_i \qquad (13)$$

For $k = 0, \hat{\delta}_i(k)$ reduces to ordinary least squares estimator.

For $k = 1, \hat{\delta}_i(k)$ reduces to 2SLS estimator.

If value of k is the smallest root of the equation

$$|W_i - \lambda W| = 0,$$

then $\hat{\delta}_i(k)$ reduces to LIML estimator.

A Family of IV Estimators

Consider system of structural equations

$$y = Z\delta + u$$

The instrumental variable (IV) estimator is defined a

$$\hat{\delta}_{IV} = (W'Z)^{-1}W'y$$

W: matrix of instruments of the same dimension and rank as Z. Further

$$plim\left(\frac{1}{n}W'u\right) = 0,$$

$$plim\left(\frac{1}{n}W'Z\right) = \Sigma_{WZ}$$

$\Sigma_{WZ} \Rightarrow$ non singular

Then

$$\text{plim } \hat{\delta}_{IV}$$

$$= \left(plim\frac{1}{n}W'Z\right)^{-1} plim\frac{1}{n}W'(Z\delta + u)$$

$$= \delta$$

$$plim\left(\hat{\delta}_{IV}-\delta\right)\left(\hat{\delta}_{IV}-\delta\right)'$$
$$=\frac{1}{n}\Sigma_{WZ}^{-1}plim\left[\frac{1}{n}W'(\Sigma\otimes I_n)W\right]\Sigma_{ZW}^{-1}$$

$\hat{\delta}_{IV}$ is a consistent estimator and the asymptotic distribution of $\sqrt{n}\left(\hat{\delta}_{IV}-\delta\right)$ is normal with mean vector 0 and covariance matrix $\Sigma_{WZ}^{-1}plim\left[\frac{1}{n}W'(\Sigma\otimes I_n)W\right]\Sigma_{ZW}^{-1}$.

For OLS estimator $d_{OLS}=(Z'Z)^{-1}Z'y$, the matrix of IV is $W=Z$.

Since $plim\left(\frac{1}{n}Z'u\right)\neq 0$, the OLS estimator is inconsistent.

For 2SLS estimator

$$d_{2SLS}$$
$$=[Z'\{I\otimes X(X'X)^{-1}X'\}Z]^{-1}Z'[I\otimes X(X'X)^{-1}X']y,$$
$$W'=Z'\{I\otimes X(X'X)^{-1}X'\}$$

And

$$plim\left(\frac{1}{n}W'u\right)=0.$$

Thus, 2SLS is consistent.

For 3SLS estimator

$$d_{3SLS}=\left[Z'\left(\hat{\Sigma}^{-1}\otimes P\right)Z\right]^{-1}Z'\left[\hat{\Sigma}^{-1}\otimes P\right]y,$$
$$W'=Z'\left(\hat{\Sigma}^{-1}\otimes P\right)$$

so that

$$plim\left(\frac{1}{n}W'u\right)=0.$$

Thus, 3SLS is consistent.

7.6 Self-Assessment Exercise

1. Write the reduced form of the following simultaneous equations model and obtain the indirect and two stage least squares estimators..

$$y_{1t} = a_0 + a_1 y_{2t} + a_2 x_{1t} + u_{1t}$$
$$y_{2t} = b_0 + b_1 y_{1t-1} + u_{2t}$$

2. For the following simultaneous equation model describe the method of indirect least squares with the help of this example.

Demand Model: $Q_t^D = \alpha_0 + \alpha_1 P_t + \alpha_2 ln C_t + u_{1t}$

Supply Model: $Q_t^S = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{1t}$

Equilibrium: $Q_t^D = Q_t^S$.

Here Q_t and P_t are endogeneous variables.

3. For the general form of simultaneous equations model, derive the indirect least squares of estimator and two stage least squares estimator. Under what conditions two-stage least squares estimator and indirect least squares estimators are identical?

7.7 Summary

This unit focuses on the estimation techniques for simultaneous equations models, which are essential in analysing systems where variables are interdependent. It begins with a review of the structural and reduced forms of these models, highlighting the challenges posed by simultaneity and endogeneity.

Key estimation methods are explored, including Indirect Least Squares (ILS), Two-Stage Least Squares (2SLS), Each method is explained in terms of its underlying assumptions, steps, and applicability to specific scenarios. Comparative analyses of these techniques are provided, focusing on their efficiency, consistency, and limitations.

This unit makes imparts knowledge about the concept of Estimators in simultaneous equation models. Additionally, it covers the concept of limited and full information estimators, indirect least squares estimators, two stage least squares estimators, three stage least squares estimators and k class estimator in depth.

7.8 References

- Gujarati, D. and Guneshker, S. (2007). *Basic Econometrics*, 4th Ed., McGraw Hill Companies.

- Johnston, J. and J. Dinardo (1997). *Econometric Methods*, 2nd Ed., McGraw Hill International.

- Koutsoyiannis, A. (2004). *Theory of Econometrics*, 2 Ed., Palgrave Macmillan Limited.

- Maddala, G.S. and Lahiri, K. (2009). *Introduction to Econometrics*, 4 Ed., John Wiley & Sons.

7.9 Further Readings

- Judge, G.G., W.E. Griffiths, R.C. Hill, Lütkepohl and T. C. Lee (1985). The theory and practice of econometrics, Wiley.

- Ullah, A. and Vinod, H.D. (1981). Recent advances in Regression Methods, Marcel Dekker.

UNIT 8 ESTIMATION IN SIMULTANEOUS EQUATION MODELS

8.1 Introduction

Simultaneous equations models (SEMs) are used in econometrics and statistics to represent systems where multiple dependent variables influence each other simultaneously. These models are common in economics, where variables like supply and demand are determined together. In SEMs, endogenous variables are those determined within the system by the equations themselves. Exogenous variables are determined outside the system and treated as given. Before estimation, it's essential to check if the model is identifiable, meaning that there is a unique solution for the parameters of the model. A model is exactly identified if the number of independent equations equals the number of parameters. It is over-identified if there are more independent equations than parameters, and under-identified if there are fewer.

Estimates the entire system of equations simultaneously and is efficient but computationally intensive are called Full Information Maximum Likelihood (FIML)

Limited Information Maximum Likelihood (LIML) is similar to 2SLS but generally provides more efficient estimates in small samples.

Simultaneous equations models are powerful tools but require careful consideration of the underlying assumptions and the methods used for estimation.

Prediction and simultaneous confidence intervals are used to quantify uncertainty when predicting outcomes in models, particularly when multiple parameters or equations are involved, as in simultaneous equations models (SEMs).

8.2 Objective

After completing this course, students should have developed a clear understanding of:

- Limited information maximum likelihood estimation
- Full information maximum likelihood estimation
- Prediction and simultaneous confidence interval.

8.3 Limited Information Maximum Likelihood (LIML) Estimators

An approach for estimating a single equation in a linear simultaneous equations model that maximises the likelihood function while adhering to the structure's constraints. When the errors are regularly distributed, the LIML estimator performs better than single equation estimators.

An alternative to the ILS and 2SLS approaches is the Limited Information Maximum Likelihood (LIML). However, the 2SLS is more frequently used because to its computational complexity. Furthermore, the LIML requires rigorous and challenging derivation.

Let us suppose that we write the i^{th} equation as

$$y_i = Y_i\beta_i + X_i\gamma_i + u_i; i = 1,2,\dots,M \quad (1)$$

$$\text{or } y_i = Z_i\delta_i + u_i;\ u_i \sim N(0, \sigma_{ii}I_n) \quad (2)$$

LIML estimator is obtained by maximizing the LF derived from the stochastic elements of single equation (2).

Here

$y_i : n \times 1,$

$Y_i : n \times (m_i - 1),$

$X_i : n \times k_i$

$\beta_i : (m_i - 1) \times 1,$

$\gamma_i : k_i \times 1,$

$Z_i : n \times (m_i + k_i - 1)$

$\delta_i = \begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix} : (m_i + k_i - 1) \times 1$

$B_i = (-1 \quad \beta_i \quad 0)', \Gamma_i = (\gamma_i \quad 0)'$ have $(m_i - 1) + k_i$ unknown elements

We write the M equations of (1) jointly as

$$y = Z\delta + u \quad (3)$$

where

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix}, \qquad Z = \begin{pmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_M \end{pmatrix}, \qquad \delta = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_M \end{pmatrix}, \qquad u = \begin{pmatrix} u_1 \\ \vdots \\ u_M \end{pmatrix}$$

$$Z : nM \times \sum_{i=1}^{M} (m_i - 1 + k_i),$$

$$\delta : \sum_{i=1}^{M} (m_i - 1 + k_i) \times 1$$

y and u are of order $nM \times 1$

The likelihood function is

$$l(\delta_*, \Sigma_*|y_*)$$

$$= (2\pi)^{-\frac{mn}{2}} |\Sigma_* \otimes I_n|^{-\frac{1}{2}} |\Gamma_*|^n \exp\left[-\frac{1}{2}(y_* - Z_*\delta_*)'(\Sigma_* \otimes I_n)^{-1}(y_* - Z_*\delta_*)\right]$$

where,

$y_* = vec(y_i \;\; Y_i): nm_i \times 1,$

$\Sigma_*: m_i \times m_i$ submatrix of Σ related to structural errors of y_*

Z_*: Submatrix of Z involving $Z_1, \ldots, Z_{m_i}$

δ_*: Subvector of δ involving $\delta_1, \ldots, \delta_{m_i}$

We maximize the LF subject to identification restriction of the i^{th} equation, it leads to the LIML estimator.

LIML estimator is efficient among the single equation estimators under the assumption of normality of disturbances.

8.3.1 Pagan's (1979) Procedure for obtaining LIML Estimator

Pagan (1979) proposed an alternative formulation of limited information maximum likelihood (LIML) estimation in terms of an iterated Seemingly Unrelated Regression (SUR) procedure, based on the fact that the limited information (LI) specification of a standard simultaneous equation system can be expressed as a triangle model. Pagan also obtained an expression for the connection between LIML and two stage least squares (2SLS) estimates in finite samples by using this method.

Let us consider the model

$$y_i = Z_i\delta_i + u_i (4)$$

Reduced form equation for Y_i

$$Y_i = X\Pi_i + V_i$$

Reduced form equation for Y_i in vector form

$$y_R = \left(I_{k_i-1} \otimes X\right)\pi_R + v_R \ (5)$$

where,

$Z_i = (Y_i \ \ X_i)$,

$y_R = vec(Y_i)$,

$\pi_R = vec(\Pi_i)$,

$v_i = vec(V_i)$

We can write (4) and (5) jointly as

$$\begin{pmatrix} y_i \\ y_R \end{pmatrix} = \begin{pmatrix} Z_i & 0 \\ 0 & I_{k_i-1} \otimes X \end{pmatrix} \begin{pmatrix} \delta_i \\ \pi_R \end{pmatrix} + \begin{pmatrix} u_i \\ v_R \end{pmatrix} (6)$$

Suppose the covariance matrix between a component of u_i and that of v_R is

$$\begin{pmatrix} \sigma_{ii} & \Psi' \\ \Psi & \Omega \end{pmatrix}$$

with

$$\begin{pmatrix} \sigma_{ii} & \Psi' \\ \Psi & \Omega \end{pmatrix}^{-1} = \begin{pmatrix} a & c' \\ c & D \end{pmatrix}.$$

Then

$a = (\sigma_{ii} - \Psi'\Omega^{-1}\Psi)^{-1}$,

$c = -(\sigma_{ii} - \Psi'\Omega^{-1}\Psi)^{-1}\Omega^{-1}\Psi$,

$D = \Omega^{-1} + (\sigma_{ii} - \Psi'\Omega^{-1}\Psi)^{-1}\Omega^{-1}\Psi'\Psi\Omega^{-1}$

From (6), we obtain the estimator

$$\begin{pmatrix} d_{iLIML} \\ \hat{\pi}_{LIML} \end{pmatrix}$$

$$= \begin{pmatrix} Z_i'(a \otimes I_n)Z_i & Z_i'(c' \otimes X) \\ (c \otimes X')Z_i & D \otimes X'X \end{pmatrix}^{-1} \times \begin{pmatrix} Z_i'(a \otimes I_n) & Z_i'(c' \otimes X) \\ c \otimes X' & D \otimes X' \end{pmatrix} \begin{pmatrix} y_i \\ y_R \end{pmatrix}$$

Hence

$$d_{iLIML} = \sigma_{ii} QZ_i' P y_i + aQZ_i' \bar{P} y_i + QZ_i'(c' \otimes \bar{P}) y_R$$

$$Q = (\sigma_{ii}^{-1} Z_i' P Z_i + aZ_i' \bar{P} Z_i)^{-1},$$

$$P = X(X'X)^{-1}X',$$

$$\bar{P} = I_n - X(X'X)^{-1}X'.$$

If we write $\hat{Z}_i = (\hat{Y}_i \;\; X_i), \widehat{W}_i = Z_i - \hat{Z}_i$, then

$$Q^{-1} = \sigma_{ii}^{-1} \hat{Z}_i' \hat{Z}_i + a\widehat{W}_i' \widehat{W}_i$$

so that

$$d_{iLIML} = d_{i2SLS} - (\hat{Z}_i' \hat{Z}_i)^{-1} F (\hat{Z}_i' \hat{Z}_i)^{-1} \hat{Z}_i' y_i + aQ \hat{Z}_i' P y_i + Q \hat{Z}_i' (c' \otimes \bar{P}) y_R$$

which gives a relationship between 2SLS and LIML estimators.

8.4 Full Information Methods

ILS, 2SLS and LIML estimators are limited information estimators in the sense that in estimation of any structural equation, all other structural equations in the model have not been considered. In full information estimator's other equations and the fact that the structural disturbances of various equations may be correlated is utilized in the estimation process. In principal information on complete system would yield estimators with improved efficiency properties.

Let the i^{th} equation be

$$y_i = Y_i \beta_i + X_i \gamma_i + u_i;\; i = 1,2, \ldots, M \quad (7)$$

Writing $Z_i = (Y_i \quad X_i), \delta_i = \begin{pmatrix} \beta_i \\ \gamma_i \end{pmatrix}$, we have

$$y_i = Z_i\delta_i + u_i;\ i = 1,2,\dots,M \quad (8)$$

The 2SLS estimator of δ_i is

$$d_{i2SLS} = [Z_i'X(X'X)^{-1}X'^{Z_i}]^{-1}Z_i'X(X'X)^{-1}X'y_i$$

Now

$$X'y_i = X'Z_i\delta_i + X'u_i; i = 1,2,\dots,M \quad (9)$$

Staking the M equations, we can write (9) as

$$\begin{pmatrix} X'y_1 \\ \vdots \\ X'y_M \end{pmatrix} = \begin{pmatrix} X'Z_1 & 0 & \cdots & 0 \\ 0 & X'Z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X'Z_M \end{pmatrix}\begin{pmatrix} \delta_1 \\ \vdots \\ \delta_M \end{pmatrix} + \begin{pmatrix} X'u_1 \\ \vdots \\ X'u_M \end{pmatrix} \qquad (10)$$

or,

$$(I_M \otimes X')y = (I_M \otimes X')Z\delta + (I_M \otimes X')u \quad (11)$$

where,

$$Z: nM \times \sum_{i=1}^{M}(m_i - 1 + k_i),$$

$$\delta: \sum_{i=1}^{M}(m_i - 1 + k_i) \times 1$$

$$Eu_iu_i' = \sigma_{ii}I_n,$$

$$Eu_iu_j' = \sigma_{ij}I_n$$

$$\Sigma = \left((\sigma_{ij})\right)_{M\times M}$$

The covariance matrix of $\begin{pmatrix} X'u_1 \\ \vdots \\ X'u_M \end{pmatrix}$ is

$$\begin{pmatrix} \sigma_{11}X'X & \sigma_{12}X'X & \cdots \sigma_{1M}X'X \\ \sigma_{21}X'X & \sigma_{22}X'X & \cdots \sigma_{2M}X'X \\ \vdots & \vdots & \ddots \quad \vdots \\ \sigma_{M1}X'X & \sigma_{M2}X'X & \cdots \sigma_{MM}X'X \end{pmatrix} = \Sigma \otimes X'X$$

Applying GLS to (11), we obtain the 3SLS estimator

$$\begin{aligned} &\tilde{\delta}_{3SLS} \\ &= [Z'(I_M \otimes X')'(\Sigma^{-1} \otimes (X'X)^{-1})(I_M \otimes X')Z]^{-1} \\ &\times Z'(I_M \otimes X')'(\Sigma^{-1} \otimes (X'X)^{-1})\,(I_M \otimes X')y \\ &= [Z'\{\Sigma^{-1} \otimes X(X'X)^{-1}X'\}Z]^{-1} \times Z'[\Sigma^{-1} \otimes X(X'X)^{-1}X']y \qquad (12) \end{aligned}$$

A consistent estimator of σ_{ij} is given by

$$s_{ij} = \frac{(y_i - Z_i d_{i,2SLS})'(y_j - Z_j d_{j,2SLS})}{n}$$

$d_{i,2SLS}$: 2SLS estimator of δ_i

In s_{ij}, in place of n we may divide by

$$(n - m_i - k_i + 1)^{\frac{1}{2}}\,(n - m_j - k_j + 1)^{\frac{1}{2}}$$

A feasible 3SLS estimator is

$$\tilde{\delta}_{3SLS} = [Z'\{S^{-1} \otimes X(X'X)^{-1}X'\}Z]^{-1} \times Z'[S^{-1} \otimes X(X'X)^{-1}X']y \qquad (13)$$

where $S = \left((s_{ij})\right)$.

Result 1: *If* $\sigma_{ij} = 0\ \forall i \neq j$*, then 3SLS reduces to 2SLS*

Proof: Let $\Sigma^{-1} = \left((\sigma^{ij})\right)$.

When $\sigma_{ij} = 0\ \forall i \neq j$,

we have $\sigma^{ij} = 0\ \forall i \neq j$.

Hence

$$Z'\{\Sigma^{-1} \otimes X(X'X)^{-1}X'\}Z$$

$$= \begin{pmatrix} \sigma^{11}Z_1'X(X'X)^{-1}X'Z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^{MM}Z_M'X(X'X)^{-1}X'Z_M \end{pmatrix}$$

$$Z'[\Sigma^{-1} \otimes X(X'X)^{-1}X']y$$

$$= \begin{pmatrix} \sum_{j=1}^{M} \sigma^{1j}Z_1'X(X'X)^{-1}X'^{y_j} \\ \vdots \\ \sum_{j=1}^{M} \sigma^{Mj}Z_M'X(X'X)^{-1}X'y_j \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^{11}Z_1'X(X'X)^{-1}X'y_1 \\ \vdots \\ \sigma^{MM}Z_M'X(X'X)^{-1}X'y_M \end{pmatrix}$$

Then

$$\tilde{\delta}_{3SLS} = \begin{pmatrix} \sigma^{11}Z_1'X(X'X)^{-1}X'Z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^{MM}Z_M'X(X'X)^{-1}X'Z_M \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \sigma^{11}Z_1'X(X'X)^{-1}X'y_1 \\ \vdots \\ \sigma^{MM}Z_M'X(X'X)^{-1}X'y_M \end{pmatrix} = \begin{pmatrix} \hat{\delta}_{1,2SLS} \\ \vdots \\ \hat{\delta}_{M,2SLS} \end{pmatrix}.$$

Hence the result follows■

Comparison of 2SLS and 3SLS Estimators:

• 3SLS is asymptotically more efficient than 2SLS estimator.

• 3SLS requires a much-detailed specification for the equation system than 2SLS estimator does.

• 3SLS requires all variables in the equations to be estimated and all predetermined variables of the system.

• If some equation of the system is mis specified, this affects the 2SLS estimator of that equation, but not those of other equations.

• For 3SLS, estimates of all equations of the system are affected by any such misspecification.

• For applying 3SLS, all equations of the system must be identified.

8.5 Full Information Maximum Likelihood (FIML) Estimation

Full Information Maximum Likelihood (FIML) Estimation is a statistical method used in econometrics to estimate the parameters of a model, particularly when dealing with incomplete data or systems of simultaneous equations. This approach leverages all available information in the model to maximize the likelihood function and produce estimates that are consistent and efficient.

FIML is especially useful in the context of simultaneous equations models, where multiple equations are estimated together. Each equation typically has a dependent variable that is also an independent variable in another equation. By estimating all equations jointly, FIML accounts for the interdependencies and correlations between the error terms of the equations. In SEM, FIML handles both measurement and structural components simultaneously, making it a preferred method when dealing with complex data structures. For example; Suppose you have a system of equations where you want to model the relationship between different economic variables like income, consumption, and investment. If these variables influence each other, using FIML allows you to estimate the entire system at once, taking into account the potential feedback loops and correlations between the error terms in each equation.

Consider the model

$$y = Z\delta + u;\ \ u \sim N(0, \Sigma \otimes I_n).$$

and the pdf of u:

$$f(u) = (2\pi)^{-\frac{Mn}{2}} |\Sigma \otimes I_n|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} u'(\Sigma \otimes I_n)^{-1} u\right]$$

For estimating $\delta = (\delta_1' \dots \delta_M')'$, the LF is

$$l(\delta, \Sigma | y, Z) = (2\pi)^{-\frac{Mn}{2}} |\Sigma \otimes I_n|^{-\frac{1}{2}} |B|^n \times \exp\left[-\frac{1}{2}(y - Z\delta)'(\Sigma^{-1} \otimes I_n)(y - Z\delta)\right]$$

Ignoring the constant term, the log LF is

$$lnl(\delta, \Sigma | y, Z) = -\frac{n}{2} ln|\Sigma| + nln|B| - \frac{1}{2}(y - Z\delta)'(\Sigma^{-1} \otimes I_n)(y - Z\delta)$$

Maximizing the log likelihood subject to the restrictions on parameters Γ, B, Σ leads to FIML estimators of parameters.

8.6 Prediction and Simultaneous confidence interval

In econometrics, prediction and simultaneous confidence intervals are tools used to assess the uncertainty around predicted values and parameter estimates, respectively. Both are important for understanding the range within which the true values of parameters or predictions are likely to lie, given a certain level of confidence.

1. Prediction Intervals

A prediction interval provides a range within which a single future observation is expected to fall, with a specified level of confidence (e.g., 95%).

2. Simultaneous Confidence Intervals

Simultaneous confidence intervals are used to provide a range of values for multiple parameter estimates (or predictions) simultaneously, while controlling the overall error rate. They ensure that the true values of all parameters fall within their respective intervals with a certain overall confidence level.

In econometrics, when making inferences about multiple coefficients (e.g., in a regression model), using individual confidence intervals for each coefficient might not be appropriate if you want to control the overall confidence level for all coefficients together. Simultaneous confidence intervals address this issue by providing intervals that account for the joint distribution of the estimates.

Application:

1. Prediction Intervals: Useful when forecasting future values, such as predicting GDP, stock returns, etc., with a specified range of uncertainty.

2. Simultaneous Confidence Intervals: Crucial when making inferences about multiple parameters, such as when analyzing the effects of multiple explanatory variables on an outcome, while maintaining overall control of the Type I error rate.

Both types of intervals are important tools in econometrics for providing a fuller picture of uncertainty in model estimates and predictions.

8.7 Self-Assessment Exercise

1. What are the advantages and disadvantages of three stage least squares over two-stage least squares estimator.

2. Give the justification of two-stage least squares estimator as an instrumental variable estimator.

3. Explain the difference between limited information and full information estimators.

4. Explain the three-stage least squares estimator. Under what conditions it reduces to the two stage least squares estimator.

8.8 Summary

This unit provides an in-depth exploration of advanced estimation methods for simultaneous equations models, focusing on Three-Stage Least Squares (3SLS), Limited Information Maximum Likelihood (LIML), and Full Information Maximum Likelihood (FIML) estimators. The unit begins by outlining the challenges of estimation in systems of

interdependent equations, such as endogeneity and simultaneity, and emphasizes the importance of choosing appropriate estimation techniques.

Three-Stage Least Squares (3SLS) is introduced as an extension of Two-Stage Least Squares (2SLS), combining system-wide efficiency with the ability to handle correlation across equations. The unit explains the derivation and application of 3SLS, highlighting its advantages and limitations in practical settings.

Limited Information Maximum Likelihood (LIML) is presented as a single-equation estimation method that addresses the identification problem while maintaining asymptotic efficiency. The discussion covers its assumptions, estimation steps, and scenarios where LIML is preferable to 2SLS.

Finally, Full Information Maximum Likelihood (FIML) is explored as a comprehensive system-wide estimator that simultaneously considers all equations in a model. The unit examines FIML's advantages in terms of efficiency and consistency, as well as its sensitivity to specification errors.

The unit also explains the prediction in simultaneous equations model.

8.9 References

- Gujarati, D. and Guneshker, S. (2007). *Basic Econometrics*, 4th Ed., McGraw Hill Companies.
- Johnston, J. and J. Dinardo (1997). *Econometric Methods*, 2nd Ed., McGraw Hill International.
- Koutsoyiannis, A. (2004). *Theory of Econometrics*, 2 Ed., Palgrave Macmillan Limited.

8.10 Further Readings

- Judge, G.G., W.E. Griffiths, R.C. Hill Lütkepohl and T. C. Lee (1985). The theory and practice of econometrics, Wiley.
- Maddala, G.S. and Lahiri, K. (2009). *Introduction to Econometrics*, 4 Ed., John Wiley

- Srivastava, V.K. and Giles D.A.E. (1987): Seemingly unrelated regression equations models, Marcel Dekker & Sons.

- Ullah, A. and Vinod, H.D. (1981). Recent advances in Regression Methods, Marcel Dekker.

UNIT 9 FORECASTING

Structure

9.1 Introduction

Forecasting in econometrics involves predicting future values of economic variables using statistical models based on historical data. This is a crucial activity in economics, finance, and related fields where future trends need to be anticipated for decision-making, policy formulation, and planning.

Time series forecasting is widely used in various fields like finance, economics, weather forecasting, and inventory management. The unique aspect of time series data is that observations are time-ordered, and this temporal structure must be accounted for in the analysis. The main application of forecasting is in different areas, like

- Macroeconomic Policy: Predicting GDP growth, inflation, unemployment rates, etc., to guide monetary and fiscal policy.
- Financial Markets: Forecasting stock prices, interest rates, exchange rates, and other financial indicators.
- Corporate Planning: Demand forecasting, sales prediction, and inventory management.
- Supply Chain Management: Demand forecasting and inventory management.
- Weather Forecasting: Predicting temperatures, rainfall, and other climatic conditions.

Forecasting combines economic theory, statistical analysis, and historical data to provide insights into future economic conditions, making it a powerful tool in both public policy and business strategy.

Exponential and adaptive smoothing processes are important techniques in time series forecasting, especially when the goal is to make predictions based on data that may exhibit trends, seasonality, or other patterns. Both methods emphasize the use of recent observations while giving progressively less weight to older data. Exponential smoothing methods forecast the future by applying exponentially decreasing weights to past observations. Simple exponential smoothing is used for series without trend or seasonality, while Holt-Winters exponential smoothing is used for series with trend and seasonality.

9.2 Objectives

After completing this unit, students should have developed a clear understanding of:

- Forecasting
- Exponential and adaptive smoothing methods
- Periodogram and correlogram analysis.

9.3 Exponential and Adoptive Smoothing Method

Exponential smoothing is a time series forecasting method where past observations are weighted using exponentially decreasing factors. This means more recent observations have a higher influence on the forecast, making it responsive to changes in the data. It gives more weight to recent observations and is suited for stable series with clear patterns. Adaptive smoothing adjusts the smoothing parameters dynamically based on the data, unlike the traditional exponential smoothing where the smoothing constant is fixed. It adjusts to changes in the data patterns over time, making it ideal for more volatile series. Both methods are valuable tools in forecasting, depending on the characteristics of the data and the specific forecasting needs.

9.3.1 Exponentially weighted moving average (EWMA)

Exponentially Weighted Moving Averages (EWMA)* is a technique used in time series analysis and data smoothing. It is a class of procedures for smoothing discrete time series. It assigns exponentially decreasing weights to older observations, giving more importance to recent data points. This makes it particularly useful in situations where the most recent data is more relevant, such as in financial analysis, quality control, and forecasting. It also applied in signal processing as low-pass filters to remove high-frequency noise.

The simple moving average uses equal weights and exponential smoothing uses exponentially decreasing weights over time. It does not require any minimum number of observations before starting exponential smoothing (unlike simple smoothing)

The EWMA series needs to start with an initial value, which can be the first observation, or it can be set to the mean of the first few observations. Let y_t; $\{y_t; t = 0, \dots, n\}$ be a given time series and y_0 *is the* initial value of y_t.

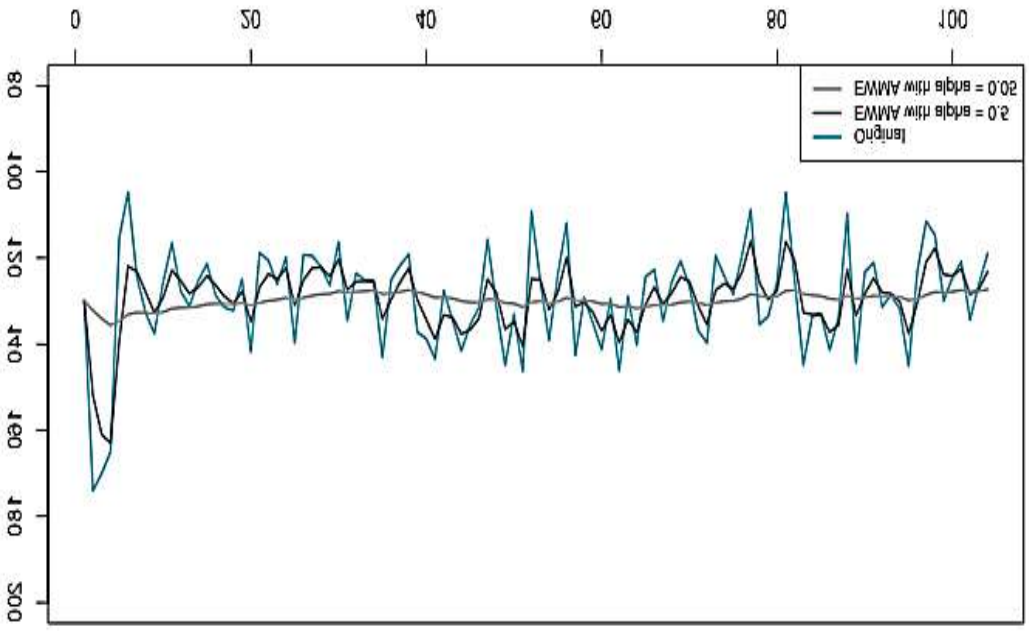
We define a new time series s_t that is a smoothed version of y_t.

$s_0 = y_0 s_t = \alpha y_t + (1 - \alpha) s_{t-1}$

where, α is the smoothing factor, $0 < \alpha < 1$.

- The smoothing parameter α controls how quickly the weights decrease for older observations.
- A higher α (close to 1) gives more weight to recent observations, making the EWMA more responsive to recent changes.
- A lower α (close to 0) smooths the data more, making the EWMA less sensitive to recent changes. The smaller the weight α, the less influence each point has on the smoothed time series.

This formulation is also known as *Brown's simple exponential smoothing.*



9.3.1.1 Brown's simple exponential smoothing

Brown's Simple Exponential Smoothing (SES) is a forecasting technique that assigns exponentially decreasing weights to past observations. The method is particularly useful for data that does not exhibit a clear trend or seasonality. Here, s_t is the smoothed statistic and it is the simple weighted average of y_t and the previous smoothed statistic s_{t-1}.

- Higher values of α give more weight to recent observations. Its larger values reduce the level of smoothing and give greater weight to recent changes in the data.
- Lower values of α smooth the data more heavily, giving more weight to older observations.
- α closer to zero have a greater smoothing effect and are less responsive to recent changes
- It is useful for short-term forecasting where data is relatively stationary.
- If $\alpha = 1$, the resulting series is $s_t = y_t$; which is the original time series.

9.3.1.2 Adoptive Forecasting Using EWMA

Forecast is constructed using exponentially weighted average of past observations. Obviously, more recent values to have greater influence on the forecast and influence of past data decreases exponentially. Adoptive forecasting using Exponentially Weighted Moving Averages (EWMA) is a method used in time series forecasting where more recent data points are given exponentially more weight than older data points. This makes the forecast more adaptive to recent changes, which can be particularly useful in environments where patterns shift over time. In adaptive forecasting, the smoothing factor α may be adjusted dynamically based on the performance of the forecast. For instance, if the forecast error increases, the model might increase α to give more weight to recent data, making the forecast more responsive to changes. This method is quite simple, computational efficiency, ease of adjusting to changes in the process being forecast with reasonable accuracy. It allows to determine influence of recent observation on forecast value.

Let $y_1, y_2, \cdots, y_n$ be the *n* observed time series values and we assume that there is neither cyclic variation nor pronounced trend.

The exponential smoothing equation is:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$$

where,

$\hat{y}_t$: Forecasted value at time t,

α: Smoothing constant, $(0 < \alpha < 1)$.

The initial forecast value $\hat{y}_1$ is unknown.

Set the first estimate $\hat{y}_1 = y_1$, this implies that, the initial value y_1 will have an unreasonably large effect on early forecasts. Use the average of the first few (10 or more) observations for the initial smoothed value.

We can write

$$\hat{y}_{t+1} - y_t$$
$$= (1 - \alpha)(\hat{y}_t - y_t)$$
$$= (1 - \alpha)e_t$$

where, $e_t = (\hat{y}_t - y_t)$ is the forecast error at time t.

By recursive substitution

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$$

$$= \alpha y_t + (1 - \alpha)[\alpha y_{t-1} + (1 - \alpha)\hat{y}_{t-1}]$$

$$= \alpha y_t + (1 - \alpha)\alpha y_{t-1} + (1 - \alpha)^2 \hat{y}_{t-1}$$

$$= \cdots$$

$$= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \cdots + \alpha(1-\alpha)^{t-1}y_1$$

The forecast equation becomes

$$\hat{y}_{t+1} = \alpha \sum_{j=0}^{t-1} (1-\alpha)^j y_{t-j}$$

where, $\hat{y}_{t+1}$ is the weighted moving average of all past observations.

The series of weights decline toward zero in an exponential fashion. As we go back in the series, each value has a smaller weight in terms of its effect on the forecast.

Notice that, α near zero allow the distant past observations to have a large influence and α near one allow the past observations to have a negligible influence.

9.3.1.3 How to select α?

In this section we consider some methods for measuring the accuracy of forecast value.

***Measuring the accuracy of forecast method*:**

1: *Mean absolute percentage error (MAPE)*

It is defined as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|e_t|}{y_t} \times 100\%,$$

where $e_t = (\hat{y}_t - y_t)$.

Note that lower the MAPE the better is the forecast.

- $MAPE$ below 10% implies highly accurate forecast.
- $MAPE$ between $11\% - 20\%$ implies good forecast.
- $MAPE$ between $21\% - 50\%$ implies reasonable forecast.

- $MAPE$ above 50% implies inaccurate forecast

2: *MSE (Mean square error) and RMSE (Root mean square error)*

It is defined as:

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

$$RMSE = \sqrt{MSE}$$

The MSE or RMSE can be used as criterion for selecting smoothing constant α. Assign values from 0.1 to 0.99 to α and select the value with smallest MSE or RMSE.

9.3.1.4 Double exponential smoothing (Holt's Method)

The Holt method, also known as double exponential smoothing, is an extension of simple exponential smoothing. It is used for forecasting time series data that exhibits both a linear trend and no seasonal pattern. it is also called Holt's trend corrected or second-order exponential smoothing. This method introduces a term to take care of trend present in the time series and is capable of capturing increase or decrease in linear trend. It is useful when the data shows a linear trend over time. If the data also exhibits seasonality, the Holt-Winters method, which extends the Holt method to include seasonal components, may be more appropriate.

