TIME SERIES REGRESSION MODELS

Dr. Panduranga Rao Repalle
Dr. Venkata Seshagiri Rao Yangala
Dr. Nagaraju Vellanki
Dr. D. Chenna Kesavaiah

Sahasra Publications **www.sahasrapublications.com**

Dr. Panduranga Rao Repalle, completed Ph.D. from Jawaharlal Nehru Technological University, Hyderabad. He is currently working as Associate professor in Vignan Institute of Technology and Science (Autonomous). He had23years of Teaching Experience in various Engineering Institutions. He had published more than 15 publications in reputed National and International Journals with Scopus and Quartile indexing along with 3 patents and also published one text book. His Area of specialization includerough sets related to the applications in Artificial Intelligence, Data Mining and Machine Learning.

Dr. Venkata Seshagiri Rao Yangala, has been associated with Vignan Institute of Technology and Science(Autonomous) as Associate professor of Mathematics in Basic Sciences and Humanities department. He was awarded with Ph.D. degree from Jawaharlal Nehru Technological University, Hyderabad. He has 22 years of teaching experience and published 25 research publications in various National, International, Scopus and SCI journals, participated in various workshops and has three patents and also published one text book. Also he worked as various positions in Academics as Head of the department. His area of interests is Lattice Theory, Fluid Dynamics and Mechanics.

Dr. Nagaraju Vellanki awarded Ph.D. from Krishna University, Machilipatnam, Andhra Pradesh, India. He has 16 years of teaching experience for both UG and PG courses. At present working as an Assistant professor in B V Raju Institute of Technology (Autonomous) Narsapur, Medak(Dist), Telangana. He published 16 research papers in various National, International, and Scopus journals, participated in various workshops. He has one patent and one Book. His Area of interest Fluid Dynamics.

Dr. D. Chenna Kesavaiah awarded a Ph.D. from Sri Venkateswara University, Tirupati. He has almost 15 years of teaching experience in various engineering colleges. He published more than 64 research papers in recognized national and international journals. He is currently working as an Associate Professor in Vignan Institute of Technology and Science (Autonomous), Hyderabad. His research specialization is Fluid Dynamics.

Sahasra Publications
Researchers Connect
www.sahasrapublications.com
Email: info@sahasrapublications.com
ISBN 978-81-19093-23-6

TIME SERIES REGRESSION MODELS

Edition-I

Dr. Panduranga Rao Repalle
Associate Professor
Vignan Institute of Technology and Science
(Autonomous)

Dr. Venkata Seshagiri Rao Yangala
Associate Professor
Vignan Institute of Technology and Science
(Autonomous)

Dr. Nagaraju Vellanki
Assistant Professor
B V Raju Institute of Technology
(Autonomous) Narsapur Medak(Dist),
Telangana

Dr. D. Chenna Kesavaiah
Associate Professor
Vignan Institute of Technology and Science
(Autonomous), Hyderabad

Published By
Sahasra Publications
203, Park view, Pushpak Co-op Housing Society
Prashanthi Nagar Industrial Estate
Kukatpally, Hyderabad, Telangana, India
Email: info@sahasrapublications.com

Author(s)

Dr. Panduranga Rao Repalle
Associate Professor
Vignan Institute of Technology and Science (Autonomous)

Dr. Venkata Seshagiri Rao Yangala
Associate Professor
Vignan Institute of Technology and Science (Autonomous)

Dr. Nagaraju Vellanki
Assistant Professor
B V Raju Institute of Technology (Autonomous) Narsapur Medak(Dist), Telangana

Dr. D. Chenna Kesavaiah
Associate Professor
Vignan Institute of Technology and Science (Autonomous), Hyderabad

Book Title: Time Series Regression Models
Publication Year: 2023
ISBN: 978-81-19093-23-6
DOI: 10.58673/sp.2023.11.15.3
Price: Rs. 780/-

ABOUT AUTHORS

Dr. Panduranga Rao Repalle, completed Ph.D. from Jawaharlal Nehru Technological University, Hyderabad. He is currently working as Associate professor in Vignan Institute of Technology and Science (Autonomous). He had23years of Teaching Experience in various Engineering Institutions. He had published more than 15 publications in reputed National and International Journals with Scopus and Quartile indexing along with 3 patents and also published one text book. His Area of specialization includerough sets related to the applications in Artificial Intelligence, Data Mining and Machine Learning.

Dr. Venkata Seshagiri Rao Yangala, has been associated with Vignan Institute of Technology and Science(Autonomous) as Associate professor of Mathematics in Basic Sciences and Humanities department. He was awarded with Ph.D. degree from Jawaharlal Nehru Technological University, Hyderabad. He has 22 years of teaching experience and published 25 research publications in various National, International, Scopus and SCI journals, participated in various workshops and has three patents and also published one text book. Also he worked as various positions in Academics as Head of the department. His area of interests is Lattice Theory, Fluid Dynamics and Mechanics.

Dr. Nagaraju Vellanki awarded Ph.D. from Krishna University, Machilipatnam, Andhra Pradesh, India. He has 16 years of teaching experience for both UG and PG courses. At present working as an Assistant professor in B V Raju Institute of Technology (Autonomous) Narsapur, Medak(Dist), Telangana. He published 16 research papers in various National, International, and Scopus journals, participated in various workshops. He has one patent and one Book. His Area of interest Fluid Dynamics.

Dr. D. Chenna Kesavaiah awarded a Ph.D. from Sri Venkateswara University, Tirupati. He has almost 15 years of teaching experience in various engineering colleges. He published more than 64 research papers in recognized national and international journals. He is currently working as an Associate Professor in Vignan Institute of Technology and Science (Autonomous), Hyderabad. His research specialization is Fluid Dynamics.

CONTENTS

CHAPTER - I INTRODUCTION

1.1. GENERAL INTRODUCTION

Statistical inference in time series regression models is a comprehensive area of statistical analysis that focuses on understanding and forecasting data that is collected over time. Time series data is unique in that it is sequentially dependent, meaning that observations in the series are dependent on previous observations. This characteristic distinguishes time series analysis from other types of statistical analysis.

Here's a general introduction to this topic:

1. **Understanding Time Series Data**: Time series data is a collection of observations recorded at regular time intervals. This could be anything from daily stock prices to yearly climate data. What makes time series data special is its chronological order.
2. **Components of Time Series**: A time series can be broken down into several components:
    - **Trend**: The long-term movement in the data over time.
    - **Seasonality**: Regular patterns that occur on a predictable cycle, like monthly or quarterly.
    - **Cyclical Components**: Fluctuations that are not of a fixed frequency.
    - **Random or Irregular Movements**: These are unpredictable and don't follow a pattern.
3. **Time Series Regression Models**: These models are used to predict or forecast future values in the series based on past values. The most basic model is the linear regression model, but time series data often require more sophisticated approaches due to factors like trend, seasonality, and autocorrelation (the correlation of a variable with itself across different time lags).
4. **Statistical Inference in Time Series**: This involves estimating the parameters of your model, testing hypotheses, and making predictions. Key steps include:
    - **Model Identification**: Determining the appropriate model to fit the data.

- **Parameter Estimation**: Estimating the model parameters using methods like Maximum Likelihood Estimation or Ordinary Least Squares.
- **Diagnostic Checking**: Assessing the adequacy of the model by looking at the residuals.
- **Forecasting**: Using the model to predict future values in the time series.

5. **Challenges and Considerations**:
    - **Autocorrelation**: In time series data, current observations might be correlated with past observations, which can violate the assumptions of standard regression models.
    - **Non-Stationarity**: If a time series is non-stationary, its statistical properties like mean and variance change over time, which can affect the model's performance.
    - **Model Selection and Overfitting**: Choosing the right model complexity is crucial to avoid overfitting.
6. **Advanced Models in Time Series Analysis**: Apart from linear regression, there are several advanced models like ARIMA (Autoregressive Integrated Moving Average), Seasonal Decomposition of Time Series (STL), and Vector Autoregression (VAR) that are widely used for more complex time series data.
7. **Applications**: Time series regression models are widely used in various fields such as economics (for forecasting GDP, unemployment rates), finance (stock prices, exchange rates), environmental science (climate modeling), and many more.

In summary, statistical inference in time series regression models is a vital tool in the analysis and forecasting of data collected over time. It requires a solid understanding of the unique characteristics of time series data and the application of specialized models and techniques to analyze and make predictions based on this data.

1.2. CLASSIFICATION OF TIME SERIES REGRESSION MODELS

The various time series regression models existing in the literature can be broadly divided into four groups namely,

i. Univariate Linear Time Series Regression Models.
ii. Univariate Non-linear Time Series Regression Models.
iii. Multivariate Linear Time Series Regression Models.
iv. Multivariate Non-linear Time Series Regression Models.

Following are some important time series regression models given in the literature:

1. Autoregressive Model with order 'p': AR (p)
2. Moving Average Model with order 'q': MA (q)
3. Autoregressive Moving Average (ARMA) Model: ARMA (p, q)
4. Autoregressive Integrated Moving Average (ARIMA) model: ARIMA (p, d, q).
5. Autoregressive Distributed Lag Models.
6. Covariance Stationary Models.
7. State-space Models.
8. Dynamic Macro Econometric Models.
9. Linear Time Series Models based on State –space processes.
10. ARMA Models based on State –space processes.
11. ARIMA Models based on State-space processes.
12. Unobserved Components Time Series Model.
13. Linear Time Series Model with Time varying coefficients.
14. Non-stationary Time Series Regression Models.
15. Time series Models of Changes in Regimes.
16. Autoregressive Conditional Heteroscedasticity (ARCH) Model.
17. Generalized ARCH (GARCH) Model.
18. Integrated GARCH (IGARCH) Model.
19. Treshold ARCH (TARCH) Model.
20. Exponential GARCH (EGARCH) Model.
21. Regime Switching Time Series Models.
22. Time Series Model with Non-stationary Regressors.
23. Vector Autoregressive (VAR) Model: VAR (p).
24. Vector Error Correction Model (VECM).
25. Simultaneous Structural Equations Time Series Regression Model.
26. Vector Moving Average (VMA) Model: VMA (q).
27. Vector ARMA Model: VARMA (p, q).

28. Vector ARIMA Model: VARIMA (p, d, q).
29. Bayesian VAR Model: BVAR (p).
30. Multivariate Time Series Models with Rational Expectations.

CHAPTER - II SOME IMPORTANT UNIVARIATE TIME SERIES REGRESSION MODELS

2.1. INTRODUCTION

In univariate time series regression models, the regressor variable is time. A linear trend relationship can be written as

$$Y = \alpha + \beta T + \in \tag{2.1.1}$$

Where T indicates time. The T variable may be specified in many ways. But each specification requires to define the origin from which time is measured and the unit of measurement that is used.

When T has zero mean, the normal equations for fitting (2.1.1) will become

$$a = \bar{Y} \text{and } b = \frac{\sum Y}{\sum T^2}$$

When variables display trends, successive values tend to be fairly close together. One way of modelling such behaviour is by means of autoregression. The simplest autoregressive model is

$$Y_t = \alpha + \beta Y_{t-1} + \in_t \tag{2.1.2}$$

This is called a first order autoregressive model and is denoted by AR(1). The order indicates the maximum lag in the equation.

From the equation (2.1.2), we make the following assumptions about $\in$ variable

$$E(\in_i) = 0 \text{ for all i}$$

$$E(\in_i^2) = \sigma^2 \text{ for all I} \tag{2.1.3}$$

$$E(\in_i \in_j) = 0 \text{ for all } i \neq j$$

These assumptions define a white noise series. Here the crucial question is, how does the Y series behave over time. Assuming that process stated a very long time ago, we take

$$Y_t = \alpha\left(1+\beta+\beta^2+.....\right)+\left(\in_t +\beta\in_{t-1}+\beta_t^2\in_{t-2}+.....\right) \quad (2.1.4)$$

$$E\left(Y_t\right)=\alpha\left(1+\beta+\beta^2+.....\right)$$

This expectation exists only when the infinite geometric series on the right hand side has limit. The necessary and sufficient conditions is $|\beta|<1$.

The expectation is then

$$E\left(Y_t\right)=\mu=\frac{\alpha}{1-\beta} \quad (2.1.5)$$

Here variance Y will be

$$Var\left(Y\right)=\sigma_y^2-\frac{\sigma^2}{1-\beta^2} \quad (2.1.6)$$

The Y series has a constant unconditional variance, independent of time.

The covariance of Y with a lagged value itself is known as autocovariance.

The lag auto covariance is defined as

$$\gamma_s=\beta^s\sigma_y^2 \quad s=0,1,2, \quad (2.1.7)$$

So that first lag autocovariance is

$$\gamma_1=\beta\sigma_y^z$$

The autocovariances depend only on the lag length and are independent of t Dividing the covariances by the variance gives the autocorrelation coefficients, also known as series correlation coefficients. These will be defined as

$$\rho_s = \gamma_s / \gamma_o , \quad s = 0, 1, 2\ldots\ldots \qquad (2.1.8)$$

Where $\gamma_o = \sigma_u^z$

Plotting the autocorrelation coefficients against the lag lengths gives the correlogram of series.

When $|\beta| < 1$ the mean, variance, and covariances of Y series are constants independent of time. The Y series is then said to be weekly or covariance stationary.

When $\beta = 1$, the AR(1) process is said to have unit root. The equation becomes.

$$Y_t = \alpha + Y_{t-1} + \in_1 \qquad (2.1.9)$$

Which is called a random walk with drift.

The conditional expectation and conditional variance are

$$E\left(Y_t / Y_0\right) = at + Y_o$$

Which increases or decreases without limit as t increases

$$Var\left(Y_t / Y_0\right) = t\sigma^2$$

Which increases without limit

In this case the unconditional mean and variance do not exist. The Y series is then said to be nonstationary.

2.2. AUTOREGRESSIVE MODEL WITH ORDER p : AR(p)

A common approach for modelling univariate time series is the autoregressive (AR) model

The AR(p) model is defined as

$$X_t = \delta + \phi X_{t-1} + \phi_2 X_{t-2} + ...\phi_p X_{t-p} + \in_t \qquad (2.2.1)$$

where X_t is the time series, $\delta = \left(1 - \sum_{i-1}^{p} \phi_i\right)\mu$, where μ is the process mean

$\in_t$ is the white noise

$\phi_1, \phi_2 .....\phi_p$ are the parameters of the model

p is the order of the AR model

Some constraints are necessary on the values of the parameters of this model so that model remains wide-sense stationary. For an AR(P) model to be wide-sense stationary, the roots of the polynomial $Z^p - \sum_{i=1}^{p} \phi_i Z^{p-i}$ must lie within the unit circle, i.e., each root Z_i must satisfy $|Z_i| > |$.

Estimation of AR Parameters

The AR(p) model is given by the equation

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \in_t \qquad (2.2.2)$$

It is based on the parameters ϕ_i where i=1, ……, p.

There is a direct correspondence between these parameters and the covariance function of the process. This correspondence can be inverted to determine the parameters from the autocorrelation function by using yule-walker equations.

$$\gamma_m = \sum_{k=1}^{p} \phi_k \gamma_{m-k} + \sigma_t^2 \delta_{m,0} \qquad (2.2.3)$$

where, m=0,1, … p. yielding p+1 equations.

γ_m is the autocorrelation function of x

$\sigma_{\in}$ is the standard deviation of the input noise process

$\delta_{m,0}$ is the Kronecker delta function

The last part of the equation is non-zero only if m=0. Hence the equation is solved by representing it as a matrix for m>0. Thus we get the equation.

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots\cdots \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots\cdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots\cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \end{bmatrix}$$

Solving all ϕ can be obtained.

For m =0,

$$\gamma_0 = \sum_{k=1}^{p} \phi_k \gamma_{-k} + \sigma_t^2$$

Which allows us to solve $\sigma_{\in}^2$

The above Yule-walker equations provide one root of estimating the parameters of an AR (p) model, by replacing the theoretical covariances with estimated valves. One way of specifying the estimated covariances is equivalent to a calculation using least squares regression of values X_t on the p previous values of the some series.

Derivation

The equation defining the AR process is

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \in_t$$

Multiplying both sides by X_{t-m} and taking expectation,

$$E\left[X_t \, X_{t-m} \right] = E\left[\sum_{i=1}^{p} \phi_i X_{t-i} X_{t-m} \right] + E\left[\in_t \, X_{t-m} \right]$$

By the definition of autocorrelation function,

$$E[X_t, X_{t-m}] = \gamma_m$$

The values of noise function are independent of each other and X_{t-m} is independent of $\in_t$ when 'm' is greater than zero.

For m>0

$$E[\in_t x_{t-m}] = 0$$

For m=0,

$$E[\in_t X_{t-m}], E\left[\in_t \left(\sum_{i=1}^{p} \phi_i X_{t-i} + \in_t\right)\right]$$

$$= \sum_{i=1}^{p} \phi_i E[\in_t X_{t-i}] + E[\in_t^2]$$

$$= 0 + \sigma_t^2$$

$$= \sigma_t^2 \qquad (2.2.4)$$

Now, we have for $m \geq 0$

$$\gamma_m = E\left[\sum_{i=1}^{p} \phi_i X_{t-i} X_{t-m}\right] + \sigma_t^2 \gamma_m$$

Further

$$E\left[\sum_{i=1}^{p} \phi_i X_{t-i} X_{t-m}\right] = \sum_{i=1}^{p} \phi_i E[X_t X_{t-m+i}]$$

$$= \sum_{i=1}^{p} \phi_i \gamma_{m-i}$$

Which yields the Yule-Walkes equations.

$$\gamma_m = \sum_{i=1}^{p} \phi_i \gamma_{m-i} + \sigma_t^2 \gamma_m \text{ for } m \geq 0 \qquad (2.2.5)$$

For m<0,

$$\gamma_m = \gamma - m$$

$$= \sum_{i=1}^{p} \phi_i \gamma_{|m|-i} + \sigma_\varepsilon^2 \delta_m \qquad (2.2.6)$$

2.3. MOVING AVERAGE MODEL WITH ORDER q : MA (q)

In time series analysis, the moving average (MA) model is common approach for modelling univariate time series models. The moving average model with order qMA(q) is

$$X_t = \mu + \in_t + \theta_1 \in_{t-1} + ...\theta_q \in_{t-q} \qquad (2.3.1)$$

Where, μ is the mean of the series,

$\theta_1, \theta_2 ....., \theta_q$ are the parameters of the model,

$\in_t, \in_{t-1}, ...$ are white noise error terms.

q is the order of the moving average model.

2.4. AUTOREGRESSIVE MOVING AVERAGE (ARMA) MODEL

Box and Jenkins popularized an approach that combines the moving average and the autoregressive approaches. ARMA models have developed in three directions– efficient identification and estimation procedures (for AR, MA and mixed ARMA processes) and model validation. This model is flexible due to the inclusion of both autoregression and moving average terms.

It is convenient to use the notation ARMA (p, q), where p is the order of the autoregressive part and q is the order of the moving average part.

The general AR (p) model was represented as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \in_t \qquad (2.4.1)$$

Multiplying both sides by X_{t-k} yields

$$X_{t-k} X_t = \phi_1 X_{t-k} X_{t-1} + \phi_2 X_{t-k} X_{t-2} + ... + \phi_p X_{t-k} X_{t-p} + X_{t-k} \in_t$$

Taking expected value both sides and assuming stationarity gives

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + ... + \phi_p \gamma_{k-p}$$

Where γ_k is the covariance between X_t and X_{t-k}

The MA (q) model is written as

$$X_k = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} ..... \theta_q e_{t-q}$$

Multiplying bothsides by X_{t-k} yields

$$X_{t-k} X_t = \left(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} ..... \theta_q e_{t-q}\right) x \left(e_{t-k} - \theta_1 e_{t-k-1} - \theta_2 e_{t-k-2} ..... \theta_q e_{t-k-q}\right) \qquad (2.4.2)$$

The expected value of the above equation will depend upon the value of k.

If k=0, and all other terms of the equation on (2.4.2) dropout because

$E\left(e_t e_{t+i}\right) = 0$ for $i \neq 0$ and $E\left(e_t e_{t+i}\right) = \sigma_e^2$ for $i = 0$.

Thus (2.4.1) becomes

$$\gamma_0 = \sigma_e^2 + \theta_1^2 \sigma_e^2 + \theta_2^2 \sigma_e^2 + ... + \theta_q^2 \sigma_e^2$$

To obtain the initial estimates for ARMA models, combine AR and MA models:

$$\gamma_k = \phi_1 E\left(x_t x_{t-k}\right) + ... + \phi_p E\left(X_{t-p} X_{t-k}\right) + E\left(e_t X_{t-k}\right)$$

$$-\theta_1 E\left(e_{t-1}X_{t-k}\right)-.....-\theta_q E\left(e_{t-q}X_{t-k}\right) \qquad (2.4.3)$$

If k>q, the terms $E\left(e_t X_{t-k}\right)=0$ which leaves

$$\gamma_k=\phi_1\gamma_{k-1}+\phi_2\gamma_{k-2}+.....+\phi_p\gamma_{k-p}$$

When k<q, the past errors and the X_{t-k} will be correlated and the autocovariances will be affected by the moving average part of the process, requiring that it will be included. The variance and autocovariances of an ARMA (1,1) process are therefore obtained as follows:

$$X_t=\phi_1 X_{t-1}+e_t-\theta_1 e_{t-1}$$

Multiplying bothsides by X_{t-k} gives

$$X_{t-k}X_t=\phi_1 X_{t-k}X_{t-1}+X_{t-k}e_t-\theta_1 X_{t-k}e_{t-1}$$

Taking expected values both sides results in

$$E\left[X_{t-k}X_t\right]=\phi E\left[X_{t-k}X_{t-1}\right]+E\left[X_{t-k}e_t\right]-\theta_1 E\left[X_{t-k}e_{t-1}\right]$$

If k=0

$$\gamma_0=\theta_1\gamma_1+E\left[\left(\phi_1 X_{t-1}+e_t-\theta_1 e_{t-1}\right)e_t\right]-\theta_1 E\left[\left(\phi_1 X_{t-1}+e_t-\theta_1 e_{t-1}\right)e_{t-1}\right] \qquad (2.4.4)$$

Since $x_t=\phi_1 k_{t-1}+e_t-\theta_1 e_{t-1}$

$$\gamma_0=\phi_1\gamma_1+\sigma_e^2-\theta_1\left(\phi_1-\theta_1\right)\sigma_e^2$$

Similarly if k=1

$$\gamma_1=\phi_1\gamma_0-\theta_1\sigma_e^2 \qquad (2.4.5)$$

Solving the equations (2.4.2) ad (2.4.3) for γ_0 and γ_1 we get

$$\gamma_0 = \frac{1+\theta_1^2 - 2\phi_1\theta_1}{1-\theta_1^2} \tag{2.4.6}$$

$$\gamma_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\theta_1^2} \tag{2.4.7}$$

Dividing (2.4.5) by (2.4.4) gives

$$\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2 - 2\phi_1\theta_1} \tag{2.4.8}$$

2.5. AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (AIMA) MODEL

A non-seasonal ARIMA model is classified as an ARIMA (p,d,q) model,

Here, p is the number of autoregressive terms,

d is the number of non seasonal differences and

q is the number of lagged forecast errors in the prediction equation

Generally ‘d’ may be taken as 0, 1 or occasssionaly 2.

To identity the appropriate ARIMA model for a time series, one has to begin by identifying the order (r) of differencing needing to stationarize the series and remove the gross features of seasonality, in conjunction with a variance–stabilizing transformation such as lagging or deflating.

The equation for the simplest case ARIMA (1, 1, 1) is as follows

$$(1-B).(1-\phi_1 B)X_t = \mu^1 + (1-\phi_1 B)e_t$$

The terms can be multiplied out and rearranged as follows.

$$\left[1 - B(1+\phi_1) + \phi_1 B^2\right]x_t = \mu^1 + e_t - \theta_1 e_{t-1}$$

$$X_t = (1+\phi_1)X_{t-1} - \phi_2 X_{t-2} + \mu^1 + e_t - \theta_1 e_{t-1} \tag{2.5.1}$$

In this form, the ARIMA model looks more like a conventional regression equation, except that there is more than one error term on the right hand side. ARIMA models are quite flexible and can represent a wide range of characteristics of time series that occur in practice.

ARIMA Model : p^{th} order autoregressive model

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ...\phi_p Y_{t-p} + \in_t \qquad (2.5.2)$$

ARIMA Model: q^{th} order moving average model

$$Y_t = \mu + \in_t - W_1 \in_{t-1} - W_2 \in_{t-2} .....W_q t_{t-q} \qquad (2.5.3)$$

ARIMA Model: ARMA (p, q) model

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \in_t$$

$$- W_1 \in_{t-1} - W_2 \in_{t-2} .....W_q \in_{t-q} \qquad (2.5.4)$$

ARIMA (0, 1, 1) Model

$$Y_t - Y_{t-1} = \in_t - w_1 \in_{t-1} \qquad (2.5.6)$$

2.6. AUTOREGRESIVE DISTRIBUTED LAG MODELS (ARDL)

In case where the variables in the long-run relation of interest are trend stationary, the general practice has been to de-trend the series and to model the detrended series as stationary distributed lag or autoregressive distributed lag (ARDL) models. Estimation and inference concerning the long-run properties of the model are then carried out using standard asymptotic normal theory. The analysis becomes more complicated when the variables are difference–stationary, or integrated of order 1.

The auto-regressive distributed lag model with p lags of dependent variable Y_t and q lags of additional regressor X_t, ADL (p, q) is defined as:

$$ADL(p,q) = Y_t = \beta_0 + \beta_1 Y_{t-1} + ... + \beta_p Y_{t-p} + \delta_1 X_{t-1} + ... + \delta_q X_{t-q} + \mu_t \qquad (2.6.1)$$

$\beta(L)Y_t = \beta_0 + \delta(L)X_{t-1}$ with lag-polynomials defined by

$\beta(L)Y_t = 1 - \beta_1 L - ..... - \beta_p L_p$

$\delta(L) = \delta_1 + \delta_2 L + ... + \delta_q L_{q-1}$ (2.6.2)

k additional predictors = ADL $(p, q_1, ..., q_k)$ (2.6.3)

$$\beta(L)Y_t = \beta_0 + \delta_1(L)X_{1,t-1} + \delta_2(L)X_{2,t-1} + ... + \delta_k(L)X_{k,t-1}$$

Model Assumptions

(1). $E\left[u_t / Y_{t-1,t-2,} ....., X_{1t-1}, X_{1t-2,} ..... X_{kt-1}, X_{kt-2} .....\right] = 0$ (2.6.4)

(2) (a). $(Y_{t,} X_{1t,} ..... X_{kt})$ are (strictly) stationary

(b). $(Y_{t,} X_{1t,} ....., X_{kt})$ are ergodic (2.65)

$(Y_{t,} X_{1t,} ....., X_{kt})$ and $(Y_{t-j,} X_{1t-j}, ..., X_{kt-j})$ becomes independent for $j \to \infty$

(3). Y_t and $X_{1t,} ....., X_{kt}$ have nonzero, finite fourth moments

(4). no perfect multicollinearity

2.7. NON STATIONARY TIME SERIES REGRESSION MODELS

A time series X_t is said to be stationary if its expected value and population variance are independent of time and if the population covariance between its values at time t and t+s depends on s but not on time.

An example of a stationary time series is an AR (1) process

$$X_t = \beta_2 X_{t-1} + \in_t \quad (2.7.1)$$

Here $-1 < \beta_2 < 1$ and $\in_t$ is white noise

If equation (2.7.1) is valid for time period t, it is also valid for time period t-1 i.e. the series is stationary.

$$X_{t-1} = \beta_2 X_{t-2} + \in_{t-1}$$

Substituting for X_{t-1} in equation (2.7.1), we get

$$X_t = \beta_2^2 X_{t-2} + \beta_2 \in_{t-1} + \in_t$$

Continuing this process of lagging and substituting, we get

$$X_t = \beta_2^t X_0 + \beta_2^{t-1} \in_1 + \ldots\ldots\ldots\ldots\ldots + \beta_2 \in_{t-1} + \in_t \qquad (2.7.2)$$

In the previous examples, if β_2 is equal to 1, the original series becomes

$$X_t = X_{t-1} + \in_t \qquad (2.7.3)$$

This is an example of a nonstationary process which is known as a random walk.

