ECO 372: Introduction to Econometrics

Spring 2025

Introduction to Time Series: History and Some Concepts

Sakib Bin Amin, Ph.D.

Associate Professor in Economics

Director, Accreditation Project Team (APT)

Outline

Our objectives for this lecture will be to learn:

- Background
- A Brief History of Time-Series Modelling
- Time-Series Modelling Branches
- Uniqueness of Time-Series Data
- Patterns in Time-Series Analysis
- White Noise
- Stationarity

Background

- In the world of econometrics, time-series stands tall since the late 1920s.
- Researchers observed that apart from the smoothing techniques, large simultaneous-equation based macroeconomic models constructed in the 1960s frequently demonstrated poorer estimation/forecasting power.
- For justifiable forecasting purposes, researchers started to think about a simple modelling approach.
  - An approach that could describe the behaviour of a variable or say set of variables in terms of past values, without the benefit of a well-developed theory.
- The thought process actually led to the emergence of time-series modelling, which was pioneered by Box and Jenkins (1984, 1970).

A Brief History of Time-Series Modelling

1920s:

- First time-series modelling came to the picture.
- This approach is known as Moving Average Smoother (MAS).
- It takes the average of the past values and attempts to predict the future value given set of clearly defined boundaries.

1937-1939:

- Introduction of OLS in time-series modelling. This introduction was criticized by Yule and Walker in 1939.
- Yule and Walker argued that in time-series, we basically have data of something (i.e. a person, a firm, a country, etc.) over a certain period of time.
- Therefore, it violates some of the core assumptions of traditional OLS.

A Brief History of Time-Series Modelling

1940s:

- Addressing the OLS issue, Yule and Walker (1940) proposed Autoregressive (AR) modelling technique.
- Slutsky introduced past shock time-series models, which are known as Moving Average (MA) modelling technique.
- After that Wald proposed combination of AR and MA modelling technique known as Autoregressive Moving Average (ARMA) modelling technique.

1970s:

- Time-series discipline observed major development.
- Box and Jenkins proposed a clear cut method for estimating class of ARMA models.
- Box and Jenkins also provided guidelines for diagnostics tests.

A Brief History of Time-Series Modelling

1980s:

- Vector Autoregressive (VAR) models for analysing system of equations and conducting forecast simulations were developed and optimised.
- There was breakthrough in 1987. Engle and Granger proposed the idea of cointegration which revolutionised the idea of time-series thinking process.
- The earlier developments of unit root tests were done in 1980s.

1990s:

- Major developments in different modelling techniques, such as Autoregressive Conditional Heteroskedasticity (ARCH), Generalised Autoregressive Conditional Heteroskedasticity (GARCH), Autoregressive Distributed Lag (ARDL), Dynamic OLS, regime switching Regression, etc.

2000s:

- Mostly theoretical level development along with some applications.

Time-Series Modelling Branches

Spectral Analysis:

- This is also known as frequency domain approach.
- Spectral analysis is a technique that allows us to discover underlying periodicities.
- Like to find an answer for a question: What is the economic cycle through periods of expansion and recession?
- A classic technique is the wavelet analysis.

Time Domain Analysis:

- In this branch, one models data by using the notion of dependence of past and present values.
- For example modelling GDP series of a country. A classic technique is the Box-Jenkins analysis.

Note: We will only focus on time domain branch for the rest of the journey.

Time-Series Modelling Branches

Within the Branches, we have two types of modelling tecniques. These are:

Univariate Modelling: A time series modelling approach that analyses single variable.

Multivariate Modelling: A time series modelling approach that analyses multiple variables.

Uniqueness of Time-Series Data

- Let us focus our attention towards two simple questions:
  - How time-series data is different from cross-section data?
  - Why OLS should not be used without having any prior information given the data series?
- By definition, cross-sectional data is random in nature.
- That is, cross-sectional variables are Independent and Identically Distributed (i.i.d.).
  - For example, households are different from each other.
- However, in time-series, we do not have i.i.d. data.
- Let us take the GDP of Bangladesh over the last 50 years. Will the year to year values of GDP be random?
- Or, will the probability of getting a value of particular year be higher and not linked with the past value?