Steps: Let $t = 0$; $s_0 = y_0$,

Initial value $b_0 = y_1 - y_0$ or

$$b_0 = \frac{y_n - y_0}{n} : \text{based on the assumption of linear trend}$$

If $t > 0$

$$s_t = \alpha y_t + (1-\alpha)(s_{t-1} + b_{t-1})$$
$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}$$

where,

s_t= Smoothed statistic

α= smoothing factor of data; $0 < \alpha < 1$

b_t= best estimate of trend at time t

β = trend smoothing factor; $0 < \beta < 1$

For forecasting beyond y_t

$\hat{y}_{t+k} = s_t + kb_t$

9.3.1.5 Brown's linear exponential smoothing (LES) or double exponential smoothing

Brown's linear exponential smoothing, also known as Brown's double exponential smoothing, is a forecasting method similar to Holt's method but with a key difference in how the trend is handled. It is a simpler method primarily used for time series that exhibit a linear trend but without seasonality. Brown's method uses a double application of exponential smoothing to handle the trend, effectively applying exponential smoothing twice to the data series. It involves two different smoothed series that are centered at different points in time. This forecasting formula is based on an extrapolation of a line through the two centers.

Steps: Let

$s_0' = y_0, s_0'' = y_0$

$s_t' = \alpha y_t + (1-\alpha)s_{t-1}'$

$s_t'' = \alpha s_t' + (1-\alpha)s_{t-1}''$

$\hat{y}_{t+k} = a_t + kb_t$

where,

a_t : Estimated level at time t

$a_t = 2s_t' - s_t''$

b_t: Estimated trend at time t

$$b_t = \frac{\alpha}{1-\alpha}(s_t' - s_t'')$$

Brown's method simplifies the estimation of trend compared to Holt's method by not requiring a separate smoothing parameter for the trend. It achieves this by smoothing the data twice. This method is best suited for time series data with a linear trend but no seasonality.

Brown's method is computationally simpler than Holt's, which can make it appealing for certain applications, but it may not be as flexible if the data requires more nuanced trend modelling.

9.3.1.6 Triple exponential smoothing (Holt-Winters Smoothing)

Holt-Winters smoothing is a time series forecasting technique that extends exponential smoothing to capture seasonality. It is particularly useful for data that exhibits both a trend and seasonal patterns. The method comes in three forms:

1. Additive Model: Used when the seasonal variation is roughly constant over time.
2. Multiplicative Model: Used when the seasonal variation increases or decreases proportionally with the level of the series.
3. Damped Model: Applies a damping factor to the trend to make it less pronounced over time.

Holt-Winters is widely used in fields like finance, economics, and inventory management for predicting future values based on historical time series data, especially when the data shows clear seasonal patterns. This model is used when time series has both trend and seasonal components.

Let

$\{y_t\}$: Sequence of observations beginning at t=0.

L: Length of the cycle of seasonal change.

N: Number of complete cycles.

Two cases of seasonality:

Seasonality is (i) Multiplicative, (ii) Additive

	1	2	⋯	N	
1	y_1	y_{L+1}	⋯	$y_{L(N-1)+1}$	c_{10}
2	y_2	y_{L+2}	⋯	$y_{L(N-1)+2}$	c_{20}
⋮	⋮	⋮	⋱	⋮	⋮
L	y_L	y_{2L}	⋯	y_{LN}	c_{L0}
	A_1	A_2	⋯	A_N	

$$c_{i0} = \frac{1}{N}\sum_{j=1}^{N}\frac{y_{L(j-1)+i}}{A_j}; i = 1, \dots, L : \text{Initial estimates of seasonal indices}$$

$$A_j = \frac{1}{L}\sum_{i=1}^{L} y_{L(j-1)+i}; j = 1, \dots, N : \text{Average values of y's in the j-th cycle}$$

Initial trend estimate b:

$$b_0 = \frac{1}{L}\left(\frac{y_{L+1} - y_1}{L} + \frac{y_{L+2} - y_2}{L} + \cdots + \frac{y_{2L} - y_L}{L}\right)$$

We denote

s_t= smoothed statistic

α= smoothing factor of data; $0 < \alpha < 1$

b_t= best estimate of a trend at time t

β = trend smoothing factor; $0 < \beta < 1$

c_t= sequence of seasonal correction factor at time t

γ= seasonal change smoothing factor; $0 < \gamma < 1.$

Triple exponential smoothing formulas for Multiplicative Seasonality:

$$s_0 = y_0$$

$$s_t = \alpha \frac{y_t}{c_{t-L}} + (1-\alpha)(s_{t-1} + b_{t-1})$$

$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}$$

$$c_t = \gamma \frac{y_t}{s_t} + (1-\gamma)c_{t-L}$$

$$\hat{y}_{t+k} = (s_t + kb_t)c_{t-L+1+(k-1) \bmod L}$$

where, *a* and *n* are positive numbers. Further, *a modulo n (a mod n)* is the remainder of division of *a* by *n*.

Triple exponential smoothing formula for additive seasonality:

$$s_0 = y_0$$

$$s_t = \alpha(y_t - c_{t-L}) + (1-\alpha)(s_{t-1} + b_{t-1})$$

$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}$$

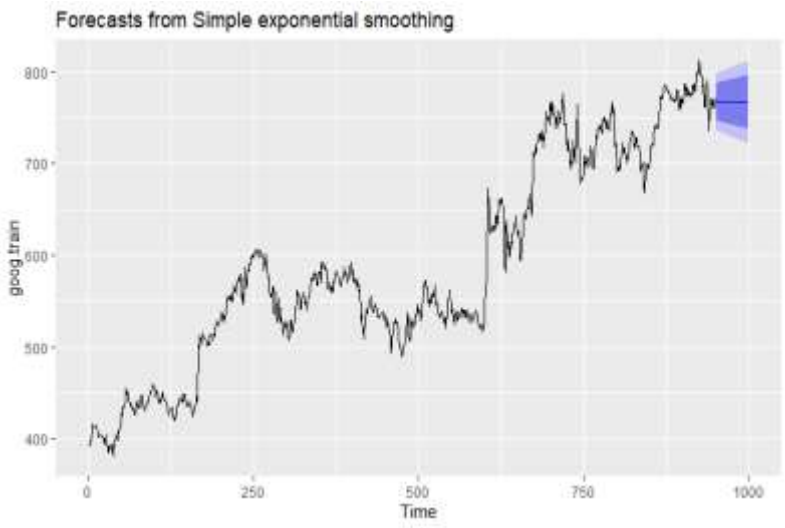
$$c_t = \gamma(y_t - s_{t-1} - b_{t-1}) + (1-\gamma)c_{t-L}$$

$$\hat{y}_{t+k} = s_t + kb_t + c_{t-L+1+(k-1) \bmod L}$$

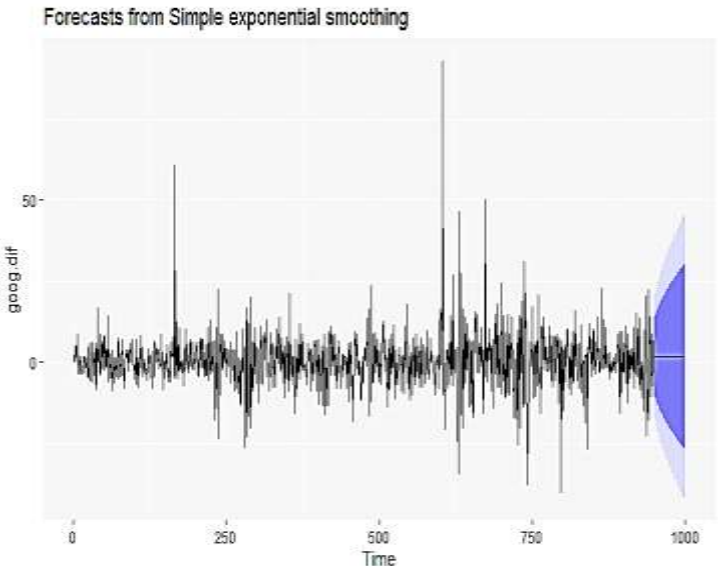
9.4 ***Numerical Examples***

We apply different exponential smoothing techniques to Google stock dataset with 1000 observations available in R-package. The dataset is divided in two groups, first 950 observations (training set) for exponential smoothing, and remaining 50 observations (test set) for checking the accuracy of forecasts.

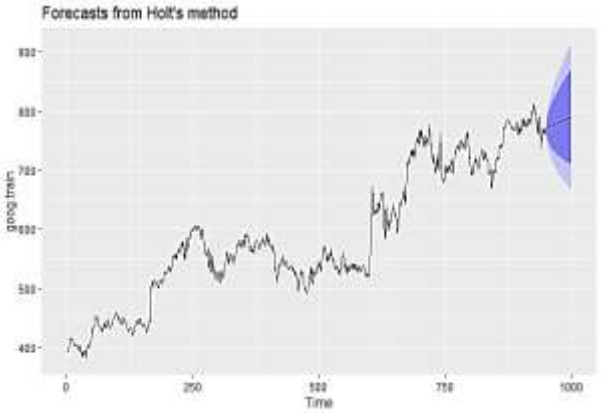
Simple Exponential Smoothing: A flatlined estimate is projected by simple exponential smoothing. The procedure is not capturing the trend present in the data.



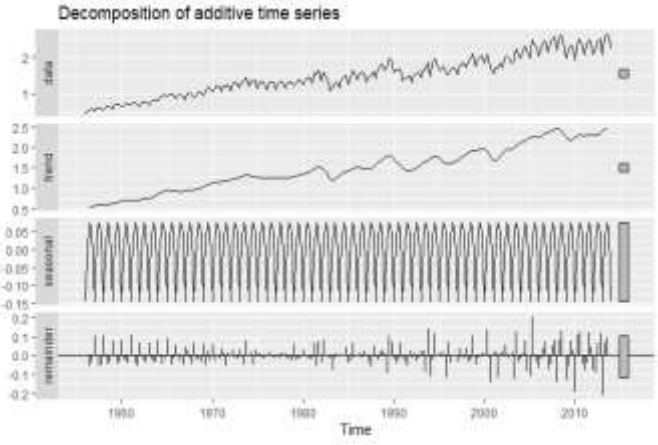
Removed Trend by differencing and used simple exponential smoothing. Best α is selected automatically. Since the data set was differenced, the procedure is forecasting differenced values.

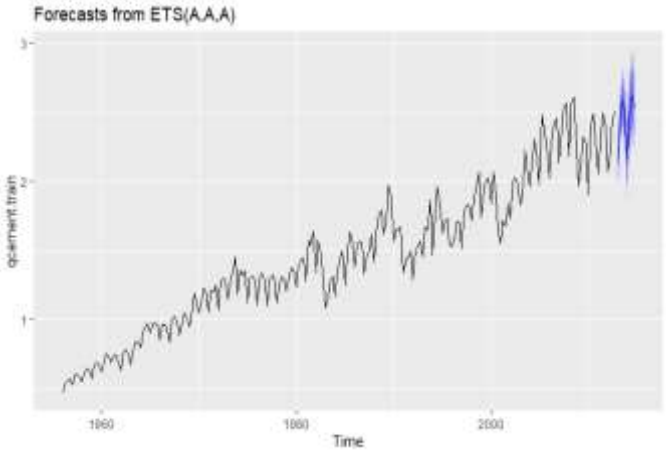


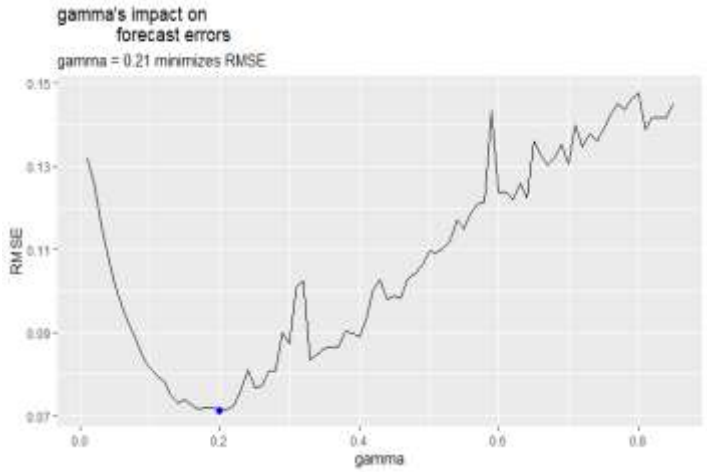
Holt Method takes care of trend. Best possible values of alpha and beta selected automatically.



Holt-Winter's Seasonal Method is used for data with both seasonal patterns and trends. Three smoothing parameters $\alpha, \beta, \text{and } \gamma$ are selected automatically.







9.4 Periodogram and Correlogram Analysis

Periodogram and correlogram analyses are techniques used in time series analysis to understand the frequency and autocorrelation properties of a dataset. Both are essential tools for identifying periodic patterns, trends, and the structure of the time series.

A periodogram is a tool used to estimate the spectral density of a time series. It provides a way to identify the dominant frequencies (or cycles) present in the data. The spectral density function reveals how the variance of the time series is distributed across different frequencies. The periodogram is computed by taking the Fourier transform of the time series data, which decomposes the data into its constituent frequencies.

Peaks in the periodogram indicate dominant frequencies in the time series. These are the frequencies at which the time series has significant periodic components. For example, if you see a peak at a frequency corresponding to one year, this suggests an annual cycle in the data.

A correlogram is a graphical representation of the autocorrelation function (ACF) of a time series. It shows the correlation of the time series with its own lagged values over different time lags. The correlogram shows the autocorrelation coefficients plotted against the lag. The significant autocorrelations at specific lags indicate that the time series has memory, meaning past values have an influence on future values.

A slowly decaying correlogram suggests a trend, while a cyclical pattern in the correlogram suggests seasonality or periodicity.

Focus:

- The periodogram focuses on the frequency domain, identifying cycles and periodic patterns in the data.
- The correlogram focuses on the time domain, showing how the data at one time point relates to data at other time points (lags).

Usage:

- Use the periodogram when you are interested in identifying specific cycles or frequencies in your data.
- Use the correlogram when you want to understand the persistence of patterns over time and the time lags at which the series is correlated with itself.

Interpretation:

- A periodogram is interpreted by identifying peaks at specific frequencies.
- A correlogram is interpreted by identifying significant autocorrelations at specific lags.

The Periodogram Analysis is ideal for detecting and understanding the frequency components within a time series whereas, the Correlogram Analysis is best for examining the autocorrelation structure and understanding the time-domain relationships within the data**.**

9.5.1 Periodogram Analysis

Consider a time series from which trend and seasonal effects have been eliminated. Let $u_t, (t = 1,2, \dots, n)$ represents the residual series. We want to know whether u_t contains a harmonic term with period μ. Consider the quantities

$$A = \frac{2}{n}\sum_{t=1}^{n} u_t \cos\frac{2\pi t}{\mu} (1)$$

And

$$B = \frac{2}{n}\sum_{t=1}^{n} u_t \sin\frac{2\pi t}{\mu} (2)$$

where *n* is the number of terms in the series. Let us write

$$R_\mu^2 = A^2 + B^2 \qquad (3)$$

R_μ^2 is known as the intensity corresponding to the trial period μ.

Let us consider a simple model, according to which u_t is composed of two components, one periodic with period λ and amplitude *a* and the other an irregular component, say b_t. Thus

$$u_t = \text{asin}\frac{2\pi t}{\lambda} + b_t \qquad (4)$$

The second component is assumed to be uncorrelated with the first or similar periodic terms.

Now,

$$A = \frac{2a}{n}\sum_t \sin\frac{2\pi t}{\lambda}\cos\frac{2\pi t}{\mu} + \frac{2}{n}\sum_t b_t \cos\frac{2\pi t}{\mu}$$

$$= \frac{2a}{n}\sum_t \sin\alpha t \cos\beta t \qquad \left(\text{putting } \alpha = \frac{2\pi}{\lambda}, \beta = \frac{2\pi}{\mu} \text{ and neglecting the second term}\right)$$

$$= \frac{a}{n} \sum_{t} \{\sin(\alpha - \beta) + \sin(\alpha + \beta)t\}$$

$$= \frac{a}{n} \left\{ \frac{\sin n \frac{(\alpha - \beta)}{2} \sin(n+1) \frac{(\alpha - \beta)}{2}}{\sin \frac{(\alpha - \beta)}{2}} + \frac{\sin n \frac{(\alpha + \beta)}{2} \sin(n+1) \frac{(\alpha + \beta)}{2}}{\sin \frac{(\alpha + \beta)}{2}} \right\}$$

Remembering that

$$\sum_{t=0}^{n-1} \sin(\alpha + \beta t) = \frac{\sin \frac{n\beta}{2}}{\sin \frac{\beta}{2}} \sin\left(\alpha + \frac{n-1}{2}\beta\right)$$

For large *n*, the second term is always small; the first term will also be small unless β tends to α, i.e. unless μ, the trial period, approaches the true period λ. Since

$$\frac{\sin \theta}{\theta} \to 1 \; as \; \theta \to 0$$

We have, if β tends to α, then

$$A = a \sin(n+1) \frac{(\alpha - \beta)}{2} \times \frac{\sin n \frac{(\alpha - \beta)}{2}}{n \frac{(\alpha - \beta)}{2}} \Big/ \frac{\sin \frac{(\alpha - \beta)}{2}}{\frac{(\alpha - \beta)}{2}}$$

$$\to a \sin(n+1) \frac{(\alpha - \beta)}{2} \qquad (5)$$

Similarly,

$$B \to a \cos(n+1) \frac{(\alpha - \beta)}{2} \; as \; \beta \to \alpha \; (6)$$

and is small otherwise, so that

$$R_{\mu}^2 \to a^2 \; when \; \beta \to \alpha \qquad (7)$$

i.e. when $\mu \to \lambda$, and is small otherwise.

We now take several trial periods μ around the true period λ, which may be guessed by plotting the data on a graph paper, and calculate R_μ^2 in each case. Finally, we draw a graph plotting R_μ^2 against μ. The diagram is called a periodogram, is a simple device for finding the true cyclical period λ in a time series by equating it to that value of μ for which R_μ^2 attains a maximum.

Similarly, if the cyclical component is composed of several periodic terms, say with periods $\lambda_1, \lambda_2, \dots, \lambda_k, R_\mu^2$ will remain small unless the trial period μ coincides with one of the true periods, in which case it attains a local maximum with value equal to the square of the amplitude of the periodic term concerned.

9.5.2 Correlogram

An autocorrelation (r_k) of order *k* is the correlation between $u_t \; and \; u_{t+k}$. From the original u_t series *n-k* pairs of values are obtained with a lag of period *k*.

Thus,

$$r_k = \frac{cov\,(u_t\,, u_{t+k})}{\{var\,(u_t)\;var\,(u_{t+k})\}^{1/2}}$$

$$= \frac{\frac{1}{n-k}\sum_{t=1}^{n-k} u_t\, u_{t+k} - \frac{1}{(n-k)^2}\sum_{t=1}^{n-k} u_t \sum_{t=1}^{n-k} u_{t+k}}{\left\{\frac{1}{n-k}\sum_{t=1}^{n-k} u_t^2 - \frac{1}{(n-k)^2}\left(\sum_{t=1}^{n-k} u_t\right)^2\right\}^{\frac{1}{2}}\left\{\frac{1}{n-k}\sum_{t=1}^{n-k} u_{t+k}^2 - \frac{1}{(n-k)^2}\left(\sum_{t=1}^{n-k} u_{t+k}\right)^2\right\}^{\frac{1}{2}}} \quad (8)$$

Obviously, we have $r_0 = 1$ and $r_{-k} = r_k$.

The diagram obtained by plotting r_k against *k* on graph paper and joining the points, each to the next, is called a correlogram. Theoretically, it can be demonstrated that the correlogram takes on significantly diverse forms in various scenarios.

(a) Correlogram of Moving Average:

When oscillatory movement is generated by an *m*-point simple moving average of a random component I_t, where $E(I_t) = 0, cov\,(I_t, I_{t'}) = 0\ and\ var\,(I_t) = \sigma^2$, we also know that

$$\rho_k = \begin{cases} 1 - \frac{k}{m} & for\ k \leq m \\ 0 & for\ k > m \end{cases} \quad (9)$$

ρ_k being the theoretical value of the serial correlation of order *k*.

Thus, the correlogram would be a straight line starting at (0,1) and ending at (m,0) and thereafter the correlogram would coincide with the *k*-axis. If the oscillations were generated by an *m*-point weighted moving average with weights $a_1, a_2, \ldots, a_m$ the correlogram would oscillate between the points (0, 1) and (m,0) and thereafter would coincide with the *k*-axis.

$$\rho_k = \begin{cases} \frac{\sum_{j=1}^{m-1} a_j a_{j+1}}{\sum_{j=1}^{m} a_j^2} & \text{for } k \leq m \\ 0 & \text{for } k > m \end{cases}$$

(b) Correlogram of oscillatory movement:

When the oscillatory movement is generated by the sum of a number of cyclical components represented by the sum of a number of harmonic terms with periods $\lambda_1, \lambda_2, \ldots$ it can be shown that ρ_k would also be the sum of a number of harmonic terms, not necessarily with same periods. If we take

$$u_t = \text{asin}\frac{2\pi t}{\lambda} + I_t, \quad (10)$$

$$E(u_t . u_{t+k}) = E(A \sin\theta t + I_t)\left(A \sin\overline{\theta t + k} + I_{t+k}\right)$$

$$= \frac{A^2}{n} \sum_{t=1}^{n} \sin\theta t \sin\overline{\theta t + k}$$

$$= \frac{A^2}{2n} \sum_{t=1}^{n} \left(\cos \theta k - \cos \theta \overline{2t + k} \right)$$

$$= \frac{A^2}{2} \cos \theta k - \frac{A^2}{2n} \frac{\cos \theta (k + n + 1) \sin n\theta}{\sin \theta}$$

$$\rightarrow \frac{A^2}{2} \cos \theta k \quad as\ n \rightarrow \infty$$

$$= B \cos \theta k, \text{ say.}$$

Similarly,

$$E(u_t^2) = \frac{A^2}{2} + var\ (I_t) = C, say.$$

So that

$$\rho_k = \frac{B}{C} \cos \theta k \qquad (11)$$

Hence the correlogram would be a strictly periodic sinusoidal curve. In this case, the correlogram will take a sinusoidal form, which will not degenerate to the *k*-axis after some fixed point and will not be damped.

(c) Correlogram of autoregressive series:

Let us consider autoregressive equation of the first and second orders. For the equation of the first order, viz. $u_{t+1} = \mu u_t + I_{t+1}$, called Markov's process

$$E\left(u_t (u_{t+1} - \mu u_t)\right) = E(u_t . I_{t+1}) = 0$$

gives

$$E(u_t . u_{t+1}) = \mu E(u_t^2)$$

so that

$$\rho_1 = \mu$$

again

$$E\{u_t(u_{t+k} - \mu u_{t+k-1})\} = E\{u_t . I_{t+k}\} = 0$$

gives

$$E(u_t . u_{t+k}) - \mu E(u_t . u_{t+k-1}) = 0$$

so that

$$\rho_k - \mu\rho_{k-1} = 0$$

or

$$\rho_k = \mu\rho_{k-1}$$
$$= \mu^k \rho_0$$
$$= \mu^k \qquad (12)$$

The correlogram would therefore now take on an exponential shape. Since μ needs to be smaller than 1 to prevent the time series from expanding to infinity, the curve would begin at (0,1), decrease quickly from there, and asymptotically gravitate towards the *k*-axis.

For the equation of second order, viz.

$$u_{t+1} = au_t + bu_{t-1} + I_{t+1}$$

called Yule's process, we have

$$u_t = au_{t-1} + bu_{t-2} + I_t$$

Multiplying both sides by u_{t-k} and taking expectations, we have

$$\rho_k - a\rho_{k-1} - b\rho_{k-2} = 0$$

The general solution of the above difference equation is given by

$$\rho_k = A_1 q_1^k + A_2 q_2^k \qquad (13)$$

A_1, A_2 being found from the initial conditions and q_1, q_2 are the roots of the equation $q^2 - aq - b = 0$, the characteristic equation of the process.

Case 1: $a^2 + 4b > 0$, roots are real

A_1, A_2 are found as follows

$$\rho_0 = 1 = A_1 + A_2$$

$$\rho_1 = a\rho_0 + b\rho_{-1} = a + b\rho$$

$$\text{or } \rho_1 = \frac{a}{1-b}$$

Again,

$$\rho_1 = A_1 q_1 + A_2 q_2$$

$$= A_1 q_1 + (1 - A_1) q_2$$

$$= A_1 (q_1 - q_2) + q_2$$

so that

$$A_1 = \frac{\frac{a}{1-b} - q_2}{q_1 - q_2}$$

$$= \frac{\left(\frac{q_1 + q_2}{1 + q_1 q_2} - q_2\right)}{q_1 - q_2}$$

$$= \frac{q_1 (1 - q_2^2)}{(1 + q_1 q_2)(q_1 - q_2)}$$

and

$$A_2 = 1 - A_1$$

$$= \frac{q_2 (1 - q_1^2)}{(1 + q_1 q_2)(q_1 - q_2)}$$

Case 2: $a^2 + 4b < 0$, roots are imaginary.

Let us write

$$q_1 = p(\cos\theta + i\sin\theta)\ and\ q_2 = p(\cos\theta - i\sin\theta)$$

$$\rho_k = p^k\{A_1(\cos\theta + i\sin\theta)^k + A_2(\cos\theta - i\sin\theta)^k\}$$
$$= p^k\{A^*\cos\theta k + B^*\sin\theta k\}$$

where

$$A^* = A_1 + A_2$$

and

$$B^* = i(A_1 + A_2)$$

$$For\ k = 0 \quad \rho_0 = 1 = A^*$$

$$k = 1 \quad \rho_1 = p(\cos\theta + B^*\sin\theta)$$

$$k = -1 \quad \rho_{-1} = \frac{1}{p}(\cos\theta - B^*\sin\theta)$$

Since

$$\rho_1 = \rho_{-1}$$

$$\Rightarrow B^* = \frac{1-p^2}{1+p^2}.\frac{\cos\theta}{\sin\theta}$$
$$= \frac{1-p^2}{1+p^2}.\cot\theta$$
$$= \cot\psi$$

Giving

$$\rho_k = p^k\{\cos\theta k + \cot\psi\sin\theta k\} = p^k\frac{\sin(\theta k + \psi)}{\sin\psi} \qquad (14)$$

In case 1 the correlogram starts at (0,1) and becomes asymptotic to the *k*-axis.

In case 2 correlogram will be oscillatory.

A correlogram is a powerful tool to detect patterns and dependencies in time series data, helping analysts choose the right models and validate their assumptions.

9.6 Self-Assessment Exercise

1. Define forecasting and explain its importance in decision-making processes.
2. Explain the concept of forecast accuracy and discuss common measures used to evaluate it (e.g., MSE, MAE, MAPE).
5. What is exponential smoothing, and how does it differ from simple moving averages?
6. Describe the concept of a smoothing constant in exponential smoothing. How does its value affect the forecast?
7. Explain the difference between single, double, and triple exponential smoothing.
8. What is adaptive smoothing? How does it address limitations of traditional exponential smoothing methods?
10. What is a periodogram, and how is it used in time-series analysis?
11. Describe the steps involved in constructing a periodogram for a given time series.
14. Define a correlogram and explain its role in time-series analysis.
15. How is the autocorrelation function (ACF) calculated, and what information does it provide?
16. Compare and contrast the correlogram with the periodogram in analysing time-series data.

9.7 Summary

This unit provides a comprehensive overview of forecasting methods used to predict future values based on historical data. The unit begins by introducing the

fundamental principles of forecasting and the importance of accurate predictions in decision-making across various domains.

A significant focus is placed on smoothing techniques, including exponential smoothing methods and adaptive smoothing methods. These approaches are explained in terms of their objectives, assumptions, and practical implementation for trend and seasonal components. The unit highlights the advantages of smoothing methods in handling time-series data with varying levels of volatility.

In addition to smoothing techniques, the unit explores frequency-domain analysis methods, such as the periodogram, for identifying dominant cycles in time-series data. The correlogram is introduced as a tool for analyzing autocorrelation, providing insights into the lag structures of data and guiding the selection of appropriate forecasting models.

By integrating these techniques, the unit equips learners with the tools to develop robust forecasts and assess the underlying patterns in time-series data, enabling informed decision-making in complex environments.

9.8 References

- Kendall, M.G. (1976*). Time Series,* 2nd Ed., Charles Griffin and Co Ltd., London and High Wycombe.
- Chatfield, C. (1980). *The Analysis of Time Series –An Introduction*, Chapman & Hall.
- Mukhopadhyay, P. (2011). *Applied Statistics*, 2nd Ed., Revised reprint, Books and Allied

1.9 Further Readings

- Goon, A. M., Gupta, M. K. and Dasgupta, B. (2003). *Fundamentals of Statistics*, 6th Ed., Vol II Revised, Enlarged.

- Gupta, S.C. and Kapoor, V.K. (2014). *Fundamentals of Mathematical Statistics*, 11th Ed., Sultan Chand and Sons.

- Montgomery, D. C. and Johnson, L. A. (1967). *Forecasting and Time Series Analysis*, 1st Ed. McGraw-Hill, New York.

UNIT 10 STRUCTURAL AND REDUCED FORM OF THE MODEL AND IDENTIFICATION PROBLEM

Structure

10.1 Introduction

A Generalized Linear Model (GLM) is a flexible framework for modeling relationships between a response variable and one or more predictor variables. It generalizes traditional linear regression by allowing the response variable to have distributions other than a normal distribution. GLMs consist of three main components:

1. Random Component

2. Systematic Component

3. Link Function

Instrumental Variables (IV) is a critical concept in econometrics used to address the problem of endogeneity in regression models. Endogeneity occurs when an explanatory variable is correlated with the error term, which can lead to biased and inconsistent parameter estimates. This typically arises from omitted variable bias, measurement error, or reverse causality. It is a powerful tool in econometrics for addressing endogeneity and obtaining unbiased parameter estimates. Understanding and applying IV methods correctly are crucial for drawing valid conclusions from econometric models. The selection of valid instruments is critical researchers must carefully justify their choices to ensure the robustness of their findings.

10.2 Objectives

After completing this course, there should be a clear understanding of:

- Review and analysis of GLM
- Generalized least square estimation
- Instrumental variables

10.3 Generalized Linear Model

Generalized Linear Model (GLM)*: *Detailed Analysis

A Generalized Linear Model (GLM) is a flexible generalization of ordinary linear regression that allows for the response variable (dependent variable) to have a non-normal distribution. It is particularly useful when the assumptions of linear regression, such as normality of residuals and homoscedasticity, do not hold.

GLM expands the framework of linear models by allowing for:

- Non-Normal Response Distributions (e.g., binomial, Poisson).

- Non-Constant Variance of Residuals.
- Link Functions that relate the linear predictor to the mean of the distribution.

Key Components of a GLM

1. Random Component: Specifies the probability distribution of the response variable. Common distributions include:

- Normal distribution (for continuous outcomes with constant variance).
- Binomial distribution (for binary or proportion data).
- Poisson distribution (for count data).
- Gamma distribution (for positive continuous data with non-constant variance).

2. Systematic Component: The systematic component of a GLM is a linear combination of the explanatory variables (predictors). It is expressed as:

$$\eta = X\beta$$

where, X is the matrix of predictor variables,

β is the vector of coefficients, and

η is the linear predictor (a linear function of the predictors).

3. Link Function: Transforms the expected value of the response variable to the linear predictor. Instead of assuming that the mean of the response is directly modeled as a linear function of the predictors. The link function relates the expected value of the response variable, $E(Y)$, to the linear predictor. Common link functions include:

- Identity link: $g(\mu) = \mu$ (used in linear regression)
- Logit link: $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$ (used in logistic regression for binary outcomes)

- Log link: $g(\mu) = \log(\mu)$ (used in Poisson regression for count data)
- Inverse link: $g(\mu) = \frac{1}{\mu}$ (used in Gamma regression for skewed continuous data)

The choice of the link function depends on the nature of the response variable.

General Framework of GLM

GLM is defined by the following:

- Response variable Y comes from an exponential family distribution.

Linear predictor $\eta = X\beta$.

- Link function $g(\mu)$ that connects the mean of the response to the linear predictor: \($g(\mu) = X\beta$.

Steps in GLM Analysis

1. Specify the Model:

- Choose the appropriate probability distribution for the response variable (e.g., binomial for binary, Poisson for counts).
- Define the systematic component (linear predictor) that includes the independent variables.
- Select the correct link function that fits the nature of the response.

2. Fit the Model:

- The parameters β are typically estimated using Maximum Likelihood Estimation (MLE). This approach maximizes the likelihood function (or equivalently, minimizes the negative log-likelihood).
- The MLE estimation process involves iteratively finding the best fit of the parameters using algorithms such as Iteratively Reweighted Least Squares (IRLS).

3. Assess Model Fit: Goodness-of-fit measures and diagnostic plots can be used to evaluate how well the model fits the data. Common measures include:

- Deviance: A generalization of the residual sum of squares for GLM. It compares the likelihood of the fitted model with that of a saturated model (a model that perfectly fits the data).
- Akaike Information Criterion (AIC): A metric that penalizes model complexity and helps select among different models.
- Pseudo: An extension of the R^2 statistic used in linear models, applicable to GLM.

4. Hypothesis Testing and Inference: Hypothesis tests are used to assess the significance of individual predictors:

- *Wald Test*: Tests whether a single coefficient is significantly different from zero.
- *Likelihood Ratio Test (LRT)*: Compares the goodness-of-fit of nested models.
- *Score Test (also called Lagrange Multiplier test)*: Tests for significance without fitting the full model.
- *Confidence intervals* for the estimated coefficients can also be computed to provide insight into the uncertainty of the estimates.

5. Model Validation: Residual analysis is crucial in checking the assumptions of the model. Common diagnostic tools include:

- Deviance residuals: Measure the difference between the observed and predicted values.
- Pearson residuals: Standardized residuals that help assess model fit.
- Leverage and Cook's Distance: To detect influential data points that may unduly affect the model.

Common GLM Applications

Some common types of GLMs and their applications include:

- Linear Regression (Normal distribution, identity link): For continuous outcomes.
- Logistic Regression (Binomial distribution, logit link): For binary outcomes, often used in classification tasks.
- Poisson Regression (Poisson distribution, log link): For count data, such as the number of occurrences in a fixed interval.
- Gamma Regression (Gamma distribution, inverse link): For positively skewed continuous data, such as response times or wait times.

Example: Logistic Regression as a GLM

Problem: Predict whether a patient has a disease based on age and smoking status.

Response Variable: Disease status (0 = no disease, 1 = disease).

Predictor Variables: Age and smoking status (1 = smoker, 0 = non-smoker).

Model:

Using a logistic regression (GLM with binomial distribution and logit link):

$$\log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1\{Age\} + \beta_2\{Smoking\ Status\}$$

Here, the coefficients β_1 and β_2 indicate how age and smoking status affect the probability of having the disease.

Advantages of GLM

- **Flexibility**: Allows modeling of different types of data (binary, count, continuous).
- **Unified Framework**: GLM provides a general framework for a variety of models.
- **Interpretability**: Coefficients in GLMs still retain interpretability similar to linear regression.

Limitations of GLM

- **Assumption of Linearity**: Even though the link function provides flexibility, the linear predictor assumes a linear relationship between the transformed mean and predictors.

- **Complexity in Model Fit**: Estimation using MLE may be computationally intensive, especially for large datasets or complicated models.

In conclusion, GLMs offer a powerful extension to linear models, accommodating a wider variety of data types and distributions. Their flexibility in handling non-normal responses and complex data structures makes them invaluable in modern statistical modeling.

10.4 Instrumental variables

Instrumental Variables (IV) is a statistical technique used primarily in econometrics and causal inference to address issues of endogeneity in regression models. Endogeneity arises when an independent variable is correlated with the error term in a regression equation, leading to biased and inconsistent estimates of the causal effect of the independent variable on the dependent variable. For example: Suppose you want to estimate the effect of education on earnings, but education is endogenous because higher earnings might motivate individuals to pursue more education.