If it starts at X_0 at time 0, its value at time t is given by

$$X_t = X_0 + \in_t + \ldots\ldots \in_t \qquad (2.7.4)$$

The key difference between this process and the corresponding A.R(1) process is that the contribution of each innovation is permanently built into the time series. By contrast, when $\beta_2 < 1$, then contribution of each shock to the series is exponentially attenuated and eventually becomes negligible. In the more general version of the autogressive process with the constant β_1, the process becomes a random walk with drift, if β_2 equals1.

$$X_t = \beta_1 + X_{t-1} + \in_t \qquad (2.7.5)$$

If the series starts at X_0 at time 0, X_t is given by

$$X_t = X_0 + \beta_{1t} + \in_1 + \ldots\ldots\ldots + \in_t$$

The expectation of X_t at time '0' ,$\left(X_0 + \beta_{1t}\right)$ is a function of 't'. Another common example of a nonstationary time series is one possessing a time trend:

$$X_t = \beta_1 + \beta_2 t + \in_t \qquad (2.7.6)$$

This type of trend is sometimes described as a deterministic trend. The expected value of X_t at time 't' $(\beta_1 + \beta_2 t)$ is not independent of t and so X_t is nonstationary.

The key difference between a deterministic trend and a random walk with drift is that in the former, the series must keep coming back to the trend line. In any given observation, X_t will be displaced from the trend line by an amount $\in_t$, but apart from transitory effect it must adhere to the trend line. By contrast, in a random walk with drift, the displacement from the underlying trend line at time t is the random walk $\Sigma \in_t$, plus the displacement at time 0.

2.8. UNOBSERVED COMPONENTS TIME SERIES MODEL

The observed series (y_t) is defined as the sum of various components $\left(u_t^i\right), i = 1.....I,$ generally unobservable.

$$\text{The moles is} \quad y_t = \sum_{i=1}^{I} U_t^i \tag{2.8.1}$$

Here we assume that the subprocesses $\left(u_t^i\right) i = 1,....., I$ are independent of each other and we specify their marginal distributions or atleast the two first moments of these distributions.

2.8.1. Seasonal Adjustment in ARIMA Components Models

When the components u^i admit ARMA or ARIMA representations, they can also be put in state-space form of the type.

$$Z_{t+1}^i = A_i Z_t^i + \in_t^i \tag{2.8.2}$$

$$u_t^i = (1, 0, ..., 0) Z_i^i$$

Where the errors $\in^i$ are independent.

The set of state variabels associated to various comparnets as state variable are

$$Z_{t+1} = \begin{bmatrix} Z_{t+1}^1 \\ : \\ Z_{t+1}^I \end{bmatrix} = \begin{pmatrix} A_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_I \end{pmatrix} Z_t + \begin{bmatrix} \in_t^1 \\ \vdots \\ \in_t^I \end{bmatrix} \qquad (2.8.3)$$

$Y_t = (1,0,\dots\dots\dots,0,1,0\dots..,1,0,\dots..0)\ Z_t$

This form can be used for estimation. After specifying the way the matrices A_i and $V\left(\in_i^i\right)$ depend on the underlying parameters θ, we use the Kalman filter to compute the likelihood function. Kalman smoothing applied to the model (2.8.3) provides some apprepriamations of the state variables such as, $E(z/Y_t \dots., Y_T)$, t=1…. T. The underned components are coordinates of this state vector. Hence the smoothing provides some approximation of the state variables such as,

$E\left(Z_t / Y_t \dots\dots., Y_T\right), t = 1....T$. The undeserved components are coordinates of this state vector. Hence the smoothing provides some approximation of the various components.

$$u_T^i = E\left(u_t^i / Y_1, \dots\dots., Y_T\right)$$

Hence we can obtain a disaggregation of the initial series at each date $Y_t = \sum_{i=1}^{I} \overline{u}_t^i$.

2.8.2. Application to Seasonal Adjustment

The approach of the seasonal adjustment problem by regression is based on a decomposition of the original series into three components: trend, seasonal, and noise. Trend is often described by a polynomial and the seasonal can be described by a strictly periodical series. Here the trend and seasonal components modelled is restrictive. Hence this type of decomposition is not very simple

By the symmetrical approach for various components, we get

$$Y_t = u_t^T + u_t^s + u_t^I \text{ for } t \geq 0 \qquad (2.8.4)$$

Here the first one is trend, second seasonal and third is irregular part. Here each component is assumed to admit a suitable ARIMA representation.

To model the trend θ we can take a representation with a unit root in the autoregressive part

$$(1-L)^{\theta} u_t^T = \frac{\theta_T(L)}{\phi(L)} t_t^{-T} \tag{2.8.5}$$

Here θ_T and ϕ_T have roots outside of the unit circle and where the noise is of variance σ_T^2

In order to model the seasonal component of period S, we can take an autoregressive part divisible by $S(L) = 1 + L + ... + L^{S-1} = \frac{1-L^S}{1-L}$

$$S(L) u_t^s = \frac{\theta_s(L)}{\phi_s(L)} \in_t^{-s}, \text{Var}(t_t^s) = \sigma_s^2 \tag{2.8.6}$$

The irregular part can be modelled by a stationary ARMA

$$u_t^I = \frac{\theta_I(L)}{\phi_I(L)} \in_T^{-I}, \text{Var}(\in_t^I) = \sigma_I^2 \tag{2.8.7}$$

The overall model relative to the initial series is

$$Y_t = \frac{1}{(1-L)^d} \frac{\theta_T^{(L)}}{\phi_T^{(L)}} \in_t^{-T} + \frac{1}{S(L)} \frac{\theta_s^{(L)}}{\phi_s^{(L)}} t_t^{-s} + \frac{\theta_I^{(L)}}{\phi_I^{(L)}} \in_T^{-I} \tag{2.8.8}$$

The ARIMA representation of the series y is deducted from the pseudo spectrum

$$f_Y(w) = \frac{\sigma_T^2}{2\pi} \frac{\left|\theta_T(\exp(iw))\right|^2}{\left|1-\exp(iw)^{2d}\right| \left|\phi_T(\exp(iw)^2\right|} + \frac{\sigma_s^2}{2\pi} \frac{\left|\theta_s(\exp(iw))\right|^2}{\left|S(\exp(iw))\right|^2 \left|\phi_s(\exp(iw))\right|^2}$$

$$+ \frac{\sigma_T^2}{2\pi} \frac{\left|\theta_I(\exp(iw))\right|^2}{\left|(\phi_I)\exp(iw)\right|^2}$$

This can be written as

$$f_Y(w) = \frac{\sigma^2}{2\pi} \frac{\left|G(\exp(iw))\right|^2}{\left|1+(\exp(iw)\right|^2}$$

Here the derivation of the variance of the noise σ^2 and of the moving average polynomial G is difficult. Their expressions depend on the initial parameters $\sigma_T^2, \sigma_s^2, \sigma_I^2, \theta_S, \theta_I, \phi_T, \phi_s, \phi_I$ in a complex way. This explains why the use of the state-space representation for the estimation simplifies this process. This representation is particularly well–suited for the estimation of the components, trend, seasonal and irregular part.

2.8.3. Optimal Moving Average

If we assume that the initial series satisfies an unobserved component model with stationary components, the best estimator of $y_t - u_t^s$ is a linear function of the observations because of model linearity.

$$Y_t^{SA} = E\left(Y_t - u_t^s / Y_{1,\ldots,} Y_T\right) = \sum_{j=1-t}^{T-t} a_{j,t,T}\, Y_{t+j} \qquad (2.8.9)$$

We get an approximate solution replacing $(y_1 \ldots y_T)$ by $(y_t, t = -\infty, \ldots\ldots +\infty)$

$$Y_t^{SA} \approx \sum_{j=-\infty}^{+\infty} a_i Y_{t+j}$$

Here the coefficients are independent of the indexes t and T. The optimal determination of the seasonal adjustment is done through the application of a moving average $\sum_{j=-\infty}^{+\infty} a_j L^j$.

CHAPTER - III SOME IMPORTANT MULTIVARIATE TIME SERIES REGRESSION MODELS

3.1. VECTOR AUTOREGRESSION MODEL

3.1.1. Introduction

In the general practice, it is not known in prior whether the time path of the dependent variable has affected the independent variable. Bass (1969) warned that advertising may be influenced by current and past sales, and should not automatically be treated as exogeneous. This means that not only marketing activities influence sales, but change in sales may also induce marketing activities. Marketing Managers may track own-brand market share or sales, and if they observe a drop in either performance measure, they may tend to compensate it with charges in marketing activities. In its basic form, multiple time series analysis treats all variables symmetrically without making reference to the issue of dependant versus independence and permits causality testing of all variables simultaneously. This is a major advantages of multiple time series analysis models.

The multiple time series analysis methodology applies iterative processes that identify basic models, the lag structure, and relationships between variables, estimate the parameters and check the estimated model. Multiple time series analysis models have been shown to be extremely flexible in capturing the dynamic interrelationships between a set of variables, to be able to treat several variables endogeneously, not to require firm prior knowledge as the nature of the different relationships, to be able to capture both short and long run inter-relationships and to outperform multivariate time series analysis models in parameter efficiency, goodness-of-fit measures as well as in forecasting performance.

Vector Autoregression (VAR) is a model used to capture the evolution and the interdependencies between multiple time series, generalizing the univariate AR models. All the variables in a VAR are treated symmetrically by including an equation for each variable, explaining its evolution based on its own lags of all the

other variables in the model. A VAR model describes the evolution of a set of k variables, called endogeneous variables, over the sample period (t = 1,……T) as a linear function of only their past evolution. The variables are collected in a kx1 vector $Y_{t.}$ Here Y_{it} is the i^{th}element of the time t observation of variable Y_i.

3.2. VAR (p) MODEL

A, p^{th} order VAR model which is denoted by VAR (P), is

$$Y_t = C + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \in_t \qquad (3.2.1)$$

Here C is K x 1 vector of constants (intercept)

A_i is a kxk matrix of error term satisfying

1) $E(\in_t) = 0$ every errors term has mean zero
2) $E(\in_t \in^1_{t-k}) = 0$ for non-zero k – there is no correlation, across time, in particular, no serial correlation individual errors.

The t-period back observation y_{t-1} is called the t^{th} lag of y. Hence, a p^{th} order VAR is also called a VAR with p lags. All the variables used have to be of the same order of integration. We have the following different cases,

- All the variables are 1 (0) (stationary): one is in the standard case.
- All the variables are 1 (d) (non-Stationary) with d > 0.
- The variables are cointegrated the error correction term has to be included in the VAR. The model becomes a vector error correction model. It can be seen as a restricted VAR.
- The variables are not cointegrated the variables have first to be differenced 'd' times and me has a VAR in difference.

A VAR (p) process with k variables:

$$Y_t = C + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + t_t + e_t) \qquad (3.2.2)$$

Where each Y_f is kx1 vector and each A_j is a kxk matrix.

Large matrix notation is

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ . \\ . \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ . \\ . \\ c_k \end{bmatrix} + \begin{bmatrix} a^1_{1,1} & a^1_{1,2} & . & . & a^1_{1,k} \\ a^1_{2,1} & a^1_{2,2} & . & . & a^1_{2,k} \\ . & . & . & . & . \\ . & . & . & . & . \\ a^1_{k,1} & a^1_{k,2} & . & . & a^1_{k,k} \end{bmatrix} + \ldots + \begin{bmatrix} a^p_{1,1} & a^p_{1,2} & . & . & a^p_{1,k} \\ a^p_{2,1} & a^1_{2,2} & . & . & a^1_{2,k} \\ . & . & . & . & . \\ . & . & . & . & . \\ a^1_{k,1} & a^1_{k,2} & . & . & a^1_{k,k} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ . \\ . \\ y_{k,t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ . \\ . \\ e_{k,t} \end{bmatrix}$$

Concise matrix notation

VAR (P) with k variables in a general way includes 'T+1' observations Y_0 through Y_T.

$$Y = BZ + U \quad \text{Where } Y = [\, y_p \; y_{p+1.....} y_T \,] \quad (3.2.3)$$

$$= \begin{bmatrix} y_{1,p} & y_{1,p+1} & . & . & y_{1,T} \\ y_{2,p} & y^1_{2,p+1} & . & . & y_{2,T} \\ . & . & . & . & . \\ . & . & . & . & . \\ y_{k,p} & a_{k,p+1} & . & . & y_{k,T} \end{bmatrix}$$

$$B = [C \; A_1 \; A_2 ... A_p]$$

$$= \begin{bmatrix} c_1 & a^1_{1,1} & a^1_{1,2} \ldots\ldots a^1_{1,k} \ldots\ldots a^p_{1,1} & a^p_{1,2} \ldots\ldots a^p_{1,k} \\ c_2 & a^1_{2,1} & a^1_{2,2} \ldots\ldots a^1_{2,k} \ldots\ldots a^p_{2,1} & a^p_{2,2} \ldots\ldots a^p_{2,k} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ c_k & a^1_{k,1} & a^1_{k,2} \ldots\ldots a^1_{k,k} \ldots\ldots a^p_{k,1} & a^p_{k,2} \ldots\ldots a^p_{k,k} \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 1 & 1 \\ y_{p-1} & y_p & y_{T-1} \\ . & . & . \\ . & . & . \\ . & . & . \\ y_0 & y_1 & y_{T-p} \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 \\ y_{1,p-1} & y_{1,p} & y_{1,T-1} \\ y_{2,p-1} & y_{2,p} & y_{1,T-2} \\ . & . & . \\ . & . & . \\ . & . & . \\ y_{k,p-1} & y_{k,p} & y_{k,T-1} \\ y_{1,p-2} & y_{1,p-1} & y_{1,T-1} \\ y_{2,p-1} & y_{2,p-1} & y_{1,T-2} \\ . & . & . \\ . & . & . \\ . & . & . \\ y_{k,p-2} & y_{k,p-1} & y_{k,T-2} \\ . & . & . \\ . & . & . \\ . & . & . \\ y_{1,0} & y_{1,1} & y_{1,T-p} \\ . & . & . \\ . & . & . \\ . & . & . \\ y_{k,0} & y_{k,1} & y_{k,T-p} \end{bmatrix}$$

$$\text{and } U = \begin{bmatrix} e_{1,p} & e_{1,p+1} \ldots\ldots e_{1,T} \\ e_{2,p} & e_{2,p+1} \ldots\ldots e_{2,T} \\ . & . & . \\ . & . & . \\ . & . & . \\ e_{k,p} & e_{k,p+1} \ldots\ldots e_{k,T} \end{bmatrix}$$

One can solve for the coefficient matrix B (using an ordinary least squares estimation).

3.2.1. Vector Auto Regressions

VAR (1); Let k =2 and P = 1 in (3.2.2)

$$Y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

$$= C + Ay_{t-1} + e_t \qquad (3.2.4)$$

In all VARs, each variable is expressed as a linear combination of lagged values of itself and lagged values of all other variables in the group.

Let the eigen values and eigen vectors of the matrix A be

$$\wedge = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad K = \begin{bmatrix} . & . \\ c_1 & c_2 \\ . & . \end{bmatrix}$$

Provided the eigen values are distinct, the eigen vectors will be linearly independent and c will be non-singular. It then follows that

$$K^{-1}AK = \wedge \text{ and } A = K \wedge K^{-1}$$

Define a new vector of variables Z_t as

$$Z_t = K^{-1}Y_t \text{ or } Y_t = KZ_t \qquad (3.2.5)$$

The process of pre-multiplying equation (3.2.4) by C^{-1} and simplifying gives

$$Z_t = C^* + \wedge z_{t-1} + \eta_t \qquad (3.2.6)$$

$C^* = K^{-1}C$ and $\eta_t = K^{-1}e_t$ which is a white noise vector.

Each Z variable follows a separate A.R.(1) scheme and is stationary, I(0), if eigen value has modulus less than1, I(1), if eigen value is 1 and is explosive if eigen value exceeders 1.

The extension of higher-order system is simple as shown below:

A second order system is

$$Y_t = C + A_1Y_{t-1} + A_2Y_{t-2} + e_t$$

Subtracting Y_{t-1} from each sides gives

$$\Delta Y_t = C + (A_1 - I)\, Y_{t-1} + A_2 Y_{t-2} + e_t \qquad (3.2.7)$$

The process of adding and subtracting $(A_1 - I)\, Y_{t-2}$ on the right hand side and simplifying results in

$$\Delta Y_t = C + (A_1 - I)\, \Delta Y_{t-1} + \pi Y_{t-2} + e_t \qquad (3.2.8)$$

Where $\pi = I - A_1 - A_2$

An alternative representation is

$$\Delta Y_t = C - A_2 \Delta Y_{t-1} - \pi Y_{t-1} + e_t \qquad (3.2.9)$$

In the first difference reformulation of a second-order system, there will be one lagged first difference term on the right-hand side. The levels term may be lagged one period or two.

If we proceed similarly VAR (P) system defined in (13.2.4) may be re-parameterized as

$$\Delta Y_t = C + B_1 + \Delta Y_{t-1} + ... + B_{p-1}\, \Delta Y_{t-p+1} - \pi Y_{t-1} + e_t \qquad (3.2.10)$$

Where Bs are functions of the Asand $\pi = I - A_1 - ..... - A_p$.

The behaviour of the Y vector depends on the values of λ that solve

$\left|\lambda^p I - \lambda^{p-1} A_1 - ... - \lambda A_{p-1} - A_1\right| = 0$. There three possibilities of roots.

1. Rank $(\pi) = k$. If each root has less than one, π will have full rank and is non-singular. All the Y variables in (3.2.1) will be I (0) and unrestricted OLS estimates of equation (3.2.1) or (3.2.10) will find identical influences about the parameters.

2. Rank $(\pi) = r < k$. This will occurs if there is a unit root will multiplicity (k-r) and the remaining r roots are numerically less than one. The Y vector will be I(1) or higher and π may be expressed as the outer product of two (k x r) matrices, each rank r. The right hand side of equation (3.2.10) then contains r integrating variables.
3. Rank $(\pi) = 0$.This is a special case. It will only occur if $A_1 + ... + A_p = I$ in which case $\pi = 0$ and equation (3.2.10) shows that VAR should be specified solely in terms of first differences of the variables.

3.3. ORDINARY LEAST SQUARES ESTIMATION FOR VAR MODEL

3.3.1. Introduction

Consider matrix notation of a VAR (p):

$$Y = BZ + U \tag{3.3.1}$$

The multivariate least square (MLS) for B yields.

$$\hat{B} = YZ^1 (ZZ^1)^{-1} \tag{3.3.2}$$

It can be written alternatively as

$$\text{Vec } (\hat{B}) (Z^1Z)^{-1} Z \otimes I_k) \text{ Vec } (Y) \tag{3.3.3}$$

Here $\otimes$ denotes the Kronecker product and Vec the vectorization of the matrix Y. This estimator is consistent and asymptotically efficient. It is equal to the conditional maximum likelihood estimator. As explanatory variables the same in each equation, the Multivariate Least Square is equivalent in the ordinary least squares estimator applied to each equation separately.

3.3.2 Estimation of the covariance matrix of the error

As in the standard case, the MLE estimator of the covariance matrix differs from the ordinary least squares estimator.

$$\text{MLE estimator } \hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t^1 \qquad (3.3.4)$$

$$\text{OLS estimator } \hat{\Sigma} = \frac{1}{T-kp-1}\sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t^1 \text{ for a model} \qquad (3.3.5)$$

With a constant, k variables and p lags.

In a matrix notation this gives

$$\hat{\Sigma} = \frac{1}{T-kp-1}(Y - \hat{B}Z)(Y - \hat{B}Z)^1 \qquad (3.3.6)$$

Estimation of the estimator's covariance matrix. The covariance matrix of the parameters can be estimates as

$$\hat{cov}(vec(\hat{B})) = (Z^1Z)^{-1} \otimes \hat{\Sigma} \qquad (3.3.7)$$

3.4. FULL INFORMATION MAXIMUM LIKELIHOOD METHOD OF ESTIMATION

3.4.1. Introduction

Let Y_t denote an (nx1) vector. Here the hypotheses is that Y_t follows a VAR (p) in levels

Any P^{th} order VAR can be written in the form

$$\Delta Y_t = \xi_1 \Delta Y_{t-1} + \xi_2 \Delta Y_{t-2} + ... + \xi_{p-1} \Delta Y_{t-p+1} + \alpha + \xi_0 Y_{t-1} + \in_t \quad (3.4.1)$$

With $E(t_t) = 0, E(t_1 t_t^1) = \begin{cases} \Omega & \text{for } t = T \\ 0 & \text{otherwise} \end{cases}$

Suppose that each individual variable Y_{it} is I (i), although h linear combinations of Y_t are stationary.

ξ_0 can be written in the for $m\, \xi_0 = -BA^1$ (3.4.2)

For B an (nxn) matrix and A^1 an (nxn) matrix. Under the hypothesis of h cointigrating relations, only h separate linear combinations of the level of Y_{t-1} (The h elements of $z_{t-1} = A^1 Y_{t-1}$) appear in (3.4.1)

Consider a sample of T + p observations on Y, denoted as

$$(Y_{-p+1,}\ Y_{-p+2}, ...Y_T)$$

If the disturbances $\in_t$ are Gaussian, then the log likelihood of $(Y_1, Y_2.....Y_T)$ conditional on $(Y_{-p+1}, Y_{-p+2}, ..., Y_0)$ is given by

$$L(\Omega, \xi_{1,}\xi_{2,}....\xi_{p-1,}\alpha, \xi_0) = \left(\frac{-T_2}{2}\right)\log(2\pi) - \left(\frac{T}{2}\right)\log 1\,\Omega\,1$$

$$-\left(\frac{1}{2}\right)\sum_{t=1}^{T}(\Delta Y_t - \xi_1\Delta Y_{t-1} - \xi_2\Delta Y_{t-2} - ..... - \xi_{p-1}\Delta Y_{t+p-1} - \alpha - \xi_0 Y_{t-1})^1 \times$$

$$\Omega^{-1}(\Delta Y_t - \xi_1\Delta Y_{t-1} - \xi_2\Delta Y y_{t-2} - ..... - \xi_{p-1}\Delta Y_{t-p+1} - \alpha - \xi_0 Y_{t-1}) \qquad (3.4.3)$$

Here the problem is to chose $(\Omega, \xi_{1,}\xi_{2,}....\xi_{p-1,}\alpha, \xi_0)$ so as to maximize (3.4.3) subject to the constraints that ξ_0 can be written in the form of (3.4.2).Johansen's algorithm can be used to calculate the maximum likelihood estimates.

Step 1: Calculate Auxiliary regressions.

The first step is to estimate a (p-1)thorder VAR for Δy_t

That is, regress the scalar Δy_{it} on a constant and all the elements of the vectors $\Delta Y_{t-1,}\ \Delta Y_{t-2,}..... \Delta Y_{t-v+1}$ by OLS. Collect the i = 1, 2,….n OLS regressions in vector from as

$$\Delta Y_t = \hat{\pi}_0 + \hat{\pi}_1 \Delta Y_{t-1} + \hat{\pi}_2 \Delta Y_{t-2} + ... + \hat{\Delta}_{p-1} Y_{t-p+1} + \hat{u}_t \quad (3.4.4)$$

Where $\hat{\pi}_i$ denotes an (nxn) matrix of OLS coefficients estimates and $\hat{u}_t$ denotes the (nx1) vector of OLS residuals.

We also estimate the regressions, regressing the scalars $y_{i,t-1}$ on a constant and $\Delta Y_{t-1}, \Delta Y_{t-2} ..... \Delta Y_{t-p+1}$ for $i = 1, 2, ..., n$.

The second set of OLS regressions can be written as

$$Y_{t-1} = \hat{\theta} + \hat{x}_1 \Delta Y_{t-1} + \hat{x}_2 \Delta Y_{t-2} + ... + \hat{x}_{p-1} \Delta Y_{t-p+1} + \hat{v}_t \quad (3.4.5)$$

Here $\hat{v}_t$ is an (nx1) vector of residuals from the second battery of regressions.

Step 2: Calculate canonical correlations

Calculatethe sample variance-covariance matrices of the OLS residuals $\hat{u}_t$ and $\hat{v}_t$

$$\hat{\Sigma}_{VV} = \frac{1}{T} \sum_{t=1}^{T} \hat{v}_t \hat{v}_t^1 \quad (3.4.6)$$

$$\hat{\Sigma}_{UU} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t^1 \quad (3.4.7)$$

$$\hat{\Sigma}_{UV} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{v}_t^1 \quad (3.4.8)$$

$$\hat{\Sigma}_{VU} = \hat{\Sigma}_{UV}^1$$

From these, find the eigen values of the matrix

$$\hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{VU} \hat{\Sigma}_{UU}^{-1} \hat{\Sigma}_{UV} \quad (3.4.9)$$

With the eigen values order $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_n$

The maximum value attained by the log likelihood function subject to the constraint that there are 'n' cointegrating relations is given by

$$L^* \equiv \left(T_n/2\right)\log(2\pi) - \left(T_n/2\right) - \left(T/2\right)\log\left|\hat{\Sigma}_{uv}\right| - \left(T/2\right)\sum_{i=1}^{n}\log(1-\hat{\lambda}_i) \quad (3.4.10)$$

Step 3: Calculate Maximum Likelihood Estimates

Let $\hat{a}_1, \hat{a}_2, ...\hat{a}_n$ denote the (nx1) eigen vectors of (3.4.9) associated with then largest eigen values. These provide a basis for the space of cointegrating vector can be written in the form

$$a = b_1\hat{a}_1 + b_2\hat{a}_2 + ... + b_n\hat{a}_k$$

For some choice of scalars $(b_1, b_2, ...b_n)$. Johansen suggested normalizing these vectors $\hat{a}_i$ so that $\hat{a}_i\,\hat{\Sigma}_{vv}\,\hat{a}_i = 1$.

For example , if the eigen vectors $\bar{a}_i$ of (3.4.9) are calculated from a standard computer program that normalizes $\bar{a}_i^{1}\,\bar{a}_i = 1$.

Jahansen's estimate is $\hat{a}_i = \bar{a}_i \div \sqrt{\bar{a}_i^{1}\,\hat{\Sigma}_{vv}\,\bar{a}_i}$.

Collect the first h normalized vectors in an (nxn) matrix $\hat{A}$

$$\hat{A} = [\hat{a}_1, \hat{a}_2, ...\hat{a}_n] \quad (3.4.11)$$

Then the MLE of ξ_0 is given by

$$\hat{\xi}_0 = \hat{\Sigma}_{uv}\,\hat{A}\hat{A}^{1}$$

The MLE of Σ_i, for i =1, 2, ….p-1 is

$$\hat{\xi}_1 = \hat{\pi}_i - \hat{\xi}_0\,\hat{x}_i$$

And the MLE of α is

$$\hat{\alpha} = \hat{\pi}_0 - \hat{\xi}_0 \hat{\theta}$$

The MLE of Ω is

$$\hat{\Omega} = \left(\frac{1}{T}\right) \sum \left[(\hat{u}_{il} - \hat{\Sigma}_0 \hat{v}_t) (\hat{u}_t - \hat{v}_t)^1 \right] \qquad (3.4.12)$$

3.5. CONDITIONAL MAXIMUM LIKELIHOOD METHOD OF ESTIMATION

Let Y_t denote an (nx1) vector containing the values that n variables assume at date t. The dynamics of Y_t are governed by a p^{th} order Gaussian vector autoregression.

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \in_t \qquad (3.5.1)$$

Where $\in_t$ i.i.d $N(0, \Omega)$

Suppose that we observe each of these n variables for (T+p) time periods. The simplest approach is to condition on the first p observations $(Y_{-p+1} + Y_{-p+2} + ... + Y_1)$ and to base estimation on the last T observations (Y_1, Y_2, …. Y_T). The objective then is to form the conditional likelihood

$$f_{Y_T, Y_{T-1} ..... Y_1 | Y_0, Y_1 ..... Y_p} \left(Y_T, Y_{T-1} ..... Y_1 \middle| Y_0, Y_{-1} ..... Y_{-p+i,}; \theta \right) \qquad (3.5.2)$$

and maximize with respect to θ.