Uniqueness of Time-Series Data

- If we think naturally, at a certain point GDP will be around the past values.
- So, there is a connection between year to year GDP values as well as there will be certainly high probability that GDP value at a particular year depends on past values.
- Thus, time-series data is not i.i.d.
- Non-technically, you can also argue that the GDP of a year say 2020 cannot just become 100 trillion USD where the value was about 350 billion USD in 2019.
- When data is not i.i.d, OLS will not be effective for estimating unknown parameters.

Uniqueness of Time-Series Data

- Let us again recall assumptions of Classical Linear Regression Model (CLRM).

$$E(\varepsilon \mid x) = 0 \rightarrow \text{No endogeneity}$$

$$Var(\varepsilon \mid x) = 0 \rightarrow \text{Constant variance (homoskedastic)}$$

$$Cov(\varepsilon_j, \varepsilon_s) = 0; j \neq s \rightarrow \text{No correlation in errors}$$

- Think of a time-series model based on the previous discussion

$$GDP_t = a_0 + a_1 GDP_{t-1} + \varepsilon_t$$

- If i.i.d. nature is violated, then do the $\{\varepsilon_t, \varepsilon_{t-i}\}$ uncorrelated? The answer is no.
- So, CLRM condition 3 is violated.
- This is because individual specific effect(s) are in the error component.
- If you recall, these are same over the years since $\{GDP_t\}$ is a process that captures data of GDP of an country.

Uniqueness of Time-Series Data

- If we go one period in the past of the model then:

$$GDP_{t-1} = a_0 + a_1 GDP_{t-2} + \varepsilon_{t-1}$$

- $\{\varepsilon_{t-1}\}$ and $\{\varepsilon_t\}$ are correlated since both of them captures same individual effect(s).
- Again, $\{GDP_{t-1}\}$ and $\{\varepsilon_t\}$ are correlated since $\{\varepsilon_t, \varepsilon_{t-1}\}$ are correlated.
- This is clearly a violation of CLRM condition 1.