Instrument: One might use the distance to the nearest college as an instrument. The idea is that those living closer to colleges may be more likely to obtain higher education, but the distance itself does not directly influence earnings.

Consider the model

$$y = X\beta + u; \qquad (1)$$

where,

$y: n \times 1$ vector of observations on dependent variable

$X: n \times k$ matrix of observations on k independent variable

$u: n \times 1,$ vector of disturbances with $u \sim (0, \sigma_u^2 I_n)$.

One of the basic assumptions is either X is non stochastic or even if stochastic, $E(X'u) = 0$.

Suppose $E(X'u) \neq 0$ (X and u correlated), then

$$E(b) = \beta + E[(X'X)^{-1}X'u] \neq \beta.$$

Further, if $plim(n^{-1}X'u) \neq 0$ (X and u correlated in limit), then

$$plim(b - \beta)$$
$$= plim[(n^{-1}X'X)^{-1}(n^{-1}X'u)] \neq \beta.$$

Hence b is a biased and inconsistent estimator of β.

Example: Let us consider the Measurement Error Model.

Suppose,

$$\tilde{y} = \beta_0 + \tilde{x}\beta_1$$

where,

Observed study variable is $\tilde{y}\ (n \times 1)$.

Explanatory variable is $\tilde{x}\ (n \times 1)$, and

$\tilde{y}$ be observed with additive measurement error

$$y = \tilde{y} + u$$

Let x be the vector of observed (proxy) values of the explanatory variables related to the true $\tilde{x}$ by

$$x = \tilde{x} + v$$

We can write the model as

$$y = \beta_0 + \tilde{x}\beta_1 + u$$
$$x = \tilde{x} + v$$

We assume that

$$E(u) = 0,$$
$$E(uu') = \sigma_u^2 I_n$$
$$E(v) = 0,$$
$$E(vv') = \omega I_n,$$
$$E(vu') = 0.$$

Then we obtain

$$y = \beta_0 + \beta_1 x + w$$

where, $w = (u - v\beta_1)$ is the composite disturbance term.

Then

$$E(x'w)$$
$$= E[x'(u - v\beta_1)]$$
$$= -E[x'v\beta_1]$$
$$= -E[(\tilde{x} + v)'v\beta_1]$$
$$= -E(v'v)\beta_1 \neq 0$$

Thus, x and w are correlated. If we apply OLS to estimate parameters of the model

$$y = \beta_0 + \beta_1 x + w$$

then the resulting estimator will be biased and inconsistent.

Example: Let us consider *Autoregressive Model:*

Consider the following autoregressive model

$$Y_t = \alpha + \beta Y_{t-1} + u_t; t = 2, \dots, n$$

where, Y_t is correlated with u_t, and Y_{t-1} is Explanatory variable and uncorrelated with u_t.

In deviation form

$$y_t = \beta y_{t-1} + u_t; t = 2, \dots, n,$$
$$y_t = Y_t - \bar{Y}$$

The OLS estimator of β is

$$\hat{\beta} = \frac{\sum_{t=2}^{n} y_t y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2}$$
$$= \beta + \frac{\sum_{t=2}^{n} u_t y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2}$$

Then

$$\mathrm{E}(\hat{\beta})$$
$$= \beta + E\left(\frac{\sum_{t=2}^{n} u_t y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2}\right) \neq \beta$$

and $y_{t-1} = Y_{t-1} - \bar{Y}$ involves $\bar{Y}$ and $\bar{Y}$ contains Y_t in it, which is correlated with u_t.

However, the estimator of β is consistent as

$$\mathrm{plim}\hat{\beta}$$
$$= \beta + \frac{Cov(Y_{t-1}, u_t)}{Var(Y_{t-1})} = \beta.$$

Example: Consider the Demand Supply Model

$$Q_t = \alpha + \beta P_t + u_t \text{ (Demand Equation)}$$

$$Q_t = \gamma + \delta P_t + v_t \text{ (Supply Equation)}$$

where,

Q_t: Quantity,

P_t: Price

Is P_t correlated with u_t (v_t)?

Now, we get P_t and Q_t by solving these two equations.

$$P_t = \frac{\gamma - \alpha}{\beta - \delta} + \frac{v_t - u_t}{\beta - \delta}$$

$$Q_t = \frac{\beta\gamma - \alpha\delta}{\beta - \delta} + \frac{\beta v_t - \delta u_t}{\beta - \delta}.$$

Hence, P_t is correlated with both u_t and v_t.

10.4.1 Instrumental variables (IV) Estimation

An instrumental variable Z is an additional variable used to estimate the causal effect of variable X on Y.

Instrumental Variables is a set of variables which are correlated with the explanatory variables in the model but uncorrelated with the composite disturbances, at least asymptotically, to ensure consistency.

Case I: Number of instrumental variables is equal to the number of explanatory variables.

Let Z be a $n \times k$ matrix of observations on k instrumental variables $Z_1, Z_2, \ldots, Z_k$ such that

(i) $plim(n^{-1}Z'u) = 0.$

(ii) $plim(n^{-1}Z'X) = \Sigma_{ZX}$ is a finite nonsingular matrix of full rank.

(iii) $plim(n^{-1}Z'Z) = \Sigma_{ZZ}$ is a finite nonsingular matrix.

If some of *X* variables are uncorrelated with u then these can be used to form some of the columns of *Z*. Pre-multiplying the model (1) by Z', we obtain

$$Z'y = Z'X\beta + Z'u$$

or

$$n^{-1}Z'y = (n^{-1}Z'X)\beta + (n^{-1}Z'u)$$

Then,

$$plim(n^{-1}Z'y) = plim(n^{-1}Z'X)\beta + plim(n^{-1}Z'u).$$

Hence

$$\beta = plim(n^{-1}Z'X)^{-1}[plim(n^{-1}Z'y) - plim(n^{-1}Z'u)]$$
$$= \Sigma_{ZX}^{-1}\Sigma_{Zy}$$

Replacing Σ_{Zy} and Σ_{ZX} by corresponding sample cross moments $n^{-1}Z'y$ and $n^{-1}Z'X$ respectively, we obtain

$$\mathrm{b_{IV}} = (Z'X)^{-1}Z'y \qquad (2)$$

Result 1*: The IV estimator b_{IV} is a consistent estimator of β. The asymptotic variance-covariance matrix of b_{IV} is given by*

$$plim\ (b_{IV} - \beta)(b_{IV} - \beta)'$$
$$= \frac{\sigma_u^2}{\mathrm{n}}\Sigma_{ZX}\Sigma_{ZZ}^{-1}\Sigma_{XZ} \qquad (3)$$

Proof: We have

$$b_{IV}$$
$$= (Z'X)^{-1}Z'(X\beta + u)$$
$$= \beta + (Z'X)^{-1}Z'u.$$

Hence

$$plim\ (b_{IV}) = \beta + plim[(n^{-1}Z'X)(n^{-1}Z'u)]$$
$$= \beta + \Sigma_{ZX}.0 = \beta$$

The asymptotic variance-covariance matrix of b_{IV} is given by

$$AsyVar(b_{IV}) = n^{-1}plim[(n^{-1}Z'X)^{-1}(n^{-1}Z'E(uu')Z)(n^{-1}X'Z)^{-1}]$$

$$= \frac{\sigma_u^2}{n}\left(plim(n^{-1}Z'X)\right)^{-1} plim(n^{-1}Z'Z)\left(plim(n^{-1}X'Z)\right)^{-1}$$

$$= \frac{\sigma_u^2}{n}\Sigma_{ZX}^{-1}\Sigma_{ZZ}\Sigma_{XZ}^{-1}.$$

We observe that as $n \to \infty$, the asymptotic variance covariance matrix of b_{IV} tends to 0. Thus, b_{IV} is a consistent estimator of β. Hence the result follows■

Case II: Number of instrumental variables is more than the number of explanatory variables:

Let Z be a $n \times l$ matrix of observations on $l(> k)$ variables such that

(i) $plim(n^{-1}Z'X) = \Sigma_{ZX}$: Finite matrix of full rank

(ii) $plim(n^{-1}Z'u) = 0$, *i.e.,* in limit Z is uncorrelated with u.

$(iii) plim(n^{-1}Z'Z) = \Sigma_{ZZ}$

Consider the model

$$y = X\beta + u$$

Pre multiplying the model by Z' gives

$$Z'y = Z'X\beta + Z'u \qquad (4)$$

Then covariance matrix of $Z'u$ is $\sigma_u^2 Z'Z$.

Applying GLS to (4), we get the IV estimator for β:

$b_{IV} = (X'PZX)^{-1}X'P_Z y;$
where $P_Z = Z(Z'Z)^{-1}Z'$■

***Result 2**: The IV estimator b_{IV} is a consistent estimator of β. Its asymptotic variance-covariance matrix is*

$$\text{AsyVar}(\text{b}_{\text{IV}}) = \frac{\sigma_u^2}{n}(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX})^{-1}$$

Proof: We have

$$b_{IV} = (X'PZX)^{-1}X'P_Z y$$
$$= \beta + (X'PZX)^{-1}X'P_Z u$$

Hence,

$$plim(b_{IV} - \beta)$$
$$= \left(plim(n^{-1}X'PZX)\right)^{-1} plim(n^{-1}X'P_Z u)$$
$$= \Big(plim(n^{-1}X'Z)\big(plim(n^{-1}Z'Z)\big)^{-1}plim(n^{-1}Z'X)\Big)^{-1}$$
$$\times plim(n^{-1}X'Z)\big(plim(n^{-1}Z'Z)\big)^{-1}plim(n^{-1}Z'u)$$
$$= 0$$

The asymptotic variance-covariance matrix of b_{IV} is obtained as

$$\text{AsyVar}(\text{b}_{\text{IV}}) = plim[(X'PZX)^{-1}X'P_Z E(uu')P_Z X(X'PZX)^{-1}]$$

$$= \frac{\sigma_u^2}{n}\left(plim(n^{-1}X'P_Z X)\right)^{-1}$$

$$= \frac{\sigma_u^2}{n}\{plim(n^{-1}X'Z)plim(n^{-1}Z'Z)^{-1}plim(n^{-1}Z'X)\}^{-1}$$

$$= \frac{\sigma_u^2}{n}(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX})^{-1}$$

As $n \to \infty$, asymptotic variance covariance matrix tends to zero. Thus b_{IV} is an unbiased and consistent estimator of β■

10.4.2 Interpretation of IV Estimator as a Two Stage Least Squares Estimator

The Instrumental Variables (IV) estimator is often implemented using the Two-Stage Least Squares (2SLS) method. This approach is particularly useful when addressing endogeneity issues in regression models. The IV estimator can be understood as a systematic way of dealing with endogeneity through 2SLS, which employs instrumental variables to extract the causal effect of an endogenous regressor on an outcome variable by addressing the correlation with the error term. The two stages highlight the separation of the estimation process into isolating the endogenous variable and then estimating the relationship of interest.

The IV estimator can be implemented using the 2SLS approach, involving two stages:

Stage 1: Consider the regression between X and Z

$$X = ZB + W$$

The OLS estimator of B is

$$\hat{B} = (Z'Z)^{-1}Z'X$$

Obtain a matrix of fitted values for X

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

Stage 2: Run regression between y and $\hat{X}$ to get the two-stage least squares estimator

$$b_{2SLS} = \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'y$$
$$= (X'^{P_Z}X)^{-1}X'^{P_Z}y$$
$$= b_{IV}$$

We obtain $\hat{X} = P_Z X$ and then run regression between y and $\hat{X}$.

IV technique allows the use of only that part of the variation in the predictor X that is not related with unobservable factors affecting both predictor and outcome.

This method allows to estimate the causal relationship between the outcome (y) and the predictor (X). The instrumental variable Z affects y only through its effect on X.

Suppose we want to investigate the relationship between depression (X) and smoking (y). Lack of job opportunities (Z) could lead to depression, but it is only associated with smoking through its association with depression. It is not direct correlated with smoking. Z can be used as an instrumental variable.

10.4.3 Choice of Instrumental Variables

We cannot use the actual data to find Instrumental Variables. One has to rely on knowledge about the model's structure and the (economic) theory behind the experiment.

(i) Z should not be affected by other variables in the system $(Cov(Z,u) = 0)$

(ii) Z should correlate with X, $(Cov(Z,X) \neq 0)$

(iii)Weak correlations lead to misleading estimates for parameters and standard errors.

Some of the X variables are uncorrelated with u and used in instrument. Partition X and Z as

$X = [X_1 \;\; X_2]$,

$Z = [X_1 \;\; Z_2]$

where,

$X_1 : n \times r \; (r < k)$, uncorrelated with u.

$X_2 : n \times (k - r)$, correlated with u.

$Z_2 : n \times (l - r)$, instrumental variables

Then

$\hat{X} = [X_1 \;\; \hat{X}_2]$,

Here

X_1: Instruments for themselves, and

$\hat{X}_2 = Z(Z'Z)^{-1}Z'X_2$: Remaining regressors

How many instrumental variables to use?

Minimum number is $k\ (l \geq k)$. Asymptotic efficiency increases with l but finite sample bias also increases. If $l = n,$ then

$$P_Z = Z(Z'Z)^{-1}Z'$$
$$= ZZ^{-1}Z'^{-1}Z'$$
$$= I_n$$

Then

$$b_{IV} = (X'X)^{-1}X'y$$
$$= b\ (\text{OLS estimator}),$$

which is biased.

Then the m^{th} moment of IV estimator exists if $m < l - k + 1$. Thus for $l = k,$ even the mean does not exist. With one more instrument, mean exists but variance does not.

10.4.4 Measurement Error Model

A measurement error model is particularly valuable for addressing bias that arises from inaccuracies in observable variables, especially in regression analysis. When there is measurement error in the independent variables, traditional regression techniques may produce biased and inconsistent estimates.

Causes behind Measurement Errors

- Taste, education, etc. are not measurable and some dummy variables are defined and observed.

- Some quantitative variables are observed with measurement error. For example, age is generally reported in complete years. Income reported in multiples of hundred.
- Some unobservable variable represented by some closely related proxy variable. For example, the level of education is measured by the number of years of schooling.
- Some qualitative variables measured by closely related quantitative variable. For example, intelligence is measured by intelligence quotient (IQ) scores.

In all the examples, the variables are observed with some error. The difference between the observed and true values of the variable is called as measurement error or errors-in-variables.

Disturbance term is defined as the i*nfluence of various explanatory variables that have not been included is the relation and Measurement errors is defined as the imperfect measure of true variables.*

True relationship between observed study variable $\tilde{y}\,(n \times 1)$ and explanatory variables $\tilde{X}(n \times k)$:

$$\tilde{y} = \tilde{X}\beta$$

Let $\tilde{y}$ be observed with additive measurement error

$$y = \tilde{y} + u$$

Let X be the matrix of observed (proxy) values of the explanatory variables related to the true $\tilde{X}$ by

$$X = \tilde{X} + V$$

Alternatively, we can write the model as

$$y = \tilde{X}\beta + u \tag{5}$$

$$X = \tilde{X} + V \tag{6}$$

We assume that

$$E(u) = 0, E(uu') = \sigma_u^2 I_n$$
$$E(V) = 0, E(V'V) = \Omega,$$
$$E(V'u) = 0.$$

Combining (5) and (6) we obtain

$$y = X\beta + w \tag{7}$$

$w = (u - V\beta)$ is called the composite disturbance term.

In model (7)

$$\begin{aligned} E(X'w) &= E[X'(u - V\beta)] \\ &= -E[X'V\beta] \\ &= -E\left[\left(\tilde{X} + V\right)'V\beta\right] \\ &= -E(V'V)\beta \neq 0 \end{aligned}$$

Thus X and w are correlated.

***Result 3**: The OLS estimator b is biased estimator of β.*

Proof: We can write the OLS estimator of β as

$$b = (X'X)^{-1}X'y$$
$$= (X'X)^{-1}X'[X\beta + w]$$
$$= \beta + (X'X)^{-1}X'(u - V\beta).$$

Hence

$$\mathrm{E}(b-\beta)$$
$$= (X'X)^{-1}[E(X'u) - E(X'V)\beta]$$
$$= -(X'X)^{-1}E(V'V)\beta$$

Hence OLS estimator b is biased. The reason is the correlation between the data matrix X and the composite disturbance term $(u - V\beta)$.

Large sample properties of OLS Estimator

We assume that

(a) The measurement errors V in $\tilde{X}$ are uncorrelated in limit with $\tilde{X}$, i.e., $plim\left(\frac{1}{n}\tilde{X}'V\right) = 0$. Hence,

$$plim(n^{-1}X'X)$$
$$= plim\left(n^{-1}(\tilde{X}+V)'(\tilde{X}+V)\right)$$
$$= \Sigma + \Omega$$

where

$$\Sigma = plim\left(\frac{1}{n}\tilde{X}'\tilde{X}\right),$$
$$\Omega = plim\left(\frac{1}{n}V'V\right),$$

(b) $plim\left(\frac{1}{n}X'u\right) = 0,\ plim\left(\frac{1}{n}\tilde{X}'u\right) = 0$.

Result 4*: The OLS estimator b is inconsistent estimator of β.*

Proof: Utilizing assumptions (a) and (b), we get

$$plim\left(\frac{1}{n}X'V\right)$$

$$= plim\left(\frac{1}{n}(\tilde{X}+V)'V\right)$$

$$= \Omega$$

Therefore

$$plim(b-\beta) = plim[(X'X)^{-1}X'u - (X'X)^{-1}X'V\beta]$$

$$= plim\left[\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'u - \left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'V\right)\beta\right]$$

$$= \left(plim\left(\frac{1}{n}X'X\right)\right)^{-1}\left\{plim\left(\frac{1}{n}X'u\right) - plim\left(\frac{1}{n}X'V\right)\beta\right\}$$

$$= -(\Sigma+\Omega)^{-1}\Omega\beta \neq 0$$

Thus, *b* is an inconsistent estimator of β ■

Here residual sum of squares $(y-X\beta)'(y-X\beta) = (u-V\beta)'(u-V\beta)$ involves β. Thus, b is not obtained by minimizing it.

Example: Consider

$$\tilde{y}_i = \beta_0 + \beta_1\tilde{x}_i, i = 1,2,\dots,n$$

where,

$$y_i = \tilde{y}_i + u_i$$

$$x_i = \tilde{x}_i + v_i \ .$$

Define

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \tilde{X} = \begin{pmatrix} 1 & \tilde{x}_1 \\ 1 & \tilde{x}_2 \\ \vdots & \vdots \\ 1 & \tilde{x}_n \end{pmatrix}, V = \begin{pmatrix} 0 & v_1 \\ 0 & v_2 \\ \vdots & \vdots \\ 0 & v_n \end{pmatrix}$$

We assume

$$plim\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\right)=\mu,$$

$$plim\left(\frac{1}{n}\sum_{i=1}^{n}(\tilde{x}_i-\mu)^2\right)=\sigma_x^2$$

Then

$$\Sigma_{xx}=plim\left(\frac{1}{n}\tilde{X}'\tilde{X}\right)$$

$$=plim\begin{pmatrix}1 & \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\\ \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i & \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i^2\end{pmatrix}$$

$$=\begin{pmatrix}1 & \mu\\ \mu & \sigma_x^2+\mu^2\end{pmatrix}.$$

Now

$$\Omega=plim\left(\frac{1}{n}V'V\right)=\begin{pmatrix}0 & 0\\ 0 & \omega\end{pmatrix}.$$

$$plim(b-\beta)=-(\Sigma_{xx}+\Omega)^{-1}\Omega\,\beta$$

$$plim\begin{pmatrix}b_0-\beta_0\\ b_1-\beta_1\end{pmatrix}$$

$$=-\begin{pmatrix}1 & \mu\\ \mu & \sigma_x^2+\mu^2+\omega\end{pmatrix}^{-1}\begin{pmatrix}0 & 0\\ 0 & \omega\end{pmatrix}\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$$

$$=-\frac{1}{(\sigma_x^2+\omega)}\begin{pmatrix}\sigma_x^2+\mu^2+\omega & -\mu\\ -\mu & 1\end{pmatrix}\begin{pmatrix}0\\ \beta\omega\end{pmatrix}$$

$$=\begin{pmatrix}\frac{\omega}{\omega+\sigma_x^2}\mu\beta\\ -\frac{\omega}{\sigma_x^2+\omega}\beta\end{pmatrix}.$$

Both b_0, b_1 are Biased and inconsistent. Measurement error in $\tilde{x}_1$ affects the estimator of intercept term also.

Different Forms of Measurement Errors:

Consider the model

$\tilde{y}_i = \beta_0 + \beta_1 \tilde{x}_i, i = 1,2, \ldots, n$

$y_i = \tilde{y}_i + u_i$

$x_i = \tilde{x}_i + v_i$.

The three forms of measurement error models:

(i) **Functional Form**: when $\tilde{x}_i's$ are unknown constants.

(ii) **Structural Form**: when $\tilde{x}_i's$ are *iid* random variables, say, with mean μ_x and variance σ_x^2. For $\sigma_x^2 = 0$, it reduces to functional form.

(iii) **Ultrastructural Form**: when $\tilde{x}_i's$ are independently distributed random variables with different means, say μ_{xi} and variance σ_x^2. Both functional form and structural form are special cases of this form.

Instrumental variables (IV) Estimation

Let Z be $n \times l$ matrix of observations on l instrumental variables such that

(i) $plim(n^{-1}Z'X) = \Sigma_{ZX}$: Finite matrix of full rank

(ii) $plim(n^{-1}Z'u) = 0$, *i.e.*, in limit Z is uncorrelated with u

(iii) $plim(n^{-1}Z'V) = 0$

(iv) $\Sigma'_{ZX} = \Sigma_{XZ} = plim(n^{-1}X'Z)$

Consider the model:

$$y = X\beta + w, \qquad w = (u - V\beta)$$

Pre multiplying the model by Z' gives

$$Z'y = Z'X\beta + Z'w \tag{8}$$

Then the variance-covariance matrix of $Z'w$ is $\sigma_u^2(Z'Z)^{-1}$. Applying GLS to (8), we get the IV estimator for β:

$$b_{IV} = (X'PZX)^{-1}X'P_Z y;\ P_Z = Z(Z'Z)^{-1}Z'.$$

Result 5*: The IV estimator b_{IV} is a consistent estimator of β. The asymptotic variance-covariance matrix of b_{IV}* is

$$AsyVar(b_{IV}) = \frac{\sigma_u^2}{n}(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX})^{-1} \tag{9}$$

Proof: We have

$$b_{IV} = (X'PZX)^{-1}X'P_Z y$$
$$= \beta + (X'PZX)^{-1}X'P_Z w$$

$$plim(b_{IV} - \beta) = plim\{(X'PZX)^{-1}X'P_Z w\}$$

$$= plim(n^{-1}X'P_Z X)^{-1}(n^{-1}X'P_Z u - n^{-1}X'P_Z V\beta)$$
$$= 0$$

$$AsyVar(b_{IV})$$
$$= plim\{(X'P_Z X)^{-1}X'Z(Z'Z)^{-1}E(Z'ww'Z)(Z'Z)^{-1}Z'X(X'P_Z X)^{-1}\}$$
$$= \frac{\sigma_u^2}{n}plim(n^{-1}X'P_Z X)^{-1}$$
$$= \frac{\sigma_u^2}{n}(\Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX})^{-1},$$

which tends to zero as $n \to \infty$■

10.4.5 Choice of instrument

Consider measurement error model with one explanatory variable:

$$y_i = \beta_0 + \beta_1 x_i + w_i,$$
$$w_i = u_i - \beta_1 v_i, i = 1,2, \dots, n.$$

(i) ***Wald's method****:*

Arrange X_i in ascending or descending order and find median. Define

$$Z_i = \begin{cases} 1 & \text{if } X_i \geq \text{median} \\ -1 & \text{if } X_i < \text{median} \end{cases}$$

and,

$$X' = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix},$$
$$Z' = \begin{pmatrix} 1 & \cdots & 1 \\ z_1 & \cdots & z_n \end{pmatrix}$$

Let two groups of $x_i's$ and $y_i's$, (i) 1^{st} group of $x_i's$ below the median and (ii) 2^{nd} group of $x_i's$ above the median. where,

$\bar{x}, \bar{y}$ are means of $x_i's$ and $y_i's$ respectively and

$\bar{x}_1, \bar{y}_1$: means for 1^{st} group,

$\bar{x}_2, \bar{y}_2$: means for 2^{nd} group

$$Z'X = \begin{pmatrix} n & n\bar{x} \\ 0 & \frac{n}{2}(\bar{x}_2 - \bar{x}_1) \end{pmatrix}$$

$$Z'y = \begin{pmatrix} n\bar{y} \\ \frac{n}{2}(\bar{y}_2 - \bar{y}_1) \end{pmatrix}$$

$$b_{IV} = \begin{pmatrix} b_{0IV} \\ b_{1IV} \end{pmatrix}$$

$$= \begin{pmatrix} n & n\bar{x} \\ 0 & \frac{n}{2}(\bar{x}_2 - \bar{x}_1) \end{pmatrix}^{-1} \begin{pmatrix} n\bar{y} \\ \frac{n}{2}(\bar{y}_2 - \bar{y}_1) \end{pmatrix}$$

$$= \frac{2}{(\bar{x}_2 - \bar{x}_1)} \begin{pmatrix} \frac{\bar{x}_2 - \bar{x}_1}{2} & -\bar{x} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{y} \\ \frac{\bar{y}_2 - \bar{y}_1}{2} \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y} - \left(\frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}\right)\bar{x} \\ \left(\frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}\right) \end{pmatrix}$$

Thus

$$b_{1IV} = \left(\frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}\right),$$

$$b_{0IV} = \bar{y} - b_{1IV}\bar{x}$$

The estimators are consistent but have large sampling variance.

(ii) ***Bartlett's method***:

Divide observations into three groups after arranging in increasing or decreasing order, say $\frac{n}{3}$ sized groups.

$$Z_i = \begin{cases} 1 & \text{if } X_i \text{ is in upper group} \\ 0 & \text{if } X_i \text{ is in middle group} \\ -1 & \text{if } X_i \text{ is in lower group} \end{cases}$$

Discard the observations in the middle group and means of $y_i's$ and $x_i's$ in bottom group is $\bar{y}_1, \bar{x}_1$ and means of $y_i's$ and $x_i's$ in top group is $\bar{y}_3, \bar{x}_3$.

Then

$$b_{1IV} = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1},$$

$$b_{0IV} = \bar{y} - b_{1IV}\bar{x}.$$

Then the estimators are consistent.

(iii) ***Durbin's method****:*

Rank $X_{ij}'s\ (j = 1, \ldots, n)$. Let r_j be the rank of X_{ij} we take $Z_{ij} = r_j$.

Then, if we have one explanatory variable

$$b_{1IV} = \frac{\sum_{i=1}^{n} Z_i(y_i - \bar{y})}{\sum_{i=1}^{n} Z_i(x_i - \bar{x})},$$

$$b_{0IV} = \bar{y} - b_{1IV}\bar{x}.$$

For more than one explanatory variable, one may choose the instrument as the rank of that variable. The estimator uses more information and expected to perform better than other grouping methods.

In general, the instrumental variable estimators may have fairly large standard errors in comparison to ordinary least square estimators which is the price paid for inconsistency. However, inconsistent estimators have little appeal.

10.5 Self-Assessment Exercise

1. What are the key components of a Generalized Linear Model (GLM)?
2. Explain the role of the link function in a GLM. Provide examples of commonly used link functions.
3. Discuss how GLMs extend ordinary linear regression to handle non-normal response variables.
6. What is the primary objective of Generalized Least Squares (GLS) estimation?
7. Explain how GLS addresses issues of heteroscedasticity and autocorrelation in regression analysis.
8. What are the key assumptions underlying GLS estimation?

9. Discuss a scenario where GLS would provide more efficient estimates than Ordinary Least Squares (OLS).

11. What is endogeneity, and why does it pose a problem in regression analysis?

12. Define instrumental variables and explain the criteria for a valid instrument.

13. Outline the steps involved in Two-Stage Least Squares (2SLS) estimation using instrumental variables.

16. Explain the concept of consistency in the context of estimators.

17. Under what conditions is an instrumental variable estimator consistent?

18. Why does the use of invalid instruments lead to inconsistent estimates?

19. How does instrument relevance and exogeneity ensure the consistency of IV estimators?

20. Derive the asymptotic variance of IV estimators.

10.6 Summary

By completing this unit, you will gain an understanding of the following concepts in econometrics:

Generalized Linear Model (GLM): A GLM is an extension of traditional linear regression that allows for response variables to have error distributions other than the normal distribution. We observe that GLMs consist of three main components: a random component specifying the distribution of the response variable (e.g., Poisson, binomial), a systematic component that includes a linear predictor, and a link function that connects the mean of the response variable to the linear predictor. GLMs are useful for modelling various types of data, including count data, binary outcomes, and other non-normally distributed variables.

Instrumental Variables (IV): Instrumental variables are used in regression models to address endogeneity issues, which occur when explanatory variables are correlated with the error term, potentially leading to biased and inconsistent estimates. An instrumental variable must satisfy two key conditions: it must be correlated with the endogenous explanatory variable

(relevance condition) and uncorrelated with the error term (exogeneity condition). Common applications of IVs include addressing omitted variable bias, simultaneity, and measurement error.

Estimation of Instrumental Variables: The most common method for estimating models with instrumental variables is Two-Stage Least Squares (2SLS). In 2SLS, the first stage involves regressing the endogenous variables on the instruments to obtain predicted values, and the second stage regresses the dependent variable on these predicted values.

Consistency Properties of Instrumental Variables Estimators: Instrumental variable estimators are consistent if the instruments are valid, meaning they meet the relevance and exogeneity conditions. Consistency implies that as the sample size grows, the IV estimator converges to the true value of the parameter.

5. Asymptotic Variance of Instrumental Variable Estimators: The asymptotic variance of an IV estimator is the variance of the estimator as the sample size approaches infinity. Understanding the asymptotic variance is important for constructing confidence intervals and conducting hypothesis tests in IV regression. IV estimators generally have larger asymptotic variance compared to ordinary least squares (OLS) estimators, reflecting the uncertainty added by the use of instruments to address endogeneity.

10.7 References

- Baltagi, Badi H. (2021): Econometric Analysis of Panel Data, Springer.
- Gujarati, D. and Guneshker, S. (2007). *Basic Econometrics*, 4th Ed., McGraw Hill Companies.
- Johnston, J. and J. Dinardo (1997). *Econometric Methods*, 2nd Ed., McGraw Hill International.

10.8 Further Readings

- Judge, G.G., W.E. Griffiths, R.C. Hill Lütkepohl and T. C. Lee (1985). The theory and practice of econometrics, Wiley.

- Koutsoyiannis, A. (2004). *Theory of Econometrics*, 2 Ed., Palgrave Macmillan Limited.

- Maddala, G.S. and Lahiri, K. (2009). *Introduction to Econometrics*, 4 Ed., John Wiley & Sons.

- Ullah, A. and Vinod, H.D. (1981). Recent advances in Regression Methods, Marcel Dekker.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 303N /MASTAT – 303N

Econometrics

Block: 3 Advance Econometrics

Unit – 11 : Autoregressive Process

Unit – 11 : Vector Autoregressive Process

Unit – 11 : Granger Causality

Unit – 11 : Cointegration

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics, D. D. U. Gorakhpur University, Gorakhpur.

Prof. Shruti **Member-Secretary**
Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Dr. Sandeep Mishra **Writer**
Department of Statistics (***Unit 11, 12, 13 & 14***)
Shahid Rajguru College of Applied Sciences for Women
University of Delhi, New Dehi

Prof. Anoop Chaturvedi **Editor**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Shruti **Course Coordinator**
School of Sciences,
U. P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 303N/ MASTAT – 303N **ECONOMETRICS**

First Edition: ***July 2024***
ISBN :

 Printed and Published by Col. Vinay Kumar, Registrar, Uttar Pradesh Rajarshi Tandon Open University, 2024.
Printed By:

Block & Units Introduction

The present SLM on ***Econometrics*** consists of fourteen units with three blocks.

The ***Block - 3 – Advance Econometrics,*** is the third block, which is divided into four units.

The ***Unit – 11 - Autoregressive Process***, deals with the Moving average (MA), Auto regressive (AR), ARMA and ARMA models, Box-Jenkins models, estimation of ARIMA model parameters, auto covariance and auto correlation function.

The ***Unit – 12 - Vector Autoregressive Process***, deals with the Multivariate time series process and their properties, vector autoregressive (VAR), Vector moving average (VMA) and vector autoregressive moving average (VARMA) process.

The ***Unit – 13- Granger Causality***, deals with the Granger causality, instantaneous Granger causality and feedback, characterization of casual relations in bivariate models, Granger causality tests, Haugh-Pierce test, Hsiao test.

The ***Unit – 14- Cointegration***, deals with the Cointegration, Granger representation theorem, Bivariate cointegration and cointegration tests in static model.

At the end of every block/unit the summary, self-assessment questions and further readings are given.

UNIT 11 AUTOREGRESSIVE PROCESS

11.1 Introduction

An autoregressive (AR) process is a type of statistical model used for analyzing and forecasting time series data. In an autoregressive model, the current value of a time series is expressed as a linear combination of its past values and a stochastic error term. Before proceeding further, we need to know the following terms.

***Stochastic Process**:* A stochastic process is a family of random variables $\{Y(t): t \in T\}$ where T denotes the time points at which the process is defined. For a particular $t \in T$, let S be the set of all possible values of $Y(t)$. Where, S is called the *state space*. $Y(t)$ is a random variable taking value in State space S. Usually we denote random variable by $Y(t)$ if T is continuous and by Y_t if T is discrete.

- Stochastic Process $\{Y(t): t \in T\}$ evolves in time according to probabilistic laws and provide a *probability model* for the analysis of time series.
- The infinite set of all possible time series is called the *ensemble.*
- Every member of ensemble is a particular *realization*

An observed time series is a particular realization of infinite set of time series, which might have been observed. *The objective is the evaluation of the statistical properties of the probability model, which generated the observed time series.*

***Descriptive Measures**:* Describing Marginal behavior of $\{Y_t; t \in T\}$ at a particular time point t

Mean function: $\mu(t) = E(Y_t)$

Variance function: $\sigma^2(t) = E[Y_t - \mu(t)]^2$

Measure of extent of dependence between Y_t *and* Y_{t+k}

- Autocovariance function (ACVF):

$$\gamma_k(t) = Cov(Y_t, Y_{t+k}) = E[\{Y_t - \mu(t)\}\{Y_{t+k} - \mu(t+k)\}]$$

$$\gamma_0(t) = \sigma^2(t).$$

- Auto correlation function (ACF) of lag k:

$$\rho_k(t) = \frac{\gamma_k(t)}{[\sigma_Y^2(t)\sigma_Y^2(t+k)]^{\frac{1}{2}}}, \rho_0(t) = 1.$$

Suppose

$\mu(t) = \mu \,\forall\, t, \sigma_Y^2(t) = \sigma_Y^2 \,\forall\, t$, $\gamma_k(t) = \gamma_k$: depends only on lag k then

ACF: $\rho_k = \frac{\gamma_k}{\gamma_0}$

ACF satisfies the following properties:

(i) For a stationary process $\rho_k = \rho_{-k}$.

(ii) $|\rho_k| \leq 1$.

(iii) Non uniqueness: A stationary normal process is completely determined by its mean, variance and ACF. It is always possible to obtain several non-normal processes with same ACF.

The ACF of purely random process or Gaussian white noise is given by

$$\rho_k = \begin{cases} 1 \; if \;\; k = 0 \\ 0 \; if \;\; k \neq 0 \end{cases}$$

Result: Let $\{Y_t\}$ *be a time series with* $E(Y_t) = \mu$. *Then ACF* ρ_k *is the value of a which minimizes*

$E[(Y_t - \mu) - a(Y_{t-k} - \mu)]^2$.