Here θ is a vector that contains the elements of $c, \phi_1, \phi_2 .... \phi_p$ and Ω. Vector autoregressions are invariably estimated on the basis of the conditional likelihood function (3.5.2) rather than the full-sample unconditional likelihood.

For brevity, we hereafter refer to (3.5.2) as the "likelihood function" and the value of θ that maximizes (3.5.2) as the maximum likelihood estimates.

Conditional on the values of y observed through date t-1, the value of y for date t is equal to a constant

$$C + \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + ... + \phi_p Y_{t-p} \text{ plus a } N(0, \Omega) \tag{3.5.3}$$

Thus $$Y \left| Y_{t-1}.Y_{t-2} .....Y_{-p+1} \sim N\left((c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p}), \Omega\right) \right. \tag{3.5.4}$$

Let X_t denote a vector containing a constant term p and p lags of each of the elements of y.

$$X_t = \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ . \\ . \\ Y_{t-p} \end{bmatrix} \tag{3.5.5}$$

Hence X_t is an [(np+1) x^1] vector.

Let π^1 denote the following [(nx(np+1) matrix

$$\pi^1 = [c, \phi_1, \phi_2 ... \phi_p] \tag{3.5.6}$$

The conditional mean (3.5.3) is equal to $\pi^1 X_t$.

The j^{th} row of π^1 contains the parameters of the j^{th} equation in the VAR. Using this notation (3.5.4) can be written as

$$Y_t \left| Y_{t-1}.Y_{t-2} .....Y_{-p+1} \sim N(\pi^1 X_t, \Omega) \right. \tag{3.5.7}$$

The conditional density of the i^{th} observation is

$$f_{Y_t, Y_{t-1}.....Y_1 \left| Y_0, Y_1 ..... Y_p \right.} \left(Y_t, Y_{t-1}.....Y_1 \left| Y_0, Y_{-1}.....Y_{-p+i,}; \theta \right. \right)$$

$$= (2\pi)^{-n/2} \left| \Omega^{-1} \right|^{1/2} \exp\left[(-1/2) (y_t - \pi^1 x_t)^1 \Omega (y_t - \pi^1 x_t) \right] \tag{3.5.8}$$

The joint density of observations through t conditioned on $Y_0, Y_{t-1}, \ldots . Y_{-p+1})$ satisfies

$$f_{Y_t, Y_{t-1} \ldots . Y_1 \mid Y_0, Y_1 \ldots \ldots Y_p} \left(Y_t, Y_{t-1} \ldots . Y_1 \middle| Y_0, Y_{-1} \ldots \ldots Y_{-p+i,}; \theta \right)$$

$$= f_{Y_t, Y_{t-1} \ldots . Y_1 \mid Y_0, Y_1 \ldots \ldots Y_p} \left(Y_t, Y_{t-1} \ldots . Y_1 \middle| Y_0, Y_{-1} \ldots \ldots Y_{-p+i,}; \theta \right)$$

$$x \, f_{Y_t / Y_{t-1} \ldots . Y-p+1} \left(Y_t, Y_{t-1} \ldots . Y_1 \middle| Y_0, Y_{-1} \ldots \ldots Y_{-p+i,}; \theta \right)$$

Applying this formula recessively, the likelihood for the full sample $Y_T . Y_{T-1} \ldots . Y_{-p+1}$ conditioned in $y_0, y_{-1} \ldots y_{-p+1}$ is the product of the individual conditional densities.

$$f_{Y_t, Y_{t-1} \ldots . Y_1 \mid Y_0, Y_1 \ldots \ldots Y_p} \left(Y_T, Y_{T-1} \ldots . Y_1 \middle| Y_0, Y_{-1} \ldots \ldots Y_{-p+i,}; \theta \right) \qquad (3.5.9)$$

$$= \pi \, f_{Y_t, Y_{t-1} \ldots . Y_1 \mid Y_0, Y_1 \ldots \ldots Y_p} \left(Y_T, Y_{T-1} \ldots . Y_1 \middle| Y_0, Y_{-1} \ldots \ldots Y_{-p+i,}; \theta \right)$$

The sample log likelihood is obtained by substituting (3.5.8) in (3.5.9) and taking logs

$$L(\theta) = \sum_{t=1}^{T} \log f_{y_t \, y_{t-1} \, y_{t-2} \ldots . y_{-p+1}} \left(y_t / y_{t-1} \; y_{t-2} \ldots . y_{-p+1}; \theta \right)$$

$$= -\left(T_n \middle/ 2 \right) \log (2\pi) + (T/2) \log \left| \Omega^{-1} \right| \qquad (3.5.10)$$

$$-\left(1 \middle/ 2 \right) \sum_{t=1}^{T} [\, (y_t - \pi^1 x_t)^1 \, \Omega^{-1} \, (y_t - \pi x_t)]$$

3.5.1. Maximum Likelihood Estimate of π

Let us consider the MLE of π which contains the constant term c and autoregressive coefficients ϕ_i.

$$\hat{\pi}^1_{[nx(np+1)]} = \left[\sum_{t=1}^{T} Y_t X_t^1\right]\left[\sum_{t=1}^{T} X_t X_t^1\right]^{-1} \tag{3.5.11}$$

The j^{th} row of $\hat{\pi}^1$ is

$$\hat{\pi}^1_{i_{[1x(np+1)]}} = \left[\sum_{t=1}^{T} Y_{jt} X_t^1\right]\left[\sum_{t=1}^{T} X_t X_t^1\right]^{-1} \tag{3.5.12}$$

Which is the estimated coefficient vector from an OLS regression of Y_{jt} on x_t. Thus, maximum likelihood estimates of the coefficients for the j^{th} equation of a VAR are found by an OLS regression of Y_{jt} on a constant term and p lag of all of the variables in the system

To verify (3.5.11), write the sum appearing in the last term in (3.5.10)

$$\sum_{t=1}^{T}[\,(Y_t - \pi^1 X_t)^1\, \Omega^{-1}\, (Y_t - \pi^1 X_t)]$$

$$= \sum_{t=1}^{T}[\,(Y_t - \hat{\pi}^1 X_t^{\,1} + \hat{\pi}^1 X_t - \pi^1 X_t)\, \Omega^{-1}\, (Y_t - \hat{\pi}^1 X_t^{\,1} + \hat{\pi}^1 X_t - \pi^1 X_t)] \tag{3.5.13}$$

$$= \sum_{t=1}^{T}[\,\hat{\in}_t - (\hat{\pi} - \pi)^1\, X_t\,]\, \Omega^{-1}\, (\hat{\in}_t + (\hat{\pi} - \pi)^1 X_t\,]$$

Where the j^{th} element of the (nx1) vector $\hat{\in}_t$ is the sample residual for observation t from an OLS regression Y_{tj} on X_t

$$\hat{\in}_t = Y_t - \hat{\pi}^1 X_t \tag{3.5.14}$$

Expression (3.5.13) can be expanded as

$$\sum_{t=1}^{T}[\,(Y_t - \pi^1 X_t)^1\, \Omega^{-1}\, (Y_t - \pi^1 X_t)]$$

$$= \sum_{t=1}^{T}[\,(\hat{\in}_t^1\, \Omega^{-1}\, \hat{\in}_t + 2\sum_{t=1}^{T} \hat{\in}_t^1 \Omega^{-1}\, (\hat{\pi} - \pi)^1 X_t + \sum X_t^1 (\hat{\pi} - \pi)\Omega^{-1}\; (\hat{\pi} - \pi)^1 X_t \tag{3.5.15}$$

Consider the middle term in (3.5.15). Since this is a scalar, it is unchanged by applying the 'trace' operator

$$\sum_{t=1}^{T} \hat{\epsilon}_t^1 \Omega^{-1} (\hat{\pi} - \pi)^1 X_t = \text{trace}\left[\sum_{t=1}^{T} \hat{\epsilon}_t^1 \Omega^{-1} (\pi^1 - \pi)^1 X_t\right]$$

$$= \text{trace}\left[\sum_{t=1}^{T} \Omega^{-1} (\hat{\pi} - \pi)^1 X_t \hat{\epsilon}_t^1\right] \qquad (3.5.16)$$

$$= \text{trace}\left[\Omega^{-1} (\hat{\pi} - \pi)^1 \sum_{t=1}^{T} x_t \hat{\epsilon}_t^1\right]$$

But the sample residuals from an OLS regression are by construction orthogonal to the explanatory variables.

$$\therefore \sum_{t=1}^{T} X_j \hat{\epsilon}_{jt}^1 = 0 \text{ for } j \text{ and so } \sum_{t=1}^{T} X_t \hat{\epsilon}_t^1 = 0 \ .$$

Hence (3.5.16) is identically zero and (3.5.15) simplifies to

$$\sum_{t=1}^{T} [(Y_t - \pi^1 X_t) \ \Omega^{-1} (Y_t - \pi^1 X_t)]$$

$$= \sum_{t=1}^{T} (\hat{\epsilon}_t^1 \Omega^{-1} \hat{\epsilon}_t + \sum_{t=1}^{T} X_t^1 (\pi^1 - \pi) \ \Omega^{-1} (\hat{\pi} - \pi)^1 X_t \qquad (3.5.17)$$

Since Ω is a positive definite matrix, Ω^{-1} is also positive definite

Thus, we define the (nx1) vector x_t^* as

$$X_t^* = (\hat{\pi} - \pi)^1 X_t$$

The last tem in (3.517) takes the form

$$\sum_{t=1}^{T} X_t^1 (\hat{\pi} - \pi) \ \Omega^{-1} (\hat{\pi} - \pi)^1 X_t = \sum_{t=1}^{T} \left|X_t^*\right|^1 \Omega^{-1} X_t^* \qquad (3.5.18)$$

This is positive for any sequence $\left[X_t^*\right]$ other than $X_t^* = 0$ for all t.

The smallest value that (3.4.17) can take achieved when x_t^*=0 or when $\pi = \hat{\pi}$. Since (3.5.17) is maximized by setting $\pi = \hat{\pi}$, it follows that (3.4.10) is maximized by setting $\pi = \hat{\pi}$. Hence OLS regressions provide the maximum likelihood estimates of the coefficients of a vector autoregression.

3.6. BAYESIAN VAR MODEL BVAR (p)

3.6.1. Introduction

Bayesian Vectors Autoregression (VAR) is an econometric model. It is the Bayesian version of simple vector autoregression (VAR) model. In Bayesian statistical inference, a prior probability distribution of an uncertain quantity p is the probability distribution that would express one's uncertainty about p before the data is taken into account. It is meant to attribute uncertainty rather than randomness to the uncertain quantity. The unknown quantity may be a parameter. In Bayesian VAR, the coefficients are assumed to have a prior distribution. This implies that after applying the data the coefficient will get prior distribution. This model provides a convenient frame work for incorporating prior information with as much weight as the analyst feels it merits.

Consider the VAR (p) model

$$Y_t = \delta + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \in_t \qquad (3.6.1)$$

or $$Y = (X \otimes I_k)\beta + \in \qquad (3.6.2)$$

When the parameter vector β has a prior multivariate normal distribution with known mean β^* and covariance matrix v_β, the prior density can be written as

$$f(\beta) = \left(\frac{1}{2\pi}\right)^{k^2 p/2} \left|v_\beta\right|^{-1/2} \exp\left[-\frac{1}{2}(\beta - \beta^*)\, v_\beta^{-1} (\beta - \beta^*)\right] \qquad (3.6.3)$$

The likelihood function for the Gaussian process becomes

$$l(\beta / Y) = \left(\frac{1}{2\pi}\right)^{KT/2} \left|I_T \otimes \Sigma\right|^{-1/2} \times \exp\left[\frac{-1}{2}(Y - (X \otimes I_k)\beta)^1 (I_T \otimes \Sigma^{-1}) (Y - (X \otimes I_k)\beta\right] \quad (3.6.4)$$

The posterior density is

$$f(\beta / Y)\ \alpha\ \exp\left[\frac{-1}{2}\left(\beta - \bar{\beta}\right)^1 \bar{\Sigma}_\beta^{-1} \left(\beta - \bar{\beta}\right)\right]$$

Where the Posterior mean is

$$\bar{\beta} = \left[V_\beta^{-1} + (X^1 X \otimes \Sigma^{-1})\right]^{-1} \left[V_\beta^{-1}\beta^* + (X^1 \otimes \Sigma^{-1})Y\right]$$

The posterior covariance matrix is

$$\bar{\Sigma}_\beta = \left[V_\beta^{-1} + (X^1 X \otimes \Sigma^{-1})\right]^{-1}$$

In practice, the prior mean β^* and the prior variance V_β need to be specified. If all parameters are considered to Shrink toward zero, the null prior mean should be specified

The prior variance can be given by

$$\upsilon_{ij}(l) = \begin{cases} (\lambda / l)^2 & \text{if} \quad i = j \\ \left(\lambda\theta\sigma_{ii} / l\sigma_{jj}\right)^2 & \text{if} \quad i \neq j \end{cases} \quad (3.6.5)$$

Here $\upsilon_{ij}(l)$ is the prior variance of the $(e_{ji})^{th}$ element of ϕ_l, λ is the prior standard deviation of diagonal elements of ϕ_l, θ is a constant in the interval (0,1), and σ_{ii}^2 is the i^{th} diagonal element of Σ. The deterministic terms have diffused prior variance. In practice, we replace σ_{ii}^2 by the diagonal element of the ML estimator of Σ in the non-constrained model.

For a bivariate BAR (2) model,

$$y_{1t} = 0 + \phi_{1,11}\, y_{1,t-1} + \phi_{1,12}\, y_{2,t-1} + \phi_{2,11}\, y_{1,t-2} + \phi_{2,12}\, y_{2,t-2} + \in_{1t} \quad (3.6.6)$$

$$y_{2t} = 0 + \phi_{1,21}\, y_{1,t-1} + \phi_{1,22}\, y_{2,t-1} + \phi_{2,21}\, y_{1,t-2} + \phi_{2,22}\, y_{2,t-2} + \in_{1t} \quad (3.6.7)$$

With prior covariance matrix

$$V_{\beta}^{1} = \text{Diag}\left(\infty_{,} \lambda_{,}^{2} (\lambda\theta\sigma_1 / \sigma_2)^2, \lambda_1^2, (\lambda\theta\sigma_1 / 2\sigma_2)^2, \infty_1 (\lambda\theta\sigma_2 / \sigma_1)^2, \lambda_1^2 (\lambda\theta\sigma_2 / 2\sigma_1)^2 (\lambda / 2)^2\right) \quad (3.6.8)$$

3.6.2. Forecasting of Bayesian VAR Modelling

The bootstrap procedure is used to estimate standard error of the forecast. Simulation are performed and in each simulation the following steps are taken.

1. The procedure generates the available number of observations, T, and uniform random integers I_t, where t=1,…. T.
2. A new observation, $\tilde{y}_t$, is obtained as a sum of the forecast based on the estimates coefficients plus the vector of residuals from the I_t;

$$\tilde{y}_t = \sum_{j=1}^{p} \hat{\phi}_j Y_{t-j} + \hat{\in}_{I_t} \quad (3.6.9)$$

A new Basian VAR model is estimated by using the most recent observations, and a prediction value is made of the most recent observations.

$$\text{MSE (l)} = \frac{1}{B} \sum_{i=1}^{B} \left(\tilde{Y}_{t+l/t}^{i} - \bar{y}_t\right)^2 \quad (3.610)$$

Here $\bar{\tilde{Y}}_t = \frac{1}{B} \sum_{i=1}^{R} \bar{\tilde{Y}}_t^{i}$ and B simulations and performed.

3.7. VECTOR ERROR CORRECTION MODEL (VECM)

3.7.1. Introduction

When variables contain stochastic trends, they must be differenced to become stationary. When it is possible to estimate a VAR in levels when the variables follow stochastic trend, it is preferable to estimate the VAR in first differences, such as

$$\Delta Y_t^1 = \alpha_0 + \alpha_1 \Delta Y_{t-1}^2 + \alpha_2 Y_{t-1}^1 + \in_t^1 \quad (3.7.1)$$

$$\Delta Y_t^2 = \beta_0 + \beta_2 \Delta Y_{t-1}^1 + \Delta Y_{t-1}^2 + \in_t^2 \qquad (3.7.2)$$

By having the initial values of y^1 and y^2, we compute the levels by successively adding the changes to the initial values. Such a system contains interactions in the short run between the variables .The changes in y depends on the changes in x in the previous period and viceversa. This system implies that there is no long-run relation between $y^1$and y^2. The reason is that the two variables will be subject to different permanent effects of the shocks. Even though both shocks will affect both variables by virtue of that they affect each other, the permant effects need not be the same. If one variable changes permanently by 5 percent in response to a given shock and the other variable by 2 percent, they permanently a part by 3 percent over time the permanent effects of additional shocks will accumulated and the gap between the two variables will tend to increase.

To insure that the two variables move together also in the long ran, the equations must be modified to include error-correction terms to make sure that the two variables are cointegrated. If there is only one cointegration relation, we would have.

$$\Delta Y_t^1 = \alpha + \alpha_1 \Delta Y_{t-1}^2 + \alpha_2 \Delta Y_{t-1}^1 - \alpha_3 (Y_{t-1}^1 - \lambda Y_{t-1}^2) + \in_t^1 \qquad (3.7.3)$$

$$\Delta Y_t^2 = \beta_0 + \beta_1 \Delta Y_{t-1}^1 + \beta_2 \Delta Y_{t-1}^2 + \beta_3 (Y_{t-1}^1 \lambda Y_{t-1}^2) + \in_t^2 \qquad (3.7.4)$$

This is a Vector Errors Correction Model (VECM). The speed of adjustment depends on the strength of the two speed-of-adjustment coefficients $(\alpha_3$ and $\beta_3)$.

3.8. STATE SPACE MODELS

The origin of statistical state space models can be traced to dynamical systems in engineering branches-communications, robotics, automatic control and aerospace systems such as space craft attitude control.

If U(t), Y(t), X(t) represent input, output and state vectors respectively, general state-space equations are non-linear equations

$$Y_{(t)} = G\left(X(t), U(t), t\right) \qquad (3.8.1)$$

$$\frac{dX_{(t)}}{dt} = F\left(X(t), V(t), t\right)$$

The corresponding special discrete-time linear case is

$$Y_{(t)} = A(t)\, X(t) + B(t)\, U(t) \qquad (3.8.2)$$

$$X_{(t+1)} = C(t)\, X(t) + D(t)\, U(t)$$

Here the state variables refer to memory variables.

The statistical adoption of (3.8.1) and (3.8.2) are widely used discrete time regression like models made of two inter connected equations the observation equation and the system equation. These equations may assume various linear and non-linear forms and referred as state space models.

3.8.1. Aspects of State Space Model

Early work on linear state space models by R.E. Kalman and others appeared in the data 1950s. The models owes their widespread use and popularity to NASA's Apollo space program, designed to achieve preeminence in space for the United States including landing human on the moon and bringing them safely back to earth. In March 1960, Kalman published a seminar paper in which he developed the "Kalman Filter" that gives the recursion formulas for filtering and prediction using the linear state space model in discrete time, thus extending the Wiener-Kolmogorov theory of filtering and prediction for stationary time series set forth in the 1940s. In the full of that year Kalman presented his paper to scientists and engineers at the Ames Research Centre (ARC) of NASA. The audience, due to notation and conceptual problem, had greater difficult at first understanding Kalman's work. But past that stage the value of the state space approach to non-linear navigation (State estimation) became apparent and a simulation study for validation of the method

took place. By early 1961 it was established that on-board optical measurements combined with the equation of motion could yield adequate estimates for navigation and guidance problems, the breakthrough that the NASA were hoping for. Subsequently an early Kalman filter application was made Circa (1961) during feasibility studies for the Apollo space program at the instrumentation labouratory of MIT. Since then the Kalman filter has been widely used in navigation and guidance systems and in many other control systems.

The application of a Kalman filter is a simple matter, the problem is formulated in terms of equations (3.8.2) and then the standard Kalman filter algorithm can be applied in a straight forward manners. Casting the problem in the right form, especially when the models are non-linear is not an easy task. Most of the systems are not fully observable and thus there are various difficulties in the successful application of the algorithm. The continuous time analog of the linear state space model and the Kalman filter have been studied in 1961 in a paper by Kalman and Bucy. In that paper the authors combined and streamlined their ideas developed independently in previous works in the late 1950s. A similar line of work during roughly the same period was also pursued in the former USSR by Russian Physicist R.I. Stratonovich, who studied a recursive algorithm for non-linear, least squares estimates of non-linear dynamical systems driven by white noise.

State space models started to permeate the statistical literature in 1960s and 1970s. Through the work of individually interested in forecasting and in particular Bayesian forecasting of nonstationary processes-where the assumption of constant coefficients is quite generous. Another reason for the interest in state space models by statisticians is the fact that general state space modelling based on recursive relations of probability densities and their integrals are useful for non-Gaussian time series with abrupt discontinuities and/or outliers.

3.8.2. Linear Gaussian State Space Models

Let Y_1, Y_2, ….. be a sequence of observations or responses and X_1, X_2, …. The

corresponding covariance sequence. Let F_t represent the available information to the observer at time t.

It is convenient to adopt the convention that

$$F_0 = 0, F_t = \{Y_t, ..., Y_{t-1}, Y_t\} = \{F_{t-1}, Y_t\}$$

The dependence on the covariates $\{x_1, ..., x_t\}$ is kept in the background in the sense that the results are interpreted as conditional on the covariates.

We have,

Observation equation

$$Y_t = z_t^1 \beta_t + \upsilon_t, \qquad \upsilon_t \sim N(0, V_t) \tag{3.8.3}$$

System equation

$$\beta_t = F_t \beta_{t-1} + w_t \qquad w_t \sim N_p(0, w_t) \tag{3.8.4}$$

Initial formation

$$\beta_0 \sim N_0, (b_0, w_0) \tag{3.8.5}$$

Here z_t is a design vector of covariates such as past observations, supposed known at time t. We first take F_t, V_t, Wt as known.

We assume that $\{v_t\}$ and $\{w_t\}$ each consist of independent random variables and that β_0, $\{v_t\}$, $\{w_t\}$ are mutually independent. The state β_t, a vector of dimension p is time dependent, random and satisfies the autoregression equations (3.8.4) by means of the time dependent transition matrix F_t.

The joint distributions of the observations and the states space are determined by the distributions of the initial state β_0 and of the error sequences $\{\upsilon_t\}, \{w_t\}$. The

system of equations (3.8.3) (3.8.4) and (3.8.5) is a regression model called linear state space model or dynamic linear model. The state space system (3.8.3)-(3.8.5) is used in three types of estimation problems at time t referred to as prediction or forecasting, filtering, and smoothing. Prediction for t>N, filtering t=N and smoothing for t<N. The estimation of the state β_t or its conditional distribution $f\left(\beta_t / F_N\right)$ is called prediction if t>N, filtering if t=N and smoothing when t<N.

3.8.3. State Space Representation for AR(p)

Let X_t be an autoregressive process of order p, not necessarily stationary

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \in_t \tag{3.8.6}$$

The equation can be expressed in matrix form to give the state equation.

$$\beta_t = \begin{pmatrix} x_{t-p+1} \\ . \\ . \\ . \\ x_{t-1} \\ x_t \end{pmatrix} = \begin{pmatrix} 0 & 1 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & . \\ \phi_p & \phi_{p-1} & . & . & . & \phi_1^1 \end{pmatrix} \begin{pmatrix} x_{t-p} \\ . \\ . \\ . \\ x_{t-2} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \\ \in_t \end{pmatrix} \tag{3.8.7}$$

or $$\beta_t = F\beta_{t-1} + w_t \tag{3.8.8}$$

and $y_t = (0, 0, ..., 0, 1)\,\beta_t$

3.8.4. State Space Representation of ARMA (p,q)

Consider the polynomials in the backward shift operator B,

$$Bx_t \equiv X_{t-1} \tag{3.8.9}$$

$$\phi(B) = 1 - \phi_1 B - ..... - \phi_p B^p \quad (3.8.10)$$

$$\phi(B) = 1 + \theta_1 B + ... + \theta_q B^q \quad (3.8.11)$$

\Where $\phi(B)$ has its roots outside the unit circle, the stationary condition.

A state equation for $\phi(B)X_t = v_t$ using matrices and an observation equation $Y_t = \theta(B)X_t$, is

$$Y_t = \theta(B)X_t$$

$$= \theta(B)\,\phi^{-1}(B)\upsilon_t \quad (3.8.12)$$

and $\theta(B)y_t = \theta(B)\upsilon_t$ is ARMA (p, q).

The role of the state component X_t is implicit in the ARMA representation but explicit in the state space representation.

The observation equation is

$$Y_t = \theta(B)X_t$$

$$= \left(\theta_{r-1}, ..., \theta_0\right)X_t$$

and the state equation is obtained by expressing

$\phi(B)X_t = v_t$ as

$$x_t = \begin{pmatrix} 0 & 1 & 0 & . & . & \theta \\ 0 & 0 & 1 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 1 \\ \phi_\gamma & \phi_{\gamma-1} & \phi_{\gamma-2} & & . & \phi_1^1 \end{pmatrix} x_{t-1} + \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \\ \upsilon_t \end{pmatrix} \qquad (3.8.13)$$

3.8.5. Estimation by Kalman Filtering and Smoothing

Consider the state space system (3.8.3)-(3.8.5) for t=1, ….,N

and Let $\beta_{t/s} = E\left[\beta^t / F_s\right]$ (3.8.14)

$$P_{t/s} = E\left[\left(\beta_t - \beta_{t/s}\right)\left(\beta_t - \beta_{t/s}\right)^1\right] \qquad (3.8.15)$$

be the conditional mean and its precision matrix. The covariance matrix between $\beta_t - \beta_{t/s}$ and y_1, …..$y_s$is zero for all t and s. Therefore by the normal assumption $\beta_t - \beta_{t/s}$ is also independent of y_1, $y_{2,\ .....,}$ y_s for all t and s which implies that $P_{t/s}$ is also the conditional covariance matrix of $\beta_{t/s}$.

Let $\beta_{0/0} = b_0, P_{0/0} = w_0$ and assume the initial condition $\beta_0 / F_0 \square N_p\left(\beta_{0/0}, P_{0/0}\right)$.

Then,

Kalman Prediction

$$\beta_{t/t-1} = F_t P_{t-1/t-1} \qquad (3.8.16)$$

$$P_{t/t-1} = F_t P_{t-1/t-1} F_t^1 + w_t$$

Kalman Filtering

$$\beta_{t/t} = \beta_{t/t-1} + k_t \left(y_t - z_t^1 \beta_{t/t-1} \right)$$

$$P_{t/t} = \left[I - k_t z_t^1 \right] p_{t/t-1} \qquad (3.8.17)$$

Where Kalman Gain k_t is given by

$$K_t = P_{t/t-1} z_t \left[z_t^1 P_{t/t-1} z_t + \upsilon_t \right]^{-1} \qquad (3.8.18)$$

Proof:

The prediction equations (3.8.18) follow from (3.7.4)

$$\beta_{t/t-1} = E\left[\beta_t / F_{t-1}\right] = E\left[F_t \beta_{t-1} + w_t / F_{t-1}\right] = F_t \beta_{t-1/t-1}$$

and $$P_{t/t-1} = E\left[\left(\beta_t - P_{t/t-1}\right)\left(\beta_t - \beta_{t/t-1}\right)^1\right]$$

$$= E\left[\left\{F_t \left(\beta_{t-1} - \beta_{t-1/t-1}\right) + w_t\right\}\left\{F_t \left(\beta_{t-1} - \beta_{t-1/t-1}\right) + w_t\right\}^1\right]$$

$$= F_t P_{t-1/t-1} F_t^1 + w_t$$

To obtain (3.8.17) consider, the initial condition $\beta_0 / F_0 \sim N_p (\beta_{0/0}, P_{0/0})$ and write

$\beta_{t-1/F_{t-1}} \sim N_p \left(\beta_{t-1/t-1}, P_{t-1/t-1}\right)$

Then $\beta_{t/F_{t-1}} = F_t \beta_{t-1} + w_t / F_{t-1} \sim N_p \left(\beta_{t/t-1}, P_{t/t-1}\right)$

from which $y_t / F_{t-1} \sim N\left(z_t^1 \beta_{t/t-1}\ z_t^1 P_{t/t-1} z_t + v_t\right)$

and $\mathrm{Cov}\left(\beta_t, y_t / F_{t-1}\right) = P_{t/t-1} z_t$

Hence $$\begin{pmatrix} \beta_t \\ y_t \end{pmatrix} \Big/ F_{t-1} \sim N_{p+1} \left[\begin{pmatrix} \beta_{t/t-1} \\ z_t^1 \beta_{t/t-1} \end{pmatrix}, \begin{pmatrix} P_{t/t-1} & P_{t/t-1} z_t \\ z_t^1 P_{t/t-1} & z_t^1 P_{t/t-1} z_t + v_t \end{pmatrix} \right]$$

From the conditional multivariate normal distribution and after some algebra,

$$\beta_t / y_t , F_{t-1} \sim \beta_{t/t-1} + k_t \left(y_t - z_t^1 \beta_{t/t-1} \right), (I - k_t z_t^1) P_{t/t-1}$$

Where k_t is given in (3.8.17)

In the above proof we used the important fact that if the p-vector $\left(x_1^1, x_2^1 \right)^1$ has a multivariate normal distribution with corresponding means $\left(\mu_1^1, \mu_2^1 \right)^1$ and covariance matrix portioned compatibly, $\Sigma = (\Sigma_{ij}), i, j = 1, 2$.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N_p \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right],$$

x_t has a multivariate normal distribution with mean μ and covariance matrix Σ_{ii} , $i = i = 1, 2$ and the conditional distribution of X_2 given x, is again multivariate normal with mean

$$E\left(x_2 / x_1 \right) = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1) \qquad (3.8.19)$$

and covariance matrix

$$\text{Cov}\left[x_2 / x_1 \right] = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \qquad (3.8.20)$$

3.8.6. Kalman Smoothing

The smoothes for obtaining $\beta_{t-1/N}$ and its covariance matrix $P_{t-1/N}$, for t=N, N-1,….,1, under normality and the initial filtering conditions $\beta_{N/N}$, $p_{N/N}$ is given by the following recursions.

$$\beta_{t-1/N} = \beta_{t-1/t-1} + \beta_t \left(\beta_{t/N} - \beta_{t/t-1} \right)$$

$$P_{t-1/N} = P_{t-1/t-1} + \beta_t \left(P_{t/N} - P_{t/t-1} \right) \beta_t^1 \qquad (3.8.21)$$

$$\beta_t \equiv P_{t-1/t-1}\, F_t^1 P_{t/t-1}^{-1}$$

Smoothing at t=N and going backward in time, the smoothing estimate $\beta_{t-1/N}$ is obtained by adjutancy the filtering estimate $\beta_{t-1/t-1}$ adding to it a weighted difference between a smoothing estimate $\beta_{t/N}$ and a prediction estimate $\beta_{t/t-1}$.

Proof: The proof of the smoothing recursions is more complicated than that of the Kalman filtering recurring of the several possible links of attacks, we follow that of maximum likelihood.

We maximize the conditional Gaussian density

$$f\left(\beta_{t-1}, \beta_t / F_N\right), t \le N \qquad (3.8.22)$$

With respect to β_{t-1}, β_t

The values of β_{t-1}, β_t that maximize (3.8.22) are the respective conditional means $\beta_{t-1/N}, \beta_{t/N}$. From (3.8.3) and (3.8.4)

$$f\left(\beta_{t-1}, \beta_t / F_N\right) \alpha\, f\left(\beta_{t-1}, F_{t-1}, Y_t, ..., Y_N\right) \qquad (3.8.23)$$

$$= f(F_{t-1})\, f(\beta_{t-1}\beta_t / F_{t-1})\, f(Y_t .....Y_N / \beta_{t-1}, \beta_t, F_{t-1})$$

$$= f(F_{t-1})\, f(\beta_{t-1} / F_{t-1})\, f(\beta_t / \beta_{t-1})\, f(Y_t, ...Y_N / \beta_t)$$

Where $f\left(\beta_{t-1} / F_{t-1}\right)$ is the density of $N_p\left(\beta_{t-1/t-1}, p_{t-1/t-1}\right)$ and $f\left(\beta_t / \beta_{t-1}\right)$ is the density of $N_p\left(F_t \beta_{t-1}, w_t\right)$.

Assume that $\beta_{t/N}$ has already been obtained.

Then $\beta_{t-1/N}$ is obtained by minimizing $-2\log f\left(\beta_{t-1},\beta_{t/N}/F_N\right)$ with respect to β_{t-1}.

This is equivalent to minimizing

$$\left(\beta_{t-1}-\beta_{t-1/t-1}\right)^1 P^1_{t-1/t-1}\left(\beta_{t-1}-\beta_{t-1/t-1}\right)+\left(\beta_{t/N}-F_t\,\beta_{t-1}\right)^1 w_t^{-1}\left(\beta_{t/N}-F_t\,\beta_{t-1}\right)$$

By differentiating with respect to $\beta_{t-1}$and equating the derivative zeo, the solution is

$$\beta_{t-1/N}=\left(P^1_{t-1/t-1}+F_t^1 w_t^{-1}F_t^1\right)^{-1}\left(P^1_{t-1/t-1}\,\beta_{t-1/t-1}+F_t^1 w_t^{-1}\beta_{t/N}\right)$$

This can be simplified using the matrix relations.

$$\left(P^{-1}+F^1w^{-1}F\right)^{-1}=P-PF^1\,(FPF^1+w)^{-1}FP$$

$$\left(P^{-1}+F^1w^{-1}F\right)^{-1}F^1w^{-1}=PF^1(FPF^1+w)^{-1}$$

Where, P, F, wstand for $P_{t-1/t-1}$, F_t , w_t respectively.

This and the smoothing expression (3.8.16) give the desired smoother.

$$\beta_{t-1/N}=\beta_{t-1/t-1}+P_{t-1/t-1}\,F_t^1\left(F_t\,p_{t-1/t-1}F_t^1+w_t\right)^{-1}\left(\beta_{t/N}-F_t\,\beta_{t-1/t-1}\right)$$

$$=\beta_{t-1/t-1}+P_{t-1/t-1}\,F_t^1P_{t/t-1}^{-1}\left(\beta_{t/N}-\beta_{t/t-1}\right)$$

$$=\beta_{t-1/t-1}+\beta_t\,\left(\beta_{t/N}-\beta_{t/t-1}\right)$$

To obtain $P_{t-1/N}=E\left[\left(\beta_{t-1}-\beta_{t-1/N}\right)\left(\beta_{t-1}-\beta_{t-1/N}\right)\right]^1$,

note that $\beta_{t-1}-\beta_{t-1/N}=\beta_{t-1}-\beta_{t-1/t-1}-\beta_t\,(\beta_{t/N}-\beta_{t/t-1})$

or by rearranging terms

$$\left(\beta_{t-1}-\beta_{t-1/N}\right)+\beta_t\beta_{t/N}=\left(\beta_{t-1}-\beta_{t-1/t-1}\right)+\beta_t F_t\,\beta_{t-1/t-1} \tag{3.8.24}$$

and the following crose terms vanish

$$E\left(\beta_{t-1}-\beta_{t-1/N}\right)\left(\beta+\beta_{t/N}\right)^1=E\left(\beta_{t-1}-\beta_{t-1/t-1}\right)\left(\beta_t F_t\,\beta_{t-1/t-1}\right)^1=0$$

and

$$E\left(\beta_{t/s}\beta_{t/s}^1\right)=E\left(\beta_t\beta_t^1\right)-P_{t/s}=F_t\,E\left(\beta_{t-1}\beta_{t-1}^1\right)F_t^1+w_t-P_{t/s}$$

Thus from (3.8.24)

$$P_{t-1/N}=P_{t-1/t-1}+\beta_t\left(P_{t/N}-F_t\,p_{t-1/t-1}\,F_t^1-w_t\right)\beta_t^1$$

which together with (3.8.16) completes the proof.

3.8.7. Estimation in the Linear Gaussian Model

Estimation of parameters in the linear Gaussian system can be done the method of maximum likelihood. Here we take a certain parameterization assumption which often occurs in practice.

We assume that the parameters b_0, w_0, F_t, V_t, W_t depend completes or in part on a vector θ of hyperperameters which do not depend on 't'.

Here, we write

$$b_0=b_0(\theta),\ w_0=w_0(\theta),\ F_t=F_t(\theta),\ V_t=V_t(\theta),\ w_t=w_t(\theta),$$

and base the inference on the joint distribution of the observations $Y_1,....,Y_N$ or equivalently the likelihood of θ. We assume that F_t, V_t, W_t in the system (3.8.4)-

(3.8.5) do not depend on t and estimate $\theta = (b_0, w_0, F, V, W)$ by maximum likelihood. In either case the likelihood is obtained from the joint distribution of the one-step prediction error or innovations.

$$\in_t = y_t - E\left[y_t / F_{t-1} \right] = y_t - z_t^1 \beta_{t/t-1}$$

$$= y_t - z_t^1 F_t \ \beta_{t-1/t-1} \ t = 1, ..., N$$

Which are independent normal random variables with mean zero and variance

$$\sigma_t^2(\theta) = z_t^1 P_{t/t-1} \ z_t + \upsilon_t$$

$$= z_t^1 \left\{ F_t \ P_{t-1/t-1} F_t^1 + w_t \right\} z_t + \upsilon_t \qquad (3.8.25)$$

The Gaussian assumption implies that $t_1, ..., t_N$ is a one-to-one linear transformation of $Y_1, ..., Y_N$ so that upto a constant the log-likelihood of base θ on $Y_1, ..., Y_N$ is given by

$$\log L_y(\theta) = -\frac{1}{2} \sum_{t=1}^{N} \log \sigma_t^2(\theta) - \frac{1}{2} \sum_{t=1}^{N} \in_t^2 / \sigma_t^2(\theta) \qquad (3.8.26)$$

Maximising the likelihood with respect to θ is the direct method.

The indirect method is based on the joint distribution of both the observed time series and the unobserved states and uses the EM algorithm to maximize the resulting likelihood.

From (3.8.4)-(3.8.5)

$$f\left(y_{t/\beta_t}, F_{t-1}, \theta \right) = f\left(y_{t/\beta_t}; \theta \right)$$

$$= f_\theta \ (y_t - z_t^1 \beta_t)$$

and $f\left(\beta_t/\beta_{t-1}, \beta_{t-2}, ..., \beta_0; \theta\right) = f\left(\beta_t/\beta_{t-1}; \theta\right)$

$= f_w\left(\beta_t - F_t\beta_{t-1}\right)$

Where $\upsilon_t \sim f_\upsilon$ and $w_t \sim f_w$

The likelihood is

$$L_{y,\beta}(\theta) = f\left(\beta_0, \beta_1, ..., \beta_N, y_1, ..., y_N, \theta\right)$$

$$= f\left(\beta_0\right)\prod_{t=1}^{N} f_w\left(\beta_t - F_t\beta_{t-1}\right) f_\upsilon\left(y_t - z_t^1\beta_t\right) \quad (3.8.27)$$

3.8.8. Non-linear and Non-Gaussian State Space Model

Predictions, filtering and smoothing can be done more generally by using the laws of conditional probability, Bayes theorem and relating linearity and the normal assumption. In the general approach the dynamics is captured directly through the conditional densities of the observations and states, without formation of any particulars system of equations. Assume that $\{Y_t\}$, t=1,….., N denotes the observed process and let the unobserved state process be $\left\{\beta_t\right\}$, $t = 0, ..., N$.

The general state space model is

General Observation equation

$$Y_t/\beta_t \sim f\left(y_t/\beta_t\right) \quad (3.8.28)$$

General System equation

$$\beta_t/\beta_{t-1} \sim f\left(\beta_t/\beta_{t-1}\right) \quad (3.8.29)$$

Initial Information

$$\beta_0 \sim f\left(\beta_r/F_0\right) \square f\left(\beta_0\right) \quad (3.8.30)$$

Here the understanding is that the responses are independent and that the sequence of unobserved states from a Markov process.

The equation (3.8.28) means that given the state sequence $\{\beta_t\}$, the observed process $\{Y_t\}$ forms an independent sequence of random variables and (3.8.29) and (3.8.30) mean that the sequence of unobserved states $\{\beta_t\}, t = 0,....,N$ is a Markov process with initial distribution $f(\beta_0)$. When the unobserved states assume discrete values then the definition of the general state space model is equivalent to that of a hidden Markov Model. The densities in (3.8.28) and (3.8.29) may depend as unknown parameters referred to as hyper parameters.

When the corresponding conditional densities are Gaussian, then the linear normal state space model (3.8.3), (3.8.4) and (3.8.5) is a special case of the system represented by (3.7.28), (3.7.29) and (3.7.30). It can be verified easily. Another special case is provided by the non-linear and non-Gaussian state space model.

$$Y_t = h_t(\beta_t, v_t) \qquad (3.8.31)$$

$$\beta_t = f_t(\beta_{t-1}, w_t)$$

Where h_t and f_t are known and suitably defined functions and υ_t, w_t are random sequences, t=1 ….N.

An example of (3.8.31) is

$$Y_t = h_t(\beta_t) + v_t \qquad (3.8.32)$$

$$\beta_t = h_t(\beta_{t-1}) + w_t$$

Here, the main problem is estimation of current, future and past states and their distributions given the data that is, filtering prediction and smoothing respectively.

CHAPTER - IV TESTS OF TIME SERIES REGRESSION MODELS

4.1. TESTS FOR STATIONARITY

4.1.1. Introduction

The stationarity or non-stationarity of a series can strongly influence its behaviour and property. For example, persistence of shocks will be infinite for non-stationary series. Some times, if two variables are trending over time, a regression of one on the other could have a high R^2 even if the two are totally unrelated. This is spurious regression. If the variables in the regression model are not stationary, then it can be proved that the standard assumptions for asymptotic analysis will not be valid. Hence, we need to test for stationarity.

4.1.2. Unit Root Test

The early and pioneering work on testing for a unit root in time series was done by Dickey and Fuller. The basic objective of the test is to test the null hypothesis that $\phi = 1 \text{ in}:$

$$y_t = \phi y_{t-1} + u_t \qquad (4.1.1)$$

Against the one-sided alternative $\phi < 1$

Hence we have,

H_0: Series contains a unit root

as H_1: Series is Stationary

We usually use the regression

$$\Delta y_t = \Psi y_{t-1} + u_t \qquad (4.1.2)$$

So that a test of $\phi = 1$ is equivalent to a test of $\Psi = 0 \,(\text{since } \phi - 1 = \Psi)$

4.1.3. Dickey Fuller Test

Dickey Fuller tests are also known as τ Test: τ, τ_μ, τ_t.

The null and alternative models in each case are

i) $H_0 : y_t = y_{t-1} + u_t$

$$H_1 : y_t = \phi y_{t-1} + u_t , \ \phi < 1 \qquad (4.1.3)$$

This is a test for random walk against a stationary autoregressive process of order 1.

ii) $H_0 : y_t = y_{t-1} + u_t$

$$H_1 : y_t = \phi y_{t-1} + \mu + \mu_t, t < 1 \qquad (4.1.4)$$

This is a test for random walk against a stationary AR(1) with drift.

iii) $H_0 : y_t = y_{t-1} + u_t$

$$H_1 : y_t = \phi y_{t-1} + \mu + \lambda t + u_t, \phi < 1 \qquad (4.1.5)$$

This is a test for random walk against a stationary AR(1) with drift and a time trend.

Computing Dickey and Fuller Test Statistic.

We write $\Delta y_t = u_t$

Where $\Delta y_t = y_t - y_{t-1},$ and alternative is

$$\Delta y_t = \Psi y_{t-1} + \mu + \lambda t + \mu_t$$

With $\mu = \lambda = 0$ in case (i)

$\lambda = 0$ in case (ii) and

$\Psi = \phi - 1$ in case (iii)

In each case the tests are based on the t-ratio on y_{t-1} term in the estimated regression of Δy_t on y_{t-1}, a constant in case (ii) and a constant and trend in case (iii).

The test statistics are defined as

$$\text{Test Statistic} = \frac{\hat{\Psi}}{S.E\left(\hat{\Psi}\right)} \qquad (4.1.6)$$

The test statistic does not follow the usual t-distribution under the null, since the null is one nonstationary, but follows a non-standard distribution.

Critical values for the Dickey Fuller Test

Significance Level	1%	5%	10%
C.V. for constant but not trend	-3.43	-2.86	-2.57
C.V. for constant and trend	-3.96	-3.41	-3.12

Critical values for DF and ADF Tests (Fuller, 1976, p.37).

The null hypothesis of a unit root is rejected in favours of the stationary alternative in each case if the test statistic is more negative than critical value.

4.1.4. Augmented Dickey Fuller Test

Dickey Fuller Test and unit root tests are valid only if u_t is white noise. In particular, u_t will be autocorrelated if there was autocorrelation in the dependent variable of the regression $\left(\Delta y_t\right)$. The solution is to augment the test using p lags of the dependent variable. The alternative model is

$$\Delta y_t = \Psi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-1} + u_t \qquad (4.1.7)$$

The same critical values from the DF tables as used before. A problem now arises in determining the optimal number of lags of the dependent variable.

There are two ways, (i) use the frequency of the data to decide, (ii) use information criteria.

Testing for Higher orders of Integration

Let us consider the simple regression

$$\Delta y_t = \Psi y_{t-1} + u_t$$

We test $H_0 : \Psi = 0 \quad H_1 : \Psi < 0$

If H_0 is rejected we conclude that y_t does not contain unit root. If H_0 is not rejected, the series contains a unit root. But if $y_t \sim I(2)$, then what it. Hence we still not have rejected. We now need to test

$$H_0 : y_t \sim I(2), \text{ as } H_1 : y_t \sim I(t)$$

We will continue to test for a further unit root until.

We reject H_0.

We now regress $\Delta^2 y_t$ on Δy_{t-1} and now we test

$H_0 : \Delta y_t \sim I(1)$ which is equivalent to $H_0 : y_t \sim I(2)$.

So in this case, if we do not reject, we conclude that y_t is at least I(2).

Main Criticism of Dickey-Fuller Test is that the process of the test is low if the process is stationary but with a root close to non-stationary boundary. If the true data generated process is $y_t = 0.95 y_{t-1} + u_t$, then the null hypothesis of a unit root should be rejected.

4.1.5. Graphical Inspection (Correlogram)

Correlogram is a graphic of autocorrelation values from several time intervals in time series data. It is also known as Auto Correlation Function (ACF). ACF represents comparison between covariant on lag k and its variant.

$$\rho_k = \frac{\sum_{t=k+1}^{T} \left(y_t - \bar{y}\right)\left(y_{t-k} - \bar{y}\right)}{\sum_{t=1}^{T} \left(y_t - \bar{y}\right)^2} \qquad (4.1.8)$$

Here ρ_k is ACF coefficient in lag k

T is the number of observations

y_t observation in t period

$\bar{y}$ is the mean

y_{t-k} is the observation in t-k period.

ACF (ρ_k) has valve started from -1 to +1. If ACF value on every lag is θ, then the data is stationary. Usually, lag length needed to analyze is one third or a quarter of the number of observations of a time series data. Another way to determine whether a time series data is stationary or not, we use the following method.

$$\rho_k = \pm 1.96 \text{ x S.E} \quad \text{or} \quad \rho_k = \pm 1.96 \sqrt{\frac{1}{T}} \qquad (4.1.9)$$

Here SE is the standard errors and the significance level is 95% with sample number equal to T.

If ACF coefficient value is in the interval with significance equal to 95%, then null hypothesis (H_0) that shows ρ_k equal to '0' cannot be rejected. It means that the data is stationary.

4.2. TEST FOR NON-STATIONARY REGRESSORS

4.2.1. Introduction

Cointegration is an econometric technique for testing the relationship between non-stationary time series variables. If two or more series each have a unit root, but liner combination of them is stationary, then the series are said to be co-integrated.

4.2.2. Johansen Cointegration Test

The Johansen test is a procedure for testing of several I(1) time series. This test permits more than one co-integrating relationship. There are two types of Johanset test. They are either with trace or with eigen value and the inference will be different.

The null hypothesis for the trace test is the number of co-integration vectors $r \leq n$, the null hypothesis for the eigen value test is $r = n$.

Like in a unit root test, there can be a constant term, a trend term, both or neither in the model. For a general VAR (p) model.

$$y_t = \mu + \phi D_t + \pi_p y_{t-p} + \ldots\ldots + \pi_1 y_{t-1} + e_t, t = 1, \ldots., T \qquad (4.2.1)$$

There are two possible specifications for error corrections, that is two VECM.

1. The long run VECM:

$$\Delta y_t = \mu + \phi D_t + \pi y_{t-p} + \gamma_{p-1}\Delta y_{t-p+1} + \gamma_1 \Delta x_{t-1} + \ldots.. + \in_t, t = 1, ..., T$$

Where $\gamma_i = \pi_1 + \ldots.. + \pi_i - I, i = 1, ..., p-1$

2. The transitory VECM:

$$\Delta y_t = \mu + \phi D_t - \gamma_{p-1}\Delta y_{t-p+1}.....\gamma_1 \Delta y_{t-1} + \pi y_{t-1} + \varepsilon_t, t = 1, ..., T$$

Where $\gamma_i = \left(\pi_{i+1} + ... + \pi_p\right),\ i = 1, ..., p-1$

In both VECM,

$$\pi = \pi_1 + ... + \pi_p - I$$

Inferences are drawn on π, and they will be the same.

4.3. COINTEGRATION REGRESSION: DURBIN-WATSON TEST

4.3.1. Introduction

This test involves a simple regression of one variable on the other, and the standard Durbin-Watson test on the residuals. Here the null hypothesis is that the residuals from a non-stationary random walk, and the alternative hypothesis is that the residuals form a stationary AR(1) process. This the most powerful test in the face of a trend. Engle and Yoo have developed critical values for the cases when upto five variables are included in the equation. Engle and Glanger point out that most economic data are integrated of order one, I(1), and are not an independent stationary process. This test can be used to provide a first or preliminary judgement as to the present of cointegration.

The null and alternate hypotheses to be tested are:

H_0: DW = 0 (Co-integration does not exist). (4.3.1)

H_1: DW ≠ 0 (Co-integration does exist).

Estimate the co-integrating regression between the series $x(t) = \alpha + \beta(t) + e(t)$.

Now, estimate the Durbin-Watson Statistic for the cointegrating regression compare the Durbin-Watson Statistic to the critical values and conclusion will be taken.

4.3.2. Tests of Model Order Selection

Tests involving forecast error and prediction errors play a major role in identification of time series models. The selection of the number of lags to include in a multivariate time series model depends largely on the use of one or more minimization criteria. Some of the these criteria measure a model's explanatory power over the sample period, some over the forecast period or a combination of both Kalyan (1988) gave a method which focuses on the method of exclusion of

variables. This relaxes the assumption of symmetry within the system and tests explicitly for the number of lags associated with each variable in the system. In the case of a symmetry in the system, Kalyan suggests alternative estimation procedures. All the above approaches are compatible with the Litterman (1986) Bayesian VAR procedures. There is no universal method for determining lag length. Hence, the fractioner must proceed with caution in using these criteria. Hence, for the sake of convention, multivariate extensions of the measure in UTM be considered. The principal difference between the univariate and multivariate cases dependent on the method of defining the error variance measure component of the equations.

The data be x, joint density $f(x;\theta,\Psi)$.

Dimension of θ is p.

Here we test $H_0 : \theta = \theta_0$

Likelihood ratio test:

Maximise log likelihood $l(\theta,\Psi)$ twice. A twice we find unrestricted MLEs $\hat{\theta},\hat{\Psi}$ by maximizing l over all possibilities. Find restricted MLEs $\theta_0, \hat{\Psi}_0$ by maximizing $l(\theta_0,\Psi)$ over Ψ.

Here the likelihood ratio statistic is $\dfrac{f\left(x,\hat{\theta},\hat{\Psi}\right)}{f\left(x,\theta_0,\hat{\Psi}_0\right)}$

Usual test statistic is 2 times log likelihood ratio

$$T = 2\left[l(\hat{\theta},\hat{\Psi}) - l(\theta_0,\hat{\theta}_0)\right] \qquad (4.3.2)$$

Large sample theory

If H_0 is true the $T \sim \Psi^2_p$.

Compare AR (P_0) and AR (P_0+P)

Take $\mu = 0$

The model is

$$X_t = a_1 x_{t-1} + a_2 x_{t-2} + \ldots\ldots + a_{p_0+p} x_{t-p_0-p} + \in_t \qquad (4.3.3)$$

Here take,

$$\Psi = (a_1, \ldots., ap_0, \sigma)$$

$$\theta = \left(ap_{0+1}, \ldots\ldots, ap_{0+p}\right)$$

$$\theta_0 = (0, \ldots\ldots, 0)$$

The likelihood is

$$f_{X_0}, \ldots\ldots, x_{T-1} = f_{X_0}, \ldots\ldots, x_{p_0+p-1} \times f_{Xp_0+p}, \ldots x_{F1} / x_0 \ldots\ldots x_{p_0+p-1}$$

Taking logs, we get

$$l(\theta, \psi) = l_M(\theta, \psi) + l_c(\theta, \psi)$$

Here subscript 'c' indicates conditional and M for marginal.

Maximize only l_c, then

$$T_c = 2\left[l_c\left(\hat{\theta}_c, \hat{\psi}_c\right) - l_c\left(\hat{\theta}_0, \hat{\psi}_{0,c}\right)\right]$$

Large sample theory still valid.

$T_c \sim \chi^2_p$ asymptotically if H_0 true.

Difficulties Asising

Which two models to compare? What p and p_0? Can compare only nested models. Must be careful to condition on same number of data points in fitting both models.

4.3.3. Final Prediction Error

Choose model order to minimize final prediction error:

$$\left(\frac{T+K}{T-K}\right)\frac{\Sigma \in^2_{K,t}}{T}$$

Here K is the number of parameters, subscript K on $\in$ means residuals from that model.

Akaike's Information Criterion

Minimize Akaike's Information Criterion

$$AIC = \log\left(\Sigma \in_{K,t}^2 / T\right) + 2K/T \tag{4.3.4}$$

$$\log\ (FPE) = \log\ \left(\Sigma \in_{K,T}^2 / T\right) + \log(1 + 2K/T) + 0(T^{-2})$$

$$= AIC + 2K/T + 0(T^{-2})$$

Compare many models with different numbers k of parameters by computing.

$$AIC_K = \log\left(\hat{\sigma}_k^2\right) + 2k/T$$

or $$AIC_K = T\log\left(\hat{\sigma}_k^2\right) + 2k$$

Latter is equivalent to -2 times log likelihood + 2k

In time series, we must make sure to use same data points compute

$$\hat{\sigma}_{p+w}^2 = \frac{\Sigma \in_t^2}{\neq \text{data points}} \tag{4.3.5}$$

Problems

1. Plot AIC_p against p for p=0,1,…P. How to select p the largest order tied.
2. The method is not consistent.

$$P_{p_0}\left(\hat{p} < p_0\right) \to 0 \text{ but } P_{p_0}\left(\hat{p} > p_0\right) \to 0$$

The final prediction error was originally proposed for AR (P) order determination and was extended to ARMA (p, q) models. This criterion was established on the basis of minimizing the one-step a head mean square forecast error after incorporating the inflating effects and estimated coefficients. The criterion to be minimized is

$$FPE = \hat{\sigma}^2\ \frac{n+p+q}{n-p-q} \tag{4.3.6}$$

Where $\hat{\sigma}^2$ is estimated variance of white noise

n is number of observation

p is order of the autoregressive component

q is the moving average component.

In 1970, Akaike found that FPE is asymptotically inconsistent and in 1973 he employed information theoretic considerations to develop the Akaike's information criterion, AIC. This was designed to be an asymptotically unbiased estimate of the Kullback-Leibler index of the fitted model relative to the true model. The AIC statistics is defined as

$$\text{AIC= -2 In likelihood } \left(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2\right) + 2(p+q+1) \qquad (4.3.7)$$

Where $\hat{\phi}$ are estimated autoregressive parameters,

θ are estimated moving average parameter, and

$\hat{\sigma}^2, n, p \text{ and } q$ are as defined in (4.3.3).