Sample ACF and ACVF of lag k: Let $y_1, y_2, \dots, y_n$: observed timeseries form (*n*-1) pairs $(y_1, y_2), (y_2, y_3), \dots, (y_{n-1}, y_n)$ then Sample autocorrelation coefficient (or serial autocorrelation coefficient) of lag one is

$$r_1 = \frac{\sum_{t=1}^{n-1}(y_t - \bar{y}_{(1)})(y_{t+1} - \bar{y}_{(2)})}{[\sum_{t=1}^{n-1}(y_t - \bar{y}_{(1)})^2 \sum_{t=2}^{n}(y_t - \bar{y}_{(2)})^2]^{\frac{1}{2}}}$$

where $\bar{y}_{(1)} = \frac{1}{n-1}\sum_{t=1}^{n-1} y_t \; ; \; \bar{y}_{(2)} = \frac{1}{n-1}\sum_{t=2}^{n} y_t$.

Let $\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$.

Since $\bar{y}_{(1)} \approx \bar{y}_{(2)} \approx \bar{y}$

$$\frac{1}{n-1}\sum_{t=1}^{n-1}(y_t - \bar{y}_{(1)})^2 \approx \frac{1}{n-1}\sum_{t=2}^{n}(y_t - \bar{y}_{(2)})^2 \approx \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2$$

r_1 is usually approximated by $r_1 = \frac{n\sum_{t=1}^{n-1}(y_t - \bar{y})(y_{t+1} - \bar{y})}{(n-1)\sum_{t=1}^{n}(y_t - \bar{y})^2}$

For large n, $n - 1 \approx n$ and r_1 can be approximated as

$$r_1 = \frac{\sum_{t=1}^{n-1}(y_t - \bar{y})(y_{t+1} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}.$$

Sample ACF of lag k: Let (*n-k*) pairs is $(y_1, y_{k+1}), (y_2, y_{k+2}), \dots, (y_{n-k}, y_n)$, then

$$c_k = \frac{1}{n-k}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y}) \approx \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})$$

Sample ACF of lag k: $r_k = \frac{\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = \frac{c_k}{c_0}$

Correlogram: The graph of r_k against k is called the correlogram. The autocorrelation function plays an important role in model identification.

Another tool, which is used in model identification, is *partial autocorrelation function*.

Partial Autocorrelation Function (PACF): The PACF of order k, say α_k, is the partial correlation coefficient between Y_t and Y_{t-k} conditional on intermediate values of the process.

- α_k is the autocorrelation between Y_t and Y_{t-k} removing the linear dependence of Y_t and Y_{t-k} on $Y_{t-1}, \dots, Y_{t-k+1}$.
- $e_t^{(k-1)}$: Least squares residual of linear regression between Y_t and $Y_{t-1}, \dots, Y_{t-k+1}$.
- $e_{*,t}^{(k-1)}$: Least squares residual of linear regression between Y_{t-k} and $Y_{t-1}, \dots, Y_{t-k+1}$.

Then PACF α_k is the correlation between $e_t^{(k-1)}$ and $e_{*,t}^{(k-1)}$.

Expression for PACF:

Since the process is (variance) stationary, Y_t's have constant variance $\forall\, t$. We assume that $\sigma_y^2 = 1$. Write $Z^{(k-1)} = (Y_{t-1}, \dots, Y_{t-k+1})'$.

We have

$$E\left(Y_t Z^{(k-1)}\right) = E\left(Y_t (Y_{t-1}, \dots, Y_{t-k+1})'\right) = (\rho_1 \quad \cdots \quad \rho_{k-1})' = \varrho^{(k-1)} \text{ (say)},$$

$$E\left(Y_{t-k} Z^{(k-1)}\right) = E\left(Y_{t-k} (Y_{t-1}, \dots, Y_{t-k+1})'\right) = (\rho_{k-1} \quad \cdots \quad \rho_1)' = \varrho_*^{(k-1)} \text{ (say)}$$

$$E\left(Z^{(k-1)} Z^{(k-1)'}\right) = E\left(\begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-k+1} \end{pmatrix} (Y_{t-1}, \dots, Y_{t-k+1})\right) = \begin{pmatrix} 1 & \cdots & \rho_{k-2} \\ \vdots & \ddots & \vdots \\ \rho_{k-2} & \cdots & 1 \end{pmatrix} =$$

$P^{(k-1)}$ (say).

Here

$$\mathrm{P}^{(k-1)} = \left((\rho_{|i-j|})\right) r_{|i-j|} = (i,j)^{th} \text{ element of } \left(\mathrm{P}^{(k-1)}\right)^{-1}$$

$$\varrho_*^{(k-1)'} \left(\mathrm{P}^{(k-1)}\right)^{-1} \varrho_*^{(k-1)} = \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} r_{|i-j|} \rho_{k-i} \rho_{k-j}$$

$= \sum_{i'=1}^{k-1} \sum_{j'=1}^{k-1} r_{|i'-j'|} \rho_{i'} \rho_{j'}\,; (i' = k-i, j' = k-j)$

$= \varrho^{(k-1)'} \left(\mathrm{P}^{(k-1)}\right)^{-1} \varrho^{(k-1)}$

$\varrho_*^{(k-1)'} \left(\mathrm{P}^{(k-1)}\right)^{-1} \varrho_*^{(k-1)} = \varrho^{(k-1)'} \left(\mathrm{P}^{(k-1)}\right)^{-1} \varrho^{(k-1)}$

where, $e_t^{(k-1)}$: Least squares residual of linear regression between Y_t and $Z^{(k-1)}$

For obtaining $e_t^{(k-1)}$ we minimize

$\eta_1 = E\left(Y_t - \alpha' Z^{(k-1)}\right)^2 = 1 + \alpha' P^{(k-1)} \alpha - 2\alpha' \varrho^{(k-1)}$

$\frac{\partial \eta_1}{\partial \alpha} = 2P^{(k-1)} \alpha - 2\varrho^{(k-1)} = 0$

$\Rightarrow \alpha = \left(P^{(k-1)}\right)^{-1} \varrho^{(k-1)}.$

Hence

$e_t^{(k-1)} = Y_t - \alpha' Z^{(k-1)} = Y_t - \varrho^{(k-1)'} \left(P^{(k-1)}\right)^{-1} Z^{(k-1)}.$

Similarly, we obtain the least squares residual $e_{*,t}^{(k-1)}$ of linear regression of Y_{t-k} on $Z^{(k-1)}$ as

$e_{*,t}^{(k-1)} = Y_{t-k} - \varrho_*^{(k-1)'} \left(P^{(k-1)}\right)^{-1} Z^{(k-1)}.$

Then variances of $e_t^{(k-1)}$ and $e_{*,t}^{(k-1)}$ are given by

$E\left(e_t^{(k-1)}\right)^2 = E\left(Y_t - \varrho^{(k-1)'} \left(P^{(k-1)}\right)^{-1} Z^{(k-1)}\right)^2$

$= 1 - \varrho^{(k-1)'} \left(P^{(k-1)}\right)^{-1} \varrho^{(k-1)} = E\left(e_{*,t}^{(k-1)}\right)^2$

Covariance between $e_t^{(k-1)}$ and $e_{*,t}^{(k-1)}$ is

$$E\left(e_t^{(k-1)}e_{*,t}^{(k-1)'}\right) = E\left[\left(Y_t - \varrho^{(k-1)'}\left(P^{(k-1)}\right)^{-1}Z^{(k-1)}\right)\left(Y_{t-k} - \varrho_*^{(k-1)'}\left(P^{(k-1)}\right)^{-1}Z^{(k-1)}\right)'\right]$$

$$= \rho_k - \varrho^{(k-1)'}\left(P^{(k-1)}\right)^{-1}\varrho_*^{(k-1)}$$

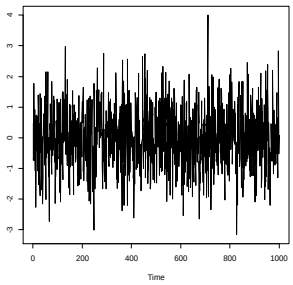
Hence the PACF of order k is given by

$$\alpha_k = \frac{\rho_k - \varrho^{(k-1)'}\left(P^{(k-1)}\right)^{-1}\varrho_*^{(k-1)}}{1 - \varrho^{(k-1)'}\left(P^{(k-1)}\right)^{-1}\varrho^{(k-1)}}.$$

***Purely Random Process or Random Shocks*:** A purely random process, often modeled as a white noise process, consists of a sequence of uncorrelated random variables with a constant mean and variance. Each observation in the time series is independent of the others. It is included the following points:

- A discrete process $\{u_t; t \in T\}$ is called a purely random process if the random variables $\{u_t; t \in T\}$ are a sequence of *iid* random variables.
- The process has constant mean and variance and γ(k)=0 for all $k = \pm 1, \pm 2, \cdots$.
- It is also refereed as shocks.
- It is also called white noise, as its spectrum is like that of white light.
- A purely random process is useful as constituents of other complicated processes.

***Gaussian White Noise Process or Gaussian Random Shocks*:** A process $\{u_t; t \in T\}$ is called a Gaussian White Noise Process if the random variables $\{u_t; t \in T\}$ are a sequence of *iid* random variables with $u_t \sim N(0, \sigma_u^2), \forall t$. We denote it by $u_t \sim GWN(0, \sigma_u^2)$.



***Stationary Process*:**

Why is Stationarity important?

- To ensure that the probabilistic mechanism which has generated the time series do not change over time.
- The way process changes is predictable.
- It becomes possible to make predictions based on stationary processes.
- It makes the process much easier to model and investigate.
- An important concept for developing several inference and analytical tools.

Definition: A time series is said to be strongly or strictly stationary if the joint distribution of $Y_{t_1}, \dots, Y_{t_n}$ is the same as the joint distribution of $Y_{t_1+k}, \dots, Y_{t_n+k}\ \forall\ n, t_1, t_2, \dots, t_n, \text{and}\ k$. Thus, if $f(y_{t_1}, \dots, y_{t_n})$ denotes the joint pdf of $Y_{t_1}, \dots, Y_{t_n}$, then the condition for strong stationarity is

$$f(y_{t_1+k}, \dots, y_{t_n+k}) = f(y_{t_1}, \dots, y_{t_n})\ \forall\ n, t_1, t_2, \dots, t_n, k$$

- For n=1 the distribution of Y_t is the same for all t so that $\mu(t) =$ μ and $\sigma_y^2(t) = \sigma_y^2\ \forall t$.
- For n=2, writing $t_1 = t, t_2 = t + k$, the joint distribution of Y_t and Y_{t+k} depends only on lag k. Thus, the ACVF $\gamma(t, t+k)$ depends only on the lag k.

$\gamma(t, t+k) = \gamma_k = E[\{Y_t - \mu(t)\}\{Y_{t+k} - \mu(t+k)\}]$: ACVF at lag k

Autocorrelation between Y_t and Y_{t+k}:

$\rho_k = \frac{\gamma_k}{\gamma_0}$: Autocorrelation function (ACF)

***Mean Stationary**:* A process is mean stationary if $E(Y_t) = \mu\ \forall\ t$.

***Variance Stationary**:* A process is variance stationary if $E(Y_t - \mu_t)^2 = \sigma_y^2\ \forall\ t$.

***Covariance Stationary**:* A process is covariance stationary if $Cov(Y_t, Y_{t+k}) = \gamma_k \; \forall \; t, k$.

***Second Order Stationarity or Weak Stationarity**:* A time series is second order stationary (or weakly stationary) if its mean and variance are constant and ACVF depends only on the lag.

We can conclude that

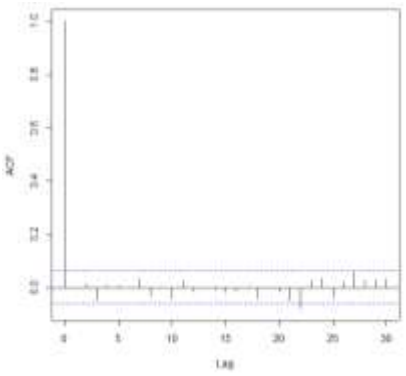
- Strict stationarity implies second order stationarity but its converse is not always true.
- When joint distribution of $Y_{t_1+\tau}, \ldots, Y_{t_{n+\tau}}$ is multivariate normal for all n, the second order stationarity implies strict stationarity.

The ACF of the purely random process is given by

$$\rho_k = \begin{cases} 1 & if \;\; k = 0 \\ 0 & if \;\; k \neq 0 \end{cases}$$

A purely random process is second order stationary as well as strictly stationary.

ACF of the purely random process



Example: Let $X \sim Poisson\ distribution\ P(\mu)$ and $\{u_t; t = 1,2, \ldots\}$ is a sequence of identically independently random variables with mean 0 and variance σ_u^2. We define a process $\{Y_t; t = 1,2, \ldots\}$ as

$Y_t = X + u_t; t = 1,2, \ldots$

Then $E(Y_t) = \mu, \sigma^2_{Y_t} = \mu, \gamma_k(t) = \mu \; \forall \; t, k$. Hence the process is second order stationary. The process is strictly stationary also.

Example: Define the process $\{Y_t; t = 1,2, \ldots\}$ as

$Y_t = \sum_{j=1}^{t} u_j,$

$\{u_t; t = 1,2, \ldots\}$: Purely random process $(0, \sigma_u^2)$

Mean, variance, ACVF:

$E(Y_t) = 0, \;\; \sigma^2_{Y_t} = t\sigma_u^2, \;\; \gamma(t, t+k) = t\sigma_u^2$

ACF: $\rho(t, t+k) = \frac{t^{\frac{1}{2}}}{(t+k)^{\frac{1}{2}}}$

The process is mean stationary but not second order stationary.

Example: Consider the process

$Y_t = ACos(\omega t + \varphi); \varphi \sim U(0, \pi)$

Then

$E(Y_t) = \frac{A}{\pi}\int_0^{\pi} Cos(\omega t + \varphi) d\varphi = -\frac{2A}{\pi} Sin(\omega t)$: Function of t

The process is not mean stationary.

(i) If $\varphi \sim U(0,2\pi)$, $E(Y_t) = 0$, and the process is mean stationary.

(ii) Is the process second order stationary?

If A is a *rv* with mean 0, the process becomes mean stationary

Ergodic Process: Let us consider an example first

Ex: Consider a process $\{Y_t; t = 1,2 \ldots.\}$: $Y_t = \mu + X + u_t$

$\{u_t; t = 1,2 \ldots .\}$: Gaussian white noise process GWN(0,1)

$X \sim$ Bernoulli distribution with *pmf*

$$p(x) = \begin{cases} \frac{1}{2}; if\ x = -1,1 \\ 0; elsewhere \end{cases}$$

Then $E(Y_t) = \mu, \sigma_{Y_t}^2 = 2.$

For $k = 1,2, \cdots$, ACVF of the process is

$$\gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)] = E[(X + u_t)(X + u_{t+k})] = 1$$

ACF: $\rho_k = \frac{1}{2} \forall k.$

- ❑ The process is stationary and achieves statistical equilibrium.
- ❑ The process revolves around -1 or 1 depending upon the initial value of X.

Thus, this statistical equilibrium state is not unique.

$\{y_1, y_2, \cdots, y_n\}$: observed time series, $\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t \cdot \mu$ is unbiasedly estimated by $\bar{y}$.

But $\bar{y}$ converges to $-1 + \mu$, if the initial value of X is -1 and to $1 + \mu$ if the initial value of X is 1. Process gets stuck away from the data generating process mean leading to inconsistent estimator. ρ_k is constant and does not diminish as $k \to \infty$. Which implies that the strength of dependence on first observation remains intact with increasing t. We can't estimate μ consistently using a single realization.

Why ergodicity matters?

- ❑ Stationarity ensures statistical equilibrium but not its uniqueness.
- ❑ Ergodicity, along with stationarity, ensures that such an equilibrium is unique.

- Ergodicity tells us that a single long time series becomes representative of the whole data-generating process, just like a large *iid* sample becomes representative of the whole population or distribution.
- The properties of ergodic process can be investigated on the basis of a single long enough observed time series.
- Ergodic processes forget the past in long run ("far apart" terms are distributed independently of each other).

***Definition**: A process $\{Y_t; t \in T\}$ is said to satisfy ergodic property with respect to a bounded function f if for the realization $\{y_t; t = 1,2, \ldots\}$, the sample average $\frac{1}{n}\sum_{t=1}^{n} f(y_t)$ converges almost everywhere as $n \to \infty$.*

If the process is stationary then $E[f(Y_t)] = \mu \; \forall \; t$. Hence, $\frac{1}{n}\sum_{t=1}^{n} f(y_t)$ converges to μ. A time series has to be stationary in order to be ergodic.

- Ergodicity is not just characteristic of the process.
- The way the experiment is conducted to collect the observed time series also effects the ergodicity of the time series.
- While conducting a study about the air pollution, suppose the data is collected daily on level of nitrogen dioxide in air at a particular time.
- On the first day, the time of measuring nitrogen dioxide level randomly and then, daily the data is collected at the same time.
- Inference drawn will be severely affected by the timing selected on the first day.
- The resulting time series won't satisfy ergodic property.

Definition. A covariance-stationary process, $\{y_t; t \in T\}$, is called (linearly) deterministic if $P[y_t | y_{t-1}, y_{t-2} \ldots,] = y_t$. For a stationary, deterministic process $\{y_t; t \in T\}$

- y_t can be predicted correctly using the entire past $y_{t-1}, y_{t-2} \cdots$.

- One-step ahead prediction error is zero
- It does not mean that y_t is non-random.

***Backward shift operator**:* Backward shift operator B is defined as

$By_t = y_{t-1}; \quad B^k y_t = y_{t-k}; (k = 1,2, \dots).$

***Purely nondeterministic process**:* If all deterministic components of a time series have been subtracted in advance, it is a purely nondeterministic process.

Example: Suppose $\{y_t; t \in T\}$ is defined by

$y_t = A\cos(t) + B\sin(t)$

A and *B* are independently distributed standard normal random variables.

Then

$y_t + y_{t-2} = A\{\text{Cos}(t) + Cos(t-2)\} + B\{Sin(t) + Sin(t-2)\}$

$= A\{2Cos(t-1)Cos(1)\} + B\{2Sin(t-1)Cos(1)\}$

$= 2Cos(1)\{ACos(t-1) + BSin(t-1)\}$

$= 2Cos(1)y_{t-1} = \frac{\sin(2)}{\sin(1)} y_{t-1}.$

Hence

$y_t = \frac{\sin(2)}{\sin(1)} y_{t-1} - y_{t-2}.$

$P[y_t | y_{t-1}, y_{t-2}, \dots] = \frac{\sin(2)}{\sin(1)} y_{t-1} - y_{t-2} = y_t$

$\{y_t; t \in T\}$ is a deterministic process.

Example: Consider the process $y_t = \sum_{j=1}^{k} R_j Cos(\omega t + \vartheta_j); \vartheta_j \sim U(0,2\pi)$

Then $y_t + y_{t-2} = \sum_{j=1}^{k} R_j\{Cos(\omega t + \vartheta_j) + Cos(\omega(t-2) + \vartheta_j)\}$

$= 2Cos(\omega)\sum_{j=1}^{k} R_j Cos(\omega_j(t-1) + \vartheta_j)$

$= 2Cos(\omega)y_{t-1}$

or

$$y_t = 2Cos(\omega)y_{t-1} - y_{t-2}$$

$P[y_t|y_{t-1}, y_{t-2}, \dots] = 2Cos(\omega)y_{t-1} - y_{t-2} = y_t$

where, $\{y_t; t \in T\}$ is a deterministic process.

Example: Let

$y_t = A + Bt; \quad A \sim N(0,1), B \sim N(0,1).$

Then

$y_{t-1} = A + B(t-1)$

$y_{t-2} = A + B(t-2)$

$\Rightarrow 2y_{t-1} - y_{t-2} = A + Bt = y_t$

$y_t = 2y_{t-1} - y_{t-2}$

$P(y_t|y_{t-1}, \dots) = 2y_{t-1} - y_{t-2} = y_t$

where, $\{y_t; t \in T\}$ is a deterministic process.

The Moving average (MA), Auto regressive (AR), ARMA and ARMA models are widely used in forecasting, econometrics, and various fields of data analysis, particularly for time series data.

- MA: Focuses on the relationship with past error terms.
- AR: Focuses on the relationship with past values in the series.

- ARMA: Combines AR and MA components for stationary data.
- ARIMA: Extends ARMA to handle non-stationary data by incorporating differencing.

11.2 Objectives

After completing this course, there should be a clear understanding of:

- Simultaneous equations model
- Concept of structural and reduced forms
- Problem of identification
- Rank and order conditions of identifiability

11.3 Moving average (MA) Model

The Moving Average (MA) Model is a type of time series forecasting model that is used to analyze and predict future values based on past data. It is particularly useful in scenarios where the underlying time series data exhibit patterns, trends, or seasonal behaviors.

- An MA(q) model specifies that the current value of the time series is a linear combination of the previous `q` white noise error terms.
- It captures the effect of past shocks over a specified number of periods (q).
- The model is called "moving average" because each forecasted value is influenced by a moving window of error terms.
- MA models are typically stationary, meaning their statistical properties do not change over time. This is a crucial condition for many time series analyses.
- Limitations: MA models are not suitable for time series with a trend or seasonality unless differencing or seasonal adjustments are applied first. They can only capture linear relationships, meaning non-linear patterns may not be effectively modeled.

The MA model is a foundational tool in time series analysis, and when combined with other models (like AR), it can provide robust forecasting capabilities. Understanding its mechanics and implications is essential for effective time series forecasting in various domains.

Now, the question is How y has evolved?

Let us consider,

$y_1, y_2, \cdots, y_{t-1}$: Time series observations up to time t-1

$u_1, u_2, \cdots, u_{t-1}$: Random shocks up to time t-1 and

u_t: Random shocks at time t

There are three possibilities:

(i) Process has memory of random noise component of where it was (random noise corresponding to past values of y) but no memory of where it (y) was.

(ii) Process has memory of where it (past values of y) was but no memory of random noise corresponding to past values of y.

(iii) Process has memory of where it (past values of y) was but as well as memory of random noise corresponding to past values of y.

***Moving Average Process*:** A process $\{y_t, t \in T\}$ is called a moving average process of order q if

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} \qquad (1)$$

We scale $u_t{}'s$ so that coefficient of u_t, say θ_0, is 1. where, $\theta_1,\ \theta_2,\ \cdots, \theta_q$ is the constants which may be positive or negative and $(\theta_1,\ \theta_2,\ \cdots \theta_q, \mu, \sigma_u^2)$ be the $(q+2)$ parameters. The process is denoted by MA(q) process. MA(q) process represents y_t against current and previous error shocks u_t and a constant μ (long-term mean). The process can be written as

$y_t - \mu = \sum_{j=0}^{q} \theta_j \, u_{t-j}$ with $\theta_0 = 1$;

which is the moving average of white noise $\{u_t; t \in T\}$. So, we have

- Process does not have memory of exactly where it was (past values of y)
- It does have memory of random noise component of where it was (random noise corresponding to past values of y).

MA(q) process can be written as:

$y_t = \mu + \sum_{j=0}^{q} \theta_j \, u_{t-j}$

Mean of the process: $E(y_t) = \mu$

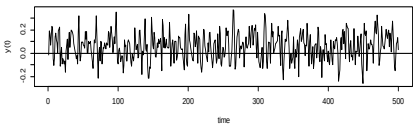
Variance of y_t: $\sigma_y^2 = E[y_t - \mu]^2 = \sigma_u^2 \sum_{j=0}^{q} \theta_q^2 \quad (\theta_0 = 1)$

MA(1) Process: For q=1

Let, $y_t = \mu + u_t + \theta_1 u_{t-1}$.

$\mu = 0.05, \theta_1 = 0.7, n = 500$

Simulated data from MA(1) process



The variance of the process:

$\sigma_y^2 = \gamma_0 = \sigma_u^2(1 + \theta_1^2)$

ACVF:

$\gamma_1 = E(u_t + \theta_1 u_{t-1})(u_{t-1} + \theta_1 u_{t-2}) = \theta_1 \sigma_u^2,$

$\gamma_2 = E(u_t + \theta_1 u_{t-1})(u_{t-2} + \theta_1 u_{t-3}) = 0$

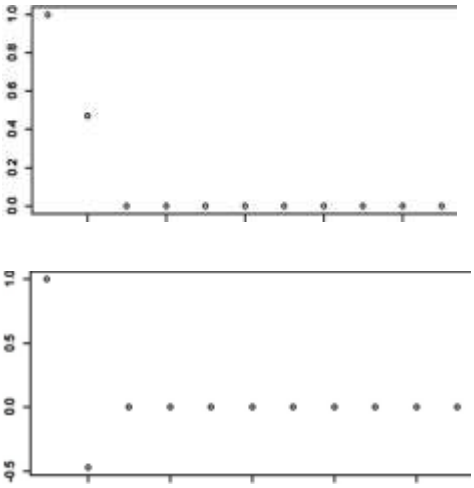
$\gamma_k = 0, \forall \, k > 1.$

Obviously $\gamma_{-k} = \gamma_k \; \forall \, k$.

ACF of MA(1) process is

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{\theta_1}{(1+\theta_1^2)} & if\ k = 1 \\ 0 & if\ k > 1 \end{cases}$$

ACF of MA(1) Process, $\theta_1 = 0.7, \theta_1 = -0.7$



***MA(2) Process**:* For q=2

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}.$$

Variance of the process:

$$\sigma_y^2 = \gamma_0 = \sigma_u^2(1 + \theta_1^2 + \theta_2^2).$$

ACVF:

$$\gamma_1 = \theta_1(1 + \theta_2)\sigma_u^2$$

$$\gamma_2 = \theta_2\sigma_u^2\gamma_k = 0\ \forall\ k \geq 3$$

Hence, ACF is given by

$$\rho_k = \begin{cases} \frac{\theta_1(1+\theta_2)}{(1+\theta_1^2+\theta_2^2)}, for\ k = 1 \\ \frac{\theta_2}{(1+\theta_1^2+\theta_2^2)}, for\ k = 2 \\ 0, \forall\ k \geq 3 \end{cases}$$

MA(q) Process: $y_t = \mu + \sum_{j=0}^{q} \theta_j \, u_{t-j}$

$$\sigma_y^2 = \gamma_0 = \left(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2\right)\sigma_u^2 = \sigma_u^2 \sum_{j=0}^{q} \theta_j^2$$

ACVF of MA(q) Process: $y_t = \mu + \sum_{j=0}^{q} \theta_j \, u_{t-j}$

$$\gamma_k = E\left[\left(u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q}\right)\left(u_{t+k} + \theta_1 u_{t+k-1} + \cdots + \theta_q u_{t+k-q}\right)\right]$$

$$= \begin{cases} \left(\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q\right)\sigma_u^2 = \sigma_u^2 \sum_{j=0}^{q-k} \theta_j \theta_{k+j} & if\ k = 1, \cdots, q \\ 0 & if\ k > q \\ \gamma_{-k} & if\ k < 0 \end{cases}$$

ACF of MA(q) Process: $y_t = \mu + \sum_{j=0}^{q} \theta_j \, u_{t-j}$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2)} = \frac{\sum_{j=0}^{q-k} \theta_j \theta_{k+j}}{\sum_{j=0}^{q} \theta_j^2} & if\ k = 1,2,\cdots q \\ 0 & if\ k \geq q+1 \end{cases}$$

ACF of MA(q) process vanishes for $k \geq q + 1$.

Example: MA(1) process

Let,

$y_t = \Theta(B)u_t;$

$where,\ \Theta(B) = (1 + \theta_1 B)$

$\gamma_0 = \sigma_u^2(1 + \theta_1^2);\ \gamma_1 = \sigma_u^2\theta_1; \gamma_k = 0\ \forall\ k > 1.$

Example: Obtain ACF using auto covariance generating function for MA(2) process:

$y_t = \Theta(B)u_t$, *where,* $\Theta(B) = (1 + \theta_1 B + \theta_2 B^2)$

$\gamma_0 = (1 + \theta_1^2 + \theta_2^2)\sigma_u^2;$

$\gamma_1 = \theta_1(1 + \theta_2)\sigma_u^2;$

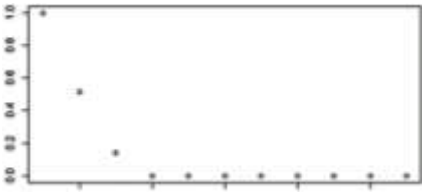
$\gamma_2 = \theta_2\sigma_u^2, \gamma_k = 0 \ \forall \ k \geq 3$.

MA(2) process:

$y_t = 10 + u_t + 0.6u_{t-1} + 0.2u_{t-2}$

ACF $\rho_1 = \frac{0.6(1+0.2)}{(1+0.6^2+0.2^2)} = 0.5143$, $\rho_2 = \frac{0.2}{(1+0.6^2+0.2^2)} = 0.1429, \rho_k = 0 \ \forall \ k \geq 3.$

Plots of ACF of MA(2) Process



***Random Walk*:** Let $\{u_t, t \in T\}$ be the purely random process with mean 0 and variance σ_u^2.

A random walk process $\{y_t, t \in T\}$ is defined as

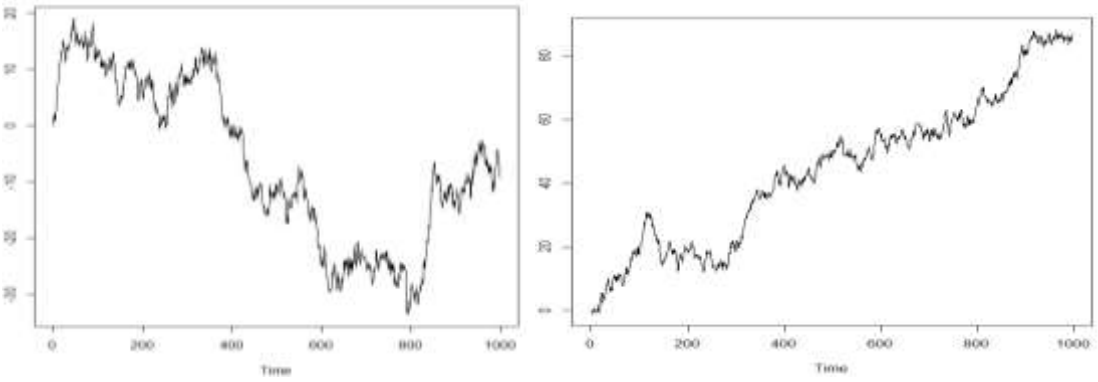
$y_t = \mu + y_{t-1} + u_t.$ (2)

Let $y_0 = 0$. By recursive substitution, after t^{th} steps, model (2) can be written as $y_t = t\mu + \sum_{i=1}^{t} u_t$. It appears that the process has linear trend. *The process is said to have stochastic trend.*

Then, $E(y_t) = t\mu; \ Var(y_t) = t\sigma_u^2$

Since mean and variance of y_t depend on t, the process is non stationary. The economic time series behaving like a random walk. For ex; Share prices, real exchange rate, GDP etc.

μ=0 μ=0.1



11.4 Auto regressive (AR) Model

Autoregressive (AR) models are a class of statistical models used for analyzing and forecasting time series data. In an autoregressive model, the current value of a time series is expressed as a linear combination of its previous values plus a stochastic (random) error term.

- Constructed by regressing current value of variable on past values.
- Uses regression of the variable against itself. Thus, it is termed as autoregression.
- AR models generally assume that the time series is stationary, meaning its statistical properties (mean, variance) do not change over time. If the series is not stationary, techniques like differencing or transformation might be necessary to stabilize its properties.
- Predicts future behavior based on past behavior. Used for forecasting when there is correlation between the current and the preceding values.
- AR models are widely used in economic forecasting, signal processing, and other fields where time-dependent data needs to be analyzed. They serve well in the context of univariate time series, especially when past values hold predictive power.
- The parameters of the AR model can be estimated using methods like the least squares method or maximum likelihood estimation.

Autoregressive models are foundational tools in time series analysis, allowing statisticians and data scientists to model and forecast trends based on historical data. Understanding their structure and when to apply them is crucial for effective time series forecasting.

An *autoregressive process of order p* (AR(p) process) is defined as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \delta + u_t \quad (3)$$

Let $\{u_t; t \in T\}$ be a purely random process with $(0, \sigma_u^2)$.

where,

$\phi_1, \phi_2, \cdots, \phi_p, \delta, \sigma_u^2$:Parameters of AR(p) process

If $E(y_t) = E(y_{t-1}) = E(y_{t-2}) = \cdots = E(y_{t-p}) = \mu$, we have

$$\mu = E(y_t) = \phi_1 \mu + \phi_2 \mu + \cdots + \phi_p \mu + \delta$$

$$\Rightarrow \mu = \frac{\delta}{(1-\phi_1-\phi_2-\cdots-\phi_p)}$$

AR(1) Process: p=1

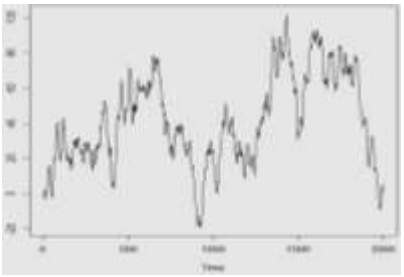
$$y_t = \phi_1 y_{t-1} + \delta + u_t$$

Given that y_{t-1}, y_t becomes independent of $y_{t-2}, y_{t-3}, \cdots$..This is called the Markovian property and the process is also called Markov process.

Mean of the process: $\mu = \frac{\delta}{1-\phi_1}$

Simulated data from AR(1) process

$\delta = 0, \phi_1 = 0.7$, n=2000



If we take $\delta = 0$, so that $\mu = 0$. Assume that the process is variance stationary. Then,

$$\gamma_0 = Var(y_t) = Var(y_{t-1}) = \cdots = \sigma_y^2 \; \forall t$$

Then

$$\sigma_y^2 = E[\phi_1 y_{t-1} + u_t]^2$$

$E[u_t y_{t-1}] = 0$, (y_{t-1} depends only on u_{t-1}, u_{t-2},…)

$$\sigma_y^2 = \frac{\sigma_u^2}{1-\phi_1^2} = \gamma_0$$

ACVF and ACF:

$$\gamma_1 = E(y_t y_{t-1}) = \phi_1 \gamma_0 = \frac{\phi_1 \sigma_u^2}{1-\phi_1^2}.$$

Further

$E(y_{t-k} u_t) = 0, \forall k = 1,2, \ldots$

$\gamma_k = E[y_t y_{t-k}] = E[(\phi_1 y_{t-1} + u_t) y_{t-k}] = \phi_1 \gamma_{k-1} = \phi_1 \phi_1 \gamma_{k-2} = \phi_1^2 \gamma_{k-2} = \cdots = \phi_1^k \gamma_0 = \frac{\phi_1^k \sigma_u^2}{1-\phi_1^2}$

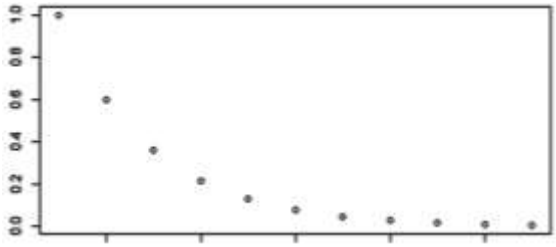
ACF of the AR(1) process:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k.$$

Since $|\phi_1| < 1$, ρ_k declines geometrically.