A criterion like AIC that penalizes the likelihood for the number of parameters in the model attempt to choose the most parsimonous model Shibata (1976) showed empirical evidence that AIC has the tendency to pick models which are over-parameterized. In view of this, Akaike applied a Bayesian modification to AIC and came up with consistent order selection criterion, known as Bayesian information criterion or BIC. If the data $\left\{x_1, \ldots, x_n\right\}$ are observations of an ARMA (p, q) process, then a Bayesian information criterion is

$$BIC = (n-p-q)\, l_n \frac{n\hat{\sigma}^2}{n-p-q} + n\,(1 + l_n 2n) + (p+q)\, l_n \left[\frac{\sum_{i=1}^{n} x_i^2 - n\hat{\sigma}^2}{p+q}\right] \qquad (4.3.8)$$

BIC is more satisfactory than the AIC as an ARMA model selection criterion since the AIC has a tendency to the pick models, which are over-parameterized.

Schwarz (1978) used a Bayesian analysis and Rissanen (1978) applied an optimal data recording scheme to independently arrive at the same criterion. This criterion later known as Schwarz-Risannen Criterion, SIC. The criterion to be minimized is given by

$$SIC = l_n\ \hat{\sigma}^2 + \left(\frac{p+q}{n}\right) l_n n \qquad (4.3.9)$$

Gemeke and Meose (1981) suggested approximating SIC by Bayesian estimation criterion, BEC,

$$BEC = \hat{\sigma}^2 + \left(p_x + q_x\right)\hat{\sigma}_x^2\, l_n \frac{n}{n - p_x - q_x} \qquad (4.3.10)$$

Where x denotes a quantity from pre-assigned high order ARMA model that includes all potential models.

Hannan and Quinn (1979) constructed Hannan-Quinn criterion from the law of iterated logarithm. It provides a penality functions, which decreases as fast as possible for a strongly consistent estimator as sample size increases. Hannan-Quinn criterion is given by

$$HQ = \ln \hat{\sigma}^2 + 2(p+q)\frac{\ln(\ln n)}{n} \qquad (4.3.11)$$

In 1989, Hurnish and Tsai found that BIC, which was modified from AIC, is not asymptotically efficient. Hence they suggested a biased corrected version of AIC, known as Akaikes information corrected criterion of AICC. AICC statistic is given by

$$\text{AICC} = -2 \ln \textit{likelihood}\left(\hat{\phi}, \hat{\theta}, \hat{\sigma}^2\right) + \left[2n\,(p+q+l)/(n-(p+q)-2\right] \qquad (4.3.12.)$$

Where $\hat{\phi}$ are estimated autoregressive parameters

$\hat{\theta}$ are estimated moving average parameters

$\hat{\sigma}^2$ is estimated variance of white noise

n is no. of observations

p is order of the autoregressive component

q is order of the moving and age component

and likelihood $\left(\hat{\phi}, \hat{\theta}, \hat{\sigma}\right)$ is the likelihood of the data under the

Gaussian ARMA model with parameters $\left(\phi, \hat{\theta}, \hat{\sigma}^2\right)$.

The penaly factors 2n (p+q+1)/[n-(p+q)-2] and 2 (p+q+1), for AICC statistic and AIC statistic respectively are asymptotically equivalent as $n \to \infty$. AICC is asymptotically efficient for autoregressive process. The AICC statistic has a more extreme penalty for large order models, which counteract the over fitting nature of the AIC. Today, the AICC statistics, as its earlies version (AIC), has has been widely used as one of the order selection criteria in ARMA time series as well as log-length selection criteria in econometric modeling process.

4.4. TESTS FOR CAUSALITY

4.4.1. Introduction

Concept of causality is closely related to the testing for the exogeneity of a variable. This concept is due to Granger (1969). Causality between two or more variables is one of the most important issues that econometricians address in their research. The term causality explains, that having knowledge of fast values of a variable x(t) does improve the ability of the model to predict another variable y(t). This relationship can be written as $x(t) \to y(t)$. Instantaneous causality can also be specified where, not only past, but also present values of x(t) improve the ability of the model to predict y(t). Similarly the causality from y(t) to x(t) can be explained. Feedback occurs in the case where x(t) causes y(t) and y(t) causes x(t) Pierce and Heaugh (1977) demonstrate instantaneous causality under the conditions that x(t) causes y(t) instantaneously, if and only if y(t) causes x(t) instantaneously.

There are three possible explanations for apparent instantaneous causality:

i) There is true instantaneous causality in the system so that elements in the system react without any measurable time delay to change in some other elements.

ii) There is no true instantaneous causality, but the finite time delay between cause and effect is small, compared to the time interval over which data is collected.

iii) There is a jointly causal variable w(t-1) that causes both x(t) and y(t). Consequently, the researcher may wish to determine if a casual relationship

exists between x(t) and y(t), if there is reverse causality, if there is instantaneous causality and if feedback occurs between the variables.

Various tests for causality have been developed. Most tests include the Granger test for causality and also a test for causality in the presence of "leads". Such tests have been developed by Sims, Geweke-Meese-Dext, Pierce-Haugh and Gewek. For all these tests it is required that the series being tested are mean-zero stationary series.

4.4.2. Granges Causality Test

This test reflects the extent to which the lag process in one variable explaining the current values of another variable. Here, we test the null hypothesis that x(t) does not Granger-Cause y(t).

$$y(t) = \sum_{i=1}^{\infty} \alpha_i y(t-i) + C_1 + v_1(t) \qquad (4.4.1)$$

$$y(t) = \sum_{i=1}^{\infty} \alpha_i y(t-i) + \sum_{j=1}^{\infty} \beta_j x(t-j) + c_2 + v_2(t) \qquad (4.4.2)$$

We test whether $\beta_j = 0$ for all lags j.

An F test would be employed where (1) is the restricted equation and (2) is the unrestricted equation. If the null hypothesis is accepted it indicates that lagged values of x(t) do not significantly explain the variation in y(t), that is, x(t) does not Granger-Cause y(t).

A similar test would be conducted to examine whether $y_{(t)}$ does not Granger-Cause x(t). The relevant equations are:

$$x(t) = \Sigma\alpha_i x(t-i) + c_1 + u_1(t) \qquad (4.4.3)$$

$$x(t) = \Sigma\alpha_i x(t-i) + \Sigma\beta_j y(t-j) + c_2 + u_2(t) \qquad (4.4.4)$$

We test whether $\beta_j = 0$ for all j. Here (4.4.3) is the restricted equation and (4.4.4) is the unrestricted equation in the F-test.

The Ganger test can also be used to test for instantaneous causality between variables, although it should be a more that this concept has lesser usefulness. Here, we wish to examine whether the zero-lag of a given variable is significant in

explaining the variation of a second variable. To test the null hypothesis that x(t) does not Granger-Cause y(t) instantaneously, we use the following:

$$y(t) = \sum_{i=1}^{\infty} \alpha_i \, y(t-i) + \sum_{j=1}^{\infty} \beta_j \, x(t-j) + v_1(t) \qquad (4.4.5)$$

$$y(t) = \sum_{i=1}^{\infty} \alpha_i \, y(t-i) + \sum_{j=0}^{\infty} \beta_j \, x(t-j) + C_2 + v_2(t) \qquad (4.4.6)$$

The only difference between these two equations is the definition of j. For the F test, equation (4.4.5) would be the restricted equation and (4.4.6) is the unrestricted equation. Similarly, to test if y(t) does Granger-Cause x(t) instantaneously, we use the following equations.

$$x(t) = \sum_{i=1}^{\infty} \alpha_i \, x(t-i) + \sum_{j=1}^{\infty} \beta_i \, y(t-j) + C_1 + u_1(t) \qquad (4.4.7)$$

$$x(t) = \sum_{i=1}^{\infty} \alpha_i \, x(t-i) + \sum_{j=0}^{\infty} \beta_j \, y(t-j) + C_2 + u_2(t) \qquad (4.4.8)$$

Here (4.4.7) is the restricted equation and (4.4.8) is the unrestricted equation for the F test. If the null hypothesis of no causality is rejected in both the tests, this indicates that a "feed back effect" exists between x(t) and y(t). This result is obvious because if x(t) instantaneously causes y(t) then y(t) must also instantaneously cause x(t). Hence rejecting the null hypothesis establishes the presence of causality. It should be stated at this point that feedback among one of the equations still implies a weak form of causality. This results from our inability to separate, whether the causality in one direction influences causality in the opposite direction. If one finds that causality does exist, then the sign as well as the magnitude of the effect of an independent on a dependent variable can be estimated by summing the coefficients in the unrestricted equation on all lags of the independent variable, that is, $\Sigma\beta_j$. The sign and significance of any particular lag of an independent variable are determined from the t statistic on the coefficient for lag j in the unrestricted equations.

An issue that assess in causality testing is the choice of i and j in the test equations. In many empirical applications, i and j are set equal to each other, that is, the same number of lags of x(t) and y(t) are used in models.

Test Procedure

The null hypothesis to be tested.

For Causality Test

H_0 : x(t) does not Granger-Cause y(t)

H_1 : x(t) does Granger-Cause y(t)

H_0 : y(t) does not Granger-Cause x(t)

H_1 : y(t) does Granger-Cause x(t)

For Instantaneous Causality Test

H_0 : y(t) does not Granger-Cause x(t) instantaneously

H_1 : y(t) does Granger-Cause x(t) instantaneously

Assure that the data series are stationary. Choose the number of lags to be used in the equations.

Causality Test

For x(t) $\rightarrow$ y(t) : Equations (4.4.1) and (4.4.2)

For y(t) $\rightarrow$ x(t) : Equations (4.4.3) and (4.4.4)

Instantaneous Causality Test

For x(t) $\rightarrow$ y(t) : Equations (4.4.5) and (4.4.6)

For y(t) $\rightarrow$ x(t) : Equations (4.4.7) and (4.4.8)

Calculate the test statistic. The formula remains same for the general causality and the instantaneous causality tests.

$$F = \frac{\left(ESS_R - ESS_{UR}\right)/q}{ESS_{UR}/(n-k)} \qquad (4.4.9)$$

Here ESS_R and ESS_{UR} are the error sum of squares for the restricted and unrestricted equations respectively; q is the number of restrictions applied, n is the total number at observations; and k is the total number of parameters in the unrestricted model (including constant). The statistic is distributed as an F with q degrees of freedom in the numerator and n-k degrees of freedom in the denominator is $F_{q,n-k}$. We accept the

alternative hypothesis when the calculated F statistic is greater than the critical value.

4.5. SIMS TEST

4.5.1. Introduction

Sims (1972) explained that a necessary condition for y(t) not to Granger Cause x(t) is for future x(t) terms to have zero coefficients in a regression of y(t) an future, current, and lagged values of x(t). As an extension of the Granges causality test, the Sims test includes future values of the independent variable in the unrestricted equation. This test is an attempt to measure unidirectional causality. This would imply that although $y(t) \rightarrow x(t)$, the relationship would be passive in the sense that x(t) would not further influence y(t). Sims begins by transforming the natural logarithms of the series x(t) and y(t) according to the filter.

$$x(t)^{*} = x(t) - 1.5\, x(t-1) + 0.562\, x(t-2) \qquad (4.5.1)$$

This process is purported to flatten the spectral density of many time series and to increase the likelihood that the regression residuals are white noise. To test the null hypothesis that x(t) does not cause y(t), the following equations would be estimated.

$$x(t)^{*} = \sum_{j=0}^{\infty} \beta_j\, y(t-j)^{*} + C_1 + v_1(t) \qquad (4.5.2)$$

$$x(t)^{*} = \sum_{j=-\infty}^{\infty} \beta_j\, y(t-j)^{*} + C_2 + v_2(t) \qquad (4.5.3)$$

Here (4.5.2) would be the restricted equation and (4.5.3) would be the unrestricted equation for F test on the future values of $y(t)^{*}$. To test the null hypothesis that y(t) does not cause x(t), the relevant equations would be:

$$y(t)^{*} = \sum_{j=0}^{\infty} \beta_j\, x(t-j)^{*} + C_1 + u_1(t)\phi \qquad (4.5.4)$$

$$y(t)^{*} = \sum_{j=-\infty}^{\infty} \beta_j\, x(t-j)^{*} + C_2 + u_2(t) \qquad (4.5.5)$$

Here (4.5.4) would be the restricted equation and (4.5.5) is the unrestricted equation. In either case, one tests whether $\beta_j = 0$ for all j<0, that is, all future values of the independent variable. Both Granger and Sims test the same null hypothesis. Hence the same F test can be employed.

4.5.2. Test Procedure

The null hypothesis to be tested is to be specified

H_0 : x(t) does not Granger-Cause y(t)

H_1 : x(t) does Granger-Cause y(t)

H_0 : y(t) does not Granger-Cause y(t)

H_1 : y(t) does Granger-Cause x(t)

Apply Sims Filter procedure to the variables by converting them to natural logarithms and using filter (4.5.1). Assume that the data are stationary. Choose the number of lags and leads to be used in models. Estimate the restricted and unrestricted models.

For x(t) $\rightarrow$ y(t) : Equations (4.5.2) and (4.5.3)

For y(t) $\rightarrow$ x(t) : Equations (4.5.4) and (4.5.5)

Calculate the test statistic where the values are defined as above.

$$F = \frac{\left(ESS_R - ESS_{UR}\right)/q}{ESS_{UR}/(n-k)} \qquad (4.5.6)$$

Accept the alternative hypothesis when the calculated F statistic is greater than F critical value. Here the placements of x(t) and y(t) in the equations is opposite that for the Granger test.

4.6. GEWEKE, MEESE-DENT TEST

4.6.1. Introduction

Geweke, Meese and Dent (1983) examined a number of forms of causality and found that the

Sims test may suffer difficulties because of its failure to correct for serially correlated residuals. They proposed a two-sided distributed lag model, then augmented it with lagged dependent variables to correct for serial correlation. It is required that the data be stationary. To test the null hypothesis that x(t) does not cause y(t), the following equations would be considered.

$$x(t) = \sum_{i=1}^{\infty} \alpha_i x(t-i) + \sum_{j=0}^{\infty} \beta_j y(t-j) + C_1 + u_1(t) \qquad (4.6.1)$$

$$x(t) = \sum_{i=1}^{\infty} \alpha_i x(t-i) + \sum_{j=-\infty}^{\infty} \beta_j y(t-j) + C_2 + u_2(t) \qquad (4.6.2)$$

Here (4.6.1) would be the restricted equations and (4.6.2) the unrestricted equations.

Similarly, for the null hypothesis that y(t) does not cause x(t), the relevant equations would be

$$y(t) = \sum_{i=1}^{\infty} \alpha_i y(t-i) + \sum_{j=0}^{\infty} \beta_j x(t-j) + C_1 + v_1(t) \qquad (4.6.3)$$

$$y(t) = \sum_{i=1}^{\infty} \alpha_i y(t-i) + \sum_{j=-\infty}^{\infty} \beta_j x(t-j) + C_2 + v_2(t) \qquad (4.6.4)$$

Here, equation (4.6.3) is the restricted equation for the F test and (4.6.4) the unrestricted equation. Here one will test if $\beta_j = 0$ for all j<0, that is, for all future values of the independent variable.

4.6.2. Test Procedure

Specify the hypothesis to be tested.

H_0 : x(t) does not Granger-Cause y(t)

H_1 : x(t) does Granger-Cause y(t)

H_0 : y(t) does not Granger-Cause x(t)

H_1 : y(t) does Granger-Cause x(t)

Assume that the data are stationary choose the numbers of lags and leads to be used in the equations. Estimate the restricted and unrestricted models.

For x(t) $\rightarrow$ y(t) : Equations (4.6.1) and (4.6.2)

For y(t) $\rightarrow$ x(t) : Equations (4.6.3) and (4.6.4)

Calculate the test statistic

$$F = \frac{\left(ESS_R - ESS_{UR}\right)/q}{ESS_{UR}/(n-k)} \qquad (4.6.5)$$

Accept the alternative hypothesis when the calculated F statistic is greater than the critical F value.

4.7. PIERCE-HAUGH TEST

4.7.1. Introduction

The causal pattern between two variables can be expressed interms of the cross-correlation functions. Define the cross correlation between a pair of series x(t) and y(t) as follows:

$$\rho_{uv}(k) = E[u(t-k)v(t)] \Big/ \left\{E\left[u^2(t)\right]E\left[v^2(t)\right]\right\}^{1/2} \qquad (4.7.1)$$

Here u(t) and $v(t)$ are white noise processor of x(t) and y(t) respectively and k is the number of cross-correlations tested. In practice, we replace (4.7.1) with the sample cross-correlation function.

$$\gamma_{uv}(k) = \Sigma u(t-k)v(t) \Big/ \left[\Sigma u^2(t)\, \Sigma v^2(t)\right]^{1/2} \qquad (4.7.2)$$

Here all terms are defined as above Pierce and Haugh (1977) transform this sample cross-correlation into the following statistic.

$$U = T \sum_{K=-N}^{M} \gamma_{u\upsilon}^2(k) \qquad (4.7.3)$$

Here T is the number of observations and K is the lag length chosen for the test. Under the null hypothesis that x(t) does not cause y(t), one estimates (4.7.3) for K=1 to M. The null hypothesis is rejected if U is greater than the selected critical value from Chi-Square distribution with N+M+1 degrees of freedom. The null hypothesis that y(t) does not cause x(t) utilizes the U statistic for the values of K=-1 to –M.

4.7.2. Test Procedure

Specify the hypothesis to be tested.

H_0 : x(t) does not cause y(t)

H_1 : x(t) does cause y(t)

H_0 : y(t) does not cause x(t)

H_1 : y(t) does not cause x(t)

Filter the series to be white noise. This can be done by obtaining the residuals for the series regressed on lags of the itself. Calculate the cross-correlations between the resulting residuals using equations (4.7.2). Calculate the test statistic, U using equations (4.7.3) for the chosen length K.

Accept the alternative hypothesis of the causality when the test statistic U is greater than $\chi^2_{(N+M+1)}$.

4.8. GEWEKE TEST

4.8.1. Introduction

Geweke (1982, 1984) further developed tests of causality to examine more explicitly the issue of reversals in linear and instantaneous linear causality and dependence between two variables. These tests attempt to decompose causality by frequency. For all three cases it is assumed that the series are stationary.

Linear Feedback

Geweke's measure of linear feedback for y(t) to x(t) and x(t) to y(t) are respectively.

$$F_{y\to x} = \ln\left(\left|\Sigma_1\right| / \left|\Sigma_2\right|\right) \qquad (4.8.1)$$

$$F_{x\to y} = \ln\left(\left|T_1\right| / \left|T_2\right|\right) \qquad (4.8.2)$$

where $\Sigma_1, \Sigma_2, T_1, T_2$ are the residual variance-covariance matrices from equations (4.8.1) and (4.8.2) respectively. Because x(t) and y(t) are one-dimensional univariate processes, the variance-covariance matrices are scalars. If the variances in the

restricted and unrestricted equations are the same, then l_n (1) will equal 0 and the statement y(t) does not cause x(t) is equivalent to $F_{y\to x} = 0$ and vice versa.

Instantaneous Linear Feedback

A measure of instantaneous linear feedback can be derived similarly as:

$$F_{x,y} = I_n \left(\left| \Sigma_2 \right| / \left| E_3 \right| \right) \qquad (4.8.3)$$

$$F_{y,x} = I_n \left(\left| T_2 \right| / \left| T_3 \right| \right) \qquad (4.8.4)$$

Such that $F_{x,y} = F_{y,y} \Sigma_2$ and T_2 are defined as above. E_3 and T_3 are the residual variance-Covariance matrices from equations (4.4.8) and (4.4.6) of Granger-Causality test respectively.

Linear Dependence

A third concept, closely related to the idea of linear feedback, is linear dependence. Geweke's measure of dependence between x(t) and y(t) can be simply derived as the sum of the three types of linear feedback presented earlier.

$$F_{x,y} = F_{y\to x} + F_{x\to y} + F_{x,y} \qquad (4.8.5)$$

That is, the level of linear dependence between x(t) and y(t) is the sum of the level of linear feedback from x(t) to y(t) and y(t) to x(t) and the level of instantaneous linear feedback between the two series.

4.8.2. Test Procedure

Consider the test procedure for each of the above. In all cases it is implicit that the series being tested are stationary. All test statistics are likelihood ratio tests that are distributed chi-squared.

4.8.3. Linear Feedback

Specify the hypothesis to be tested.

Feedback from x(t) to y(t).

$H_0 : F_{x\to y} = 0$ [No linear feedback from x(t) to y(t)]

$H_1 : F_{x \to y} \neq 0$ [Linear feedback from x(t) to y(t)]

4.8.4. Feedback from y(t) to x(t)

$H_0 : F_{y \to x} = 0$ [No linear feedback from y(t) to x(t)]

$H_1 : F_{y \to x} \neq 0$ [Linear feedback from y(t) to x(t)]

Note that acceptance of the null hypothesis is the same as finding unidimentional causality in the sense of Granger and Sims.

Estimate the relevant equations

For Feedback from x(t) to y(t) : Equations (4.8.1(and (4.8.2).

For Feedback from y(t) to x(t): Equations (4.8.3) and (4.8.4)

Calculate the relevant test statistic

For Feedback from x(t) to y(t)

$$n F_{x \to y} \sim \chi^2_{(klp)}$$

Here n is the number of observations

k is the number of variables comprising the vector x(t)

l is the number of variables comprising the vector y(t)

p is the number of lags on the independent variable in the equation tested under null hypothesis

The common case where x(t) and y(t) are univariate series, k=*l*=1. Compare the text statistic to the critical value of χ^2 with klp degrees of freedom for a given acceptable level of probability. The alternative hypothesis of linear feedback is accepted when the test statistic $> \chi^2_{(klp)}$.

4.8.5. Instantaneous Linear Feedback

Specify the hypotheses to be tested

$H_0 : F_{x.y} = 0$ [No instantaneous feedback between x(t) and y(t)]

$H_1 : F_{x.y} \neq 0$ [Instantaneous feedback between x(t) and y(t)]

or $H_0 : F_{y.x} = 0$ [No instantaneous feedback between x(t) and y(t)]

$H_1 : F_{y.x} \neq 0$ [Instantaneous feedback between x(t) and y(t)]

Testing only one of these pairs of hypotheses is necessary because $F_{x.y} = F_{y.x}$

Estimate the relevant equations

for $F_{x.y}$: Equations (4.4.4) and (4.4.8)

$F_{y..x}$: Equations (4.4.2) and (4.4.6)

Calculate the test statistic. Compare the test statistic to the criticals χ^2 value with klp degrees of freedom for the acceptable level of probability. Accept the alternative hypothesis if instantaneous linear feedback when test statistic $> \chi^2_{(klp)}$.

4.8.6. Linear Dependence

Specify the hypothesis to be tested.

$H_0 : F_{x.y} = 0$ [No Linear Dependence between x(t) and y(t)]

$H_1 : F_{x.y} \neq 0$ [Linear Dependence between x(t) and y(t)]

Calculate the measure of linear dependence as specified in equation (4.8.5). Calculate the test statistic.

$$nF_{x,y} \sim \chi^2_{(kl(2p+1)}$$

Where the values are defined as above. Compare the test statistic to the critical χ^2 acceptance value with kl (2p+1) degrees of freedom for acceptance level of probability. Accept the alternative hypothesis of linear dependence when the test statistic $> \chi^2_{[kl(2p+1)]}$.

4.8.7. General Guidelines

Several tests mentioned in this section, which test is the most desirable for testing the presence of causality? At this point there is no irrevocable evidence that can point to one test over the other. The Granger test seems to dominate the empirical

studies because of its case of implementation. Other tests, attempt to overcome some of the underlying testing problems. Judge et al (1985), suggest that testing based on present and past values of a variable does not include all possible relevant information. In addition, these are concerned with employing linear forecasts instead of nonlinear as well as using the mean square error criterion as a measure of forecast accuracy.

4.9. TESTS FOR INDEPENDENCE

4.9.1. Introduction

Testing for independence can be divided into two categories. Category (1) refers to independence in each elements of the vectors x(t). Category (2) refers to independence among the variables in x(t). For the tests of independence of category, we can utilize the methods of VTM for testing the null hypothesis of independence. Testing for category (2) type of independence requires examining the correlation structure among series. This can be accomplished in a bivariate fashion as in Pierce and Haugh (1977) or in a multivariate fashion as in the Portmantea u-test of Hosking (1981). Examining Category (2) type of independence in the bivariate case is problematic, because it ignores the cross-covariance's of the other variables. If these cross-covariance's are negligible then are can ignore them, which is often the case in empirical work.

4.9.2. Portmanteau-Test

Chitturi (1974) and Hosking (1981) developed and refined a multivariate test of independence based on the univariate version of the Portmanteau Test. In order to test null hypothesis of independence the following test statistic is proposed

$$Q = T \sum_{i=1}^{k} tr\left(\Sigma_i^1 \Sigma_0^{-1} \Sigma_i \Sigma_0^{-1}\right) \qquad (4.9.1)$$

Here T is the sample size

k is the number of lags

tr is the trace operator

Σ_i is the i^{th} covariance matrix of the variables in the jointly,

stational vector x(t).

Test Procedure

Form the null hypothesis of independence for the vector x(t).

H_0 : x(t) is independent

H_1: x(t) is not independent

for a given number of lags k, calculate the test statistic Q.

for a given significance level, α, obtain the critical value τ from the chi-square distribution with m^2k degrees of freedom. Reject H_0 if $Q > \tau$.

4.9.3. Engle-Granger Test

The Engle-Granger tests include two variations of a vector autoregressive (VAR) model. It is a restricted and unrestricted model tests. The restricted VAR test specifies the following two equations VAR (1) model.

$$\left.\begin{aligned} \Delta y(t) &= \beta_1 e(t-1) + \upsilon_1(t) \\ \Delta x(t) &= \beta_2 e(t-1) + \gamma \Delta y(t) + \upsilon_2(t) \end{aligned}\right\} \qquad (4.9.2)$$

Here, $\upsilon_1(t)$ and $\upsilon_2(t)$ are white noise processes and e(t) are the residuals from the equation

$$x(t) = a + \beta u(t) + e(t) \qquad (4.9.3)$$

To perform the test, one estimates (4.9.2) and (4.9.3) and calculate the test statistic

$$H_1 = t^2(\beta_1) + t^2(\beta_2) \qquad (4.9.4)$$

Here, $t(\beta_1)$ and $t(\beta_2)$ are the t statistics for β_1 and β_2 respectively from equations (4.9.3).

The null hypothesis is that co-integration does not exist. Critical values for the test can be found in Engle and Granger (1987). As an extension of the above test, an augmented restricted VAR test can be employed in the case where one includes upto p lags of the dependent variable in each equation of (4.9.3). The test statistic remains that of (4.9.4) and the null hypothesis is also the same.

The unrestricted VAR test is similar and specifies the following two equation VAR (1) model.

$$\left.\begin{aligned}\Delta y(t) &= \beta_1 y(t-1) + \beta_2 x(t-1) + C_1 + \upsilon_1(t)\\ \Delta x(t) &= \beta_3 y(t-1) + \beta_4 x(t-1) + \gamma \Delta y(t) + C_2 + \upsilon_2(t)\end{aligned}\right\} \quad (4.9.5)$$

Here C_1 and C_2 are constants and $\upsilon_1(t)$ and $\upsilon_2(t)$ are white noise processes. Under the null hypothesis of no co-integration, the test statistic is defined as

$$H_2 = 2(F_1 + F_2) \quad (4.9.6)$$

Here F_1 and F_2 and the F-statistics for the first and second equations in (4.9.5) respectively. Alternatively, one can alter (4.9.5) similarly to that of the restricted VAR case. The augmented unrestricted VAR is same as (4.9.5); except that p lags of the dependent variable are again included in each equation. The test statistic is same as in (4.9.6) and the critical values for both of these tests are provided in Engle and Granger (1987). The critical values for the restricted VAR, augmented VAR, unrestricted VAR, and the augmented unrestricted VAR depend on the assumption that the true model is VAR (1) with I(1) variables. If the true model is not the first order, then these tests should be discarded in favour of CRDW, DF and ADF tests. The simulation results of Engle and Granger (1987) and Engle and Yoo (1987) indicate that DF and ADF tests are quite powerful relative to the other tests.