ACF of the AR(1) process: $y_t = 0.6 y_{t-1} + u_t$

ACF: $\rho_k = (0.6)^k$



Example: Let us consider AR(2) process with $\delta = 0$. So the model is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$

We have

$$\gamma_0 = E(y_t^2)$$

$$= E\{(\phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t) y_t\} = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_u^2 \quad (E(u_t y_t) = \sigma_u^2)$$

$$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1$$

$$\gamma_2 = \phi_1 \gamma_1 + \phi_2 \gamma_0.$$

Hence, we obtain

$$\gamma_1 = \frac{\phi_1}{1-\phi_2} \gamma_0,$$

$$\gamma_2 = \left(\frac{\phi_1^2}{1-\phi_2} + \phi_2\right) \gamma_0,$$

$$\gamma_0 = \frac{1-\phi_2}{1+\phi_2} \frac{\sigma_u^2}{(1-\phi_2)^2 - \phi_1^2}$$

Example: AR(2) process

$$(1 - 0.5B - 0.06B^2) y_t = u_t$$

or

$$(1 - 0.6B)(1 + 0.1B) y_t = u_t.$$

Then

$$\rho_k = a_1 (0.6)^k + a_2 (-0.1)^k.$$

For AR(2) process, $\rho_0 = 1$ and $\rho_1 = \frac{\phi_2}{1-\phi_1} = \frac{0.06}{1-0.5} = 0.12$, we have for $k = 0,1$

$1 = a_1 + a_2; \qquad 0.12 = 0.6a_1 - 0.1a_2$

Hence $a_1 = \frac{11}{35}, a_2 = \frac{24}{35}$.

ACF of the process: $\rho_k = \frac{11}{35}(0.6)^k + \frac{24}{35}(-0.1)^k$.

Note: For MA processes the correlogram vanishes after a certain point and for AR process it declines but never vanishes.

11.4.1 Yule-walker Equations

The Yule-Walker equations are a set of equations that relate the autocovariance function of a stationary time series to the parameters of an autoregressive (AR) model. These equations are particularly useful in the field of time series analysis for estimating the parameters of an AR process from sample data. This equation is fundamental in time series analysis, bridging the gap between the statistical properties of time series data and the mathematical representation of autoregressive models.

Let us consider the AR(p) process and the model is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + u_t$$

Multiplying by y_{t-k} and taking expectation, we obtain

$$E[y_t y_{t-k}] = \phi_1 E[y_{t-1} y_{t-k}] + \phi_2 E[y_{t-2} y_{t-k}] + \cdots + \phi_p E[y_{t-p} y_{t-k}] + E[u_t y_{t-k}] \qquad (4)$$

Substituting $k = 0,1,2, \ldots, p$ in (4)

$$E[y_t^2] = \phi_1 E[y_{t-1} y_t] + \phi_2 E[y_{t-2} y_t] + \cdots + \phi_p E[y_{t-p} y_t] + E[u_t y_t]$$

$$E[y_t y_{t-1}] = \phi_1 E[y_{t-1}^2] + \phi_2 E[y_{t-2} y_{t-1}] + \cdots + \phi_p E[y_{t-p} y_{t-1}] + E[u_t y_{t-1}]$$

$$E[y_t y_{t-2}] = \phi_1 E[y_{t-1} y_{t-2}] + \phi_2 E[y_{t-2}^2] + \cdots + \phi_p E[y_{t-p} y_{t-2}] + E[u_t y_{t-2}]$$

$$\vdots$$

$$E[y_t y_{t-p}] = \phi_1 E[y_{t-1} y_{t-p}] + \phi_2 E[y_{t-2} y_{t-p}] + \cdots + \phi_p E[y_{t-p}^2] + E[u_t y_{t-p}] \quad (5)$$

We observe that

$$E[y_t u_t] = E[(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + u_t) u_t] = E[u_t^2] = \sigma_u^2$$

$$E[y_{t-k} u_t] = 0 \; \forall \; k = 1,2, \ldots, p.$$

Hence, set of equations (5) reduces to

$$\left.\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma_u^2 \\ \gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1 + \cdots + \phi_p \gamma_{p-1} \\ &\vdots \\ \gamma_p &= \phi_1 \gamma_{p-1} + \phi_2 \gamma_{p-2} + \cdots + \phi_p \gamma_0 \end{aligned}\right\} \quad (6)$$

Dividing each of $p+1$ equations of (6) by γ_0 gives the following p equations:

$$\rho_0 = \phi_1 \rho_1 + \phi_2 \rho_2 + \cdots + \phi_p \rho_p + \frac{\sigma_u^2}{\gamma_0}$$

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_1 + \cdots + \phi_p \rho_{p-1}$$

$$\vdots \quad (7)$$

$$\rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \cdots + \phi_p \rho_0$$

Here $\rho_0 = 1.$ The equations in (7) jointly determine the p values of ACF and called *Yule-Walker equations*.

Let us write

$$\rho_{(p)} = (\rho_1 \;\; \rho_2 \cdots \rho_p)'; \; \phi = (\phi_1 \;\; \phi_2 \;\; \cdots \phi_p)';$$

$$P_{(p)} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & & \rho_{p-2} \\ \vdots & & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{pmatrix}.$$

We can write Yule-Walker equations as

$$1 = \rho'_{(p)}\phi + \frac{\sigma_u^2}{\gamma_0}\rho_{(p)} = P_{(p)}\phi.$$

Partial autocorrelation function for AR processes:

PACF of order *k*, a_{kk}, is the Correlation coefficient between y_t and y_{t+k} after eliminating the effect of $y_{t+1}, \cdots, y_{t+k-1}$. a_{kk} is obtained by solving

$$\Gamma_k \alpha_k = \gamma_k, \tag{8}$$

with

$$\alpha_k = (a_{k1}, \ldots, a_{kk})'; \gamma_k = (\gamma_1, \ldots, \gamma_k)'$$

$$\Gamma_k = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & & \gamma_{k-2} \\ \vdots & & \ddots & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{pmatrix}$$

So, we can write (8) as

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & & \gamma_{k-2} \\ \vdots & & \ddots & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{pmatrix} \begin{pmatrix} a_{k1} \\ \vdots \\ a_{kk} \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_k \end{pmatrix} \tag{9}$$

Dividing each row of (9) by $\gamma(0)$, we obtain

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & & \rho_{k-2} \\ \vdots & & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_{k1} \\ \vdots \\ a_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_k \end{pmatrix} \tag{10}$$

And the equations are

$$a_1x + b_1y + c_1z = d_1$$

$$a_2x + b_2y + c_2z = d_2$$

$$a_3x + b_3y + c_3z = d_3$$

Solution for z:

$$z = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}$$

Hence,

$$a_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 \cdots & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} \cdots & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_{k-2} \\ \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} \cdots & 1 \end{vmatrix}} \qquad (11)$$

Example: Consider the AR(1) process and the model is

$$y_t = \phi_1 y_{t-1} + u_t;$$

ACF: $\rho_k = \phi_1^k$

PACF: $a_{11} = \rho_1$

Hence,

$$\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} a_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \phi_1 \\ \phi_1 & \phi_1^2 \end{vmatrix}}{\begin{vmatrix} 1 & \phi_1 \\ \phi_1 & 1 \end{vmatrix}} = 0;$$

$$a_{kk} = \frac{\begin{vmatrix} 1 & \phi_1 \dots & \phi_1 \\ \phi_1 & 1 & \phi_1^2 \\ \vdots & \ddots & \vdots \\ \phi_1^{k-1} & \phi_1^{k-2} \dots & \phi_1^k \end{vmatrix}}{\begin{vmatrix} 1 & \phi_1 \dots & \phi_1^{k-1} \\ \phi_1 & 1 & \phi_1^{k-2} \\ \vdots & \ddots & \vdots \\ \phi_1^{k-1} & \phi_1^{k-2} \dots & 1 \end{vmatrix}} = 0 \ (k \geq 2)$$

AR(1) Process we can conclude that

- Markov Property: Given y_{t-1}, y_t becomes independent of $y_{t-2}, y_{t-3,\dots}$
- $a_{kk} = 0 \ \forall \ k \geq 2$ supports this Markovian property.

Note: In general, for AR(p) process, PACF of order higher than p are zero.

11.5 Autoregressive Moving Average (ARMA) Model

The Autoregressive Moving Average (ARMA) model is a popular statistical tool used for analyzing and forecasting time series data. It combines two components: autoregression (AR) and moving average (MA).

- The AR part of the model uses the dependency between an observation and several lagged observations (previous data points).
- The MA part models the error of the series as a linear combination of error terms (also known as shocks) from previous time points.
- ARMA models are widely used in various fields such as finance, economics, and environmental science for time series forecasting and analysis.
- Non-stationary data must be transformed through differencing or other means before fitting an ARMA model.
- ARMA models assume linear relationships and may not perform well if the underlying process is nonlinear.

The ARMA model is a foundational method in time series analysis that can be used to understand and forecast time-dependent data effectively, as long as its assumptions are satisfied.

Mixed Autoregressive-Moving Average (ARMA) Process of order (p, q)

It is denoted by $ARMA(p, q)$. The model contains p autoregressive and q moving average terms. The process is defined as

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} \cdots + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} \cdots + \theta_q u_{t-q} \quad (12)$$

$E(y_t) = \mu \ \forall t$.

Taking expectation of (12), we have

$$\mu = \phi_1 \mu + \phi_2 \mu + \phi_3 \mu + \cdots + \phi_p \mu + \delta \Rightarrow \mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \phi_3 - \cdots - \phi_p}$$

For $\delta = 0$, μ is also zero.

In terms of backward shift operator "B":

$$\left(1 - \phi_1 B - \phi_2 B^2 \cdots - \phi_p B^p\right) y_t = \delta + \left(1 + \theta_1 B + \theta_2 B^2 \cdots + \theta_q B^q\right) u_t \quad (13)$$

L$^{\text{et}}$

$$\Phi(B) \equiv \left(1 - \phi_1 B - \phi_2 B^2 \cdots - \phi_p B^p\right)$$

$$\Theta(B) \equiv \left(1 + \theta_1 B + \theta_2 B^2 \cdots + \theta_q B^q\right).$$

Then (13) can be represented as

$$\Phi(B) y_t = \delta + \Theta(B) u_t \quad (14)$$

Wold Representation for ARMA process

The infinite lag polynomial of the Wold decomposition can be approximated by the ratio of two finite-lag polynomials:

$$\Psi(B) = \frac{\Theta(B)}{\Phi(B)}$$

(15)

where $\Theta(B)$ is a polynomial of order q in backward shift operator B and $\Phi(B)$ is a polynomial of order p in B.

ARMA Representation

- Approximates the dynamic of any purely nondeterministic weakly stationary process.
- Describes a weakly stationary process in terms of two polynomials, one for AR and the other for MA.
- Higher order AR or MA processes with large number of parameters can be approximated with lower order ARMA processes with lesser number of parameters.

$ARMA(1,1)$ Process:

Consider $ARMA(1,1)$ process with δ=0

$$y_t = \phi_1 y_{t-1} + u_t + \theta_1 u_{t-1} \qquad (16)$$

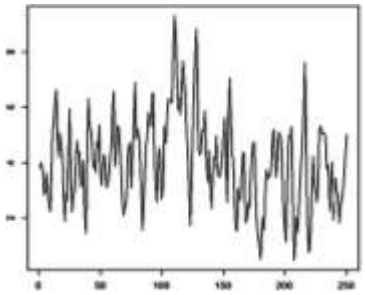
The variance of the process is given by

$$\gamma_0 = E(y_t^2)$$

$$= E[\phi_1 y_{t-1} + u_t + \theta_1 u_{t-1}]^2$$

$$= E[\phi_1^2 y_{t-1}^2 + u_t^2 + \theta_1^2 u_{t-1}^2 + 2\phi_1 y_{t-1} u_t + 2\phi_1 \theta_1 y_{t-1} u_{t-1} + 2\theta_1 u_t u_{t-1}] \qquad (17)$$

Simulated sample from ARMA(1,1) process with $\phi_1 = 0.6, \theta_1 = 0.5$



Notice that

$$E[y_{t-1}u_t] = 0 = E[u_t u_{t-1}]$$

$$E[y_{t-1}u_{t-1}] = E[(\phi_1 y_{t-2} + u_{t-1} + \theta_1 u_{t-2})u_{t-1}] = \sigma_u^2.$$

Therefore, (17) reduces to

$$\gamma_0 = \phi_1^2\gamma_0 + \sigma_u^2 + \theta_1^2\sigma_u^2 + 2\theta_1\phi_1\sigma_u^2 \tag{18}$$

$$\Rightarrow \quad \gamma_0 = \frac{\sigma_u^2(1+\theta_1^2+2\theta_1\phi_1)}{(1-\phi_1^2)} \tag{19}$$

Further

$$\gamma_1 = E[y_t y_{t-1}]$$

$$= E[(\phi_1 y_{t-1} + u_t + \theta_1 u_{t-1})y_{t-1}]$$

$$= (\phi_1\gamma_0 + \theta_1\sigma_u^2)$$

$$\Rightarrow \gamma_1 = \frac{\sigma_u^2(\phi_1+\theta_1+\phi_1\theta_1^2+\theta_1\phi_1^2)}{(1-\phi_1^2)}$$

$$= \frac{\sigma_u^2(\phi_1+\theta_1)(1+\phi_1\theta_1)}{(1-\phi_1^2)} \tag{20}$$

$$\gamma_2 = E[y_t y_{t-2}] = E[(\phi_1 y_{t-1} + u_t + \theta_1 u_{t-1})y_{t-2}] = \phi_1\gamma_1 \tag{21}$$

$$\gamma_k = E\,[y_t y_{t-k}] = E[(\phi_1 y_{t-1} + u_t + \theta_1 u_{t-1})y_{t-k}] = \phi_1\gamma_{k-1};\ \ (k \geq 3) \tag{22}$$

ACF of ARMA(1,1) process:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\phi_1+\theta_1)(1+\phi_1\theta_1)}{(1+\theta_1^2+2\phi_1\theta_1)} \tag{23}$$

For $k \geq 2$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1 \frac{\gamma_{k-1}}{\gamma_0} = \phi_1\rho_{k-1} = \phi_1^2\rho_{k-2} = \cdots = \phi_1^{k-1}\rho_1 \tag{24}$$

$$\rho_1 = \frac{(\phi_1+\theta_1)(1+\phi_1\theta_1)}{(1+\theta_1^2+2\phi_1\theta_1)};\ \rho_k = \phi_1^{k-1}\rho_1; k \geq 2$$

The following is observed:

i. If $\phi_1 + \theta_1 > 0$, $\phi_1 > 0$, $\rho_k > 0\ \forall k$

ii. $\phi_1 + \theta_1 < 0$, $\phi_1 > 0$, $\rho_k < 0\ \forall k$

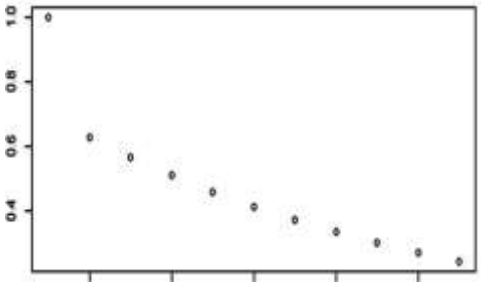
iii. If $\phi_1 + \theta_1 > 0$, $\phi_1 < 0$, ρ_k oscillates with $\rho_1 > 0$

iv. $\phi_1 + \theta_1 < 0$, $\phi_1 < 0$, ρ_k oscillates with $\rho_1 < 0$

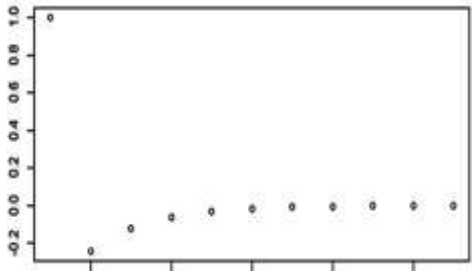
v. $\forall\ k \geq 2, \rho_k$ decays exponentially in magnitude.

ACF of ARMA(1,1) Process $y_t = \phi y_{t-1} + u_t + \theta u_{t-1}$

i. $\phi = 0.9, \theta = 0.5$

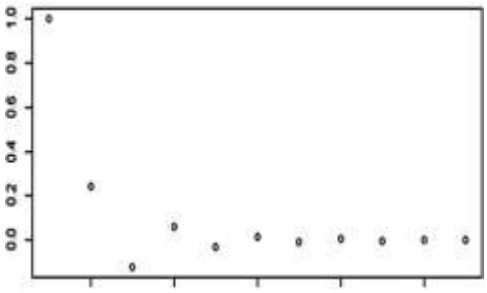


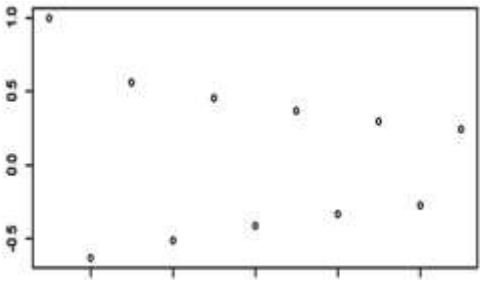
ii. $\phi = 0.5, \theta = 0.9$



iii. $\phi = -0.5, \theta = -0.9$

iv. $\phi = -0.9, \theta = -0.5$





ACVF and ACF of $ARMA(p,q)$ *Process:*

$\Phi(B)y_t = \Theta(B)u_t$

If the process is stationary, we can write

$y_t = \frac{\Theta(B)}{\Phi(B)}\mathrm{u_t} = \Psi(B)u_t = \sum_{j=0}^{\infty}\psi_j B^j u_t = \sum_{j=0}^{\infty}\psi_j u_{t-j}$

Variance:

$\gamma_0 = \sigma_y^2 = E(y_t^2)$

$= E\left[y_t\{(\phi_1 y_{t-1} + \phi_2 y_{t-2}\cdots + \phi_p y_{t-p}) + (u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}\cdots + \theta_q u_{t-q})\}\right]$

$= \phi_1 E(y_t y_{t-1}) + \phi_2 E(y_t y_{t-2}) + \cdots + \phi_p E(y_t y_{t-p}) + E(y_t u_t) + \theta_1 E(y_t u_{t-1}) + \cdots + \theta_q E(y_t u_{t-q})$

We have

$E(y_t u_{t-l}) = E\left[\left(\sum_{j=0}^{\infty}\psi_j u_{t-j}\right)u_{t-l}\right] = \sigma_u^2\psi_l \;\forall l = 0,1,\ldots,q$

$E(y_t y_{t-l}) = E\left(\sum_{j=0}^{\infty}\psi_j u_{t-j}\sum_{j'=0}^{\infty}\psi_{j'} u_{t-j'-l}\right) = E\left(\sum_{j=l}^{\infty}\psi_j\psi_{j-l}\,u_{t-j}^2\right) = \sigma_u^2\sum_{j=l}^{\infty}\psi_j\psi_{j-l}$

Hence $\gamma_0 = \sigma_u^2\left[\sum_{l=1}^{p}\sum_{j=l}^{\infty}\phi_l\psi_j\psi_{j-l} + \psi_0 + \sum_{l=1}^{q}\theta_l\psi_l\right]$

ACVF:

$\gamma_k = E(y_t y_{t-k})$

$= E\left[(\phi_1 y_{t-1} + \phi_2 y_{t-2}\cdots + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}\cdots + \theta_q u_{t-q})y_{t-k}\right]$

For $1 \leq k \leq q$

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p} + \sigma_u^2(\theta_k\psi_0 + \theta_{k+1}\psi_1 + \cdots)$$

$$= \sum_{l=1}^{p} \phi_l\gamma_{k-l} + \sigma_u^2 \sum_{l=k}^{q} \theta_l\, \psi_{l-k}\ (\theta_l = 0 \forall l \geq q+1)$$

For $k \geq q+1$

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p} = \sum_{l=1}^{p} \phi_l\gamma_{k-l}$$

Hence, for $k \geq q+1$

$$\rho_k = \phi_1\rho_{k-1} + \cdots + \phi_p\rho_{k-p}.$$

For $k \geq q+1$ ACF behaves like that of an AR(p) process.

11.6 Stationarity and Invertibility of the Processes

In time series analysis, stationarity and invertibility are critical concepts that help in understanding the properties of time series models, particularly in the context of autoregressive moving average (ARMA) models.

Stationarity refers to the property of a time series that its statistical properties do not change over time. A stationary time series has constant mean, variance, and covariance across different time periods.

Invertibility is a property related to the moving average (MA) component of an ARMA model. A time series model is said to be invertible if its MA representation can be expressed as an AR representation.

Stationarity of a Process

Consider **AR(1)** process:

$$y_t = \phi_1 y_{t-1} + u_t$$

Using successive substitution, we can write the process as

$$y_t = \phi_1 y_{t-1} + u_t = \phi_1^r y_{t-r} + u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots + \phi_1^{r-1} u_{t-r-1}$$

If $|\phi_1| < 1$, as $r \to \infty$, we have

$$y_t = u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots,$$

If $|\phi_1| \geq 1$, the process explodes to infinity.

Theorem: For the general linear process to be stationary, the series $\Theta(B) = \sum_{j=0}^{\infty} \theta_j B^j$ *must converge for* $|B| \leq 1$*, i.e., on or within the unit circle.*

(Proof is beyond the scope of this book)

For **MA(q)** process, ACVF is

$$\gamma_k = \sigma_u^2 [\theta_k + \theta_1 \theta_{k+1} + \cdots + \theta_{q-k} \theta_q],$$

which is a function of *k* only and is independent of t. So, we can conclude that

- A process is stationary if it can be written as a moving average process of finite or infinite order with $\Theta(B) = \sum_{j=0}^{\infty} \theta_j B^j$ converging for $|B| \leq 1$.
- No conditions are required for MA process to be stationary

Invertibility

To illustrate the motivation behind invertibility, consider **MA(1)** model

$$y_t = (1 - \theta B) u_t; u_t \sim N(0, \sigma_u^2). (\theta_1 = -\theta)$$

Expressing u_t in terms of y_t

$$y_t = -\theta y_{t-1} - \theta^2 y_{t-2} - \cdots - \theta^k y_{t-k} + u_t - \theta^{k+1} u_{t-k-1}$$

If $|\theta| < 1$, as $k \to \infty$, we obtain AR process of infinite order

$$y_t = -\theta y_{t-1} - \theta^2 y_{t-2} - \cdots + u_t; \quad (\phi_j = -\theta^j)$$

If $|\theta| > 1$, y_t depends upon $y_{t-1}, y_{t-2} \cdots$ with increasing weights. We avoid this situation and assume that $|\theta| < 1$. We say that the series is invertible if this condition is satisfied.

Theorem (without proof): The general linear process is invertible if weights ϕ_j*'s (*θ_j*'s) are such that* $\Phi(B) = [\Theta(B)]^{-1}$ *converges on or within the unit circle* $|B| \leq 1$.

Condition for the Stationarity of an ***AR(p)*** *Process*

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) y_t = u_t, \tag{25}$$

$$\Phi(B) y_t = u_t \tag{26}$$

where,

$$\Phi(B) \equiv \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right). \tag{27}$$

Transfer function for AR(p) process is $\Theta(B) = [\Phi(B)]^{-1}$.

Theorem*: For the Stationarity of AR(p) process, roots of the equation* $\Phi(B) = 0$ *must be greater than 1 in magnitude.*

Proof*:* For stationarity $\Theta(B) = [\Phi(B)]^{-1}$ must converge for $|B| \leq 1$. Let $g_1^{-1}, g_2^{-1}, \ldots, g_p^{-1}$ be the roots of the equation

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) = 0. \tag{28}$$

Then we can write

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) = \prod_{i=1}^{p} (1 - g_i B).$$

We can write (assuming all roots to be distinctive)

$$\Theta(B) = \prod_{i=1}^{p} (1 - g_i B)^{-1} = \sum_{i=1}^{p} \frac{K_i}{1 - g_i B}. \tag{29}$$

Hence, for convergence of $\Theta(B)$ for all $|B| \leq 1$, $|g_i| < 1 \; \forall \; i = 1,2, \dots, p$. This implies that for stationarity, for all the roots $g_1^{-1}, g_2^{-1}, \dots, g_p^{-1}$,

$\left|g_i^{-1}\right| > 1 \; \forall \; i = 1,2, \dots, p$■

Example: Consider AR(1) process $(1 - \phi_1 B) y_t = u_t$.

The root of $(1 - \phi_1 B) = 0$ is $B = \phi_1^{-1}$. The process is stationary whenever $|\phi_1^{-1}| > 1$ or $|\phi_1| < 1$.

Example: AR(2) process $(1 - \phi_1 B - \phi_2 B^2) y_t = u_t$. Let g_1^{-1}, g_2^{-1} be the roots of $(1 - \phi_1 B - \phi_2 B^2) = 0$.

We can write the process as

$(1 - g_1 B)(1 - g_2 B) y_t = u_t$

For stationarity $|g_1| < 1, |g_2| < 1$.

Further

$\phi_1 = g_1 + g_2, \phi_2 = -g_1 g_2$

Suppose g_1, g_2 is a conjugate pair $Ae^{\pm i2\pi\omega} = A(cos2\pi\omega \pm isin2\pi\omega)$. Then $\phi_1 = g_1 + g_2 = 2ACos(2\pi\omega)$,

$\phi_2 = -g_1 g_2 = -A^2$, with $-\frac{1}{2} < \omega < \frac{1}{2}, 0 < A < 1$.

The conditions on A and ω are required to ensure $|g_1| < 1, |g_2| < 1$.

Stationarity of $ARMA\ (p, q)$ process:

Let us write the model as

$\Phi(B) y_t = \Theta(B) u_t$

$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p,$

$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q.$

Process is stationary whenever roots of $\Phi(B) = 0$ lie outside the unit circle.

Example: Consider the $ARMA(1,1)$ process

$y_t = \phi y_{t-1} + u_t + \theta u_{t-1}$

Root of the model is $1 - \phi B = 0$, i$s\ B = \phi^{-1}$. So, the process is stationary whenever $|\phi| < 1$.

Invertibility Conditions

- ❑ The process is invertible if it can be written as an autoregressive process (of finite or infinite order).
- ❑ No conditions are required for Autoregressive processes to be invertible.

Consider the MA(q) process with μ=0:

$y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} = \Theta(B) u_t$

$\Theta(B)$ is a finite polynomial. So, no condition required on the parameters for stationarity.

Theorem*: For invertibility of MA(q) process, roots of equation* $\Theta(B) = 0$ *must be greater than 1 in magnitude.*

Proof: Let us suppose that $h_1^{-1}, h_2^{-1}, \ldots, h_q^{-1}$ roots of the MA(q) process $\left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right) = 0.$

Write $\left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right) = \prod_{i=1}^{q}(1 - h_i B).$

Then, $\Phi(B) = [\Theta(B)]^{-1} = \prod_{i=1}^{q}(1 - h_i B)^{-1} = \sum_{i=1}^{q} \frac{H_i}{1 - h_i B}.$

We can easily verify that the process is invertible if $|h_i^{-1}| > 1 \; \forall \; i = 1,2, \ldots, q.$ So,

- For the stationarity of an $\text{ARMA } (p, q)$ process all roots of the equation $\Phi(B) = 0$ are greater than 1 in magnitude.
- For the Invertibility of $\text{ARMA}(p, q)$ process, roots of the equation $\Theta(B) = 0$ must be greater than 1 in magnitude.

11.7 Estimation of parameters

Suppose $y_1, y_2, \cdots, y_n$ be the observed time series, and

Sample mean= $\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$

Sample ACF= $c_k = \frac{1}{n-k}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})$

$$\approx \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})$$

where, c_0: Sample variance

The sample ACF of lag k is $r_k = \frac{c_k}{c_0}; k = 1,2, \ldots$

Empirical Version of Yule-Walker Equations for AR(p) Processes

Let the model is:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t,$$

Then the Yule-Walker equations become

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_1 + \cdots + \phi_p \rho_{p-1}$$

$$\vdots$$

$$\rho_p = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \cdots + \phi_p \rho_0$$

Again let $\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p$ be the estimators of $\phi_1, \phi_2, \cdots, \phi_p$. Replacing ρ_k by r_k (γ_k by c_k) and ϕ_k by $\hat{\phi}_k$ leads to empirical version of these equations then we get

$$r_1 = \hat{\phi}_1 r_0 + \hat{\phi}_2 r_1 + \cdots + \hat{\phi}_p r_{p-1}$$
$$\vdots$$
$$r_p = \hat{\phi}_1 r_{p-1} + \hat{\phi}_2 r_{p-2} + \cdots + \hat{\phi}_p r_0$$

or

$$c_0 = \hat{\phi}_1 c_1 + \cdots + \hat{\phi}_p c_p$$
$$c_1 = \hat{\phi}_1 c_0 + \cdots + \hat{\phi}_p c_{p-1}$$
$$\vdots$$
$$c_p = \hat{\phi}_1 c_{p-1} + \cdots + \hat{\phi}_p c_0$$

11.7.1 Estimation of parameters for AR Process

Let us consider the AR(1) process. The model is

$$y_t = \mu + \phi_1 y_{t-1} + u_t$$

***Method of Moments**:*

Compare parameters by corresponding sample estimates. Solve empirical Yule-Walker equations for $\Phi's$. Let us consider that the $\hat{\mu}, \hat{\phi}_1, \hat{\sigma}_u^2$ are the Moment Estimators of μ, ϕ_1, σ_u^2.

For estimating μ

$$\bar{Y} = \hat{\mu} + \hat{\phi}_1 \bar{Y} \Rightarrow \hat{\mu} = \left(1 - \hat{\phi}_1\right)\bar{Y}$$

and Yule Walker Equations for AR(1) is

$$\gamma_0 = \phi_1 \gamma_1 + \sigma_u^2 \gamma_1 = \phi_1 \gamma_0$$

Replacing parameters by estimators

$$c_0 = \hat{\phi}_1 c_1 + \hat{\sigma}_u^2; c_1 = \hat{\phi}_1 c_0$$

Hence we get

$$\hat{\phi}_1 = \frac{c_1}{c_0} = r_1, \ \ \hat{\sigma}_u^2 = (1 - \hat{\phi}_1^2)c_0$$

Method of Least squares:

Let $\hat{\mu}, \hat{\phi}_1$ be the least square estimators of $\mu, \phi_1,$ then minimizing S i.e. residual sum of squares, we get

$$S = \sum(y_t - \hat{\mu} - \hat{\phi}_1 y_{t-1})^2$$

$$= \sum[(y_t - \bar{Y}) - \hat{\phi}_1(y_{t-1} - \bar{Y}) - \{\hat{\mu} - (1 - \hat{\phi}_1)\bar{Y}\}]^2$$

$$= nc_0 + n\hat{\phi}_1^2 c_0 - 2n\hat{\phi}_1 c_1 + n\{\hat{\mu} - (1 - \hat{\phi}_1)\bar{Y}\}^2$$

$$= nc_0(\hat{\phi}_1^2 - 2\hat{\phi}_1 r_1) + nc_0 + n\{\hat{\mu} - (1 - \hat{\phi}_1)\bar{Y}\}^2$$

$$S = nc_0(\hat{\phi}_1^2 - 2\hat{\phi}_1 r_1 + r_1^2 - r_1^2) + nc_0 + n\{\hat{\mu} - (1 - \hat{\phi}_1)\bar{Y}\}^2$$

$$= nc_0(\hat{\phi}_1 - r_1)^2 + nc_0(1 - r_1^2) + n\{\hat{\mu} - (1 - \hat{\phi}_1)\bar{Y}\}^2$$

S is minimum when

$$\hat{\mu} = (1 - \hat{\phi}_1)\bar{Y}$$

$$\hat{\phi}_1 = \frac{c_1}{c_0} = r_1$$

Alternatively, we can minimize $S = \sum(y_t - \hat{\mu} - \hat{\phi}_1 y_{t-1})^2$ using differentiation.

Estimation of parameters for AR(2) Process:

The model of AR(2) process is, $y_t = \mu + \phi_1 y_{t-1} + + \phi_2 y_{t-2} + u_t$

Method of Moments:

The empirical Yule-Walker equations is

$r_1 = \hat{\phi}_1 + \hat{\phi}_2 r_1 r_2 = \hat{\phi}_1 r_1 + \hat{\phi}_2$

$\Rightarrow \hat{\phi}_1 = \frac{r_1 - r_1 r_2}{1 - r_1^2}$; $\hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$ and

$\bar{Y} = \hat{\mu} + \hat{\phi}_1 \bar{Y} + \hat{\phi}_2 \bar{Y}$

$\Rightarrow \hat{\mu} = \bar{Y}(1 - \hat{\phi}_1 - \hat{\phi}_2)$

$\Rightarrow$Least Squares estimators ≈ Method of moments estimators

Estimation of Parameters of AR(p) process:

The model for AR(p) process is

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t \quad (30)$$

Let $\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p$ be the estimators of parameters $\phi_1, \phi_2, \cdots, \phi_p$.

The empirical version of Yule-Walker equations is

$$\begin{aligned} r_1 &= \hat{\phi}_1 r_0 + \hat{\phi}_2 r_1 + \cdots + \hat{\phi}_p r_{p-1} \\ &\vdots \\ r_p &= \hat{\phi}_1 r_{p-1} + \hat{\phi}_2 r_{p-2} + \cdots + \hat{\phi}_p r_0 \end{aligned} \quad (31)$$

By solving these equations, the estimators $\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p$ can be obtain.

Estimate μ by

$$\hat{\mu} = \bar{Y}\left(1 - \hat{\phi}_1 - \hat{\phi}_2 - \cdots - \hat{\phi}_p\right). \quad (32)$$

From the above we can conclude that

- Method of moments estimators are very close to least squares estimators.
- If roots of the AR polynomial are close to unit circle, this method will give estimates which are far from actual parameter values.

- The estimates based on Yule Walker equations can be used as initial values for running numerical optimization method for calculating MLE.

Suppose we write

$$r_{(p)} = \begin{pmatrix} r_1 \\ \vdots \\ r_p \end{pmatrix};\ \hat{\phi} = \left(\hat{\phi}_1 \quad \hat{\phi}_2 \cdots \hat{\phi}_p\right)';$$

$$R_{(p)} = \begin{pmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & \cdots & r_0 \end{pmatrix} = \left(\left(r_{|i-j|}\right)\right)$$

In the first Yule-Walker equation, replacing parameter ϕ and σ_u^2 by their estimators, we obtain

$$c_0 = \hat{\phi}_1 c_1 + \cdots + \hat{\phi}_p c_p + \hat{\sigma}_u^2 = c_0 \hat{\phi}' r_{(p)} + \hat{\sigma}_u^2$$

or

$$\hat{\sigma}_u^2 = c_0\left(1 - \hat{\phi}' r_{(p)}\right). \qquad (33)$$

If $c_0 > 0$, then the matrix $R_{(p)}$ is nonsingular and

$$\hat{\phi} = R_{(p)}^{-1} r_{(p)} = \hat{\phi}_{(p)}\ (say). \qquad (34)$$

***Maximum likelihood estimation**:*

Let $y = (y_1, y_2, \cdots, y_n)$. Since AR process has Markovian structure, the joint density of the data is

$$f(y|\phi, \sigma_u^2) = f\left(y_1, y_2, \ldots, y_p \middle| \phi, \sigma_u^2\right) \times \prod_{t=p+1}^{n} f(y_t | y_{t-1}, \ldots, y_{t-p}, \phi, \sigma_u^2)$$

Ignoring initial *p* terms $y_1, y_2, \cdots, y_p$, LF becomes

$$f(y|\phi, \sigma_u^2) \propto \prod_{t=p+1}^{n} f\left(y_t \middle| y_{t-1}, \ldots, y_{t-p}, \phi, \sigma_u^2\right)$$

We can write

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

$$= z_t'\phi + u_t$$

where,

$z_t' = (y_{t-1}, \ldots, y_{t-p}); t = p+1, \ldots, n$ and

$Z: p \times (n-p)$ matrix with rows z_t'

$x = (y_{p+1}, \ldots, y_n)' : (n-p) \times 1$

Then LF is

$f(y|z_t, \phi, \sigma_u^2) \propto \prod_{t=p+1}^{n} f(y_t|z_t, \phi, \sigma_u^2)$

If $u_t \sim N(0, \sigma_u^2)$, the LF conditional on the initial values is multivariate normal $N_{n-p}(x|Z\phi, \sigma^2 I_{n-p})$.