4.9.4. Test Procedure

Specify the hypotheses to be tested

Restricted VAR test

$H_0: H_1 = 0$ (Cointegration does not exist)

$H_1: H_1 \neq 0$ (Cointegration does exist)

4.9.5. Unrestricted VAR test

$H_0: H_2 = 0$ (Cointegration does not exist)

$H_1: H_2 \neq 0$ (Cointegration does exist)

Estimate the cointegrating regression (4.9.3) and retain the residuals. Estiamte the relevant VAR test equations.

Restricted VAR Test

Equation (4.9.2)

Unrestricted VAR Tests

Equation (4.9.4)

Calculate the relevant test statistic

Restricted VAR Test

H_1 in equation (4.9.5)

Unrestricted VAR Test

H_2 in equation (4.9.6)

Compare the test statistic to the critical values in Engle and Granger (1987). We accept the alternative hypothesis of co-integration if the test statistic is greater than the critical values.

4.10. TESTS FOR FORECAST ACCURACY

4.10.1. Tests Evaluating Forecast Accuracy

Parametric Tests	**Non-parametric Tests**
Coefficient Tests	Single Point Criteria (MAPE, MSE, RMSE)
Pseudo Forecasts	Interval Criteria
Sensitivity Analysis	Error Cost Analysis
Direction and Turning Point Analysis	Turning Point Analysis Sensitivity Analysis

4.10.2. Comparative Accuracy Across Models

Parametric Tests

Granger-New Bold Tests

Steckler Test

In selecting a useful series of tests the goals suggested by Granger and NewBold (1986).

1. How good in some sense is a forecasting model or particular set forecasts?
2. Is one forecasting model or a set of forecasts better than its competitors?

Several possible tests that have been proposed to evaluate those goals have been summarized in (4.10.1). The tests relate to the accuracy of a single time series model or to comparative accuracy across models. The tests have also been classified as to whether they are parametric or non-parametric. Most of the tests presented so for to identify the various time series models have been parametric. They require same form of normal, t, F, or other statistical distribution to perform hypothesis tests. Such distribution methods can be applied here to obtain further information as to how a time series model might perform in the out-of-sample period. Non-parametric tests constitute a variety of tests whose application does not depend on hypothesis tests based on a known statistical distribution. These tests are performed primararely. In the out-of-sample period.

4.11. ACCURACY OF INDIVIDUAL MODELS

4.11.1. Parametric Tests

Coefficient tests

A traditional approach for evaluating forecast performance has been to determine the statistical properties of the regression coefficients in a give equation, where a particular functional form is assumed or given Direct tests of these coefficients often meet with difficulty in time series models and instead one must apply goodness-of-fit tests, such as the coefficient of determination, the standard error of the estimate and the chi-square or F statistics.

4.11.2. Pseudo Forecasts

Normally one has available a set of t observations on each of the independent and dependent variables of a model, some of these data can be saved for comparison purposes. Model estimation can be based on the shortened original series and

forecast evaluation can be made by comparing the forecast values and the actual values of the saved series.

4.11.3. Sensitivity Analysis

It is also important to observe the sensitivity of equation solutions or forecasts to variations in the parameters of that equation. In the case of multivariate models, parameters of an independent variable can simply be varied systematically and variations in the solutions evaluated in terms of deviations from the base solution. F tests are often used to validate such response surfaces. In the case of the multivariate VAR models, such sensitivity analysis can also be performed by examining the responses in the impulse functions.

4.12. DIRECTION AND TURNING POINT ANALYSIS

4.12.1. Introduction

The Henriksson-Merton (1983) test, which evaluates the ability of the time-series model to predict directional changes in the forecast variable, meets the first goal set by Granger and Newbold. Merton proposes that if a forecast has any value, it must cause a rational observer to modify prior beliefs about the distribution of subsequent movements in the variable being forecast. Because of the complexity of such a test, Cumby and Modest (1987) suggest the following simpler versions. Let $\gamma_i(t) = 1$ if the forecast change for a particular series is non-negative, otherwise $\gamma_i(t) = 0$. Under the null hypothesis that the forecast has no value, Hensiksson and Merton showed that the following condition must hold

$$\rho\left[\gamma_i(t) = 0 / \Delta x_i(t) < 0\right] + \rho\left[\gamma_i(t) = 1 / \Delta x_i(t) \geq 0\right] = 1 \qquad (4.12.1)$$

Here $\Delta x_i(t)$ is the actual change in the i^{th} variable of x(t).

4.12.2. Test Procedure

Let $x_i(t)$ be the actual value of the i^{th} variable in the vector x(t), and $\hat{x}_i(t)$ be the forecast value of $x_i(t)$.

Perform the following regressions:

$$\left.\begin{aligned}\Delta x_i(t) &= \alpha_1 + \beta_1 z_i(t) + \in_1(t) \\ \Delta x_i(t) &= \alpha_2 + \beta_2 \Delta\hat{x}_i(t) + \in_2(t)\end{aligned}\right\} \qquad (4.12.2)$$

Here, $z_i(t) = 1$ if $\Delta\hat{x}_i(t) > 0$ and

$z_i(t) = 0$ if $\Delta\hat{x}_i(t) < 0$ and

$\Delta\hat{x}_i(t) = \hat{x}_i(t) - \hat{x}_i(t-1)$

Form the null and alternative hypothesis

H_0: Forecast has no value, $\beta_1 = 0$ or $\beta_2 = 0$

H_1: Forecast has value, $\beta_1 > 0$ or $\beta_2 > 0$

From the $t(\beta_i)$ statistic for the parameter β_i based on the regressions in (4.12.2). For a given significance level α, obtain the critical value τ from the student t distribution. Reject H_0 if the $|t_p| > |\tau|$.

The Henriksson-Meston test also can provide information about the number of times each model correctly and incorrectly predicts both upward and downward directional changes in the variables of interest. For a series of N observed out-of-sample forecasts for the variables, N_1 is the number of observations with positive revisions, N_2 is the number of observations with non-positive revisions, total revisions one $N=N_1+N_2$, n_1 is the number of successful predictions given a positive revision, n_2 is the number of unsuccessful predictions given a positive revisions, and a the total is $n = n_1+n_2$. One then tests to determine whether the observed number of successful predictions is unlikely under the null hypothesis of n forecasting ability. Let υ represents the number of correct predictions. The null hypothesis of no forecasting ability is rejected when the probability of observing n_1 or more correct signs is unacceptably small. For a given significance level (α), the null hypothesis of no value in the forecasts is rejected when $n_1 > \upsilon^*$ is the solution to

$$HM = \sum_{v=v^T}^{n_1} \binom{N_1}{v}\binom{N_1}{n-v} \Big/ \binom{N}{n} = 1-\alpha \qquad (4.12.3)$$

Here n_1 is equal to the minimum of N_1 and n.

4.12.3. Test Procedure

Obtain N out-of-sample forecasts for each variables. Define N_1=the number of positive changes, N_2= the number of negative changes and $N=N_1+N_2$.

Count the number of successful predictions given positive changes n_1 and the number of successful predictions given negative changes n_2, where $n=n_1+n_2$. Let υ represent the number of correct predictions.

Form the null and alternative hypothesis

H_0: No forecast ability

H_1: Some forecast ability

Choose v^* and form test statistics (3) where n_1 = min (N, n).

For a given significance level α, reject H_0 if HM=1-α.

4.13. NON-PARAMETRIC TESTS

4.13.1. Single Point Criteria

The most popular tests that have been traditionally applied to measure the forecast errors between actual and forecast observations are single point criteria, such as the mean absolute percentage error, the mean squared error, and the root mean squared error.

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|x(t)-x^*(t)\right|}{x(t)}.100 \qquad (4.13.1)$$

$$MSE = \frac{1}{T}\sum_{t=1}^{T}\left[x(t)-x^*(t)\right]^2 \qquad (4.13.2)$$

$$RMSE = \left[\frac{1}{T}\Sigma\left[x(t)-x^*(t)^2\right]\right]^{1/2} \qquad (4.13.3)$$

Here x(t) = actual observation values and

$x^*(t)$=Forecast or estimated values in the out-of-sample period, t=1, ..., T.

These tests can also be made in the context of the pseudo forecasts. Although the error criteria relate to period-by-period error, they can also be applied to analysis of multiperiod forecast analysis.

4.13.2. Interval Criteria

If one wants or interested in predicting an intervals or range of values, then a test of confidence intervals can be developed and applied for this purpose.

4.13.3. Error Cost Analysis

Although the above criteria are the most practical for evaluating performance with a historical sample or a small post-sample data set, they fail to define the surrounding probabilistic conditions in a way that would be useful with a large post-sample data set. There are two ways to improve this situations. First, an informative forecast could accompany the point forecasts based on some mathematical statement regarding the probability distributions surrounding these forecasts. This amounts to an interval forecast in which the point forecast is none presented along with an appropriate confidence interval. Second, a decision forecast could be prepared that recommends that the forecast be accepted in relation to some alternative consequence. Granger and Newbold (1986) consider the case where the policy makes must decide on a certain policy that depends upon the future value $x^*(t)$ of a dependent variable x(t). The future value is not known and the policy maker could make in correct decision. The loss or consequence the policy maker must undergo an such a case is given by the loss function L $[D_i, x(t)]$, which describes the loss of selecting decision D_i. When x(t) turns out to be the true value of $x^*(t)$.

4.13.4. Turning Point Analysis

A major characterization of a models performance is its ability to explain the turning points of fluctuations in values of a dependent variable. There are a number of descriptive variables or statistics that can be used to evaluate turning point errors. Most often this pertain to the number of turning points missed, the number of turning points falsely predicted, the number of under and over predictions, rank correlations of the predicted and actual changes and various tests of randomness in prediction. Naik and Leathold (1986) have suggested a method of evaluating turning

point performance based on the use of contingency tables. Henriksson and Merton (1983) also provide a parameter tests.

4.14. COMPARATIVE ACCURACY ACROSS MODELS

4.14.1. Introduction

4.14.2. Granger-Newbold Test

Consider the forecast errors generated from two different models

$$e_1(t) = x_1(t) - x_1^*(t) \quad \text{from model 1} \tag{4.14.1}$$

$$e_2(t) = x_2(t) - x_2^*(t) \quad \text{from model 2}$$

Assume that $l_1(t)$ and $l_2(t)$ constitute a random sample from a bivariate normal distribution with means zero, variances σ_1^2 and σ_2^2 and correlation coefficient ρ.

$$E\left[e_1(t)\right] = E\left[e_2(t)\right] = 0$$

In this case the individual forecasts are unbiased and the forecast errors non-correlated. Form a pair of random variables given by $l_1(t) + l_2(t)$ and $l_1(t) - l_2(t)$.

In this case

$$E\left\{\left[e_1(t) + e_2(t)\right]\left[e_1(t) - e_2(t)\right]\right\} = \sigma_1^2 - \sigma_2^2 \tag{4.14.2}$$

Given the assumption of unbiasedness, the two error variances and the two expected square errors will be equal if and only if this pair of random variables is uncorrelated. The test Granger and Newbold propose to evaluate zero correlations is based on the sample correlation coefficient.

$$\gamma = \frac{\sum_{t=1}^{T}\left[e_1(t) + e_2(t)\right]\left[e_1(t) + e_2(t)\right]}{\left[\sum_{t=1}^{T}\left[e_1(t) + e_2(t)\right]^2 \sum_{t=1}^{T}\left[e_1(t) - e_2(t)\right]^2\right]^{1/2}} \tag{4.14.3}$$

This coefficient is used to form a test statistic that compares the equality of the squared forecast errors from each model.

$$Z = \left[\text{In}(1+r) - \text{In}(1-r)\right]\frac{(T-3)^{1/2}}{2} \qquad (4.14.4)$$

Where γ is the sample correlation obtained from (4.14.3) and T is the number of out of sample predictions. Under the null hypothesis of no correlation, Z is approximately distributed N(0,1).

4.14.3. Test Procedure

Obtain the forecast errors, $e_1(t)$ and $e_2(t)$ from two competing models, Model 1 and Model 2. Form the null and alternative hypothesis. Form the test statistic.

$$Z = \left[In(1+r) - In(1-r)\right](T-3)^{1/2}/2 \qquad (4.14.5)$$

Here, γ is the correlation coefficient from (4.14.3)

For a given significance level α, obtain the critical valve z from the standard normal distribution. Reject H_0 if $|Z| > |z|$.

4.14.4. Stickler Test

Another approach compares the accuracy of forecasts generated by different models employing a ranking procedure. Here, each model is ranked according to its forecast accuracy, the latter being measured by the root mean squared error. A score equal to the ranking is assigned to each variable. Aggregate scores are obtained for each of the series by summing of the rankings across the given forecast horizon. If the time series models have equal forecast accuracy, the scores would have the same expected value for each model. A χ^2 goodness of fit statistic is used to test for differences in forecast accuracy by examining whether the aggregate score differs significantly from the expected score, assuming the models had equal forecast accuracy. This criterion explicitly compares the complete set of forecasts over each period for each model.

4.15. TESTS FOR NORMALITY

4.15.2. Skewness and Kurtosis Test

Mardia (1970) has developed multivariate extensions to the univariate measures of skewness and kurtosis in univariate time series models. These can be computed by

the following formulae.

$$b_{1,m} = T^{-2} \sum_i \left[\left(x_i - \mu \right)^1 \Sigma^{-1} \left(x_i - \mu \right) \right]^3 \text{ Skewness} \quad (4.15.1)$$

$$b_{2,m} = T^{-2} \sum_i \left[\left(x_i - \mu \right)^1 \Sigma^{-1} \left(x_i - \mu \right) \right]^2 \text{ Kurtosis} \quad (4.15.2)$$

Here μ is the mean vector of x(t) and i=1, ….. T.

In the case when the vector x(t) is distributed multivariate normally, the following expectation equalities hold:

$$E[b_1, m] = 0 \quad (4.15.3)$$

$$E[b_s, m] = m(m+2)$$

In order to test the null hypothesis of multivariate normality, Mordia (1970) proposes testing each moment individually.

4.15.3. Test Procedure

Estimate the skewness and Kurtosis coefficients

$$b_{1,m} = T^{-2} \sum_i \left[\left(x_i - \mu \right)^1 \Sigma^{-1} \left(x_i - \mu \right) \right]^3 \text{ Skewness} \quad (4.15.4)$$

$$b_{2,m} = T^{-1} \sum_i \left[\left(x_i - \mu \right)^1 \Sigma^{-1} \left(x_i - \mu \right) \right]^2 \text{ Kurtosis} \quad (4.15.5)$$

Here μ is the mean vector of x.

Form the null hypothesis for each measure.

$H_0 : x(t)$ is distributed multivariately normally, $N(\mu, \Sigma)$

$H_1 : x(t)$ is not distributed normally

Calculate the test statistics, A and B.

$$A = b^{-1}_{Tb_1, m} \quad \text{Skewness} \quad (4.15.6)$$

$$B = b_{2,m} - m(m+2) / \left[8m(m+2)/T \right]^{1/2} \text{ Kurtosis} \quad (4.15.7)$$

For a given significance level, α, choose the critical value τ from a chi-square distribution with f degrees of freedom, where $f = b^{-1}m(m+1)(m+2)$ to test the skewness and choose the critical value z from the standard normal distributions to test the Kurtosis. Reject H_0 if $A > \tau$ skewness; $|\beta| > |Z|$ Kurtosis.

4.16. DIAGNOSTIC CHECKING FOR LINEAR (STATIONARY) TIME SERIES MODELS

4.16.1. Introduction

After a model being identified and parameters estimated, the problem is to decide whether the model is adequate. The model should be modified when there be an evidence of serious inadequacy. No system of diagnostic checks can ever be comprehensive. Hence it is always possible that characteristics in the data of an unexpected kind can be overlooked. If thoughtfully devised diagnostic checks applied to a model fitted to a reasonably large body of data fail to show serious discrepancies, it is reasonable to feel more comfortable about using that model. If any inadequacies are revealed, they may suggest alternatives that appear more appropriate.

Box and Jenkins suggest two ways to perform diagnostic checking the method of over fitting and diagnostic checks applied to residuals. The method of over fitting of makes use of the assumption that at a particular ARMA model of order (p,q) implicitly imposes the restrictions that in the more general model

$$\left(1 + a_1 L + ... + a_p L^p + a_{p+1} L^{p+1} + ... + a_{p+r} L^{p+r}\right) y_t \qquad (4.16.1)$$

$$= \left(1 + \theta_1 L + ... + \theta_q L^q + \theta_{q+1} L^{q+1} + ... + \theta_{q+s} L^{q+s}\right) \in_t$$

The coefficients a_{p+j}, j=1, ……, γ and d_{q+j}, j=1, …… s are zero. This is a testable hypothesis. The model identified can be extended by adding extra coefficients. The augmented model can then be estimated and the standard deviations of the estimates of the added coefficients will indicates whether or not the additional coefficients differ significantly from zero by usual t-test. Because of the issue of model multiplicity, as a practical consideration, in fitting extra coefficients one should to add terms simultaneously to both sides of the ARMA model. It would be more

appropriate not add such terms to the autoregressive operator and moving average operator sequentially.

Diagnostic checks applied to residuals are under the assumptions that if a time-series model is correctly specified, the error $\in_t$ constitute a white noise process. If the series $\in_1, ..., \in_n$ are available, natural checks on model adequacy could be based on the sample autocorrelations of this series Anderson (1942) has shown that the sample autocorrelations of white noise are asymptotically independently normally distributed with zero means and standard errors $n^{-1/2}$. Hence the statistic $n\sum_{k=1}^{k}\hat{\tau}_k^2$ is asymptotically distributed as Chi-square with k degrees of freedom. Here the $\hat{\tau}_k$'s are sample autocorrelations. This statistic can be used to test whether the $\in_t$'s are uncorrelated.

In practice, the $\in_t$'s are unknown and only the residuals from the fitted ARMA (p,q) models, $\hat{\in}_t$ are available. For a univariate time-series model, the fitted value is the one-step ahead forecast error. For example, with an AR(1) model where $\hat{a}_1$ is the estimate, the fitted value at time t is $-\hat{a}_1 y_{t-1}$ and the estimated residual $\hat{\in}_t$ is $\hat{\in}_t = y_t + \hat{a}_1 y_{t-1}$.

When sample autocorrelations are calculated from $\hat{\in}_t$, as

$$\gamma_k = \frac{\left[1/(n-k)\right]\sum_{t=k+1}^{n}\hat{\in}_t\hat{\in}_{t-k}}{(1/n)\sum_{t=1}^{n}\hat{\in}_{t-k}^2} \qquad (4.16.2)$$

The statistic

$$Q(k) = n\sum_{k=1}^{k}\gamma_k^2 \qquad (4.16.3)$$

is better approximated by a Chi-Square distribution with k-p-q degrees a freedom as shown is Box and Pierce (1970). However, for n<100, the approximation can be rather poor. Ljung and Box (1978) suggest using, instead of Q(k).

$$Q^{*}(k) = n(n+2)\ \sum_{k=1}^{k} \frac{\gamma_k^2}{n-k} \qquad (4.16.4)$$

If several models are estimated and all pass the diagnostic checks, the model selection criteria, say, Akaike information criterion could be used to select a specific model representation.

In a diagnostic checking we look at the residuals $e_t = y_t - \hat{y}_t$, they express the variation that the regression model has not been able to explain. We consider the residual e_t as an estimate of the error $\in_t$. If the fitted model is correct, the residual should confirm the assumptions we have made about the error terms. A histogram of the residuals and plots of the residuals e_t against the fitted values $\hat{y}_t$, against each independent variable, and against time are useful at this stage of the analysis. If the assumptions in the regression model are satisfied, the histogram of the residuals should resemble a normal distribution. The residuals, when plotted $\hat{y}(t)$ against each independent variable or against time, should vary in a horizontal band around zero. Any departure from such a horizontal band will be taken as an indication of model inadequacy.

A plot of e_t against y_t would be meaningless, since the residuals and the dependent variable are always correlated, even if the model is adequate. The sample correlation γ_{ey} between e and y is given by $(1\text{-}R)^{1/2}$.

4.16.2. Incorrect Functional Form

Residual plots will indicate whether the functional form of the regression model is misspecified. For example, if the true model is described by a quadratic relationship but only a linear model is fitted to the data, the residual e_t, when plotted against $\hat{y}_t$ or the independent variable, will exhibit a curvilinear pattern. If a constant or a linear term in the regression model is incorrectly omitted, the residual plot will show a linear relationship between e_t and $\hat{y}_t$ or the independent variable.

4.16.3. Lack of Fit Tests

Residual plots are important tools in model diagnostic checking. They provide a visual indication of whether the considered model form is adequate and suggest

modification of the model if lack of fit is found. These diagnostic tools are quite general, since they do not assume a specific alternative hypothesis.

A more formal test of lack of fit can be obtained if we have genuine replications at some of the predictor levels. Genuine replications have to be uncorrelated with all other observations. It would not usually be sufficient to take two measurements from the same experiment, since in such a case the measurements would probably be correlated. The replications can be used to partition the error sum of squares into a part that is due to pure error, SSPE, and one that is due to lack of fit, SSLF. Let us assume that we have observed the responses at k different settings of the p predictor variables, x_1, x_2, …., x_k. At each level x_i we observe n_i responses is $y_1^{(i)}, y_2^{(i)}, \ldots\ldots, y_{ni}^{(i)}$, where $\sum_{i=1}^{k} n_i = n$. Then the SSPE contribution at level x_i is given by $\sum_{t=1}^{n_i}\left(y_t^{(i)} - \bar{y}^{(i)}\right)^2$.

$$\text{Where } \bar{y}^{(i)} = \frac{1}{ni}\sum_{t=1}^{n_i} y_t^{(i)} \qquad (4.16.5)$$

is the average at level x_i. Since one parameters the mean is estimated, this sum of squares contribution has $n_i - 1$ degrees of freedom. Overall, the pure error sum of squares is given by

$$SSPE = \sum_{i=1}^{k}\sum_{t=1}^{n_i}\left(y_t^{(i)} - \bar{y}^{(i)}\right)^2 \qquad (4.16.6)$$

and has $\sum_{i=1}^{k}\left(x_i - 1\right) = n - k$ degrees of freedom. The lack of fit sum of squares is given by

SSLF = SSE-SSPE

and has $n - p - 1 - (n - k) = -k - p - 1$ degrees of freedom.

The following is the concerned ANOVA table.

Source	SS	df	M.S	F
Regression	SSR	p	MSR	$F_{LF} = \frac{MSLF}{MSPE}$
Error	SSE	n-p-1	MSE	
Lack of Fit	SSLF	k-p-1	MSLF	
Pure Error	SSPE	n-k	MSPE	
Total (corrected for mean)	SSTO	n-1	MSPE	

If the model is adequate, MSE, MSLF and MSPE all estimate the variances σ^2. If there is lack of fit, the mean square error MSE estimates a combination of the variance and lack of fit. To test for lack of fit, we look at the ratio of the lack of fit and the pure error mean square, $F_{LF} = MSE/MSPE$. If there is no lack of fit, their ratio has an F distribution with k-p-1 and n-k degrees of freedom. In lack of fit situations, this ratio will be larger than the values that can be expected from this distribution. Thus lack of it is indicated if $F_{LF} > F_{\alpha}(k-p-1, n-k)$. In such a case we must modify the original model. If no lack of fit is indicated, we can pool the variance estimates from MSLF and MSPE and use the mean square errors MSE in the subsequent significance tests.

4.16.4. Non-constant variance

If the scatter plot of e_t against $\hat{y}_t$ does not full within two horizontal bands around zero but exhibits a 'funnel' shape, we can conclude that the equal variance assumption is violated. In such a case we use weighted least squares or use transformations to stabilize the variance.

Consider the general regression model

$$y_t = f(x_t;\beta) = \eta_t + \in_t \qquad (4.16.7)$$

Where $\eta_t = f(x_t;\beta)$

Let us assume that the variance of the errors is functionally related to the mean level η_t as

$$Var\ (y_t) = Var\ (\in_t) = h^2(\eta_t)\sigma^2 \qquad (4.16.8)$$

Where h is some known function.

Our objective is to find a transformation of the data, $g(y_t)$, that will stabilize the variance, that is, the variance of the transformed variable $g(y_t)$ should be constant. Expanding the function $g(y_t)$ in a first-orders Taylor series around η_t.

$$g(y_t) = g(\eta_t) + (y_t - \eta_t)\, g^1(\eta_t) \tag{4.16.9}$$

Where $g^1(\eta_t)$ is the first derivative of $g(y_t)$ evaluated at η_t.

The variance of the transformed variable can be approximated as

$$\text{Var}\left[g(y_t)\right] = \text{Var}\left[g(\eta_t) + (y_t - \eta_t)\, g^1(\eta_t)\right]$$

$$= \left[g^1(\eta_t)\right]^2 V(y_t) - \left[g^1(\eta_t)\right]^2 \left[h(\eta_t)\right]^2 \sigma^2$$

In order to stabilize the variance, we have to choose the transformation g(.) such that

$$g^1(\eta_t) = \frac{1}{h(\eta_t)} \tag{4.6.11}$$

These transformations not only stabilize the variance, but also lead to simplifications in the functional representation of the regression model.

4.16.5. Serial Correlation Among the Errors

In ordinary regression models are under the assumption that the errors $(\ldots\ldots, \in_{t-1}, \in_1 \ldots..)$ are uncorrected, when a regression model is fitted on time series data, it is likely that the errors are serially correlated. Thus, diagnostic checks that test for correlation among the errors are of particular importance. If the correlated errors are ignored in the regression models, then the consequences be quite serious. For example, let there be a positive serial correlation among the errors, the true standard errors of the regression coefficients can be considerably underestimated by the usual standard errors $s\hat{\beta}_t = s\sqrt{c_{ii}}$. This means that if the usual least squares procedures are employed in the presence of serially correlated errors, the parameter estimates may appear significantly different from zero when infact they are not. This phenomenon is called spurious regression. [Box and Newbold (1971), Granger and Newbold (1974)].

4.16.6. Serial Correlation in a Time Series

The presence of serial correlation among a time-order sequence of random variables $\{....., z_{t-1}, z_t, z_{t+1}.....\}$ indicates. That the random variables at different time periods all correlated. Serial correlation is measured by auto-covariances γ_k or autocorrelations ρ_k.

The lag k autocovariance is defined by

$$\gamma_k = \text{Cov}(z_t, z_{t-k}) = E(z_t - \mu)(z_{t-k} - \mu) \quad k = 0,1,2,.... \tag{4.16.12}$$

Here we have assumed that the mean $\mu = E(z_t)$ is constant over time and further more that the autocovariance depends only on the time difference k, but not on time t. This assumption requires that the first two moments (mean μ, variance γ_0 and auto-covariance γ_k) are invariant with respect to changes along the time axis. This is called stationarity condition.

From (1) and the stationarity assumption, it follows that

$$\gamma_{-k} = E(z_t - \mu)(z_{t+k} - \mu) = E(z_{t+k} - \mu)(z_t - \mu) = \gamma_k \tag{4.16.13}$$

The auto-correlations ρ_k are defined as

$$\rho_k = \frac{\text{cov}(z_t, z_{t-k})}{\left[\text{Var}(z_t)\,\text{Var}(z_{t-k})\right]^{1/2}} = \frac{\gamma_k}{\gamma_0} \quad k = 0,1,2,..... \tag{4.16.14}$$

The set of autocorrelations ρ_k, considered as a function of the lag k, is called the correlogram or the autocorrelation function. Since $\rho_{-k} = \rho_k$, only non-negative k have to be considered. Further $\rho_0 = 1$. Estimates of the auto-covariance γ_k from a sample series $(z_1, z_2, ..., z_n)$ are given by

$$C_k = \frac{1}{n}\sum_{t=k+1}^{n}(z_t - \bar{z})(z_{t-k} - \bar{z}) \quad k = 0,1,2,..... \tag{4.16.15}$$

Where $\bar{z}$ is the sample mean and

$$\bar{z} = \frac{1}{n}\sum_{t=1}^{n} z_t$$

Estimates of the auto-correlations ρ_k are given Δy the sample auto-correlations

$$r_k = \frac{C_k}{C_0} = \frac{\sum_{t=k+1}^{n} (z_t - \bar{z})(z_{t-k} - \bar{z})}{\sum_{t=1}^{n} (z_t - \bar{z})^2} \qquad (4.16.16)$$

Basic results for the distribution theory of sample auto-correlations were derived by Bartlett (1946). He showed, among other results, that if there is no correlation among observations that are more than q steps apart $(\rho_k = 0 \text{ for } k > q)$, the variance of r_k can be approximated by

$$\text{Var}(r_k) = \frac{1}{n}\left(1 + 2\sum_{k=1}^{q} \rho_k^2\right) \text{ for } k > q \qquad (4.16.17)$$

In the special case when all observations are uncorrelated, ($\rho_k = 0$ for k>0), this equation reduces to

$$\text{Var}(r_k) = n^{-1} \text{ for } k > 0 \qquad (4.16.18)$$

For large n and $\rho_k = 0$, the distribution of r_k will be approximately normal. Hence, one can test the null hypothesis $(H_0 : \rho_k = 0)$ by comparing r_k with its standard error $n^{-1/2}$ and reject H_0 at the common significance level $\alpha = 0.05$ if $\frac{|r_k|}{n^{-1/2}} = \sqrt{n}\,|r_k| > 1.96$

CHAPTER - V ESTIMATION IN TIME SERIES REGRESSION MODELS

5.1. INTRODUCTION

The classical linear regression mode is

$$Y = XB + U \tag{5.1.1}$$

$$E(U) = 0 \tag{5.1.2}$$

$$E(UU^{1}) = \sigma_u^2 I \tag{5.1.3}$$

X is fixed and has rank (k+1)<T (5.1.4)

Here Y is a Tx1 vector of observations on the dependent or endogenous variable.