MLE's of ϕ and σ_u^2 are

$\hat{\phi} = (Z'Z)^{-1}Z'x$

$\hat{\sigma}_u^2 = \frac{1}{n-p}(x - Z\hat{\phi})'(x - Z\hat{\phi})$

For an unbiased estimator, we use

$s^2 = \frac{1}{n-2p}(x - Z\hat{\phi})'(x - Z\hat{\phi})$

Result (without proof): When n is large

$\hat{\phi} \sim Asymptotic\ N\left(\phi, \frac{\sigma_u^2}{n\gamma_0} P_{(p)}^{-1}\right),\ \hat{\phi}_h \sim Asymptotic\ N\left(0, \frac{1}{n}\right), \forall h > p,$

and $\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$.

11.7.2 Estimating parameters of Moving Average Processes

Does method of moments work to fit MA processes?

Example: Let the model for MA(1) process is

$$y_t = u_t - \theta u_{t-1}, \tag{35}$$

We have

$$\gamma_0 = \sigma_u^2(1+\theta^2)\,,\ \gamma_1 = -\sigma_u^2\theta.$$

Hence,

$$\frac{\gamma_1}{\gamma_0} = -\frac{\theta}{1+\theta^2} \Rightarrow \gamma_1\theta^2 + \theta\gamma_0 + \gamma_1 = 0$$

Replacing γ_0 and γ_1 by their sample estimates c_0 and c_1 and θ by $\hat{\theta}$, we obtain

$$c_1\hat{\theta}^2 + c_0\hat{\theta} + c_1 = 0.$$

The solution for $\hat{\theta}$ is

$$\hat{\theta} = -\frac{c_0}{2c_1} \pm \sqrt{\left(\frac{c_0}{2c_1}\right)^2 - 1}\,. \tag{36}$$

If one of the two solutions in (3) give invertible process, select that value as an estimate of θ.

- If $c_0 = 1,\ c_1 = 0.4$, the two solutions are $\hat{\theta} = -0.5, -2$. Process is invertible for $\hat{\theta} = -0.5,$ we use it as estimate of θ.
- If $c_0 = 1,\ c_1 = 0.5$, then $\hat{\theta}_1 = -1$. Process is not invertible for $\hat{\theta} = -1$.
- If $c_0 \leq 2c_1$, we have $\hat{\theta} = -\frac{c_0}{2c_1} \pm i\sqrt{1 - \left(\frac{c_0}{2c_1}\right)^2}$. (37)
- Hence $|\hat{\theta}| = 1$ and the process is not invertible.

Fitting MA and mixed ARMA models require numerical optimization techniques.

***Estimation of Moving Average Processes using Innovation Algorithm**:*

Let fitted MA(m) model is

$y_t = u_t + \hat{\theta}_{m1} u_{t-1} + \cdots + \hat{\theta}_{mm} u_{t-m},$

where, $\{u_m\}$: White noise process $(0, v_m)$

the innovation estimate of MA parameter is

$\hat{\theta}_{m,m-k} = v_k^{-1}\left[c_{m-k} - \sum_{j=0}^{k-1} \hat{\theta}_{m,m-j}\hat{\theta}_{k,k-j} v_j\right], k = 0, \ldots, m-1$ and

$v_m = c_0 - \sum_{j=0}^{m-1} \hat{\theta}_{m,m-j}^2 v_j$

***Hannan-Rissanen Algorithm for Fitting ARMA(p,q) Process**:*

Take $\delta = 0$

$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$

Steps of Algorithm:

***Step 1**:* Fit a high order AR(m) model using Yule-Walker equations. Choose m=maximum integer at which ACF (or PACF) is significant. The estimated coefficients be $\left(\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_m\right)$.

Estimate u_t by $\hat{u}_t = y_t - \hat{\phi}_{m1} y_{t-1} - \cdots - \hat{\phi}_{mm} y_{t-m};\ \ t = m+1, \ldots, n.$

***Step 2**:* Regress y_t over $y_{t-1}, \ldots, y_{t-p}, \hat{u}_{t-1}, \ldots \hat{u}_{t-q}$. $\beta = (\theta', \phi')'$ can be estimated by minimizing

$S(\beta) = \sum_{t=m+1}^{n} \left(y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} - \theta_1 \hat{u}_{t-1} - \cdots - \theta_q \hat{u}_{t-q}\right)^2$ with respect to β.

LS estimator of β: $\hat{\beta} = (Z'Z)^{-1} Z' X_n$,

$X_n = (y_{m+1}, \ldots, y_n)'$

$$Z = \begin{pmatrix} y_m & y_{m-1} & \cdots & y_{m-p+1} & \hat{u}_m & \hat{u}_{m-1} & \cdots & \hat{u}_{m-q+1} \\ y_{m+1} & y_m & \cdots & y_{m-p+2} & \hat{u}_{m+1} & \hat{u}_m & \cdots & \hat{u}_{m-q+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \ddots \\ y_{n-1} & y_{n-2} & \cdots & y_{n-p} & \hat{u}_{n-1} & \hat{u}_{n-2} & \cdots & \hat{u}_{n-q} \end{pmatrix}$$

where, Z: $(N-m)\times(p+q)$

$\hat{u}_t$ is computed for $t = m+1, \ldots, n.$ Take $\hat{u}_t = 0$ for t≤m.

Estimated error variance: $\hat{\sigma}_u^2 = \frac{(x_n - z\tilde{\beta})'(x_n - z\tilde{\beta})}{n-m}$.

Step 3*:* Repeat step 1 with ARMA model, obtain new set of residuals and carry out step 2. Repeat these steps unless reduction in error variance becomes insignificant. For obtaining the MLE's of the parameters of ARMA models, iterative optimization procedures like innovations algorithm are required.

11.7 Self-Assessment Exercise

Example: Consider the process

$y_t = \sum_{j=1}^{k} R_j Cos(\omega_j t + \vartheta_j)$

(i) If $\{R_j\}$ are *iid rv's* following $N(0, \sigma^2)$, then the process is mean stationary. Is the process variance stationary?

(ii) If $\vartheta_j \sim U(0, 2\pi)$, then the process is mean stationary.

Example: Process with linear trend: $Y_t = \alpha + \beta t + u_t; u_t \sim WN(0, \sigma_u^2)$.

Is the process (i) mean stationary, (ii) variance stationary?

Example*:* The PACF for AR(2) process vanishes for all $k \geq 3$.

1.8 Summary

This chapter contain the concept of the autoregressive process, which have models and techniques form the backbone of time series analysis The choice of model depends on the

characteristics of the dataset, particularly whether it is stationary or non-stationary. Box-Jenkins methodology helps streamline the process of model selection and validation. Understanding autocovariance and autocorrelation is essential for diagnosing model fit and performance. This encapsulates the essential elements and principles that govern the AR, MA, ARMA, ARIMA, and Box-Jenkins models and their related concepts.

11.9 References

- Kendall, M.G. (1976*). Time Series*, 2nd Ed., Charles Griffin and Co Ltd., London and High Wycombe.
- Chatfield, C. (1980). *The Analysis of Time Series –An Introduction*, Chapman & Hall.
- Mukhopadhyay, P. (2011). *Applied Statistics*, 2nd Ed., Revised reprint, Books and Allied

11.10 Further Readings

- Goon, A. M., Gupta, M. K. and Dasgupta, B. (2003). *Fundamentals of Statistics*, 6th Ed., Vol II Revised, Enlarged.
- Gupta, S.C. and Kapoor, V.K. (2014). *Fundamentals of Mathematical Statistics*, 11th Ed., Sultan Chand and Sons.
- Montgomery, D. C. and Johnson, L. A. (1967). *Forecasting and Time Series Analysis*, 1st Ed. McGraw-Hill, New York.

UNIT 12 VECTOR AUTOREGRESSIVE PROCESS

Structure

12.1 Introduction

Christopher Sims is often credited with pioneering the VAR model in the 1980s. Sims introduced VAR models as an alternative to the large-scale simultaneous equation models commonly used in econometrics. He argued that structural models imposed too many restrictions and often led to biased results due to unverified assumptions. VAR models, in contrast, treat all variables symmetrically without requiring prior causal ordering.

Building on univariate AR models, Christopher Sims expanded autoregressive modeling to handle multiple time series in Vector Autoregressive (VAR) models. VAR models became essential for analyzing the relationships among multiple interdependent time series. In finance, AR processes were further adapted by Robert Engle with the development of ARCH (Autoregressive Conditional Heteroskedasticity) and later GARCH models, to handle volatility clustering in time series. Sims applied VAR models to study how different

economic indicators interact dynamically. For example, he used VARs to analyze the relationships between GDP, interest rates, inflation, and other key economic variables. This made VAR a popular tool for economic forecasting and policy analysis.

A Vector Autoregressive (VAR) process is a statistical model used to capture the linear interdependencies among multiple time series. A VAR model consists of a system of equations where each variable is expressed as a linear combination of its own past values and the past values of all other variables in the model. This makes VAR suitable for capturing dynamic interdependencies among multiple time series. It is particularly useful in econometrics and time series analysis, where the relationships between different variables need to be examined simultaneously. The main features of this process are:

1. Multiple Time Series: VAR models treat each time series in a system symmetrically. This differs from single-equation models, where only one variable is typically treated as endogenous, allowing VARs to capture complex, bidirectional relationships among variables.

2. Stationarity: This is a crucial aspect because if the time series are non-stationary, spurious relationships might appear, leading to misleading interpretations. Stationarity can sometimes be addressed by transforming the series (e.g., differencing), but when variables are non-stationary and cointegrated, adjustments or alternative models (like the Vector Error Correction Model, or VECM) might be required.

3. Estimation via OLS: OLS estimation for each equation in a VAR model is simple and computationally efficient since each equation can be treated separately. However, the independence of equations might introduce inefficiency when the errors are correlated across equations, so some methods might incorporate more sophisticated estimation techniques to address this.

4. Applications: VAR models are extensively used in econometrics to explore dynamic interdependencies, forecast future values, and simulate policy impacts. Impulse response functions and variance decomposition are two common tools derived from VARs to analyze how shocks in one variable propagate through the system.

A big advantage of VAR process is its flexibility. Unlike structural models, VARs do not require specifying a strict causal order among variables, which can be advantageous in

exploratory research or in complex systems where theoretical guidance on causal order is limited.

Main limitations of VAR processes are the following:

1. Data Requirements: VAR models can require a large amount of data, especially as the number of variables and lags increases.

2. Complexity: As the number of variables and lags increases, the model can become complex and may lead to overfitting.

3. Interpretability: While VAR models can capture relationships, the interpretation of coefficients may not be straightforward, particularly if the variables are cointegrated.

The Vector Autoregressive process is a versatile tool in time series analysis, allowing researchers and practitioners to model and analyze the relationships among multiple time-dependent variables. It is essential to ensure assumptions like stationarity are met for the model to yield reliable results.

12.2 Objectives

After completing this unit, there should be a clear understanding of:

- Multivariate time series process and their properties
- Vector autoregressive (VAR)
- Vector moving average (VMA)
- Vector autoregressive moving average (VARMA) process

12.3 Multivariate Time Series Process

Multivariate time series analysis involves the study of more than one time-dependent variable simultaneously. It allows for the examination of how multiple variables interact over time, and how they may influence each other. Multivariate Time Series is a collection of time series data where multiple variables are recorded over the same time intervals. For example, stock prices of several companies over time. On the other hand, vector autoregression (VAR)

is a common model used in multivariate time series analysis. It generalizes the univariate autoregressive model by allowing for multiple interdependent time series, where the current value of each series depends on its own past values and on past values of other series.

Multivariate processes/Vector processes emerges when several related time series processes are observed simultaneously over time. One may be interested in investigating the cross relationships between the series. The objectives for jointly analyzing and modeling the series is

- To understand the dynamic relationships over time among the series
- To improve accuracy of forecasts for individual series by utilizing the additional information available from the related series.

Here are some key properties:

1. Stationarity

- Weak Stationarity: The mean and variance are constant over time, and the covariance between variables depends only on the lag between them.
- Strict Stationarity: The joint distribution of any collection of observations is invariant to shifts in time.

2. Correlation and Covariance

- Cross-Correlation: Measures the correlation between different time series at different lags, capturing the lead-lag relationships.
- Covariance Matrix: Represents the covariance between multiple series, providing insight into how series move together.

3. Causality

- Granger Causality: Tests whether one time series can predict another, implying a directional relationship between the variables.

- Transfer Function Models: Capture the influence of one variable on another using input-output relationships.

4. Seasonality

- Seasonal patterns may exist in one or more of the time series, requiring adjustments or specific seasonal modeling techniques.

5. Trends

- Long-term trends may be present in individual series or across the multivariate set, requiring differencing or detrending.

6. Asymptotic Properties

- Behavior of estimators as the sample size approaches infinity, which affects the reliability of conclusions drawn from the data.

7. Homoscedasticity vs. Heteroscedasticity

- Homoscedasticity: Constant variance across time.
- Heteroscedasticity: Variance changes over time, which can complicate modeling and forecasting.

8. Multicollinearity

- Occurs when two or more time series are highly correlated, complicating the estimation and interpretation of relationships.

9. Integration

- A time series is said to be *integrated* of order d (I(d)) if it becomes stationary after differencing d times. Multivariate frameworks might include considerations for cointegration.

10. Dependence Structures

- Some series might be conditionally dependent based on the values of other series, requiring copula approaches or vector autoregressions.

11. Modeling Frameworks

- Vector Autoregression (VAR): A common model for multivariate time series that captures the linear interdependencies among multiple time series.
- Vector Autoregressive Moving Average (VARMA) and VARIMA: Extensions that incorporate moving averages and integrated processes.
- State Space Models: Useful for handling latent variables and nonlinear relationships.

12. Forecasting Techniques

a) Various approaches such as dynamic factor models, Bayesian methods, and machine learning techniques to predict future observations based on historical data.

13. Conditional Expectations

b) Multivariate processes often involve the computation of conditional expectations to understand how one series might change given the values of others.

14. Error Structures

c) The error terms in multivariate models may be correlated, indicating that shocks to one series can affect others.

Understanding these properties helps in the selection and application of appropriate models for analysis and forecasting in multivariate time series contexts.

The techniques using for this are:

1. Vector Moving Average (VMA): Used to model the influence of current and past innovations on multiple time series.

2. Dynamic Factor Models: Simplifies the analysis by assuming that multiple series can be driven by a few unobservable latent factors.

3. State Space Models: This framework is useful for handling multivariate time series with unobserved components, giving flexibility in how data is structured and modeled.

4. Bayesian Approaches: Bayesian methods help to incorporate prior knowledge and manage uncertainties in parameter estimation.

5. Longitudinal Data Analysis: Often used in multivariate time series to investigate relationships over time in different subjects or groups.

Vector autoregressive moving average (VARMA) time series models have been developed keeping these objectives in mind. Vector processes are of considerable interest in several fields like:

Economics: One may be interested in simultaneous behavior of interest rate, inflation, money supply, unemployment etc. Focus may be on simultaneous study of time series of GDP, percentage of people below the poverty line, unemployment rate, female-headed household, crime, average income, minimum wages etc.

Environmental sciences and Agriculture: Joint study of time series observations of maximum and minimum temperatures, rainfall, atmospheric humidity, wind speed and direction, etc. and the total production of wheat.

Health and Environment related studies: Joint study of air pollution level, number of asthma patients visiting the hospitals, number of registered cars in a city, monitoring and analyzing multiple health indicators or biomarkers simultaneously, etc.

Multivariate time series processes represent a rich area of study, with numerous applications across fields. By considering multiple time-dependent variables together, analysts can gain insights into complex interactions and improve predictive accuracy. Understanding the fundamental properties and techniques is essential for effective modeling and analysis.

12.3.1 Vector Autoregressive Moving Average (VARMA) Process

The Vector Autoregressive Moving Average (VARMA) process is a multivariate time series model that combines the characteristics of both Vector Autoregressive (VAR) and Vector Moving Average (VMA) processes. It is used to capture the linear interdependencies among

multiple time series variables. A VARMA(p, q) model integrates both the autoregressive and moving average components.

In this process we assume that the error terms are typically assumed to be normally distributed and uncorrelated with each other and Stationarity is another critical assumption; the time series should exhibit constant mean and variance over time.

Wold's infinite MA representation of Stationary processes is

$$Y_t - \mu = \sum_{j=0}^{\infty} \Psi_j u_{t-j}$$

Suppose $\Psi(\mathrm{B})$ can be expressed (at least approximately) as

$$\Psi(\mathrm{B}) \approx \Phi(B)^{-1}\Theta(B),$$

where,

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p : \text{Matrix polynomial}$$

$$\Theta(B) = 1 + \Theta_1 B + \Theta_2 B^2 + \cdots + \Theta_q B^q : \text{Matrix Polynomial}$$

and,

$$\Phi_j ; j = 1, \ldots, p : k \times k \text{ autoregressive coefficients matrices}$$

$$\Theta_j ; j = 1, \ldots, p : k \times k \text{ moving average coefficients matrices}$$

$Y_t : k \times 1$

$u_t : k \times 1$ vector of white noise with

$$E(u_t) = 0,$$
$$E(u_t u_t') = \Sigma$$

It leads to the following vector autoregressive moving average process of order (p, q) $(VARMA(p, q) \text{ process})$.

Suppose,

$$\Phi(B)(Y_t - \mu) = \Theta(B)u_t$$

or

$$Y_t = \delta + \sum_{j=1}^{p} \Phi_j Y_{t-j} + u_t + \sum_{j=1}^{q} \Theta_j u_{t-j}$$

where,

$$\delta = \left(1 - \Phi_1 - \Phi_2 - \cdots - \Phi_p\right)\mu$$

The Process Mean is

$$E(Y_t) = \mu = \left(1 - \Phi_1 - \Phi_2 - \cdots - \Phi_p\right)^{-1}\delta$$

Stationarity and Invertibility of the process:

The process is *invertible* then it can be represented as a convergent vector autoregressive process of infinite order.

Let

$$Y_t = \delta + \sum_{j=1}^{\infty} \Pi_j Y_{t-j} + u_t$$

or

$$\Pi(B)Y_t = \delta + u_t; \sum_{j=1}^{\infty} \|\Pi_j\| < \infty$$

$$\Pi(B) = 1 + \sum_{j=1}^{\infty} \Pi_j B^j$$

$$= \Theta(B)^{-1}\Phi(B)$$

The process is *stationary* if it can be represented as a convergent vector moving average process of infinite order

$$Y_t = \mu + u_t + \sum_{j=1}^{\infty} \Psi_j u_{t-j}$$

or $Y_t = \mu + \Psi(\mathrm{B})u_t;$

$$\sum_{j=1}^{\infty} \|\Psi_j\| < \infty$$

$$\Psi(\mathrm{B}) = 1 + \sum_{j=1}^{\infty} \Psi_j B^j$$

$$= \Phi(B)^{-1}\Theta(B)$$

Let $|A|$ denotes the determinant of a matrix A, and $adj(A)$ denoted adjoint of A. Then

$A^{-1} = \det(A)^{-1}\, adj(A).$

We can write

$$\Phi(z)^{-1} = \det\big(\Phi(z)\big)^{-1}\, \Phi^*(z),$$

where,

$$\Phi^*(z) = adj\big(\Phi(z)\big)$$

Now, we can write the process as

$$Y_t = \mu + \det\big(\Phi(B)\big)^{-1}\, \Phi^*(B)\Theta(B)u_t$$

$\Phi^*(z)\Theta(z)$ is a polynomial of finite order in z. The process is stationary if $\{\det(\Phi(z))\}^{-1}$ is convergent for $|z| < 1$. Further, $\det(\Phi(z))$ is a polynomial of degree p in z. If $\xi_1^{-1}, \dots, \xi_p^{-1}$ are roots of $\det(\Phi(z)) = 0$, then

$$\det\big(\Phi(z)\big) = (1 - \xi_1 z)(1 - \xi_2 z) \dots (1 - \xi_p z)$$

and $\{\det(\Phi(z))\}^{-1}$ is convergent $\forall |z| < 1,$ if roots $\xi_1^{-1}, \dots, \xi_p^{-1}$ lie outside the unit circle.

The process is invertible if we can write the process as an autoregressive process of infinite order. Along the similar lines as we proved the condition for stationarity, it can be shown that the process is invertible if all roots of $\det(\Theta(B)) = 0, \text{say}, \eta_1, \eta_2, \dots, \eta_q$ lie outside the unit circle.

In $VARMA(p,q)$, let $\mu = 0$. The model becomes

$$\Phi(B)Y_t = \Theta(B)u_t \tag{1}$$

Let $k = k_1 + k_2$ and we partition $Y_t, \Phi(B), \Theta(B),$ and u_t as

$$Y_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} \begin{matrix} k_1 \\ k_2 \end{matrix},$$

$$\Phi(B) = \begin{pmatrix} \Phi_{11}(B) & \Phi_{12}(B) \\ \Phi_{21}(B) & \Phi_{22}(B) \end{pmatrix},$$

$$\Theta(B) = \begin{pmatrix} \Theta_{11}(B) & \Theta_{12}(B) \\ \Theta_{21}(B) & \Theta_{22}(B) \end{pmatrix},$$

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

Suppose $\Phi_{12}(B) = 0, \Theta_{12}(B) = 0.$ We also take $\Theta_{21}(B) = 0$.

Then (1) can be written as

$$\Phi_{11}(B)Y_{1t} = \Theta_{11}(B)u_{1t}$$

$$\Phi_{22}(B)Y_{2t} = -\Phi_{21}(B)Y_{1t} + \Theta_{22}(B)u_{2t}$$

T$\text{hen } Y_{1t}$ are said to cause Y_{2t}, but Y_{2t} do not cause Y_{1t}.

where, Y_{1t} denotes the *exogenous variables.*

Then, the model referred to as an ARMAX model (X stands for exogenous). Suppose Y_{2t} denotes the vector of output variables and Y_{1t} denotes the vector of input (exogenous) variables.

Future values of the process Y_{1t} are influenced by its own past, not by the past of Y_{2t} and future values of the process Y_{2t} are influenced by both Y_{1t} and Y_{2t}.

Again, consider the $VARMA(p,q)$ *Process* for forecasting. Let

$$\Phi(B)Y_t = \delta + \Theta(B)u_t$$

or

$$Y_t = \delta + \sum_{j=1}^{p} \Phi_j Y_{t-j} + u_t + \sum_{j=1}^{q} \Theta_j u_{t-j}$$

where

$$\Phi(B) = \left(I_k - \Phi_1 B - \cdots - \Phi_p B^p\right)$$

$$\Theta(B) = \left(I_k + \Theta_1 B + \cdots + \Theta_q B^q\right)$$

The process is said to be

a) Stationary if roots of $\det\left(\Phi(B)\right) = 0$ lie outside the unit circle.

b) Invertible if roots of $\det\left(\Theta(B)\right) = 0$ lie outside the unit circle.

The VARMA model is a powerful tool for modelling multivariate time series data, allowing for richer interpretations of the relationships between multiple time series. Proper use and understanding of this model can lead to insights in various domains, particularly in analysing complex systems where multiple interrelated factors are at play.

12.3.2 Vector Moving Average Processes of order q

Vector Moving Average (VMA) processes are a generalization of univariate moving average processes for multidimensional data. While a univariate moving average (MA) process models a single time series, a vector moving average process allows for modelling multiple time series simultaneously, capturing potential interactions and dependencies between them.

Let us assume the model,

$$Y_t = \mu + u_t + \sum_{j=1}^{q} \Theta_j u_{t-j}$$

$$= \mu + \Theta(B)u_t$$

where,

$$\Theta(B) = I_k + \Theta_1 B + \Theta_2 B^2 + \cdots + \Theta_q B^q$$

Now, consider the *Vector* $MA(1)$ *process* i.e. $\mu = 0$.

By recursive substitution

$$Y_t = u_t + \Theta_1 u_{t-1}$$

$$= u_t + \Theta_1(Y_{t-1} - \Theta_1 u_{t-2})$$

$$= u_t + \Theta_1 Y_{t-1} - \Theta_1^2 u_{t-2}$$

$$= \cdots$$

$$= \Theta_1 Y_{t-1} + \Theta_1^2 Y_{t-2} + \cdots + \Theta_1^j Y_{t-j} + u_t + (-\Theta_1)^{j+1} u_{t-j-1}$$

$$= \Theta_1 Y_{t-1} + \Theta_1^2 Y_{t-2} + \cdots + u_t$$

The above process is a $VAR(\infty)$ process provided $\Theta_1^j \to 0, \text{as } j \to \infty$.

This requires that the eigenvalues of Θ_1 are all less than 1 in modulus, i.e., $\det(I_k + \Theta_1 z) \neq 0 \text{ for } |z| \leq 1$. Notice that, if $\gamma_1, \cdots, \gamma_q$ are eigen values of Θ_1, then

$$\det(I_k + \Theta_1 z) = \prod_{j=1}^{q} (1 + \gamma_j z).$$

The autocovariance matrix of the process is

$$\Gamma(l) = E(Y_t Y_{t+l})$$

$$= \begin{cases} \Sigma + \Theta_1 \Sigma \Theta_1'; & if\ l = 0 \\ \Sigma \Theta_1'; & if\ l = 1 \\ 0; & if\ l \geq 2 \end{cases}$$

Again, consider the *Vector* $MA(q)$ *process*

$$Y_t = \Theta(B) u_t$$

The process is invertible, if it can be written as

$$\Theta(B)^{-1} Y_t = u_t$$

or

$$Y_t - \sum_{j=1}^{\infty} \Pi_j Y_{t-j} = u_t,$$

$$\Theta(z)^{-1} = \left(I_k - \sum_{j=1}^{\infty} \Pi_j z^j \right),$$

$$\sum_{j=1}^{\infty} \|\Pi_j\| < \infty$$

For invertibility

$$\det\big(\Theta(z)\big) = \det\big(I_k + \Theta_1 z + \cdots + \Theta_q z^q\big) \neq 0, \forall\ |z| < 1$$

Thus, roots of $\det\big(\Theta(z)\big) = 0$ lie outside the unit circle.

Now define $\Pi_j' s$ in terms of $\Theta_j' s$:

We have

$$\big(I_k + \Theta_1 z + \cdots + \Theta_q z^q\big) \left(I_k - \sum_{j=1}^{\infty} \Pi_j z^j \right) = I_k$$

Comparing the coefficients of different powers of z we obtain the following recursive relations for $\Pi_j's$

$\Pi_1 = \Theta_1,\ \Pi_j = \Theta_j - \sum_{i=0}^{j-1} \Theta_i \Pi_{j-i}\,; j = 2,3,\ldots$

$\Theta_j = 0\ \forall j > q,\ \Pi_0 = -I_k, \Pi_j = 0\ \forall j < 0$

Covariance Matrices of Vector $MA(q)$ Processes:

We have

$E(Y_t) = \mu = 0$

$\Gamma(l) = Cov\,(Y_t, Y_{t+l})$

$= E(Y_t Y'_{t+l})$

$$= E\left\{\left(u_t + \sum_{j=1}^{q} \Theta_j u_{t-j}\right)\left(u_{t+l} + \sum_{j=1}^{q} \Theta_j u_{t+l-j}\right)'\right\}$$

$$= \begin{cases} \Sigma + \sum_{j=1}^{q} \Theta_j \Sigma \Theta'_j = \sum_{j=0}^{q} \Theta_j \Sigma \Theta'_j\,; if\ l = 0 \\ \Sigma \Theta'_l + \sum_{j=1}^{q-l} \Theta_j \Sigma \Theta'_{j+l} = \sum_{j=0}^{q-l} \Theta_j \Sigma \Theta'_{j+l}\,; if\ l = 1, \ldots, q \\ 0; if\ l \geq q+1 \end{cases}$$

Example: Consider bivariate $MA(1)$ process:

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} - \begin{pmatrix} 0.6 & 0.4 \\ -0.3 & 0.8 \end{pmatrix} \begin{pmatrix} u_{1t-1} \\ u_{2t-1} \end{pmatrix};$$

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$$

Roots of $\det \begin{pmatrix} 0.6-\lambda & 0.4 \\ -0.3 & 0.8-\lambda \end{pmatrix} = 0$ are $0.7 \pm i\sqrt{0.11}$, with absolute value 0.6. Hence process is invertible. Further

$$\Gamma(0) = \begin{pmatrix} 0.6 & 0.4 \\ -0.3 & 0.8 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} 0.6 & -0.3 \\ 0.4 & 0.8 \end{pmatrix}$$

$$= \begin{pmatrix} 1.84 & 1.28 \\ 1.28 & 2.26 \end{pmatrix}$$

$$\Gamma(1) = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} 0.6 & -0.3 \\ 0.4 & 0.8 \end{pmatrix}$$

$$= \begin{pmatrix} 1.6 & 0.2 \\ 2.2 & 2.9 \end{pmatrix}$$

$$\Gamma(l) = 0 \forall l \geq 2$$

$$V = \begin{pmatrix} 1.84 & 0 \\ 0 & 2.26 \end{pmatrix}$$

Autocorrelation matrix

$$\rho(0) = V^{\frac{1}{2}}\begin{pmatrix} 1.6 & 0.2 \\ 2.2 & 2.9 \end{pmatrix}V^{\frac{1}{2}}$$

$$= \begin{pmatrix} 1 & \dfrac{1.28}{(1.84 \times 2.26)^{\frac{1}{2}}} \\ \dfrac{1.28}{(1.84 \times 2.26)^{\frac{1}{2}}} & 1 \end{pmatrix}$$

$$\rho(1) = V^{\frac{1}{2}}\begin{pmatrix} 1.6 & 0.2 \\ 2.2 & 2.9 \end{pmatrix}V^{\frac{1}{2}}$$ ■

Vector Moving Average processes provide a powerful tool for modelling and understanding relationships in multivariate time series data. They allow researchers and practitioners to analyse complex dynamics that exist in various fields. Proper specification and estimation of VMA models are crucial for accurate analysis and forecasting.

12.3.3 Vector Autoregressive Process $\mathbf{VAR(p)}$

The Vector Autoregressive (VAR) process is a statistical model used to capture the linear interdependencies among multiple time series. In a VAR model, each variable in the system is modeled as a linear function of past values of itself and past values of all the other variables in the system.

A vector autoregression of order p, VAR (p), can be described as

$$Y_t = \delta + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} + u_t \quad (2)$$

where, $\Phi_i, (i = 1 \ldots p)$ is a *k*-dimensional square matrices;

Let u_t be a k-dimensional vector of residuals (vector of purely random processes) at time t and δ is a vector of constant terms.

We can write (2) as

$$\Phi(B) Y_t = \delta + u_t, \quad (3)$$

$$\Phi(B) = I_k - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p,$$

Now, $\forall\, t$ and $\forall\, l \neq 0$

$$E[u_t] = 0,$$
$$E[u_t u_t'] = \Sigma_u,$$
$$E[u_t u_{t+l}'] = 0$$

This system is stable if and only if all included variables are weakly stationary, i.e., if all roots of the characteristic equation of the lag polynomial are outside the unit circle. Hence

$$det\,(I_k - \Phi_1 z - \Phi_2 z^2 - \ldots - \Phi_p z^p) \neq 0 \text{ for } |z| \leq 1 \quad (4)$$

We can write (2) as

$$(Y_t - \mu) = \Phi_1 (Y_{t-1} - \mu) + \Phi_2 (Y_{t-2} - \mu) + \ldots + \Phi_p (Y_{t-p} - \mu) + u_t$$

Where

$$\delta = \Phi(I)\mu,$$

or

$$\mu = \Phi^{-1}(I)\, \delta = \Psi(I)\, \delta.$$

Under this condition, system (3) has the MA representation

$$Y_t = \Phi^{-1}(B)\delta + \Phi^{-1}(B)u_t$$

$$= \mu + u_t + \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + \cdots$$

$$= \mu + \Psi(B)u_t,$$

with

$$\Psi(\mathrm{B}) = \sum_{j=0}^{\infty} \Psi_j B^j$$

$$\equiv \Phi^{-1}(B),$$

$$\Psi_0 = I_k.$$

Notice that

$$\Phi(B)\mu = (I_k + \Phi_1 + \cdots + \Phi_k)\mu$$

$$= \Phi(I_k)\mu$$

The *autocovariance matrices* are defined as:

$$\Gamma_y(l) = E[(Y_t - \mu)(Y_{t+l} - \mu)']$$

Set $\delta = 0$ so that $\mu = 0$. Due to (2), it holds that

$$\Gamma_y(l) = E[Y_t Y'_{t+l}]$$

$$= \Phi_1 E[Y_t Y'_{t+l-1}] + \Phi_2 E[Y_t Y'_{t+l-2}] + \ldots + \Phi_p E[Y_t Y'_{t+l-p}] + E[u_t Y'_{t+l}]$$

This leads to the equations determining the autocovariance matrices for $l = 1,2, \ldots, p$.

$$\Gamma_y(l) = \Phi_1 \Gamma_y(l-1) + \Phi_2 \Gamma_y(l-2) + \ \ldots \ + \ \Phi_p \Gamma_y(l-p),$$

$$\Gamma_y(0) = \ \Phi_1 \Gamma_y(-1) + \Phi_2 \Gamma_y(-2) + \ \ldots \ + \ \Phi_p \Gamma_y(-p) + \Sigma_u$$

$$= \Phi_1 \Gamma_y(1)' + \Phi_2 \Gamma_y(2)' + \ldots + \Phi_p \Gamma_y(p)' + \Sigma_u \qquad (5)$$

Thus Yule-Walker Equations for $VAR(p)$ process are

$$\Gamma_y(0) = \Phi_1\Gamma_y(-1) + \Phi_2\Gamma_y(-2) + \ \dots \ + \ \Phi_p\Gamma_y(-p) + \Sigma_u$$

$$\Gamma_y(1) = \Phi_1\Gamma_y(0) + \Phi_2\Gamma_y(-1) + \ \dots \ + \ \Phi_p\Gamma_y\big(-(p-1)\big)$$

$$\vdots$$

$$\Gamma_y(p) = \Phi_1\Gamma_y(p-1) + \Phi_2\Gamma_y(p-2) + \ \dots \ + \ \Phi_p\Gamma_y(0)$$

$\gamma_{ij}(l) =$ $(i,j)^{th}$ element of $\Gamma_y(l) = Cov(Y_{i,t}, Y_{i,t+l})$.

Since

$$\gamma_{ij}(l) = \gamma_{ji}(-l) \forall\ i,j,l,$$

we have

$$\Gamma_y(l) = \Gamma_y(-l)'.$$

The individual correlation coefficients is

$$\rho_{ij}(l) = \frac{\gamma_{ij}(l)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}, \forall\ i,j \ = \ 1,2,\dots,k.$$

Autocorrelation matrices are given by

$$\rho_y(l) = V^{-\frac{1}{2}}\Gamma_y(l)V^{-\frac{1}{2}},$$

where

$$V^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\gamma_{11}(0)}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\gamma_{22}(0)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\gamma_{kk}(0)}} \end{bmatrix}$$

Consider $VAR(1)$ *Process:*

$$Y_t = \Phi Y_{t-1} + u_t,$$
$$E(u_t u_t') = \Sigma_u$$

Then, by recursive substitution

$$Y_t = u_t + \Phi u_{t-1} + \Phi^2 u_{t-2} + \cdots + \Phi^j u_{t-j} + \Phi^{j+1} Y_{t-j-1}$$

Let $\lambda_1, \dots, \lambda_p$ be the eigen values of $\Phi \text{ (or } \Phi')$.

As $j \to \infty, \Phi^j \to 0$ and we can write the process as a MA process of infinite order and the process is stationary (stable).

$$\Gamma(0) = E\{Y_t Y_t'\}$$
$$= \Phi \Gamma_y(1) + \Sigma_u$$

$$\Gamma_y(1) = E\{Y_t Y_{t+1}'\}$$
$$= \Gamma_y(0)\Phi'$$

$$\Rightarrow \Gamma_y(0) = \Phi \Sigma_u \Phi' + \Sigma_u$$

$$\Gamma_y(1) = \Gamma_y(0)\Phi'$$

$$\Gamma_y(l) = E(Y_t Y_{t+l}')$$
$$= E\{Y_t(\Phi Y_{t+l-1} + u_{t+l})'\}$$
$$= \Gamma_y(l-1)\Phi'$$
$$= \cdots$$
$$= \Gamma_y(0)\Phi'^{l}, l \geq 1.$$

Let $\Lambda = diag(\lambda_1, \dots, \lambda_p)$ be the matrix of eigen values of Φ' and P be a lower triangular nonsingular matrix such that $\Phi' = P\Lambda P^{-1}$. Then

$$\Phi'^{l} = P\Lambda^l P^{-1}$$

Hence

$$\Gamma_y(l) = \Gamma_y(0) P \Lambda^l P^{-1}$$

Further

$$\rho_y(l) = V^{-\frac{1}{2}} \Gamma_y(0) V^{-\frac{1}{2}} V^{\frac{1}{2}} P \Lambda^l P^{-1} V^{-\frac{1}{2}}$$

$$= \rho_y(0) V^{\frac{1}{2}} P \Lambda^l P^{-1} V^{-\frac{1}{2}}$$

Hence for vector $AR(1)$ model, the correlations will exhibit a mixture of damped exponentials and damped harmonics depending upon whether the roots are real or complex.

Example: consider the VAR (1) model as

$$Y_{1,t} = 0.8 Y_{1,t-1} - 0.4 Y_{2,t-1} + u_{1,t}$$

$$Y_{2,t} = -0.2 Y_{1,t-1} + 0.6 Y_{2,t-1} + u_{2,t}$$

We can write the model as the following AR process

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} 0.8 & -0.4 \\ -0.2 & 0.6 \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

with

$$\Sigma_u = \begin{pmatrix} 1.00 & 0.70 \\ 0.70 & 1.49 \end{pmatrix}.$$

To check whether the system is stable, we calculate roots

$$\det(I_2 - \Phi z) = \begin{vmatrix} 1 - 0.8z & 0.4z \\ 0.2z & 1 - 0.6z \end{vmatrix} = 0$$

This gives roots $z_1 = 1,, z_2 = 2.5$. Since one of the roots is 1, the model is not stable.

Example: let us consider the VAR (1) model

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} 0.2 & -0.4 \\ -0.2 & 0.6 \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

$$\Sigma_u = \begin{pmatrix} 1.00 & 0.50 \\ 0.50 & 1 \end{pmatrix}$$

The roots of

$$\begin{vmatrix} 1 - 0.2z & 0.4z \\ 0.2z & 1 - 0.6z \end{vmatrix} = 0$$

are $5(2 \pm \sqrt{3})$ and both roots are larger than one in modulus. Thus, the system is stable.

Variance-covariance matrix:

$$\Gamma_y(0) = \Phi \Gamma_y(0) \Phi' + \Sigma_u$$

$$\Gamma_y(1) = \Gamma_y(0) \Phi'$$

For obtaining the variances $\gamma_{11}(0)$ and $\gamma_{22}(0)$ for Y_1 and Y_2 as well as their covariance $\gamma_{12}(0) (= \gamma_{21}(0),)$ we solve the following linear equation system:

$$0.96\, \gamma_{11}(0) + 0.16\, \gamma_{12}(0) - 0.16\, \gamma_{22}(0) = 1.00$$

$$0.08\, \gamma_{11}(0) + 0.72\, \gamma_{12}(0) + 0.12\, \gamma_{22}(0) = 0.50$$

$$-0.16 \gamma_{11}(0) + 0.48\, \gamma_{12}(0) + 0.64\, \gamma_{22}(0) = 1.00$$

Solving which, we obtain

$$\gamma_{11}(0) = 1.485, \gamma_{12}(0) = -1.057, \gamma_{22}(0) = 2.132.$$

Thus, the instantaneous correlation between Y_1 and Y_2 is -0.594.

Forecasting using VAR(p) models:

The forecasts for VAR processes are obtained as:

$$\hat{Y}_t(1) = E_t[Y_{t+1}]$$
$$= \delta + \Phi_1 Y_t + \Phi_2 Y_{t-1} + \ \ldots + \Phi_p Y_{t-p+1}$$

$$\hat{Y}_t(2) = \delta + \Phi_1 \hat{Y}_t(1) + \Phi_2 Y_t + \ \ldots + \Phi_p Y_{t-p+2}$$

and so on.

Alternatively, for the MA representation, we get

$$\hat{Y}_t(1) = \mu + \Psi_1 u_t + \Psi_2 u_{t-1} + \cdots$$

Since

$$Y_{t+1} = \mu + u_{t+1} + \Psi_1 u_t + \Psi_2 u_{t-1} + \cdots,$$

we have the forecast error as

$$Y_{t+1} - \hat{Y}_t(1) = u_{t+1}$$

The autoregressive representation is used to generate forecasts and MA representation is used for calculating the corresponding forecast errors.

The VAR model is a powerful tool for understanding and predicting the dynamics of multivariate time series data.

12.4 Self-Assessment Exercise

1. What distinguishes multivariate time series from univariate time series?
2. Explain the concept of stationarity in the context of multivariate time series processes.
3. Define and differentiate between weak stationarity and strong stationarity in multivariate time series.
4. What is the role of cross-correlation in analysing multivariate time series data?
5. What is a Vector Autoregressive (VAR) model, and how is it mathematically represented?

6. Describe the conditions for stability in a VAR model.
7. What are the limitations of VAR models, and how can they be addressed in practice?
8. What is a Vector Moving Average (VMA) model, and how is it represented mathematically?
9. Compare VMA models to univariate Moving Average (MA) models in terms of complexity and interpretation.
10. What is a Vector Autoregressive Moving Average (VARMA) model, and how does it generalize VAR and VMA models?
11. Discuss the conditions for stationarity and invertibility in a VARMA model.
12. What are the advantages of using VARMA models over VAR or VMA models alone?

12.5 **Summary**

In multivariate time series analysis, understanding the interdependencies among multiple variables is essential, especially in fields like economics, finance, and engineering. This unit discusses the three primary models addressing these relationships: Vector Autoregressive (VAR), Vector Moving Average (VMA), and Vector Autoregressive Moving Average (VARMA) processes.

The VAR process models each variable as a function of its own past values and the past values of other variables, making it highly suitable for analyzing interconnected time series. The VMA process models each variable based on past shocks across variables, focusing on the effect of these shocks over time. Lastly, the VARMA process combines both autoregressive and moving average components, offering a more comprehensive approach to capture complex dependencies among variables. These models allow for dynamic forecasting and policy analysis, providing insights into how variables interact over time. We also discuss the stationarity and invertibility conditions for these models. Together, VAR, VMA, and VARMA processes serve as essential tools for multivariate time series modeling, and decision-making in various applications. After the completion of this unit, you have a clear concept of Multivariate time series process. And also you will be able to understand the

different multivariate time series processes like; vector autoregressive (VAR), Vector moving average (VMA) and vector autoregressive moving average (VARMA) process

12.6 References

- Brockwell, P.J. and Davis, (2009) R.A.. Time Series: Theory and Methods (Second Edition), Springer-Verlag.
- Kirchgassner, G. and Wolters, J. (2007). Introduction to Modern Time Series Analysis, Springer.
- Montgomery, D.C. and Johnson, L.A. (1977): Forecasting and Time Series Analysis, McGraw Hill.

12.7 Further Readings

- Box, George E. P., Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung. (2015) Time Series Analysis, Forecasting and Control, Wiley.
- Chatfield, C. (1980). *The Analysis of Time Series –An Introduction*, Chapman & Hall.
- Granger, C.W.J. and Newbold (1984): Forecasting Econometric Time Series, Third Edition, Academic Press.

UNIT 13 GRANGER CASUALITY

Structure

13.1 Introduction

Granger causality is a statistical hypothesis test used to determine whether one time series can predict another time series. The concept was developed by the economist Clive Granger, who won the Nobel Prize in 2003 for his work in time series analysis. The key points are:

1. Causality vs. Correlation: Granger causality does not imply true causation in the philosophical sense. It simply indicates that past values of one variable can provide information about future values of another variable.

2. Time Series Data: The technique is typically applied to time series data, which means the data points are collected or recorded at multiple time intervals.

3. Statistical Test: To determine whether one series Granger-causes another, you can use regression models. If the lagged values of time series X contribute significantly to predicting time series Y, then X is said to Granger-cause Y.

4. Lag Length: The analysis typically involves choosing the number of lags to include in the model. This can be determined using information criteria such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

5. Assumptions: Stationarity: The time series should be stationary, meaning that its statistical properties such as mean and variance are constant over time. If not, transformations (like differencing) may be required.

Granger causality tests are a powerful tool for exploring predictive relationships in time series. Careful consideration of assumptions, testing procedures, and result interpretations is essential for drawing meaningful insights from such analyses. Granger causality tests are statistical methods used to determine whether one time series can predict another time series based on their past values.

13.2 Objectives

After completing this unit, there should be a clear understanding of:

- Granger causality

- Instantaneous Granger causality and feedback
- Characterization of casual relations in bivariate models
- Granger causality tests

13.3 Granger Causality

Granger causality is a powerful tool for exploring predictive relationships in time series data, but it should be used with caution and in conjunction with other methods and analyses to gain a comprehensive understanding of the dynamic relationships between variables.

Suppose we have more than one time series. The question is whether data generating processes of these time series are independent of each other or dependent on each other. If yes, then the dynamic mechanism of dependence.

Steps to Perform Granger Causality Test:

1. Preprocess Data: Ensure your data is stationary.

2. Choose the Lag Length: Use statistical criteria to determine how many lags of time series X to include when predicting Y.

3. Fit Models: Fit both a restricted model (only Y's own lags) and an unrestricted model (Y's own lags plus X's lags).

4. Conduct Hypothesis Test: Perform an F-test to compare the two models. The null hypothesis is that X does not Granger-cause Y.

5. Interpret Results:

- If you reject the null hypothesis, it suggests that past values of X provide statistically significant information about future values of Y.
- If you fail to reject, it suggests that X does not help predict Y.

Limitations of the test:

1. Not True Causality: Just because one variable Granger-causes another does not mean there is a direct cause-and-effect relationship.

2. Omitted Variable Bias: If a confounding variable affects both time series, the results may be misleading.

3. Sensitivity to Specification: The results can be sensitive to the choice of lags and model specification.

Application of the test:

Granger causality is widely used in various fields, including economics, finance, neuroscience, and other areas that deal with time series data to analyze relationships between variables, forecast future movements, and inform decision-making.

The two major challenges of test are:

1. Correlation does not imply causality. It is important but difficult task to distinguishing between these two.

2. The causal relationship among variables might disappear when the previously ignored common causes are considered.

The two basic assumptions of the test are:

1. The future cannot cause the past. The past causes the present or future.

2. A cause contains unique information about an effect not available elsewhere.

Definition: A variable x is causal to variable y if x could be interpreted as a cause to y or y as effect of x.

Let I_t be the total information set available at time t. This information set includes, above all, the two-timeseries x and y.

Let $\tilde{x}_t = \{x_t, x_{t-1} \dots x_{t-k} \dots\}$: Set of all current and past values of x;

$\tilde{y}_t = \{y_t, y_{t-1} \cdots y_{t-k} \cdots\}$: Set of all current and past values of y and

$\sigma^2(\cdot)$: Variance of the corresponding forecast error.

Then x_t is said not to Granger cause y_t if for any h>0,

Let $F(y_{t+h}|I_t) = F(y_{t+h}|I_t - \tilde{x}_t)$

Let F denotes the conditional distribution and $I_t - \tilde{x}_t$is all the information except $\tilde{x}_t$. Thus x_t is said not to Granger cause y_t if $\tilde{x}_t$ does not help to predict future $y_\cdot$ The whole distribution F is generally difficult to handle empirically and we turn to conditional expectation and variance.

Definition: x is (simply) Granger causal to y if and only if the application of an optimal linear prediction function leads to

$\sigma^2(y_{t+1}|I_t) < \sigma^2(y_{t+1}|I_t - \tilde{x}_t)$,

i.e., if current and past values of x are used future values of y can be predicted better.

13.3.1 Instantaneous Granger Causality

Instantaneous Granger Causality is a concept in time series analysis that involves assessing whether one time series can predict another, specifically in the context of their simultaneous relationships. Traditionally, Granger causality tests determine if past values of one variable can predict future values of another. However, instantaneous Granger causality extends this by examining whether the current values of one time series can provide predictive information about the current values of another. This evaluates whether the current value of one time series affects the current value of another, alongside the historical values. It recognizes that some influences may act immediately rather than unfolding over time.

Instantaneous Granger causality is a useful tool in understanding the immediate interdependencies between time series data. While it offers powerful insights, researchers should critically evaluate their results, considering potential confounding factors and the limitations inherent in statistical testing.

Here, x is instantaneously Granger causal to y if and only if the application of an optimal linear prediction function leads to

$$\sigma^2(y_{t+1}|\{I_t, x_{t+1}\}) < \sigma^2(y_{t+1}|I_t)$$

i.e., the future value of y, y_{t+1}, can be predicted with a smaller forecast error variance, if the future value of x, x_{t+1}, are used in addition to the current and past values of x.

13.3.2 Feedback

There is feedback between x and y if x is causal to y and y is causal to x. There are eight different, exclusive possibilities of causal relations between two time series:

(i) x and y are independent: $(x \perp y)$ or (x, y)

(ii) There is only instantaneous causality: $(x - y)$

(iii) x is causal to y, without instantaneous causality: $(x \rightarrow y)$

(iv) y is causal to x, without instantaneous causality: $(x \leftarrow y)$

(v) x is causal to y, with instantaneous causality: $(x \Rightarrow y)$

(vi) y is causal to x, with instantaneous causality: $(x \Leftarrow y)$

(vii) There is feedback without instantaneous causality: $(x \leftrightarrow y)$

(viii) There is feedback with instantaneous causality: $(x \Leftrightarrow y)$

Now, consider the Bivariate $AR(1)$Process

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

Four possible causal directions between x and y are:

1. Feedback

$$H_0: x \leftrightarrow y,\ H_0 = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

2. Independence,

$$H_1: x \perp y,\ H_1 = \begin{pmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{pmatrix}$$

3. H_2: $x \rightarrow y,\ y \nrightarrow x,\ H_2 = \begin{pmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{22} \end{pmatrix}$

4. H_3: $y \rightarrow x,\ x \nrightarrow y,\ H_3 = \begin{pmatrix} \phi_{11} & 0 \\ \phi_{21} & \phi_{22} \end{pmatrix}$

Now for Two-stage testing procedure

1. Test H_1 (null) against H_0, H_2 against H_0, and H_3 against H_0.

2. If necessary, test H_1 against H_2, and H_1 against H_3.

Equivalent definition

For an r-dimension stationary process, Y_t, there exists a canonical MA representation

$$Y_t = \mu + \Psi(L)u_t$$

$$= \mu + \sum_{i=1}^{\infty} \Psi_i u_{t-i}\ ,\ \Psi_0 = I_r$$

Let $Ly_t = y_{t-1}$ is the Lag operator (same as backward shift operator "B") and $\Psi_{jk,i}$ is the $(j,k)^{th}$ element of Ψ_i. Then

i. A necessary and sufficient condition for variable k not Granger-cause variable j is that $\Psi_{jk,i} = 0\ \forall\ i = 1,2,\ldots$

ii. If the process is invertible, then

$$Y_t = \delta + \Pi(L)Y_{t-i} + u_t$$

$$= \delta + \sum_{i=1}^{\infty} \Pi_i Y_{t-i} + u_t$$

$\Pi_{jk,i}$ is the $(j,k)^{th}$ element of Π_i.

iii. If there are only two variables, or two-group of variables, j and k, then a necessary and sufficient condition for variable k not to Granger-cause variable j is that $\Pi_{jk,i} = 0, \forall\, i = 1, 2, \cdots$.

iv. For a $VAR(1)$ process with dimension equal or greater than 3, $\Pi_{jk,i} = 0$, for i = 1, 2,⋯,is sufficient for non-causality at h = 1 but it is insufficient for non-causality at h > 1.

v. Variable k might affect variable j in two or more period in the future via the effect through other variables. For example, in the AR model

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} 0.4 & 0 & 0 \\ 0.2 & 0.1 & 0.2 \\ 0 & 0.3 & 0.3 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}$$

Then

$$y_0 = \begin{pmatrix} u_{10} \\ u_{20} \\ u_{30} \end{pmatrix},$$

$$y_1 = \Phi y_0\,,$$

$$y_2 = \Phi^2 y_0,$$

and

$$\Phi = \begin{pmatrix} 0.4 & 0 & 0 \\ 0.2 & 0.1 & 0.2 \\ 0 & 0.3 & 0.3 \end{pmatrix},\ \Phi^2 = \begin{pmatrix} 0.16 & 0 & 0 \\ 0.10 & 0.07 & 0.08 \\ 0.06 & 0.12 & 0.15 \end{pmatrix}$$

Observe that

y_{23} is not influenced by y_{11} but influenced by y_{01}.

Causality is defined for all lags h > 0 and not just for h = 1.

Causality for a particular h is neither necessary nor sufficient for some other lag. Then, it is important to understand the causal mechanisms by which $Y_t's$ are produced.

Two main tasks in causal inference:

(a) Defining the Set of Hypothesis or counterfactuals that based on some economic theory.

(b) Identifying causal models from data which require estimation and hypothesis testing theory.

13.4 Causal analysis using Bivariate VAR and Bivariate MA representations

Causal analysis using Bivariate Vector Autoregression (VAR) and Bivariate Moving Average (MA) models involves understanding how two time series interact over time.

In a Bivariate VAR model, each of the two-time series is modeled as a linear function of their own past values and the past values of the other series.

In a Bivariate MA model, the current value of a series is expressed as a linear combination of past error terms from both series.

Causal interpretation requires careful consideration of the data structure, stationarity, and potential confounding variables. The presence of endogeneity or omitted variable bias can complicate causal claims. While both approaches offer valuable insights into bivariate time series data, VAR is typically favored for causal analysis due to its explicit modeling of dynamic interactions, whereas MA is more about understanding the influence of shocks.

Let the AR representation of the processes is

$$\mathrm{A}(L)\begin{pmatrix} y_t \\ x_t \end{pmatrix}$$

$$= \begin{pmatrix} \alpha_{11}(L) & \alpha_{12}(L) \\ \alpha_{21}(L) & \alpha_{22}(L) \end{pmatrix}\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} u_t \\ v_t \end{pmatrix} \qquad (1)$$

where, $A(L)$ is a matrix polynomial.

$$\alpha_{jk}(L) = \sum_{i=0}^{\infty} \alpha_{jk,i} L^i ;\ j,\ k\ =\ 1,2$$

$\alpha_{jk,i}$ is the coefficient of B^i in $A_{jk}(L)$.

To normalize the system, we take $\alpha_{11,0} = \alpha_{22,0} = 1$.

$\alpha_{jk}(L) \equiv 0$ if all their coefficients, $\alpha_{jk,i}$ are zero.

Also, u_t, and v_t are the white noise which might be correlated with each other.

Then x_t does not Granger-cause y_t if $\alpha_{12}(L) = 0$ or $\alpha_{12,i} = 0$, $\forall\ i\ =\ 1,\ 2, \cdots$.

The instantaneous causality exists iff u and v are contemporaneously correlated. Then in this case forecast error of y is reduced if current value of x is included in the forecast regression implying that either $\alpha_{12,0} \neq 0$ or $\alpha_{21,0} \neq 0$.

Now, let us consider the MA representation of the process

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = B(L) \begin{pmatrix} u_{t-1} \\ v_{t-1} \end{pmatrix}$$

$$= \begin{pmatrix} \beta_{11}(L) & \beta_{12}(L) \\ \beta_{21}(L) & \beta_{22}(L) \end{pmatrix} \begin{pmatrix} u_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix} \qquad (2)$$

Here, $B(L)$ is a Matrix polynomial with elements $\beta_{jk}(L),\ j, k = 1,2, \ldots$

Further

$$\beta_{jk}(L) = \sum_{i=0}^{\infty} \beta_{jk,i} L^i$$

$$\beta_{11,0} = \beta_{22,0} = 1$$

For the identifiability of the model

$$\beta_{12,0} = \beta_{21,0} = 0$$

The x_t does not Granger-cause y_t if

$$\beta_{12}(L) = 0$$

or

$$\beta_{12,i} = 0, \forall\, i\, =\, 1,\, 2, \cdots$$

Now

$$\begin{pmatrix} \beta_{11}(L) & \beta_{12}(L) \\ \beta_{21}(L) & \beta_{22}(L) \end{pmatrix} = \begin{pmatrix} \alpha_{11}(L) & \alpha_{12}(L) \\ \alpha_{21}(L) & \alpha_{22}(L) \end{pmatrix}^{-1}$$

$$\Rightarrow \beta_{11}(L) = \frac{\alpha_{22}(L)}{\eta(L)},$$

$$\beta_{12}(L) = -\frac{\alpha_{12}(L)}{\eta(L)},$$

$$\beta_{22}(L) = \frac{\alpha_{11}(L)}{\eta(L)},$$

$$\beta_{21}(L) = -\frac{\alpha_{21}(L)}{\eta(L)}$$

and

$$\eta(L) = \alpha_{11}(L)\alpha_{22}(L) - \alpha_{12}(L)\alpha_{21}(L)$$

Then

$$\beta_{12}(L) \equiv 0$$

$$\Rightarrow\ \alpha_{12}(L) = 0,$$

$$\beta_{21}(L) = 0$$

$$\Rightarrow \alpha_{21}(L) = 0$$

Hence x_t does not Granger-cause y_t if $\forall\, i\, =\, 1,\, 2, \cdots$.

$\beta_{12}(L) = 0$

or $\beta_{12,i} = 0$,

The hypothesis that all cross-lags coefficients are zero can be tested using F test for testing significance of coefficients.

13.4.1 Characterizing Causal Relations using Residuals of individual Univariate Processes of x and y

Using Wold representation, we express $\{y_t\}$ and $\{x_t\}$ by two separate MA processes of infinite order. Consider that,

$$\left.\begin{aligned} y_t &= \Psi_{11}(L)a_t \\ x_t &= \Psi_{22}(L)b_t \end{aligned}\right\} (3)$$

and

$$\Psi_{jk}(L) = \sum_{i=0}^{\infty} \psi_{jk,i} L^i$$

We can write (3) as

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \Psi(L) \begin{pmatrix} a_t \\ b_t \end{pmatrix} \qquad (4)$$

where,

$$\Psi(L) = \begin{pmatrix} \Psi_{11}(L) & 0 \\ 0 & \Psi_{22}(L) \end{pmatrix}$$

Further

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \Psi(L)\Psi(L)^{-1}B(L)\begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

$$= \Psi(L)\mathrm{H}(\mathrm{L})\begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

$$= \Psi(L)\begin{pmatrix} \eta_{11}(L) & \eta_{12}(L) \\ \eta_{21}(L) & \eta_{22}(L) \end{pmatrix}\begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

where,

$$\mathrm{H}(\mathrm{L}) = \Psi(L)^{-1}B(L)$$

Hence

$$\begin{pmatrix} a_t \\ b_t \end{pmatrix} = \Psi(L)^{-1}\begin{pmatrix} y_t \\ x_t \end{pmatrix}$$

$$= \Psi(L)^{-1}\mathrm{B}(\mathrm{L})\begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

$$= H(L)\begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

Now

$$\eta_{jk}(L) = \frac{\beta_{jk}(L)}{\Psi_{jj}(L)}; j,k = 1,2$$

Thus

$$\alpha_{12}(L) = 0$$

$$\Rightarrow \beta_{12}(L) = 0$$

$$\Rightarrow \eta_{12}(L) = 0$$

Hence x_t does not Granger-cause y_t if $\forall\, i$

$$\eta_{12}(L) = 0$$

or

$$\eta_{12,i} = 0.$$

The cross covariance between a_t and b_t is

$$\gamma_{ab}(k) = E(a_t b_{t-k})$$
$$= E[(\eta_{11}(L)u_t + \eta_{12}(L)v_t)(\eta_{21}(L)u_{t-k} + \eta_{22}(L)v_{t-k})]$$
$$= E[\eta_{11}(L)u_t.\eta_{21}(L)u_{t-k}] + E[\eta_{21}(L)u_{t-k}.\eta_{12}(L)v_t] + E[\eta_{11}(L)u_t.\eta_{22}(L)v_{t-k}]$$
$$+ E[\eta_{12}(L)v_t.\eta_{22}(L)v_{t-k}]$$

Assuming no instantaneous causality, u_t and $v_{t'}$ are uncorrelated $\forall\, t,\ t'$. Hence

$$\gamma_{ab}(k) = E[\eta_{11}(L)u_t.\eta_{21}(L)u_{t-k}] + E[\eta_{12}(L)v_t.\eta_{22}(L)v_{t-k}]$$

When $(x \nrightarrow y)$, $\eta_{12}(L) = 0$ and $a_t = \eta_{11}(L)u_t$. u_t and a_t are white noise, thus both the representations of y_t

$$y_t = \beta(L)u_t,$$
$$y_t = \Psi_{11}(L)a_t$$

imply that $\eta_{11}(L) = 1$. Hence $a_t = u_t$.

Thus, $\forall\, k$

$$\gamma_{ab}(k) = E[u_t.\eta_{21}(L)u_{t-k}]$$
$$= 0$$

Hence, cross correlation

$$\rho_{ab}(k) = 0\ \forall\, k$$

Similarly, when $(y \nrightarrow x)$,

$$\rho_{ab}(k) = 0\ \forall\, k.$$

When $(x \perp y)$,

$$\rho_{ab}(k) = 0\ \forall\, k.$$

13.5 Causality Tests

Causality tests are statistical methods used to determine whether one variable influence or causes changes in another variable. Establishing causation is more complex than correlation, as correlation does not imply causation.

Granger causality tests are a statistical hypothesis test used to determine whether one time series can predict another. It is based on the premise that if a variable (X) Granger-causes another variable (Y), then past values of (X) should contain information that helps to predict future values of (Y).

13.5.1 The Direct Granger Procedure (Thomas SARGENT, 1976)

The Direct Granger Procedure, introduced by Thomas Sargent in 1976, is a statistical method used primarily for testing and estimating causal relationships in time series data. It builds on the concept of Granger causality, which posits that if a time series (X) Granger-causes another time series (Y), then past values of (X) contain information that helps predict (Y) beyond the information contained in past values of (Y) alone.

The Direct Granger Procedure is widely used in various fields, such as economics, finance, and social sciences, to understand complex relationships among time-dependent variables, assess shock transmission, and inform policy decisions.

Let $\{x_t\}, \{y_t\}$ be the two stationary time series.

Test for simple causality from x *to* y *is equivalent to examining whether the lagged values of* x *in the regression of* y *on lagged values of* x *and* y *significantly reduce the error variance.*

Consider the models

$$M1: y_t = \alpha_0 + \sum_{k=1}^{k_1} \alpha_{11}^k y_{t-k} + u_t;\ t = 1,2,\dots,n$$

$$M2: y_t = \alpha_0 + \sum_{k=1}^{k_1} \alpha_{11}^k y_{t-k} + \sum_{k=k_0}^{k_2} \alpha_{12}^k x_{t-k} + u_t;\ t = 1,2,\dots,n$$

We estimate both the models using least squares with $k_0 = 1$.

The Granger causality index is defined as

$$GCI = \log\left(\frac{\hat{\sigma}_{u1}^2}{\hat{\sigma}_{u2}^2}\right)$$

where $\hat{\sigma}_{u1}^2$ and $\hat{\sigma}_{u2}^2$ are estimated error variances of models M1 and M2 respectively.

Testing the null hypothesis, H_0: $\alpha_{12}^1 = \alpha_{12}^2 = \cdots = \alpha_{12}^k = 0$ is equivalent to testing

H_0: True model is M1 against

H_1: True model is M2 .

Let

RSS_1= Residual sum of squares for the fitted model M1.

RSS_2= Residual sum of squares for the fitted model M2.

For testing the significance of difference between these variances, the Fisher F-statistic is

$$F = \frac{\frac{RSS_1 - RSS_2}{k_2}}{\frac{RSS_2}{n - k_1 - k_2 - 1}}$$

Here, H_0 is the hypothesis that x does not cause y. Under H_0, statistic F follows the F-distribution with $F(k_2, n - k_1 - k_2 - 1)$. Interchanging x and y in M1 and M2, a simple causal relation from y to x can be tested. If the null hypothesis is rejected in both directions, we conclude that there is a feedback relation. For testing instantaneous causality, set $k_0 = 0$, in M2 and test H_0: $\alpha_{12}^0 = 0$ using t test.

Shortcomings:

The results are strongly dependent on the number of lags of the explanatory variable $k2$.

There is a trade-off that the more lagged values we include, the better the influence of this variable can be captured but the power of the test decreases as more lagged values are included. Use different values of $k2$ (and $k1$) to inspect the sensitivity of the results to the number of lagged variables.

Note: Use different information criteria (AIC, BIC) to select $k2$ (and $k1$).

Sargent's contribution through the Direct Granger Procedure significantly advanced the framework for analysing dynamic relationships in time series, allowing researchers to derive insights about causality and influence among multiple variables.

13.5.2 The Haugh-Pierce Test

The Haugh-Pierce test is a statistical method used for assessing Granger causality in time series data. It is specifically designed to test whether one time series can predict another, extending the classic Granger causality framework. The method extends traditional Granger causality tests by allowing for non-linear relationships between the variables. Unlike standard tests, it can account for situations where the relationship may not be strictly linear.

The Haugh-Pierce test is useful in various fields, including economics, finance, and social sciences, where understanding interdependencies and predictive relationships among time series is critical.

The steps for the Haugh-Pierce test are as fllows:

1. Fit a univariate ARMA models for $\{x_t\}, \{y_t\}, t = 1,2, \dots, n$.
2. Obtain estimated residuals, say, $\{\hat{b}_t\}$ and $\{\hat{a}_t\}$.
3. Obtain the cross correlations between $\{\hat{b}_t\}$ and $\{\hat{a}_t\}$, say $\hat{\rho}_{ab}(k)$.
4. Use Box-Pierce Q statistic (or the Box-Ljung Q statistic) to test the null hypothesis that the estimated residuals are white noise. If hypothesis not accepted, calculate the following statistic:

$$S = n \sum_{k=k_1}^{k_2} \hat{\rho}_{ab}^2(k)$$

Under the null hypothesis $H_0: \rho_{ab}(k) = 0$ for all k with $k1 \leq k \leq k2$, S is asymptotically χ^2 distributed with $(k2 - k1 + 1)$ *d.f.*.

5. Check for $k_1 < 0$ and $k_2 > 0$, whether there is any causal relation. If this hypothesis can be rejected, causal relation from x to y can be checked for $k_1 = 1$ and $k_2 \geq 1$. In the reverse direction, for $k1 \leq -1$ and $k2 = -1$, it can be checked whether there is a simple causal relation from y to x. It can be tested by using $\rho_{ab}(0)$ whether there exists an instantaneous relation.

The Haugh-Pierce test adds depth to traditional Granger causality analysis by accommodating non-linear relationships, making it a powerful tool for researchers analysing time-dependent data.

13.5.3 Hsiao Procedure

The Hsiao procedure is a method used for testing Granger causality in time series analysis. Developed by Hsiao (1981), it is particularly useful for examining the causal relationships between two or more time series variables. It determines the lag lengths and then estimate parameters. The procedure is divided into six steps:

(i) Optimal lag length k_1^* of univariate AR process of y is determined.

(ii) Fixing k_1^*, the optimal lag length k_2^* of x is determined.

(iii)Given k_2^* optimal lag length of y, k_1^*, is again determined.

(iv)If the value of the information criterion in the third step is smaller than that of the first step, x has a significant impact on y. Otherwise, the univariate representation of y is used. This gives a preliminary model for y.

(v) Steps (i) to (iv) are repeated by exchanging the variables x and y. We get a (preliminary) model for x.

(vi) Estimate the two specified models simultaneously taking into account the correlation between their residuals.

The Hsiao procedure captures the simple causal relations between the two variables. The instantaneous relation is reflected by the correlation between the residuals. The Hsiao procedure is a valuable tool for researchers looking to establish causal relationships in time series data. Its systematic approach helps in identifying whether one variable can be used to predict another, which is essential in fields like econometrics, finance, and social sciences.

13.5.4 Direct Granger Procedure for Testing the Causality of More than two Time Series

The Direct Granger Causality procedure for testing the causal relationships among more than two time series is an extension of the classical Granger causality test, which is originally designed for two time series. The Haugh-Pierce test uses cross correlations between two time series and cannot be applied for more than two series. We can apply direct Granger procedure for more than two series.

Let $\{z_{j,t}\}, j = 1, \dots, m$ is m time series. Estimating equation can be extended to

$$y_t = \alpha_0 + \sum_{k=1}^{k_1} \alpha_{11}^k y_{t-k} + \sum_{k=1}^{k_2} \alpha_{12}^k x_{t-k} + \sum_{j=1}^{m} \sum_{k=1}^{k_{j+2}} \beta_j^k z_{j,t-k} + u_t; t = 1,2, \dots, n.$$

where, β_j^k : Coefficients corresponding to additional variables.

After determining $k_1, k_2, k_3, \dots, k_{m+2}$, estimate the model using least squares. Using F-test, we can test hypothesis such as whether different coefficients of x variables are significantly different from 0 or not. It can be tested if there exists a simple causal relation from x to y or feedback. The significance of coefficients of other time series $z_j's$ can also be tested by using the F-test.

The Direct Granger causality procedure allows researchers to explore complex interrelationships among multiple time series. While powerful, it is essential to interpret the results within the context of the data and consider potential confounding variables. Further techniques, such as Structural VAR (SVAR), may be employed for more nuanced causal analysis.

13.5.5 Systems with more than two variables

Granger causality is a statistical hypothesis test for determining whether one time series can predict another time series. While the classic Granger causality test is generally applied to pairs of time series, it can certainly be extended to systems with more than two variables.

If there are more than two time series then the question arises here is that; Are the inferences drawn from Bivariate Tests misleading?

Then,

- Instantaneous relations detected with the direct Granger procedure or Haugh-Pierce test using bivariate tests are only preliminary. Definite evidence about whether these relations are real or spurious can be drawn in a complete model using additional information.
- If third variables are ignored, conclusion regarding feedback relation might be spurious.
- Inclusion of other relevant variables might reduce it to a one sided relation.
- If the relation between x and y is only one-sided in the bivariate model, there are no third variables which are Granger causal to x and y.

While causality tests provide valuable insights, interpreting the results requires careful consideration of context, methodology, and assumptions underlying each test. Employing a combination of methods often strengthens causal inferences.

13.6 Self-Assessment Exercise

1. What is Granger causality, and how does it differ from traditional causality in time-series analysis?

2. Giving the underlying assumptions, explain Granger's causality. Discuss the concept of causality with the help of following example:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} .5 & 0 & 0 \\ .1 & .1 & .3 \\ 0 & .2 & .3 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}, \begin{bmatrix} y_{10} \\ y_{20} \\ y_{30} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

3. Define error correction representation of a bivariate VAR process. How it helps in testing long term and short-term relationship between two nonstationary time series?

4. How does Granger causality account for the directionality of causation in time-series data?

6. Define instantaneous Granger causality and explain how it differs from standard Granger causality.

7. What is feedback in the context of causality analysis, and how can it be identified?

8. Discuss the implications of finding bidirectional causality in a bivariate time-series model.

9. How can instantaneous Granger causality be detected using statistical methods?

11. Explain how causal relations are characterized in bivariate time-series models.

12. Why is it important to assess the stationarity of time series before testing for Granger causality?

16. What are the key steps involved in performing a Granger causality test?

17. What are the assumptions underlying Granger causality tests, and how can violations affect results?

21. What is the Haugh-Pierce test, and how does it relate to Granger causality? Explain the role of residual cross-correlation functions in the Haugh-Pierce test.

22. Discuss the advantages and limitations of the Haugh-Pierce test in identifying causality.

13.7 Summary

This unit explores the concept of causality in time-series analysis, with a focus on Granger causality and its extensions. Granger causality is introduced as a statistical framework for determining whether one time series can predict another, distinguishing between correlation and predictive causation. The unit also examines instantaneous Granger causality, where causal relationships may occur without time lag, and feedback, a bidirectional causality between variables.