X is a Tx(1+k) matrix of observations on the k independent or exogenous variables and the intercept vector (of is)

B is a (1+k)x1 vector of coefficients

U is a Tx1 vector of disturbances

E denotes the "expected value of", $E(u) = \Sigma u_i \text{ Prob}(u_i) = \text{Mean of } U$.

The first assumption of the classical linear regression model specifies that each observation on the endogenous variable y can be expressed as a linear function of the exogenous variables x plus the disturbance U. This can be expressed as

$$Y_t = \beta_0 + \sum_k \beta_k X_{tk} + u_t$$

The second property of the model implies that the disturbances have no systematic components and therefore each has zero expectation. The third assumption says that expected value of u_t^2 is σ_u^2 for all t and that the covariance of u_t with $u_t \pm \theta$ zero where $(\theta \neq 0)$.

The final property of the classical model specifies that the X_{tk} are fixed in repeated sampling, uncorrelated with any omitted variables and thus independent of the disturbance u_t.

Given the sample data on y and x and the prior assumptions of equations (5.1.1) to (5.1.4) the researcher seeks to estimate the model

$$Y = X\hat{\beta} + \hat{c} \tag{5.1.5}$$

Such that the sum of the squared error $\hat{U}^1\hat{U}$ is minimized. By definition $\hat{U}$ is expressed.

$$\hat{U} = \left(Y - X\hat{B}\right) \tag{5.1.6}$$

ordinary least-squares estimation involves minimizing

$$\begin{aligned}\hat{U}^1\hat{U} &= \left(Y - X\hat{\beta}\right)^1 \left(Y - X\hat{\beta}\right) \\ &= Y^1Y - Y^1X\hat{B} - \left(X\hat{B}\right)^1 Y + \left(X\hat{B}\right)^1 X\hat{B} \\ &= Y^1Y - 2\hat{B}^1X^1Y + \hat{B}^1X^1X\hat{B}\end{aligned} \tag{5.1.7}$$

Solving for the parameter $\hat{B}$ so that $\hat{U}^1\hat{U}$ is minimized requires differentiating equation (5.1.7) with respect to $\hat{B}$.

$$\begin{aligned}\frac{\partial\hat{U}^1\hat{U}}{\partial\hat{B}} &= 0 - 2X^1Y + 2X^1X\hat{B} \\ &= -XY + X^1X\hat{\beta}\end{aligned} \tag{5.1.8}$$

By setting equation (5.1.8) equation to zero and rearrange terms to isolate $\hat{B}$, we obtain ordinary least squares (OLS) estimator.

$$-X^1Y + X^1X\hat{B} = 0$$

$$X^1X\hat{B} = X^1Y$$

$$\left(X^1X\right)^{-1} X^1X\hat{B} = \left(X^1X\right)^{-1} X^1Y$$

$$\hat{\beta} = \left(X^1X\right)^{-1} X^1Y \tag{5.1.9}$$

In case of simple bivariate regression the OLS estimator in equation (5.1.9) has the familiar algebraic expression,

$$\hat{\beta} = \frac{\sum_t \left(X_t - \bar{X}\right)\left(Y_t - \bar{Y}\right)}{\sum_t \left(X_t - \bar{X}\right)^2} \tag{5.1.10}$$

A few additional preliminary results need to be established. First it must be demonstrated that the OLS estimator $\hat{B}$ is unbiased, that is an the average it hits the target parameter B^2. This is shown by substituting the expression for y in equation (5.1.1) into the expression for $\hat{B}$ in equation (5.1.9)

$$\hat{B} = \left(X^1X\right)^{-1} X^1 \left(XB + U\right)$$

$$= \left(X^1X\right)^{-1} X^1XB + \left(X^1X\right)^{-1} X^1U \tag{5.1.11}$$

$$= B + \left(X^1X\right)^{-1} X^1U$$

The unbiasedness of $\hat{B}$ is proved by taking the expected value of equation (5.1.11).

$$E(\hat{B}) = B + E\left[\left(X^1X\right)^{-1} X^1U\right] \tag{5.1.12}$$

$$= B + \left[\left(X^1X\right)^{-1} X^1E(U)\right] \text{ by equation } \quad (5.1.4) – (5.1.12)$$

$$E(\hat{B}) = B \text{ by equation (5.1.2)}$$

Finally the variance of $\hat{B}$ in the classical model must be generated because greater parts of the discussion evaluate the precision and significance of parameter estimates when disturbances are autocorrelated. The variance-covariance matrix of the OLS estimator is defined as

$$\text{Var } (\hat{B}) = E\left[(\hat{B} - \beta)(\hat{B} - \beta)^1\right] \tag{5.1.13}$$

It is convenient to use the expression for $\hat{B}$ in equations (5.1.11) and to rewrite the equation (5.1.13) as

$$\text{Var } (\hat{B}) = E\left[B + (X^1X)^{-1}X^1U - B\right]\left[B + (X^1X)^{-1}X^1U - B\right]^1$$

$$= E\left[(X^1X)^{-1}X^1U\right]\left[(X^1X)^{-1}X^1U\right]^1$$

$$= E\left[(X^1X)^{-1}X^1UU^1\ X(X^1X)^{-1}\right] \qquad (5.1.14)$$

$$= \left[(X^1X)^{-1}X^1\ E(UU^1)\ X(X^1X)^{-1}\right] \qquad \text{by equation (5.1.4)}$$

$$= E(X^1X)^{-1}X^1\left(\sigma_u^2 I\right)X(X^1X)^{-1} \qquad \text{by equation (5.1.3)}$$

$\text{Var}(\hat{B}) = \sigma_u^2(X^1X)^{-1}$ (Since σ_u^2 is a scalar quantity and the identity matrix may be suppressed).

In the more familiar bivariate case, the variance of the OLS estimator may be expressed.

$$\text{Var}(\hat{\beta}) = \sigma_u^2 \Big/ \sum_t \left(X_t - \bar{X}\right)^2 \qquad (5.1.15)$$

Autocorrelated Disturbances and Generalized Least-Squares Estimation

The consequences of Autocorrelation

What are consequences of serially dependent disturbances for statistical estimation, hypothesis testing and casual inference? The classical model is clearly no longer appropriate. In particular the assumption (3) must be revised as follows.

$$E(UU^1) = \sigma_u^2\Omega \qquad (5.1.16)$$

Where Ω is a TxT, symmetric, positive definite matrix. The Ω specification in the revised, generalized linear regression model allows for both heteroscedasticity (non-constant diagonal elements) and autocorrelations (non-zero off-diagonal elements). Here we are concerned with problems of time-series estimation and heteroscadasticity is commonly a cross-sectional problem. For our purpose, the Ω matrix is considered to have is in the diagonal and autocorrelation parameters in the off-diagonal cells. Hence in scalar notation equation (5.1.16) implies,

$$E(u_t u_t \pm \theta) = \sigma_u^2 \qquad \text{for } \theta = 0 \qquad \text{(homoscadasticity)} \qquad (5.1.17)$$

$$= \gamma_\theta \qquad \text{for } \theta \neq 0 \qquad \text{(autocovariance)}$$

Where γ_θ is the lag θ autocovariance.

The first point to be made concerning the impact of autocorrelated disturbances is that OLS estimates remain unbiased. From equation (5.1.11)

$$\text{Var}\,(\hat{\beta}) = \sigma_u^2 \Big/ \sum_t \left(X_t - \bar{X}\right)^2 \qquad (5.1.18)$$

Since in the revised, autocorrelation model x remain fixed or $E\,(x^1U)$ remains zero, the OLS estimates still has expectation B.

$$E\,(\hat{B}) = B + E\left[\left(X^1X\right)^{-1} X^1U\right]$$

$$= B + E\left[\left(X^1X\right)^{-1} X^1E(U)\right] \qquad (5.1.19)$$

$$E\,(\hat{B}) = \beta$$

Hence it is possible to estimate a regression model in the conventional (OLS) manners without danger of bias even if the disturbances are serially correlated.

However, the variance of $\hat{B}$ in the presence of autocorrelated disturbances is no longer that of the classical model in equation (5.1.14), but is

$$\text{Var}\,(\hat{B}) = E\left[\left(\hat{B}-\beta\right)\left(\hat{B}-\beta\right)^1\right] \qquad \text{see equation (5.1.14)}$$

$$= E\left[\left(X^1X\right)^{-1} X^1(UU^1)\, X\left(X^1X\right)^{-1}\right] \qquad (5.1.20)$$

$$= E\left[\left(X^1X\right)^{-1} X^1(\sigma_u^2\Omega)\, X\left(X^1X\right)^{-1}\right] \quad \text{by equation (5.1.16)}$$

$$\text{Var}\,(\hat{B}) = \sigma_u^2\left(X^1X\right)^{-1} X^1\Omega\, X\left(X^1X\right)^{-1}$$

Thus, when disturbances are interdependent, which frequently is true in time series models, OLS regression yields biased estimates of the coefficient variances. Since the bias is generally negative, the estimated variances and standard errors understate the true variances and standard errors. This produces inflated t-ratios, a false sense of confidence in the precision of the parameter estimates, and often leads to spurious attributions of significance to independent variables. Moreover the OLS estimate of the disturbances covariance σ_u^2 is also biased and since the bias is typically negative, R^2 as well as t-and F-statistics tend to be exaggerated.

This result is straight forwardly demonstrated. The true disturbances are never observed directly but must be derived from the fitted model and hence are "filtered"

though the x's. Thus the residuals $\hat{u}_t$ are generated,

$$\hat{U} = Y - X\hat{B}$$

$$= Y - X(X^1X)^{-1}X^1Y \quad \text{by equation (5.1.9)}$$

$$= (XB + U) - X(X^1X)^{-1}X^1(XB + U) \quad \text{by equation (1)}$$

$$= (XB + U) - X(X^1X)^{-1}X^1XB - X(X^1X)^{-1}X^1U$$

$$= (XB + U) - XB - X(X^1X)^{-1}X^1U$$

$$= U - X(X^1X)^{-1}X^1U$$

$$= \left[I_T - X(X^1X)^{-1}X^1\right]U$$

$$= MU$$

Where $M = \left[I_T - X(X^1X)^{-1}X^1\right]$

Equations (5.1.21) establishes that the residuals of the classic model are a linear function of the unknown disturbances. The sum of the squared residuals the quantity minimized by least-squares regression can therefore be expressed,

$$\hat{U}^1\hat{U} = U^1M^1MU$$

$$= U^1M^2U \qquad (5.1.22)$$

$$= U^1MU \text{ (Since M is symmetric and idempotent).}$$

$$= U^1\left[I_T - X(X^1X)^{-1}X^1\right]U \text{ by equation (5.1.21)}$$

The expected value of equation (5.1.22) yields the classical estimator of the disturbance sum of squares in terms of the true disturbance variance σ_u^2,

$$E\left(\hat{U}_1\hat{U}\right) = E\left(U^1MU\right)$$

$$= Etr\left(U^1MU\right) \text{(Since } U^1MU \text{ is scalar and therefore equal to its trace)}$$

$$= Et_r\left(MUU^1\right) \text{ (Since tr AB = tr BA)}$$

$$= \sigma_u^2 \text{trM} \qquad (5.1.23)$$

$$= \sigma_u^2 \text{tr}\left[I_T - X(X^1X)^{-1}X^1\right] \quad \text{by equations (5.1.21) and (5.1.22)}$$

$$= \sigma_u^2 \text{tr}\,(I_T) - t_r\left[(X^1X)\,(X^1X)^{-1}\right]$$

$$= \sigma_u^2 t_r\,(I_T) - \text{tr}(I_{k+1})$$

$$= \sigma_u^2\,(T-k-1)$$

Where k denotes the number of exogenous variables and tr denotes trace-the sum of the diagonal elements of a matrix. Thus an unbiased sample estimate of the disturbance variance in the classical case is given by

$$\sigma_u^2 = \hat{U}^1\hat{U}/(T-k-1) \qquad (5.1.24)$$

If scalar algebra this is expressed as

$$\hat{\sigma}_u^2 = \sum_t \hat{u}_t^2 \Big/ (T-k-1) \qquad (5.1.25)$$

When the disturbances are autocorrelated, the expectation of $\hat{U}^1\hat{U}$ is no longer σ_u^2 tr M, but rather

$$E\left(\hat{U}^1\hat{U}\right) = \text{tr}\,(MUU^1) \quad \text{by equation (5.1.23)}$$

$$= \sigma_u^2 \text{tr}\,(M\Omega) \quad \text{by equation (5.1.16)}$$

$$= \sigma_u^2 \text{tr}\left[\Omega - X(X^1X)^{-1}X^1\Omega\right] \qquad (5.1.26)$$

$$= \sigma_u^2 \text{tr}\,\Omega - \text{tr}\,(X^1X)^{-1}X^1\Omega X$$

$$= \sigma_u^2 T - \text{tr}\left[(X^1X)^{-1}X^1\Omega X\right] \text{ by the specification in}$$

equation (5.1.16) that Ω is TxT with is in the diagonal.

Hence the classical OLS estimator of the disturbances variance is biased to the extent that $\text{tr}\left[(X^1X)^{-1}X^1\Omega X\right]$ differs from $\text{tr}\left[(X^1X)^{-1}X^1X\right]$. Further more this bias is negative (towards zero) whenever positive autocorrelation predominates in

regressors and disturbances, which is generally the case for socioeconomic and political time series.

This result has two implications that are of interest. First the $\text{Var}(\hat{B})$ is biased not only because $(X^1X)^{-1}X^1\Omega X(X^1X)^{-1} \neq (X^1X)^{-1}$, but also because $E(\hat{\sigma}_u^2) \neq \sigma_u^2$. This means that if an equation is estimated via ordinary least squared when regressors and disturbances are (positive) autocorrelated, we will obtain a spurious underestimate of the error variance and an inflation of the R^2. The model will appear to provide a much better fit to the empirical data than is actually the case.

Nevertheless OLS regression in the presence of seriously correlated disturbances is not necessarily disastrous-especially when the functional form of a model is not in question because of well established theory, prior empirical results and so on. More problematic and more typical is the situation where the researcher analyzes many equations in the process of evaluating competing hypotheses and equally plausible alternative functional forms.

5.2. SOME MODELS FOR TIME-DEPENDENT DISTURBANCES

5.2.1. First-Order Autoregressive Process

The time-dependence model that has received the most attention in the econometric literature is the first-order autoregressive process [AR(1)]. Here each disturbance u_t depends only on its own previous value (the Markov Property) and a random, "white noise" component. The basic model is as follows:

$$Y_t = \beta_0 + \sum_k \beta_k X_{tk} + \mu_t \qquad \text{(equation to be estimated)} \qquad (5.2.1)$$

$$u_t = \theta_1 u_{t-1} + \upsilon_1 \quad \text{(disturbance time-dependence process)} \qquad (5.2.2)$$

$$-1 < \phi_1 < 1 \quad \text{(Stationary Condition)} \qquad (5.2.3)$$

$$\left.\begin{aligned} &E(u_t) = E(u_t) = E(u_{t-\theta}\, \upsilon_t) = 0 \qquad \theta > 0 \\ &E\left(\upsilon_t \upsilon_t \pm \theta\right) = \sigma_u^2 \qquad\qquad \theta = 0 \\ &\qquad\qquad = 0 \qquad\qquad\qquad \theta \neq 0 \end{aligned}\right\} \qquad (5.2.4)$$

The variance of u_t in the AR(T) process is derived as follows:

$$\sigma_u^2 = E(u_t^2)$$

$$= E\left[\left(\phi_1 u_{t-1} + \upsilon_t\right)\left(\phi_1 u_{t-1} + \upsilon_t\right)\right] \quad (5.2.5)$$

$$= \phi_1^2\, E(u_{t-1}^2) + 2\phi_1 E(u_{t-1}\, \upsilon_t) + E(\upsilon_t^2) \quad (5.2.6)$$

By assumption (5.2.1) that u_t follow a stationary stochastic process and therefore have identical variance for all t, and by assumption (5.2.4) that the lagged u_t are independent of the random

υ_t components, the result in equation (5.2.5) becomes

$$\sigma_u^2 = \phi_1^2\, \sigma_u^2 + \sigma_\upsilon^2$$

$$\sigma_u^2 - \phi_1^2 \sigma_u^2 = \sigma_\upsilon^2$$

$$\sigma_u^2\,(1-\phi_u^2) = \sigma_\upsilon^2 \quad (5.2.7)$$

$$\sigma_u^2 = \sigma_\upsilon^2 / (1-\phi_1^2) = \gamma_0$$

Where γ_0 denotes the variance u. The autocovariance and autocorrelation function for the AR(1) model (denoted γ_θ and ρ_θ respectively) are similarly generated. Autocovariances are derived by multiplying through the expression for u_t (equation (5.2.2) by $\text{lag}\, u_{t-\theta}$ and taking expectations. Hence the autocovariance of $u_t$and u_{t-1} is obtained.

$$\gamma_1 = E\left[u_{t-1}\,(\phi_1 u_{t-1} + u_t)\right]$$

$$\gamma_1 = \phi_1 E(u_{t-1}^2) + E(u_{t-1}\, \upsilon_t) \quad (5.2.8)$$

$$= \phi_1\, \sigma_u^2 \text{ by equations (5.2.3) and (5.2.4)}$$

Parallel operations give the autocovariance of u_t and u_{t-2}.

$$\gamma_2 = E\left[u_{t-2}\,(\phi_1 u_{t-1} + \upsilon_t)\right]$$

$$= \phi_1 E\left(u_{t-2}\, u_{t-1}\right) + E\left(u_{t-2}\, \upsilon_t\right)$$

$$= \phi_1 E\left(u_{t-2}\, u_{t-1}\right) \text{ by equation (5.2.4)} \quad (5.2.9)$$

$= \phi_1 \gamma_1$ by equation (5.2.3)

$= \phi_1^2 \sigma_u^2$ by equation (5.2.7)

Successive operations of the same sort would show the general autocovariance function for a first-order autoregressive model to be

$$\gamma_\theta = \phi_1^{|\theta|} \sigma_u^2 \qquad (5.2.10)$$

$$= \phi_1^{|\theta|} \gamma_u$$

It is now easy to desire the general autocorrelation function for AR(1) process. The autocorrelation of u_t and $u_{t\pm\theta}$ is conventionally defined,

$$\rho_\theta = \frac{\text{Cov}\,(u_{t\pm\theta})}{\sqrt{V(u_t)}\,\sqrt{V(u_{t\pm\theta})}} = \rho_{-\theta} \qquad (5.2.11)$$

$$= \frac{r_\theta}{\sigma_u^2}$$

$$= \frac{r_\theta}{r_u} \text{ by equation (5.2.3)} \qquad (5.2.12)$$

It is now directly follows from equation (5.2.11) that the lag autocorrelation function of the u_t is

$$\rho_\theta = \phi_1^{|\theta|} \qquad (5.2.13)$$

Equation (5.2.12) is an important because it describes the autocorrelation function when the disturbance is infact generated by a first-order autoregressive mechanism.

5.2.3. Second and Higher Order Autoregressive Processes

Although AR(1) process have received the most attention in the literature, there is no reasons to expect a priori that autocorrelation of disturbances in time-series regression models will be generated by such simple mechanism. It is probably accurate to say that autoregressive processes of order higher than two are relatively uncommon-unless the data have cyclical or seasonal variability, in which case appropriate dummy variables should appear in the model.

Consider a regression model where the disturbance follows a second-order autoregressive scheme [AR(2)] such that u_t depends on u_{t-1}, u_{t-2} and a random perturbations,

$$y_t = \beta_0 + \sum_k \beta_k X_{tk} + u_k \qquad \text{(equation to be estimated)} \qquad (5.2.14)$$

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + v_t \quad \text{(disturbance time-dependence-process)} \qquad (5.2.15)$$

$$\left.\begin{array}{l} \phi_2 + \phi_1 < 1 \\ \phi_2 - \phi_1 < 1 \\ -< \phi_2 < 1 \end{array}\right\} \quad \text{(stationary conditions)} \qquad (5.2.16)$$

$$\left.\begin{array}{ll} E(u_t) = E(\upsilon_t) = E(u_{t-\theta}\ \upsilon_t) = 0 & \theta > 0 \\ E(\upsilon_t\ \upsilon_{t\pm\theta}) = \sigma_u^2 & \theta = 0 \\ \quad = 0 & \theta \neq 0 \end{array}\right\} \quad \text{for all t} \qquad (5.2.17)$$

The variance of u_t in the AR (2) model is derived as follows.

$$\sigma_u^2 = E(u_t)^2$$

$$= E\left[u_t + \left(\phi_1 u_{t-1} + \phi_2 u_{t-2} + \upsilon_t\right)\right] \text{ by equation (5.2.12)} \qquad (5.2.18)$$

$\phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_u^2$ (Since the only part of u_t correlated with u_t is the current perturtation υ_t)

Dividing equation (5.2.18) by σ_u^2 allows the variance of the process to be expressed in terms of ϕs, ρs and σ_υ^2,

$$\frac{\sigma^2}{\sigma_\upsilon^2} = \phi_1\rho_1 + \phi_2\rho_2 + \sigma_\upsilon^2/\sigma_u^2 \qquad \text{by equation (5.2.10)}$$

$$1 - \phi_1\rho_1 - \phi_2\rho_2 = \sigma_\upsilon^2/\sigma_u^2 \qquad (5.2.19)$$

$$\sigma_u^2 = \sigma_\upsilon^2 / (1 - \rho_1\rho_1 - \phi_2\rho_2) = \gamma_0$$

The autocovariance and autocorrelation functions are generated analogously by multiplying the equation for u_t by $u_{t\pm\theta}$ and taking expectations. The autocovariance of u_t and u_{t-1} is

$$\gamma_1 = E\left[u_{t-1}\left(\phi_1 u_{t-1} + \phi_2 u_{t-2} + \upsilon_t\right)\right]$$

$$= \phi_1 E(u_{t-1}^2) + \phi_2 E(u_{t-1}\, u_{t-2}) + E(u_{t-1}\, \upsilon_t) \qquad (5.2.20)$$

$$= \phi_1\sigma_u^2 + \phi_2\gamma_1 \text{ by equations (5.2.16) and (5.2.17)}$$

Similarly the autocovariance of u_t and u_{t-2} is derived θ,

$$\gamma_2 = E\left[u_{t-2}\left(\phi_1 u_{t-1} + \phi_2 u_{t-2} + \upsilon_t\right)\right]$$

$$= \phi_1 E(u_{t-2}\, u_{t-1}) + \phi_2 E(u_{t-2}^2) + E(u_{t-2}\, \upsilon_1)$$

$$= \phi_1\gamma_1 + \phi_2\sigma_u^2 \text{ by equations (5.2.16) and (5.2.17)}$$

Recall that γ_0 denotes σ_u^2, successive operations would shows the autocovariance function of the second-order autoregressive process to be

$$\gamma_\theta = \phi\gamma_{\theta-1} + \theta_2\gamma_{\theta-2} \qquad \theta > 0 \qquad (5.2.21)$$

The autocorrelation function for this model follows straight forwardly ρ_θ is defined as follows,

$$\rho_\theta = \left(\phi\gamma_{\theta-1} + \phi_2\gamma_{\theta-2}\right)/\gamma_0$$

$$\rho_\theta = \phi_1\rho_{\theta-1} + \phi_2\rho_{\theta-2} \quad \theta > 0 \qquad (5.2.22)$$

Our interest is primarily on the autcorrelation function because it specifies the behaviour of the ρ_θ when the disturbances in a particular equation do follows an AR(2) process.

Autoregressive processes of order greater than two are likely to be less common empirically. The essentials involve minor extensions of previous results. Here the disturbance is generated by a ρ^{th} order autoregressive scheme,

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + ... + \phi_\rho u_{t-\rho} + \upsilon_t \qquad (5.2.23)$$

The variance of u_t is easily shown to be

$$\sigma_u^2 = \frac{\sigma_\upsilon^2}{1 - \rho_1\phi_1 - \rho_2\phi_2 ..... \rho_p\phi_b} \qquad (5.2.24)$$

The autocovariance functions is

$$\gamma_\theta = \phi_1\gamma_{\theta-1} + \phi_2\gamma_{\theta-2} + ... + \phi_p\gamma_{\theta-\rho} \qquad \theta > 0 \qquad (5.2.25)$$

Finally the autocorrelation function of a p^{th} order autoregressive model is given by

$$\rho_\theta = \phi_1\rho_{\theta-1} + \phi_2\rho_{\theta-2} + ... + \phi_p\rho_{\theta-p} \qquad \theta > 0 \qquad (5.2.26)$$

These general expressions and particulars equation (5.2.26) can be used to deduce the empirical behaviour of disturbances generated by an autoregressive process of any order.

5.2.4. First-Order Moving Average Processes

An alternative class of models for time-dependent disturbances is provided by moving average processes. In contrast to autoregressive models, disturbances generated by moving average schemes depend only on a moving linear combination of random variables υ_t with coefficients $(1, -\phi_1, ..., -\phi_p)$. Hence a random shock υ_t enters the system at time t and disturbs the equilibrium level of u_t for (p+1) periods before dissipation. The autocorrelation functions and correlograms produced by moving-average. Processes differ sharply from those of autoregressive models.

In the first-orders moving-overage process [MA(1)] we have

$$Y_t = \beta_0 + \sum_k \beta_k X_{tk} + v_t \qquad \text{(equation to be estimated)} \qquad (5.2.27)$$

$$u_t = \upsilon_t - \phi_1 u_{t-1} \qquad \text{(disturbance time-dependence)} \qquad (5.2.28)$$

$$-1 < \phi_1 < 1 \qquad \text{(invertibility condition)} \qquad (5.2.29)$$

$$\left.\begin{aligned} &E(u_t) = E(\upsilon_t) = E(u_{t-\theta}\ \upsilon_t) = 0 \\ &E(\upsilon_t\ \upsilon_t \pm \theta) = \sigma_\upsilon^2 \qquad \theta > 0 \\ &\qquad\qquad = 0 \qquad \theta \neq 0 \end{aligned}\right\} \quad \text{for all t} \qquad (5.2.30)$$

The variance of the first-orders moving –average model may be derived as follows,

$$\sigma_u^2 = E(u_t^2)$$

$$= E\left[\left(\upsilon_t - \phi_1\upsilon_{t-1}\right)\left(\upsilon_t - \phi_1\upsilon_{t-1}\right)\right]$$

$$= E(\upsilon_t^2) - 2\phi_1 E\left[\left(\upsilon_{t-1}\,\upsilon_t\right) + \phi_1^2\, E\left(\upsilon_{t-1}^2\right)\right] \qquad (5.2.31)$$

$$= \sigma_\upsilon^2 + \phi_1^2\,\sigma_\upsilon^2 \qquad (5.2.32)$$

$$\sigma_u^2 = \sigma_\upsilon^2\ (1+\phi_1^2) = \gamma_0 \qquad (5.2.33)$$

The MA(1) autocovariance function is similarly generated. Thus the autocovariance of u_t and u_{t-1} is

$$\gamma_1 = E(u_t\ u_{t-1})$$

$$= E\left[\left(\upsilon_t - \phi_1\phi_{t-1}\right)\left(\upsilon_{t-1} - \phi_1\upsilon_{t-2}\ \right)\right]$$

$$= E(\upsilon + \upsilon_{t-1}) - \phi_1\ E(\upsilon + \upsilon_{t-2}) - \phi_1\ E(\upsilon_{t-1}^2) + \phi_1^2\ E(\upsilon_{t-1}\,\upsilon_{t-2}) \qquad (5.2.34)$$

$$= -\phi_1\ \sigma_u^2 \qquad \text{equation (5.2.32)}$$

Parallel operations yield autocovariances of greater lag.