The characterization of causal relationships in bivariate models is discussed, including how lag structures and model specifications influence causality interpretation. Emphasis is placed on practical techniques for testing Granger causality, including the Granger causality test, the Haugh-Pierce test for identifying cross-correlation between residuals, and Hsiao's test, which combines Granger causality with model selection procedures for improved robustness.

13.8 References

- Brockwell, P.J. and Davis, (2009) R.A.. Time Series: Theory and Methods (Second Edition), Springer-Verlag.
- Kirchgassner, G. and Wolters, J. (2007). Introduction to Modern Time Series Analysis, Springer.
- Montgomery, D.C. and Johnson, L.A. (1977): Forecasting and Time Series Analysis, McGraw Hill.

13.7 Further Readings

- Box, George E. P., Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung. (2015) Time Series Analysis, Forecasting and Control, Wiley.
- Chatfield, C. (1980). *The Analysis of Time Series –An Introduction*, Chapman & Hall.

- Granger, C.W.J. and Newbold (1984): Forecasting Econometric Time Series, Third Edition, Academic Press.

UNIT 14 COINTEGRATION

Structure

14.1 Introduction

Causal analysis for nonstationary processes:

Causality tests are based upon the assumption that the underlying processes X_t is stationary. The existence of unit root might make the traditional asymptotic inference invalid.

Cointegration is a statistical concept used in time series analysis to establish whether two or more non-stationary series share a common long-term relationship. In simpler terms, even if the individual time series are not stationary (i.e., they have trends or seasonality), they can

still be said to be cointegrated if a linear combination of them results in a stationary series. The main concepts added here is:

1. Non-stationary Series: Time series data that do not have constant mean and variance over time.

2. Stationary Series: Time series data that have constant mean and variance over time; they are easier to model.

3. Linear Combination: A mathematical combination of the series, often involving coefficients. For instance, if (Y_t) and (X_t) are two time series.

4. Error Correction Model (ECM): When time series are cointegrated, they tend to adjust towards equilibrium in the long run; the ECM captures how short-term deviations from this equilibrium affect the changes in the series.

Testing for Cointegration

Several tests exist to assess whether series are cointegrated:

1. Engle-Granger Test: A two-step method that involves regressing one series on the other and examining the residuals for stationarity.

2. Johansen Test: A more advanced method that allows testing for multiple cointegration relationships among several time series.

Applications

Economics: Cointegration is often applied in economics to explore relationships among macroeconomic indicators.

Finance: It is useful for pairs trading strategies, where traders look for cointegrated pairs of stocks to exploit price inefficiencies.

Example

Consider two time series, (e.g., stock prices of two related companies). While both may trend upward over time, if the difference between them remains stable (mean-reverting), they could be considered cointegrated.

Cointegration is a powerful tool for understanding the long-term relationships between non-stationary time series. It helps analysts and researchers model and predict economic and financial phenomena more accurately when direct analysis of non-stationary data would lead to spurious results.

14.2 Objectives

After completing this course, there should be a clear understanding of:

- **Error Correction Model**
- Cointegration
- Granger representation theorem
- Bivariate cointegration
- Cointegration tests in static model

14.3 Error Correction Model

Spurious Regression

Spurious regression refers to a misleading statistical relationship that appears significant in a regression analysis, but arises purely due to the non-stationary nature of the variables involved. It often occurs when time series variables with trends are regressed against each other without accounting for their non-stationarity, leading to erroneous conclusions about the relationship.

The key features of the spurious regression are:

1. High R^2 and Significant t-statistics: The regression may show high goodness-of-fit and statistically significant coefficients, even when no genuine relationship exists.

2. Nonsensical Relationships: The results often suggest correlations between variables that are logically unrelated.

3. Non-Stationary Variables: The underlying issue is that the variables are not stationary (i.e., they have trends, unit roots, or other forms of time dependence).

The modeling of two or more time series using traditional regression methods requires all variables to be I(0). However, the question arises, do the statistical results of regression hold if some or all of the variables are I(1)?

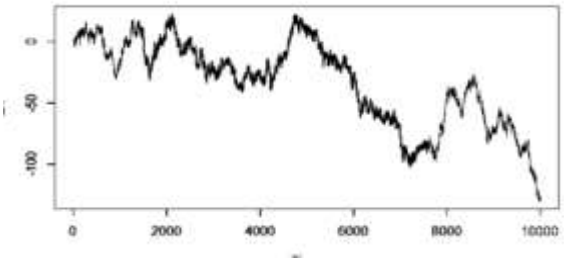
Let us illustrate with the help of simulated data examples, the problems one may face when regressors are I(1).

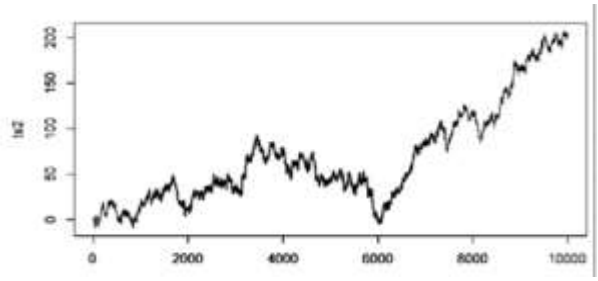
Example: Consider two independent and I(1) processes $\{x_t\}$ and $\{y_t\}$ generated by

$x_t = x_{t-1} + v_t,\ v_t \sim WN(0,1) \rightarrow time\ series\ 1$

$y_t = y_{t-1} + u_t,\ u_t \sim WN(0,1) \rightarrow time\ series\ 2$

10000 observations are simulated from each series. A visual inspection shows that the levels of the two series are negatively related.





The regression of y on x gives the following results

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	27.91741	0.50138	55.68	<2e-16
x	-1.20342	0.01055	-114.07	<2e-16

Residual standard error: 35.82

Multiple R-squared: 0.5655, Adjusted R-squared: 0.5654

F-statistic: 1.301e+04 on 1 and 9999 DF,

p-value: < 2.2e-16

The estimated slope coefficient is negative and significant, R^2 is moderate.

These statistics are representative of the spurious regression as both the time series are independently drawn.

Example: Let processes $\{x_t\}$ and $\{y_t\}$ be generated by

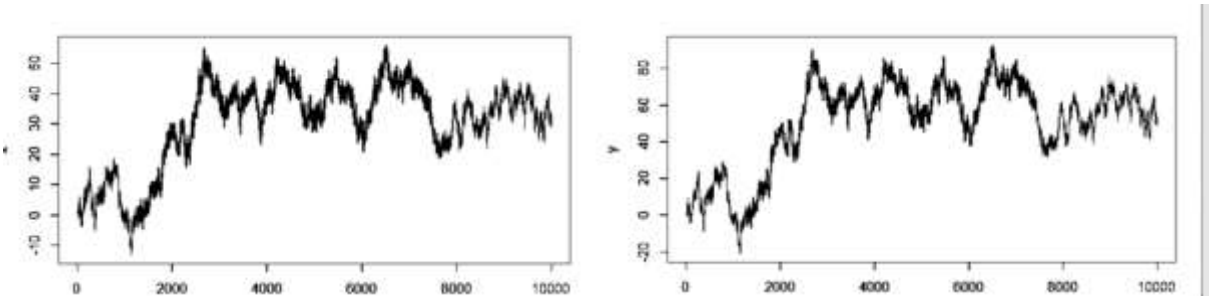
$x_t = 0.6I_t + v_t,\ v_t \sim WN(0,1),$

$y_t = I_t + u_t,\ u_t \sim WN(0,1)$

$I_t = I_{t-1} + w_t,\ w_t \sim WN(0,1)$

The processes $\{x_t\}$ and $\{y_t\}$ involve common stochastic trend (I(1) process) $\{I_t\}$ in them.

10000 observations are simulated from each series and their plots indicate that the two processes are highly positively related.



We run a regression of y_t on x_t and the results are as follows:

Residual standard error: 1.933 on 9998 degrees of freedom

Multiple R-squared: 0.9932,

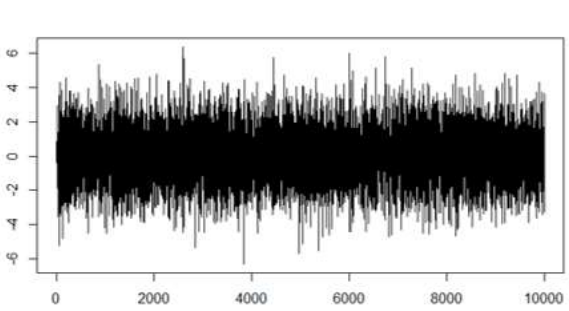
Adjusted R-squared: 0.9932

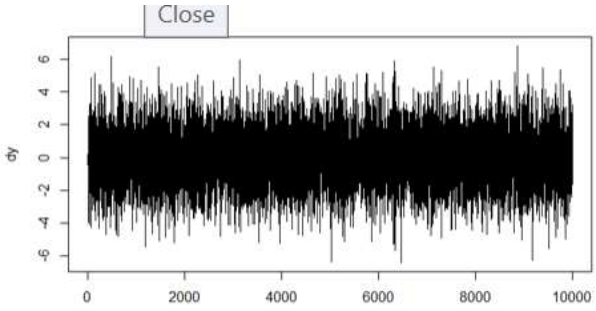
F-statistic: 1.465e+06 on 1 and 9998 DF,

p-value: < 2.2e-16

The regression results are highly significant with the large value of R^2. Is it because of the common stochastic trend in the two processes?

Since both y_t and x_t are I(1) processes, to remove the stochastic trend, we take the first difference and run regression between Δy_t on Δx_t. The plots of the first differences are like purely random processes.





	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-0.004409	0.016721	-0.264	0.792
Δx	0.257132	0.010847	23.706	<2e-16

Residual s.e.: 1.672,

Multiple R^2: 0.053,

Adjusted R^2: 0.053,

F-statistic: 562 on 1 and 9997 DF

The regression is insignificant with low value of R^2.

An interesting feature of the process is

$$y_t - (0.6)^{-1}x_t = u_t^*,$$

$$u_t^* = u_t - (0.6)^{-1}v_t \sim WN(0,1 + (0.6)^{-2})$$

Although both x_t and y_t are I(1), $\exists\, \beta = (1,\ -(0.6)^{-1})'$ such that

$$\beta' \begin{pmatrix} y_t \\ x_t \end{pmatrix} \sim I(0)$$

Such processes are called cointegrated processes and β is called the cointegration vector.

It is important to develop statistical tools suited for capturing the relations between nonstationary time series properly.

To overcome the spurious relations problem, instead of using the original series, the series should be transformed so that they can be considered as realizations of weakly stationary processes.

In the previous example, the transformation

$$\beta' \begin{pmatrix} y_t \\ x_t \end{pmatrix}$$

leads to the stationary process, which can be estimated using usual statistical techniques.

Causal analysis for nonstationary processes:

Causality tests assume that the underlying processes Y_t is stationary.

The existence of the unit root implies that the traditional asymptotic inference might be invalid.

Error Correction Representation of VAR(p) process:

The vector autoregressive process (VAR) of order p is defined as

$$Y_t = \delta + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + U_t \qquad (1)$$

MA representation of the above process is

$$Y_t = A^{-1}(L)\delta + A^{-1}(L)U_t$$
$$= \mu + U_t + B_1U_{t-1} + B2U_{t-2} + \cdots$$
$$= \mu + B(L)U_t \qquad (2)$$

$$B\,(L) = I + \sum_{j=1}^{\infty} B_j L^j \equiv A^{-1}(L),$$

$$B_0 = I_k$$

Error correction representation:

An **error correction model (ECM)** is a statistical tool often used in econometrics to model relationships between time series variables that are integrated and may have a long-run equilibrium relationship. It is widely applied in situations where:

1. **Variables are non-stationary**: The variables individually have trends or unit roots, meaning their means and variances change over time.

2. **Cointegration exists**: Despite being non-stationary, a linear combination of the variables is stationary, indicating a long-term equilibrium relationship.

The ECM captures both the short-term dynamics and long-term equilibrium adjustments of the variables.

The key components of an ECM are as follows:

1. **Error Correction Term**: Represents the deviation from the long-term equilibrium. The model "corrects" this error over time.

2. **Short-term Dynamics**: Explains immediate changes in the dependent variable due to changes in explanatory variables.

3. **Long-term Equilibrium**: Ensures that the system converges to a stable relationship over time.

Every stationary VAR of order p

$$Y_t = \delta + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + U_t$$

$$\text{or } \left(\mathrm{I} - \mathrm{A_1 L} - \mathrm{A_2 L^2} - \cdots - A_p L^p\right) Y_t = \delta + U_t \qquad (3)$$

can be written as

$$A^*_{p-1}(L)\Delta Y_t = \delta - A(1) Y_{t-1} + U_t \qquad (4)$$

with

$$A^*_{p-1}(L) = I - A^*_1 L - \cdots - A^*_{p-1} L^{p-1};\ A^*_i = -\sum_{j=i+1}^{p} A_j \ ,\ i = 1, 2 \ldots p-1.$$

Proof: We have

$$A_{i+1} = A^*_{i+1} - A^*_i,\ i = 1, \ldots, p-1,$$

$$A^*_p = 0,$$

$$A^*_0 = -\left(A_1 + A_2 + \cdots + A_p\right)$$

$$I + A^*_0 = I - A_1 - A_2 + \cdots - A_p = A(1)$$

We can write

$$\mathrm{I} - \mathrm{A_1 L} - \mathrm{A_2 L^2} - \cdots - A_p L^p$$

$$= I - (A^*_1 - A^*_0)L - (A^*_2 - A^*_1)L^2 - \cdots - \left(A^*_p - A^*_{p-1}\right)L^p$$

$$= (I + A^*_0)L + I(1-L) - \sum_{j=1}^{p-1} A^*_j (1-L)L^j$$

Hence, we can write (3) as error correction representation.

$$\left\{(I + A^*_0)L + I(1-L) - \textstyle\sum_{j=1}^{p-1} A^*_j (1-L)L^j\right\} Y_t = \delta + U_t$$

or

$$\left\{I - \sum_{j=1}^{p-1} A_j^* L^j\right\} \Delta Y_t = \delta - A(I) Y_{t-1} + U_t$$

or

$$A_{p-1}^*(L) \Delta Y_t = \delta - A(I) Y_{t-1} + U_t \qquad (5)$$

We can write (5) as

$$\Delta Y_t = \delta - A(I) Y_{t-1} + \sum_{j=1}^{p-1} A_j^* \Delta Y_{t-j} + U_t \qquad (6)$$

Components of Y_t are I(1) variables. Hence each component of $\Delta Y_t, \Delta Y_{t-1}, \Delta Y_{t-p+1}$ is stationary. Further, each component of $Y_t \sim I(1)$.

Since $\Delta Y_t \sim I(0)$, as t increases, ΔY_t and U_t approach to zero. Then the process approaches to an equilibrium state.

$$\delta - \Pi Y_{t-1} = 0, \Pi = A(I)$$

Here Π represents the matrix of the long-run equilibrium relations and can be estimated directly in the framework of a linear model.

Granger's Representation Theorem: The entries of the $I(1)$ vector Y_t are cointegrated if and only if they have an ECM representation.

Example: Consider a general dynamic model of a single equation, and one explanatory variable, which is assumed to be exogenous:

$$\alpha_p(L) y_t = \delta + \beta_q(L) x_t + u_t.$$

In the long-run equilibrium, it holds that

$$y_t = y_{t-1} = \cdots = y_{t-p} = \cdots = \bar{y},$$

$$x_t = x_{t-1} = \cdots = x_{t-p} = \cdots = \bar{x},\ u_t = 0$$

Thus, for the long-run equilibrium, we get:

$$\alpha p(1)\bar{y} = \delta + \beta_q(1)\bar{\bar{x}}\,.$$

$$\bar{y} = \frac{\delta}{\alpha_p\,(1)} + \frac{\beta_q(1)}{\alpha_p(1)}\bar{x}$$

$$= \mu + \beta\bar{x}.$$

with

$$\mu = \frac{\delta}{\alpha_p\,(1)},$$

$$\beta\ = \frac{\beta_q(1)}{\alpha_p(1)}$$

If y and x are weakly stationary (or nonstationary but co-integrated), we get the following alternative representation:

$$\alpha^*_{p-1}(L)(1 - L)y_t = \delta + \beta^*_{q-1}(L)(1\ -\ L)x_t - \gamma_0\, y_{t-1}\ +\ \gamma_1 x_{t-1} + u_t, \qquad (7)$$

With

$$\alpha^*_{p-1}(L) = 1\ -\ \alpha^*_1\, L - \cdots - \alpha^*_{p-1}\, L^{p-1}\alpha^*_i$$

$$=\ -\ \sum_{j=i+1}^{p} \alpha_j\,,\ i\ =\ 1,\ 2\ldots p-1,$$

$$\beta^*_{q-1}(L) =\ 1\ -\ \beta^*_1 L - \cdots - \beta^*_{q-1}L^{q-1}$$

$$\beta^*_i = -\ \sum_{j=i+1}^{p} \beta_j\ \ ,\ i =\ 1,\ 2\ldots q-1,$$

$$\gamma_0\ =\ \alpha p(1),$$

$$\gamma_1\ =\ \beta q(1).$$

In equilibrium $\Delta y_t\ =\ \Delta x_t =\ 0, u_t = 0$ and, therefore, $\mathrm{y_t} = \bar{y}, x_t = \bar{x}\ \forall\ t$. Hence

$$- \gamma_0 \bar{y} + \delta + \gamma 1 \, \bar{x} = 0$$

or

$$- \alpha p(1)\bar{y} + \delta + \beta q(1)\bar{x} = 0 \,,$$

Thus, representation (5) has a long-run equilibrium. Here, the short-term and long-run effects are separated and can be directly estimated.

Cointegration (Engle and Granger, 1987):

When the linear combination of two I(1) processes becomes an I(0) process, then these two series are cointegrated.

Why do we care about cointegration?

a. Cointegration implies the existence of long-run equilibrium

b. Cointegration implies the common stochastic trend

c. With cointegration, we can separate the short- and long- run relationship among variables

d. Cointegration can be used to improve long-run forecast accuracy

e. Cointegration implies restrictions on the parameters and proper accounting of these restrictions could improve estimation efficiency

Definition: The elements of a k-dimensional vector Y are cointegrated of order (d, c), denoted as $Y \sim CI(d, c)$, if all elements of Y are $I(d)$ and there exists at least one non-trivial linear combination z of these variables, which is $I(d - c)$, where $\mathrm{d} \geq \mathrm{c} > 0$ holds, i.e.

$$\beta' Y_t = z_t \sim I(d - c).$$

Here β is termed as a cointegration vector.

The number of linearly independent cointegration vectors is known as cointegration rank r. The cointegration matrix B is formed by cointegration vectors as the columns, with $B'Y_t = Z_t$.

14.4 Cointegration and Granger Representation Theorem

Cointegration and the Granger representation theorem are key concepts in time series analysis, particularly in the context of non-stationary data.

Cointegration refers to a statistical property of a collection of time series variables which, while individually non-stationary (typically following a stochastic trend), can exhibit a stable long-term relationship when combined. Specifically, two or more non-stationary series are said to be cointegrated if there is a linear combination of them that is stationary.

The Granger representation theorem extends the concept of cointegration by showing how cointegrated time series can be described using an error correction model (ECM). The theorem states that if two or more time series are cointegrated, then there exists a dynamic model that describes the short-term behavior of the series in relation to their long-term relationship.

Understanding cointegration and the Granger representation theorem is crucial for:

- Modeling Economic Relationships: Many economic variables are believed to be related in the long run (e.g., GDP and consumption).
- Forecasting: Improved predictions by considering both short-term dynamics and long-term relationships.
- Policy Analysis: Assessing the effects of interventions in economic models where variables are cointegrated.

Cointegration highlights the existence of long-term relationships among non-stationary time series, while the Granger representation theorem provides a framework for modeling the dynamics of these relationships through error correction models.

14.4 Cointegration for Bivariate Processes

Bivariate cointegration specifically refers to the analysis of the cointegration relationship between two time series. Bivariate cointegration specifically refers to the analysis of the cointegration relationship between two time series. Bivariate cointegration is a fundamental concept for understanding the long-run relationships between two economic or financial time series. By establishing cointegration, analysts can more accurately model and predict future behaviors of these related series, taking into account both their long-term equilibrium and short-term dynamics.

Here the notation $x_t \sim I(q)$ denotes that the time series x_t is integrated of order q. Thus $I(0)$ represents a stationary process and $I(1)$ represents an integrated process of order one.

Let us consider that x_t and y_t be two $I(1)$ processes. If there exists a parameter b such that

$$y_t - bx_t = z_t + a \tag{1}$$

is stationary then x_t and y_t are said to be cointegrated. Here z_t is a stationary process. Then Corresponding equilibrium relation is

$$y = a + bx \tag{2}$$

where,

a: level of equilibrium relation

$\beta' = [1 \quad -b]$: Cointegration vector and

z: Deviations from the equilibrium, *i.e.*, *equilibrium error*.

Since z has finite variance, the deviations from the equilibrium are bounded and the system is always returning to its equilibrium path. In this sense, relation (2) is called an *attractor*. Cointegration of x and y implies that both variables follow a common stochastic trend. Now, we model it as a random walk,

$$w_t = w_{t-1} + u_t, \tag{3}$$

where,

u_t: White noise process

The two cointegrated $I(1)$ processes can be represented as

$$y_t = b\, w_t + \tilde{y}_t \; with \; \tilde{y}_t \sim I(0) \tag{4}$$

$$x_t = w_t + \tilde{x}_t \; with \; \tilde{x}_t \sim I(0). \tag{5}$$

The linear combination

$$y_t - b\, x_t = \tilde{y}_t - b\, \tilde{x}_t = z_t \tag{6}$$

is a linear combination $\tilde{y}_t - b\, \tilde{x}_t$ of two $I(0)$ processes. Thus, it is also $I(0)$ (or stationary). Hence, (6) is a cointegrating relation.

The error correction representation for k-variables case is

$$\Delta Y_t = \delta - A(I)Y_{t-1} + \sum_{j=1}^{p-1} A_j^* \, \Delta Y_{t-j} + U_t$$

For bivariate case, we write

$$\Delta Y_t = \begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix},$$

$$\delta = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix},$$

$$A(I) = \begin{pmatrix} \gamma_y & -b\gamma_y \\ \gamma_x & -b\gamma_x \end{pmatrix},$$

$$U_t = \begin{pmatrix} u_{yt} \\ u_{xt} \end{pmatrix}$$

and

$$A_j^* = \begin{pmatrix} a_{yj} & a_{xj} \\ b_{yj} & b_{xj} \end{pmatrix}$$

Then, in the bivariate case, the reduced form can be written as

$$\Delta y_t = a_0 - \gamma_y (y_{t-1} - b x_{t-1}) + \Sigma_{j=1}^{p_x} a_{xj} \Delta x_{t-j} + \Sigma_{j=1}^{p_y} a_{yj} \Delta y_{t-j} + u_{y,t}$$

$$\Delta x_t = b_0 - \gamma_x (x_{t-1} - b y_{t-1}) + \Sigma_{j=1}^{q_x} b_{xj} \Delta x_{t-j} + \Sigma_{j=1}^{q_y} b_{yj} \Delta y_{t-j} + u_{x,t}$$

(7)

Equation (7) involves p_x $[q_x]$ and p_y $[q_y]$ terms involving lagged differences Δx_{t-j} and Δy_{t-j} respectively. The representation contains stationary variables although the underlying relation

is between nonstationary $I(1)$ variables. It separates short-run adjustments from long-term equilibrium term. Model fitted only for first differences misses the equilibrium term $y_{t-1} - b x_{t-1}$ or $(x_{t-1} - b y_{t-1})$.

If x and y are cointegrated, at least one γ_i, $i = x, y$, is different from zero. If the variables are cointegrated the traditional statistical procedures can be applied. The system reacts to the deviations from the equilibrium relations which is lagged by one period. If $b > 0$, (7) is stable whenever $0 \leq \gamma_y < 2$ and $0 \leq \gamma_x < 2$, and at least one of the two parameters is different from zero. If $\gamma_x = 0$, second equation of (7) becomes

$$\Delta x_t = b_0 + \Sigma_{j=1}^{q_x} b_{xj} \Delta x_{t-j} + \Sigma_{j=1}^{q_y} b_{yj} \Delta y_{t-j} + u_{x,t}$$

The adjustment is only possible via changes in y. Then development of x is independent of the equilibrium error, it is the stochastic trend driving the system. In this situation, x is called *weakly exogenous*.

If $\gamma_x > 0$ and $\gamma_y < 0$, or vice-versa, the system might be stable depending on the other parameters. Thus, the following two situations can occur:

(i) The two variables are not cointegrated, i.e. $\gamma_x = \gamma_y = 0$. Then the system contains two stochastic trends.

(ii) The two variables are cointegrated, i.e. at least one $\gamma_i, i = x, y$, is positive. Then the system contains one cointegrating relation and one common stochastic trend.

14.4.1 The advantage of error correction representation

- ECMs measure the correction from disequilibrium of the previous period.
- ECMs are formulated in terms of first differences and eliminate stochastic trends from the variables involved. It resolves the problem of spurious regressions.
- They can easily fitted as the variables involved are stationary.
- The disequilibrium error term is a stationary variable (by definition of cointegration). The cointegrated variables imply that there is some adjustment process preventing the errors in the long-run relationship from becoming larger and larger.

Example: Consider an ARIMA(1,1,0) process $\{x_t\}$ as

$$\Delta x_t = \alpha \Delta x_{t-1} + u_t,\ |\alpha| < 1$$

$$\text{or } (1 - \alpha L)\Delta x_t = u_t,\ u_t \sim WN \tag{8}$$

$$\text{and } y_t = b x_t + z_t; b \neq 0, \tag{9}$$

$$\text{with } z_t = \rho z_{t-1} + v_t,\ v_t \sim WN \tag{10}$$

The x and y are cointegrated for $|\rho| < 1$. If $\rho = 1$, $y_t - b x_t \sim I(1)$ and there is no cointegration. For $\rho = 1$, the development of y is determined by two stochastic trends (both x_t and z_t are (1)). For error correction representation, we write

$$y_t - \rho\, y_{t-1} = b\, x_t - \rho\, b\, x_{t-1} + v_t$$

or

$$\Delta y_t = -(1-\rho)y_{t-1} + b\,(1-\rho)x_{t-1} + b\Delta x_t + v_t \quad (11)$$

Then, substituting $\Delta x_t = \alpha\,\Delta x_{t-1} + u_t$ in (11), the reduced form of the system is

$$\Delta x_t = \alpha\,\Delta x_{t-1} + u_{x,t} \quad (12)$$

and $\Delta y_t = -(1-\rho)(y_{t-1} - b\,x_{t-1}) + b\alpha\Delta x_{t-1} + u_{y,t}$, (13)

where $u_{x,t} = u_t$ and $u_{y,t} = v_t + bu_t$.

The equation (12) of x does not involve the equilibrium error $y - bx$. Thus, x is weakly exogenous and drives the whole system.

If $-1 < \rho < 1$, then $0 < \gamma_y = (1-\rho) < 2$. Thus, the system is stable; y is adjusting to the long-run equilibrium. For $\rho = 1$, i.e., there is no cointegration, the error correction term vanishes from (13) and the system contains two stochastic trends.

The error correction model only contains stationary variables, the differences of $I(1)$ *variables and the stationary equilibrium error.*

14.5 Cointegration tests in static model

In the context of time series analysis, cointegration tests are essential for assessing whether a long-run equilibrium relationship exists between non-stationary time series variables. In a static model, we typically imply a model where the relationships do not change over time, and the focus is solely on the long-term relationship between variables without considering dynamics like lags. Here's some cointegration tests that can be applied in a static modeling framework:

- Engle-Granger Two-Step Approach
- Johansen Cointegration Test
- Phillips-Ouliaris Cointegration Test

- Kao Cointegration Test
- Cointegration in Error Correction Models (ECM)

14.5.1 Engle and Granger two-step procedure for cointegration analysis

The Engle-Granger two-step procedure is a widely used method for testing for cointegration between two or more time series. Cointegration implies that even though the individual time series may be non-stationary, a linear combination of them is stationary. This two-step procedure is a straightforward yet powerful tool for investigating the long-run relationship between non-stationary time series. Its simplicity makes it popular in empirical studies. However, users should ensure the assumptions and conditions for validity are met, particularly regarding the stationarity of residuals.

(i) Estimate the long-run (equilibrium) equation:

$$y_t = \delta_0 + \delta_1 x_t + u_t \qquad (14)$$

$\hat{u}_t = y_t - \hat{\delta}_0 - \hat{\delta}_1 x_t$: OLS residuals are a measure of disequilibrium

A test of cointegration is a test of whether $\hat{u}_t$ is stationary.

Apply ADF tests on the residuals, (critical values are given by MacKinnon, 1991). If cointegration holds, the OLS estimator of (14) is said to be super-consistent.

Implications: As $T \to \infty$ there is no need to include I(0) variables in the cointegrating equation.

(ii) Estimate Error Correction Model

$$\Delta y_t = \phi_0 + \Sigma_{j=1}\phi_j \Delta y_{t-j} + \Sigma_{h=0}\theta_h \Delta x_{t-h} + \alpha \hat{u}_{t-1} + \epsilon_t$$

by OLS. This equation has only $I(0)$ variables and standard hypothesis testing using t ratios and diagnostic testing of the error term is appropriate.

Suppose we consider the special case:

$$\Delta y_t = \phi_0 + \phi_1 \Delta y_{t-1} + \theta_1 \Delta x_{t-1} + \alpha(y_{t-1} - \hat{\delta}_0 - \hat{\delta}_1 x_{t-1}) + \epsilon_t$$

ECM describes how y and x behave in the short run consistent with a long run cointegrating relationship.

Dynamic approach to ECM and cointegration

As residuals of (14) often have serial correlation, the least squares estimates can be substantially biased in small samples. For reducing the bias, one may allow for some dynamics in the model. For this purpose:

(i) Apply least squares to estimate autoregressive distributed lag (ADL) model

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \gamma y_{t-1} + \epsilon_t \qquad (15)$$

Solve (15) for the long run equation

$$y_t = \frac{\alpha}{1-\gamma} + \frac{\beta_0 - \beta_1}{1-\gamma} x_t + u_t$$

$\hat{u}_t = y_t - \frac{\alpha}{1-\gamma} - \frac{\beta_0 - \beta_1}{1-\gamma} x_t$: Estimated residuals from (15).

where, $\hat{u}_t$ are measure of disequilibrium.

A test of cointegration is a test of whether $\hat{u}_t$ is stationary or not. The ECM model can be estimated using the residuals from (15). If cointegration holds, the OLS estimators of (15) are super-consistent.

14.5.2 Johansen Methodology for Cointegration Test (Johansen, 1995)

The Johansen methodology is a statistical technique used to test for cointegration among multiple time series. Cointegration is a concept from econometrics that refers to the existence of a long-term equilibrium relationship among non-stationary time series. If two or more time series are individually non-stationary but a linear combination of them is stationary, they are said to be cointegrated. This methodology is widely used in econometrics to analyze and model economic relationships, such as those between macroeconomic variables (e.g., GDP,

inflation, interest rates) and It's particularly useful in establishing relationships that can inform policy analyses and forecasting.

It is a powerful tool for assessing the existence of cointegration between multiple time series and understanding long-term economic relationships.

Write (7), without deterministic term, as

$$\Delta Y_t + \Gamma B' Y_{t-1} = \sum_{j=1}^{p-1} A_j^* \Delta Y_{t-j} + U_t$$

Given r, MLE of B defines the combination of y_{t-1} giving r largest canonical correlations of Δy_t with y_{t-1} after correcting for lagged differences and deterministic variables. Johansen proposes two different likelihood ratio tests for the significance of these canonical correlations and thereby the reduced rank of the Π matrix, (i) trace test and (ii) maximum eigenvalue test.

Let $\hat{\lambda}_i$ be the i^{th} largest canonical correlation. Then,

(i) For testing the null hypothesis of r cointegrating vectors against the alternative hypothesis of m cointegrating vectors, the trace statistic is

$$J_{tr} = -n \sum_{i=r-1}^{m} \ln(1 - \hat{\lambda}_i)$$

$$J_{max} = -nln(1 - \hat{\lambda}_{r+1})$$

(ii) For testing the null hypothesis of r cointegrating vectors against the alternative hypothesis of $r + 1$ cointegrating vectors, the maximum eigenvalue test statistic is $J_{max} = -nln(1 - \hat{\lambda}_{r+1})$

The critical values are given by Johansen and Juselius (1990).

Cointegration tests in the static model context are vital for confirming long-run relationships among non-stationary time series. The choice of the test depends on the specific characteristics of the data, such as the number of time series being analysed and whether they

are in a panel structure. Understanding the underlying assumptions of each test and the implications of findings is crucial for sound empirical analysis.

14.6 Self-Assessment Exercise

1. What is spurious regression, and why does it occur in time-series analysis?
2. Explain the role of non-stationarity in causing spurious regression results.
3. Discuss the implications of spurious regression for hypothesis testing in econometric models.
4. Define cointegration and explain its significance in time-series analysis.
5. What are the key conditions for two time-series variables to be cointegrated?
6. Explain the concept of a cointegrating vector and its role in cointegration analysis.
7. What is an Error Correction Model (ECM), and how is it derived?
8. How does an ECM capture both short-term dynamics and long-term equilibrium relationships?
9. Explain the role of the error correction term in an ECM.
10. Describe the steps involved in estimating an ECM for a pair of cointegrated variables.
11. What is the Granger Representation Theorem, and why is it important in time-series analysis? How does the Granger Representation Theorem establish a link between cointegration and ECMs?
12. In a multivariate system, how does the Granger Representation Theorem guide the modeling of relationships among variables?
13. What is bivariate cointegration, and how is it analyzed?
14. Describe the Engle-Granger two-step procedure for testing cointegration in bivariate models.

15. What are the limitations of the Engle-Granger cointegration test?

16. What are the main tests used to detect cointegration in static models?

17. Explain the concept of residual-based tests for cointegration.

18. How does the Johansen cointegration test extend the Engle-Granger framework for multivariate systems?

19. Compare the performance of Engle-Granger and Johansen tests in detecting cointegration.

20. Define error correction representation of a bivariate VAR process. How it helps in testing long term and short term relationship between two nonstationary time series? Describe Engle-Granger procedure for the cointegration analysis.

14.7 Summary

This unit delves into key concepts and methods for analysing relationships among non-stationary time series. It begins with a discussion on spurious regression, highlighting the misleading inferences that arise when non-stationary variables are regressed without proper adjustments. The unit introduces cointegration as a solution, explaining its significance in identifying long-term equilibrium relationships between time-series variables.

The Error Correction Model (ECM) is presented as a framework for integrating short-term dynamics with long-term cointegrating relationships, emphasizing its relevance in econometric modelling. The Granger Representation Theorem is discussed as a theoretical foundation linking cointegration and ECMs, demonstrating how cointegrated systems naturally lead to error correction representations.

Specific focus is given to bivariate cointegration, outlining methods for identifying and interpreting cointegrated relationships in pairs of variables. Cointegration tests in static models, including methods like the Engle-Granger two-step procedure, are explained in detail, with attention to their assumptions, applications, and limitations.

14.8 References

- Brockwell, P.J. and Davis, (2009) R.A.. Time Series: Theory and Methods (Second Edition), Springer-Verlag.
- Kirchgassner, G. and Wolters, J. (2007). Introduction to Modern Time Series Analysis, Springer.
- Montgomery, D.C. and Johnson, L.A. (1977): Forecasting and Time Series Analysis, McGraw Hill.

14.9 Further Readings

- Box, George E. P., Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung. (2015) Time Series Analysis, Forecasting and Control, Wiley.
- Chatfield, C. (1980). *The Analysis of Time Series –An Introduction*, Chapman & Hall.
- Granger, C.W.J. and Newbold (1984): Forecasting Econometric Time Series, Third Edition, Academic Press.