For example, γ_2 is obtained,

$$\gamma_2 = E(u_t\ u_{t-2})$$

$$= E\left[\left(\upsilon_t - \phi_1\upsilon_{t-1}\right)\ \left(\upsilon_{t-2} - \phi_1\upsilon_{t-2}\right)\right]$$

$$= E(\upsilon_t\ \upsilon_{t-2}) - \phi_1 E(\upsilon_t\ \upsilon_{t-3}) - \phi_1 E(\upsilon_{t-1}\,\upsilon_{t-2}) + \phi_1^2 E(\upsilon_{t-1}\,\upsilon_{t-3}) \qquad (5.2.35)$$

= 0 by equation (5.2.30)

It should now be apparent that in first-orders moving-average processed all autocovariances beyond lag 1 are zero. Thus the generals autocovariance function for the MA(1) model is written,

$$\left.\begin{aligned}\gamma_\theta &= -\phi,\ \sigma_u^2 \quad \theta = \pm 1\\ &= 0 \qquad |\theta| > 1\end{aligned}\right\} \qquad (5.2.36)$$

Obviously, the autocorrelation function of the MA(1) process, defined by $\rho_\theta = \gamma_\theta / \gamma_0$, shares the same property.

Thus,

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-\phi_1 \sigma_u^2}{(1+\phi_1^2)\,\sigma_u^2} = \frac{-\phi_1}{(1+\phi_1^2)} \tag{5.2.37}$$

and successive lag ρ_θ are zero. Therefore the general autocorrelation function of the MA (1) process, defined by $\rho_\theta = \frac{\gamma_\theta}{\gamma_0}$, shares the same property. Thus

$$\rho_1 = \frac{\gamma_1}{\gamma_0}$$

$$= \frac{-\phi_1 \sigma_\upsilon^2}{(1+\phi_1^2)\,\sigma_\upsilon^2}$$

$$= \frac{-\phi_1}{(1+\phi_1^2)}$$

and successive log ρ_θ are zero. Therefore the general autocorrelation function is expressed

$$\left.\begin{aligned} \rho_\theta &= \frac{-\phi_1}{1+\phi_1^2} \qquad \theta = \pm 1 \\ &= 0 \qquad |\theta| > 1 \end{aligned}\right\} \tag{5.2.38}$$

The results in equations (5.2.37) and (5.2.38) establish that the autocovariances and autocorrelations of the first-order moving-average process have a cutoff after log 1.

Second and Higher-Order Moving-Average Processes

The essential results of Second and Higher-Order Moving-average process follows from those developed above consider a regression model where the disturbance follows a Second-Order Moving Average Scheme MA(2),

$$Y_t = \beta_0 + \sum_k \beta_k X_{tk} + u_t \text{ (equation to be estimated)} \tag{5.2.39}$$

$$u_t = \upsilon_t - \phi_1 \upsilon_{t-1} - \phi_2 \upsilon_{t-2} \text{ (disturbance time dependence process)} \tag{5.2.40}$$

$$\left.\begin{aligned}&\phi_2+\phi_1<1\\&\phi_2-\phi_1<1\\&-1<\phi_2<1\end{aligned}\right\}\text{(invertibility conditions)} \qquad (5.2.41)$$

$$\left.\begin{aligned}&E(u_t)=u_{t-\theta}\,\upsilon_t=0\\&E(u_t\,\upsilon_t\pm\theta)=\sigma_\upsilon^2 \quad \theta>0\\&\qquad\qquad =0 \quad \theta\neq 0\end{aligned}\right\} \qquad (5.2.42)$$

The variance of the MA(2) model is

$$\sigma_u^2 = E(u_t^2)$$

$$= E\left[\left(\upsilon_t-\phi_1\,\upsilon_{t-1}-\phi_2\,\upsilon_{t-2}\right)\left(\upsilon_t-\phi_1\,\upsilon_{t-1}-\phi_2\,\upsilon_{t-2}\right)\right]$$

$$= E(\upsilon_t^2)-2\phi_1\,E(\upsilon_t\,\upsilon_{t-1})-2\phi_2\,E(u+\upsilon_{t-2})$$

$$+2\phi_1\phi_2\,E(\upsilon_{t-1}\,\upsilon_{t-2})+\phi_1^2\,E(\upsilon_{t-1}^2)+\phi_2^2\,E(\upsilon_{t-2}^2) \qquad (5.2.43)$$

$$\sigma_u^2=\sigma_\upsilon^2(1+\phi_1^2+\phi_2^2)$$
$$=\gamma_0$$

The autocovariance function is derived analogously,

$$\gamma_1 = E(u_t\,u_{t-1})$$

$$= E\left[\left(\upsilon_t-\phi_1\upsilon_{t-1}-\phi_2\upsilon_{t-2}\right)\left(\upsilon_{t-1}-\phi_1\upsilon_{t-2}-\phi_2\upsilon_{t-3}\right)\right]$$

$$= E(\upsilon_t\upsilon_{t-1})-\phi_1E(\upsilon_t\upsilon_{t-2})-\phi_2E(\upsilon_t\upsilon_{t-3})-\phi_1E(\upsilon_{t-1}^2)+\phi_1^2E(\upsilon_{t-1}\,\upsilon_{t-2})$$

$$+\phi_1\phi_2\,E(\upsilon_{t-1}\,\upsilon_{t-3})-\phi_2E(\upsilon_{t-2}\,\upsilon_{t-1})+\phi_1\phi_2\,E(\upsilon_{t-2}^2)+\phi_2^2E(\upsilon_{t-2}\,\upsilon_{t-3}) \qquad (5.2.44)$$

$$=(-\phi_1+\phi_1\phi_2)\,\sigma_\upsilon^2 \quad \text{by equation (5.2.42)}$$

$$\gamma_2 = E(u_t\,u_{t-2})$$

$$= E\left[\left(\upsilon_t - \phi_1\upsilon_{t-1} - \phi_2\upsilon_{t-2}\right)\left(\upsilon_{t-2} - \phi_1\upsilon_{t-3} - \phi_2\upsilon_{t-4}\right)\right]$$

$$= E(\upsilon_t\upsilon_{t-2}) - \phi_1 E(\upsilon_t\upsilon_{t-3}) - \phi_2 E(\upsilon_t\upsilon_{t-4}) - \phi_1 E(\upsilon_{t-1}\upsilon_{t-2}) + \phi_1^2 E(\upsilon_{t-1}\upsilon_{t-3})$$

$$+ \phi_1\phi_2 E(\upsilon_{t-1}\upsilon_{t-4}) - \phi_2 E(\upsilon_{t-2}^2) + \phi_2\phi_1 E(\upsilon_{t-2}\upsilon_{t-3}) + \phi_2^2 E(\upsilon_{t-2}\upsilon_{t-4})$$

(5.2.45)

$= \phi_2\ \sigma_u^2$ by equation (52.2.42)

$\gamma_3 = E(u_t\ u_{t-3})$

$$= E\left[\left(\upsilon_t - \phi_1\upsilon_{t-1} - \phi_2\upsilon_{t-2}\right)\left(\upsilon_{t-3} - \phi_1\upsilon_{t-4} - \phi_2\upsilon_{t-5}\right)\right]$$

$$= E(\upsilon_t\upsilon_{t-3}) - \phi_1 E(\upsilon_t\upsilon_{t-4}) - \phi_2 E(\upsilon_t\upsilon_{t-5}) - \phi_1 E(\upsilon_{t-1}\upsilon_{t-3}) + \phi_1^2 E(\upsilon_{t-1}\upsilon_{t-4})$$

$$+ \phi_1\phi_2 E(\upsilon_{t-1}\upsilon_{t-5}) - \phi_2 E(\upsilon_{t-2}\upsilon_{t-3}) + \phi_2\phi_1 E(\upsilon_{t-2}\upsilon_{t-4}) + \phi_2^2 E(\upsilon_{t-2}\upsilon_{t-5})$$

(5.2.45)

$= \phi_3\ \sigma_u^2$ by equation (52.2.42)

In generalized notation, the autocovariance function for the second order moving-average process is therefore,

$$\gamma_\theta = (-\phi_\theta + \phi_1\ \phi_{\theta+1})\ \sigma_u^2 \quad \theta = \pm 1, 2$$

$$= 0 \qquad \theta > \pm 2 \qquad (5.2.47)$$

Since $\rho_\theta = \dfrac{\gamma_\theta}{\gamma_0}$, the general autocorrelation function for the MA (2) mechanism follows directly,

$$\rho_\theta = \frac{\left(1 - \phi_\theta + \phi_1\ \phi_{\theta+1}\right)\sigma_\upsilon^2}{\left(16\phi_1^2 + \phi_2^2\right)\sigma_\upsilon^2} \quad \theta = \pm 1, 2$$

$$\rho_\theta = \frac{-\phi_\theta + \phi_1\ \phi_{\theta+1}}{1 + \phi_1^2 + \phi_2^2} \qquad \theta = \pm 1, 2$$

$$= 0 \qquad |\theta| > 2$$

Hence in the second-order moving-average process the autocovariance and autocorrelation functions are zero after lag 2. Moving-Average models of higher order [MA(p1)] have analogous properties, that is, autocovariances and autocorrelations exhibit a cutoff beyond lag p.

In the p^{th} order moving-average process we have

$$u_t = \upsilon_t - \phi_1 \upsilon_{t-1} - \phi_2 \upsilon_{t-2} - ..... - \phi_p \upsilon_{t-p} \tag{5.2.48}$$

The u_t in this model have variance

$$\sigma_u^2 = \sigma_u^2 (1 + \phi_1^2 + \phi_2^2 + ... + \phi_p^2) = \gamma_0 \tag{5.2.49}$$

and covariance

$$\gamma_\theta = \left(-\phi_\theta + \phi_1\ \phi_{\theta+1} + +\phi_2\ \phi_{\theta+2} + ... + \phi_{p-\upsilon}.\phi_p\right) \sigma_\upsilon^2 \quad \theta = \pm 1, 2, ... p \tag{5.2.50}$$

$$= 0 \qquad\qquad |\theta| > p$$

Finally, and most importantly, the autocorrelation function of the MA(p) process is

$$\rho_\theta = \frac{-\phi_\theta + \phi_1\phi_{\theta+1} + ... + \phi_{p-\theta}\ \phi_p}{1 + \phi_1^2 + ... + \phi_p^2} \quad \theta = \pm 1, 2, ... p$$

$$= 0 \qquad\qquad |\theta| > p \tag{5.2.51}$$

Like higher-orders autoregressive models, moving average processes of order greater than two are likely to be rare in practice.

5.3.2. Estimation of an ARIMA Model

The parameters $\phi_j, \theta_j, \sigma^2$ are generally estimated by the maximum likelihood method or a least squares technique. The methods are applied assuming the orders p_j, d_j, q_j as fixed. ARIMA (p, d, q) model is

$$\phi(L)\, \Delta^d x_i = \theta^* + \theta(L) \in_t \tag{5.3.2}$$

Where $\in_t$ t_t 's are nonzero, uncorrelated variables with the same variance σ^2. The roots of the polynomials ϕ and θ have modulus greater than 1. The process

$\omega_t = \Delta^d x_t$ is then an ARMA (p, q) process such as

$$\phi(L)\omega_t = \theta^* + \theta(L)\omega_t \qquad (5.3.3)$$

If we have x_1, ……. x_{T0} as available observations of the process $\{x_t\}$ we can compute the d^{th} order differences

$$\omega_{d+1} = \Delta^d x_{d+1}$$
$$\bullet \qquad \bullet$$
$$\bullet \qquad \bullet$$
$$\bullet \qquad \bullet$$
$$\bullet \qquad \bullet$$
$$w_{T0} = \Delta^d x_{T0}$$

We may suppress the constant θ^* so long as we replace ω_t by $\hat{\omega}_t = \omega_t - \mu$ with $\mu = \theta^* / (1 + \phi_1 + ... + \phi_p)$. We suppose $\theta^* = 0,$ but results can be extended to the case $\theta^* \neq 0$ by replacing ω_t by $\hat{\omega}_t$. The additional parameter μ can be estimated in the same way as the other. Where there is a constant θ^*, another frequently used method to suppress it is replace the process $\{\omega\}$ with the centered process $\{\omega_t - \bar{\omega}\}$ and to estimate the other parameters afterwards. Finally, we have to solve the estimation problem of the parameters of an ARMA (p, q) without a constant, of the parameters $\phi_1, ..., \phi_p, \theta_1 .....\theta_q, \sigma^2,$ from the information contained in the observations denoted by $Z_{1, ...}Z_T$, $T=T_0$-d. We assume that $\in_t$ are normal in order to derive the asymptotic properties of the estimators more easily.

5.3.3. The case of ARMA (o, q) ≡ MA(q)

Likelihood or joint density of the model for the observations z_1, z_2 ...., z_T. We have

$z_T = \theta(L) \in_t, t = 1, ..., T$

$$\text{or } z_t = \in_t + \sum_{i=1}^{q} \theta_i \in_{t-i} \qquad (5.3.4)$$

Here $\in_t$ are assumed normal. The vector $z^1 = (z_1, ..., z_T)$ can be expressed as a linear transformation of the vector $\in^1 = \left(\in_{1-qi} ...., \in_T\right)$, therefore it is normally distributed. We write a matrix $M(\theta)$ such that $z = M(\theta)_t$ and derive from this the density of z. This will be the density of the normal distribution $N\left(0, \sigma^2 M(\theta)\ M(\theta)^1\right)$. But, this method is not the most suitable to solve the likelihood maximization problem, since it implies the computation of the matrix $\left(M(\theta)\ M(\theta^1)\right)^{-1}$, the dimension of which is T. Let us start from the system

$$\in_{1-q} = \in_{1-q}$$

.....

$$\in\text{-}1 = \in\text{-}1$$

$$\in_1 = z_1 - \theta_1 \in_0 ..... - \theta_q \in_{1-q} \qquad (5.3.5)$$

$$\in_2 = z_2 - \theta_1 \in_1 ..... - \theta_q t_{2-q}$$

.....

$$\in_T = z_t - \theta_1 \in_{T-1} - ..... - \theta_q \in_{T-q}$$

Let us imagine that we substitute $\in_1$ in the expressions for $\in_2$ with a function of $t_{0,} ....., t_{1-q}$ then $\in_1$ and $\in_2$ in the expression for $\in_3$ with a function of $t_{0,} ....., t_{1-q}$ and so on. The we get

$$\in = NZ + x \in_x \qquad (5.3.6)$$

Here N is a $(T+q)\times T$ matrix equal to

$$\begin{pmatrix} U \\ A_1(\theta) \end{pmatrix}$$ and

$$\in_* = \left(\in_{1-q}, ..., \in_0\right),$$

x is a $(T+q)\times q$ matrix of the type $\begin{pmatrix} I_q \\ A_2(\theta) \end{pmatrix}$,

$A_1(\theta)$ is a TxT lower triangular matrix with unit elements along the main diagonal and $A_2(\theta)$ is a Txq matrix.

Equation (5.3.6) can also be written as

$$\in = (X\ N)\begin{pmatrix} \in_x \\ z \end{pmatrix} \tag{5.3.7}$$

The square matrix (x N) is lower triangular with unit elements along the main diagonal. Hence its determinant is equal to one. We get the density of the random vector $\left(\in_*^1 z^1\right)^1$, by replacing the vector $\in$ by $Nz+ \in_*^t$ in density of $\in$.

We obtain

$$\frac{1}{(2\pi\sigma^2)^{(T+q)/2}} \exp\left[\frac{-1}{2\sigma^2}(Nz + x\in_*)^1 (Nz + x\in_*)\right] \tag{5.3.8}$$

We compute the marginal distribution of z from the distribution of $\begin{pmatrix} c^* \\ z \end{pmatrix}$. The orthogonal projection of Nz on the subspace of R^{T+q} spanned by the columns of x is given by $X\in_x$.

Here $\in_* = -(X^1X)^{-1}X^1Nz$

By Pythagoras's theorem we write (5.3.8) in the form

$$\frac{1}{(2\pi\sigma^2)^{(T+q)/2}} \exp\left[\frac{-1}{2\sigma^2}(Nz + X\in_*)^1 (Nz + X\in_*) + \left(\in_* - \bar{\in}_*\right)^1 X^1X\left(\in_* - \bar{\in}_*\right)\right] \tag{5.3.9}$$

We derive the density of z by integrating with respect to $\in_*$.

We have

$$L_T = \frac{1}{\left(2\pi\sigma^2\right)^{T/2}} \left(\det\,(X^1X)\right)^{-1/2} \exp\left[\frac{-1}{2\sigma^2}(Nz+\bar{\in}_X)^1(Nz+X\,\bar{\in}_X)\right] \qquad (5.3.10)$$

Let us denote $S(\theta) = (Nz + X\bar{t}_*)^1\,(Nz + X\bar{t}_*)$

We get

$$\text{In}\; L_T = \frac{-T}{2}\text{In}2\pi - \frac{T}{2}\text{In}\sigma^2 - \frac{1}{2}\text{In}\left(\det\,(x^1x)\right) - \frac{S(\theta)}{2\sigma^2}$$

This log likelihood can be maximized with respect to the parameters θ and σ^2 to determine the maximum likelihood estimators.

Finding partial derivative of $\text{In}\; L_T$ with respect to σ^2, we get

$$\frac{\partial\,\text{In}\; L_T}{\partial\sigma^2} = -\frac{T}{2\sigma^2} + \frac{-S(\theta)}{2\sigma^4}$$

From which we get

$$\sigma^2 = \frac{S(\theta)}{T} \qquad (5.3.11)$$

Substituting it in the log-likelihood, we get

$$\text{In}\; L_T^* = \frac{-T}{2}\text{In}\; 2\pi - \frac{T}{2}\text{In}\;\; \frac{S(\theta)}{T} - \frac{1}{2}\text{In}\;\left(\det\,(x^1x)\right) - \frac{T}{2}$$

Finally, the function to be minimized is

$$I_T^* = T\;\text{In}\; S(\theta) + \text{In}\left(\det\,(x^1x)\right) \qquad (5.3.12)$$

Here there are two kinds of methods to find an estimator of θ.

1. The exact methods aimed at minimizing l_T^* directly, using numerical methods.
2. The least squares methods aimed at minimizing $S(\theta)$, on the basis that the second term of l_T^* becomes negligible with respect to the first as T increases. They allow the use of simple algorithms to compute the value of the objective function. These algorithms are generally linked to recursive

methods of forecast computations. One of such procedures, proposed by Box and Jenkins is back forecasting.

Equation (5.3.9) shows that $\bar{\in}_*$ is conditional expectation of $\in_*$ given z. Denoting by $\bar{\in}$ the conditional expectation of $\in$ given z. We use relationship (5.3.6) to get

$$\bar{\in} = Nz + x\,\bar{\in}_* \text{ and}$$

$S(\theta)$ can be written as

$$S(\theta) = \left\|\bar{\in}\right\|^2 = \sum_{i=1-q}^{T} \bar{\in}_i^{-2} \qquad (5.3.13)$$

This interpretation of $S(\theta)$ allows the deviation of a simple computation rule for a given θ. We first determine $\bar{\in}_*$ by back forecasting. We have

$$\bar{\in}_{1-q} = \bar{z}_{1-q} - \theta_1\,\bar{\in}_{1-q-i} - \ldots\ldots\ldots - \theta_q\,\bar{\in}_{1-2q}$$

That is

$$\bar{\in}_{1-q} = \bar{z}_{1-q},$$

Since $\bar{\in}_{1-q-i} = 0, i = 1, \ldots, q$

We have $\bar{\in}_{2-q} = \bar{z}_{2-q} - \theta_1\,\bar{\in}_{1-q}$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\bar{\in}_0 = \bar{z}_0 - \theta_1\,\bar{\in}_1 \ldots\ldots\theta_{q-1}\,\bar{\in}_{1-q}$$

Thus, $\bar{\in}_*$ is known when the forecasts $\bar{z}_i = E(z_i / z), i = 1-q, \ldots, 0$ are known. The computation of $\bar{z}_i$ is easily done by using the forward representation of the process z.

$$z_t = \theta_{(F)}\eta_t$$

$$\bar{z}_0 = -\sum_{i=1}^{+q} \pi_i z_i \approx -\sum_{i=1}^{T} \pi_i z_i \qquad (5.3.14)$$

Here the π's are the same as the ones used in the forward forecast, that is the coefficient of the long division of 1 by θ. The $\bar{\epsilon}_i, i = 1,....,T$ are computed recursively using the expressions (5.3.5). This fast computations of $S(\theta)$ translates into the possibility of using numerical optimization algorithms.

5.3.4. ARMA (p, q) Estimation

The least squares methods can be generalized to the case of an ARMA (p, q) extending the case of an MA(q) process described above. The basic intuition behind the treatment of an ARMA (p, q) is that is that it can be written in a moving average form

$$\sum_{i=0}^{\infty} \Psi_i \in_{t-i}$$

Which can be approximated by an MA(q) process

$$\sum_{i=0}^{\infty} \Psi_i \in_{t-i}$$

Where Q is large enough. For all practical purposes we need to solve the problem of the choice of Q. we have seen that for an MA(q), the $\bar{z}_i$ are o for $i \leq -q$. In the ARMA (p, q) case we will stop the backward forecast when the $\bar{z}_i$'s become negligible. They we have a way of computing.

$$S(\phi, \theta) + \sum_{i=1-q}^{T} \bar{\epsilon}_i^2$$

for given $\phi^1 = (\phi_1, ..., \phi_p)$

$$\theta^1 = (\theta_1, ..., \theta_p)$$

which can be minimized numerically.

Initial Values

The optimization algorithms used in the various estimation methods need initial values for the parameters. The choice of the initial values is important, since, as the

one hand it reflects on the number of needed iterations to reach the optimum, on the otherhand it results in the attainment of a local or global optimum.

5.5. GENERALIZED LEAST SQUARES METHOD OF ESTIMATION

5.5.1. Introduction

The standard linear model is

$$Y = X\beta + \in \tag{5.5.1}$$

Here Y is the nx1 response vector

X is annxp model matrix

β is a px1 vector of parameters to estimate

$\in$ is an nx1 vector of errors

Let us assume that $\in \sim N_n(0, \sigma^2 I_n)$, the familiar ordinary least squares (OLS) estimator of β,

$$b_{OLS} = (X^1X)^{-1}X^1Y \tag{5.5.2}$$

with covariance matrix

$$V(b_{OLS}) = \sigma^2(X^1X)^{-1} \tag{5.5.3}$$

More generally, we can assume that $\in \sim N_n(0, \Sigma)$, where the error-covariance matrix Σ is symmetric and positive definite. Different diagonal entities in Σ correspond to non-constant error variances, while non-zero off-diagonal entries correspond to correlated errors. Suppose that Σ is known. Then, the log-likelihood for the model is

$$\log_e^{L(\beta)} = \frac{-n}{2}\log_e(2\pi) - \frac{1}{2}\log_e(\det \Sigma) - \frac{1}{2}(Y - X\beta)^1 \Sigma^{-1}(Y - X\beta) \tag{5.5.4}$$

Which is maximized by the generalized least squares (GLS) estimator of β,

$$b_{GLS} = (X^1\Sigma^{-1}X)^{-1}X^1\Sigma^{-1}Y \tag{5.5.5}$$

with covariance matrix

$$V_{(b_{GLS})} = (X^1\Sigma^{-1}X)^{-1} \qquad (5.5.6)$$

For example, when Σ is a diagonal matrix of unequal error variances, then b_{GLS} is just the weighted least squares (WLS) estimator. In a real application, the error covariance matrix Σ is not known, and must be estimated from the data along with the regression coefficients β. There are too many elements in $\Sigma - \frac{n(n+1)}{2}$ distinct elements to estimate the model without further restrictions with a suitably restrictive parameterization of Σ, the model can be estimated by maximum likelihood or another appropriate method.

5.5.2. Serially Correlated Errors

When the observations represent different moments or intervals of time, usually equally spaced, then the errors from a regression model are unlikely to be independent is in time series data. Let us assume that the process generating the regression errors is stationary. That is, all of the errors have the same expectation (assumed to be 0) and the same variance (σ^2), and the covariance of two errors depends only upon their separation 's' in time.

$$C(\in_t, \in_{t+s}) = C(\in_t, \in_{t-s}) = \sigma^2\rho_s \qquad (5.5.7)$$

Where ρ_s is the error autocorrelation at lag 's'.

Here, the error-covariance matrix has the following structure.

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_1 & \ldots\ldots\ldots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \ldots\ldots\ldots & \rho_{n-2} \\ . & . & . & \ldots\ldots\ldots & . \\ . & . & . & \ldots\ldots\ldots & . \\ . & . & . & \ldots\ldots\ldots & . \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \ldots\ldots\ldots & i \end{bmatrix} = \sigma^2\rho \qquad (5.5.8)$$

If we know the values of σ^2 and ρ's then we could apply this result to find the GLS estimator of β in a time series regression, but, these are generally unknown parameters. Further, while they are many fever than the number of elements in the unrestricted error-covariance matrix Σ, the large number (n-1) of different ρ's makes their estimation impossible without specifying additional structure for the autocorrelated errors. The first order autoregressive process AR(1).

$$\in_t = \phi \in_{t-1} + \upsilon_t \tag{5.5.9}$$

Here, the random shocks υ_t are assumed to Gaussian white noise, $N, D(0, \sigma_\upsilon^2)$. Under this model $\rho_1 = \phi, \rho_s = \phi^s$ and $\sigma^2 = \sigma_\upsilon^2/(1-\phi^2)$. As a correlation, $|\phi| < 1$, the error autocorrelations ρ_s delay exponentially towards 0 as increases. Higher order autoregressive models are a direct generalization of the first order model, the second order autoregessive model AR(2) is

$$\in_t = \phi_1 \in_{t-1} + \phi_2 \in_{t-2} + \upsilon_t \tag{5.5.10}$$

In contrast, in the first-order moving average process, MA(1), the current error, depends upon the random shock from the current and previous periods,

$$\in_t = \upsilon_t + \Psi \upsilon_{t-1} \tag{5.5.11}$$

and higher-order MA(q) process are similarly defined. AR and MA terms are combined in ARMA (p, q) processes, AR MA(1,1) errors follow the process

$$\in_t = \phi \in_{t-1} + \upsilon_t + \Psi \upsilon_{t-1} \tag{5.5.12}$$

Examining the residual autocorrelations from a preliminary OLS regression can suggest a reasonable from for the error-generating process. The lag-s residual autocorrelation is

$$\gamma_s = \frac{\sum_{t=s+1}^{n} e_t \; e_{t-s}}{\sum_{t=1}^{n} e_t^2} \tag{5.5.13}$$

If the residuals were independently distributed, the standard error of each γ_s would be approximately $1/\sqrt{n}$, a quantity that can be used as a rough guide to the statistical significance of the residual autocorrelations. A more accurate approach is to calculate the Durbin-Watson statistics,

$$D_s = \frac{\sum_{t=s+1}^{n} \left(e_t - e_{t-s}\right)^2}{\sum_{t=1}^{n} e_t^2} \qquad (5.5.14)$$

Which have a known sampling distribution that depends upon the model matrix x. When the sample size is large, $D_s \cong 2(1-\gamma_s)$ and so Durbin-Watson statistics near 2 are indicative of small residual autocorrelation, those below 2 of positive autocorrelation, and those above 2 of negative autocorrelation.