$$GDP_t = a_0 + a_1 GDP_{t-1} + \varepsilon_t$$

$$GDP_{t-1} = b_0 + b_1 GDP_{t-2} + \varepsilon_{t-1}$$

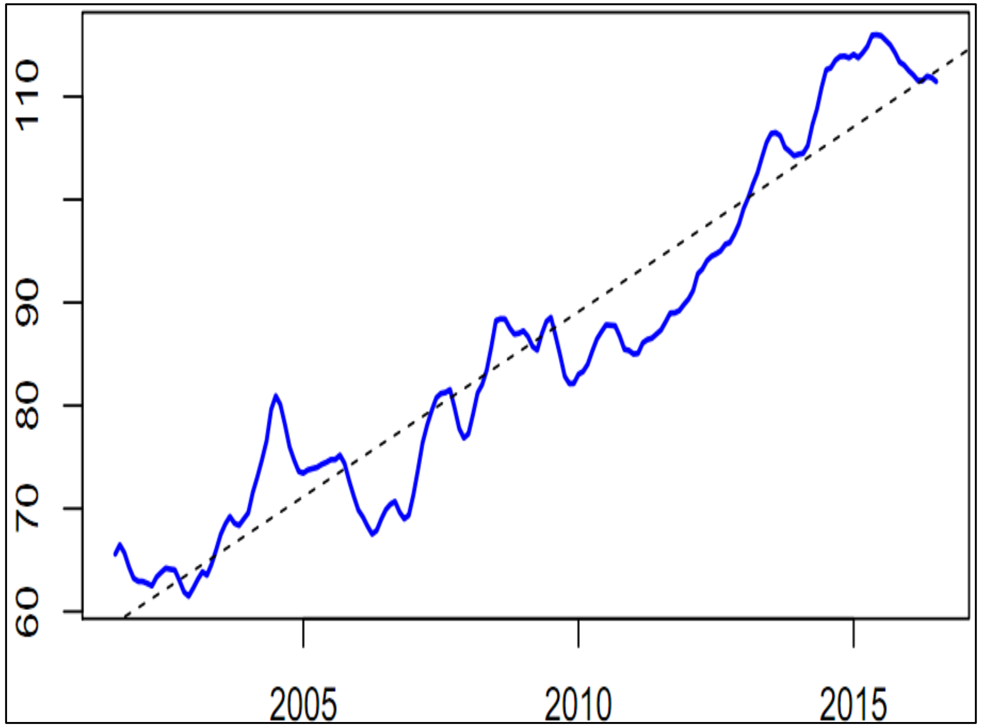
Patterns in Time-Series Analysis

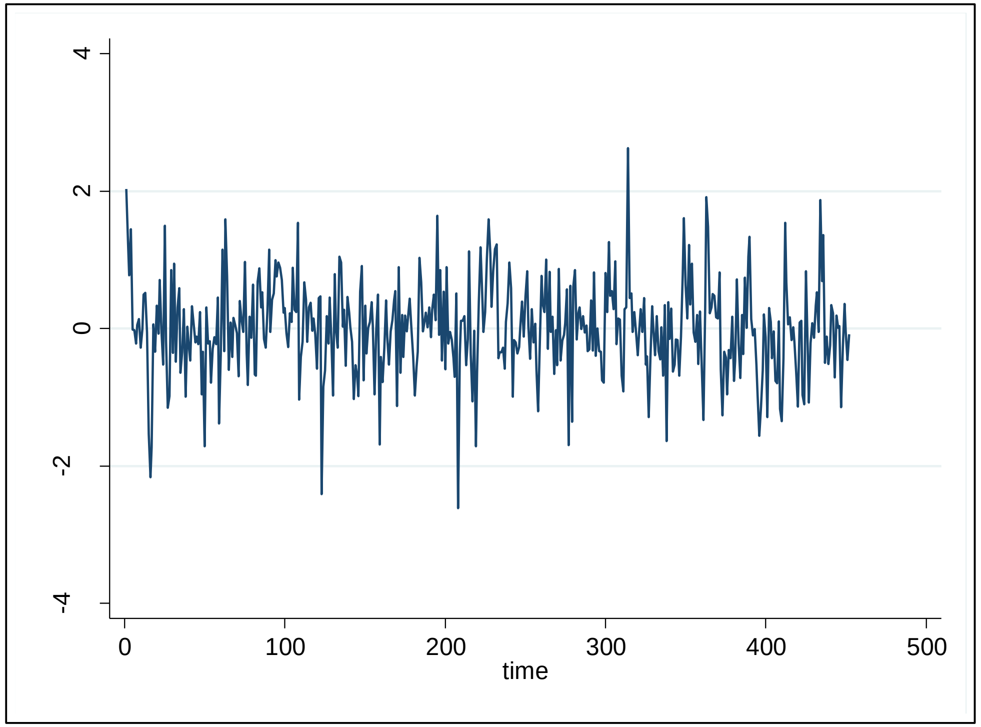
- The most important aspect of time-series analysis is patterns.
- Without patterns, a series loses its ingenuity.
- For example, without patterns, we will not understand the notion of stationarity.
- Mainly, we will discuss three patterns that are mostly observed in the time-series data. These are as follows:
  1. Trend
  2. Seasonality
  3. Cycle
- We will discuss them briefly.

Patterns in Time-Series Analysis

Trend:

- In simple words, a trend is a long-run pattern. If the average after a particular period changes, there is the presence of a trend (which could be positive or negative).
- If there is no trend, then the averages do not change. But if they do so, the change is very minimal as well as the change does not sustain for a longer time.





Patterns in Time-Series Analysis

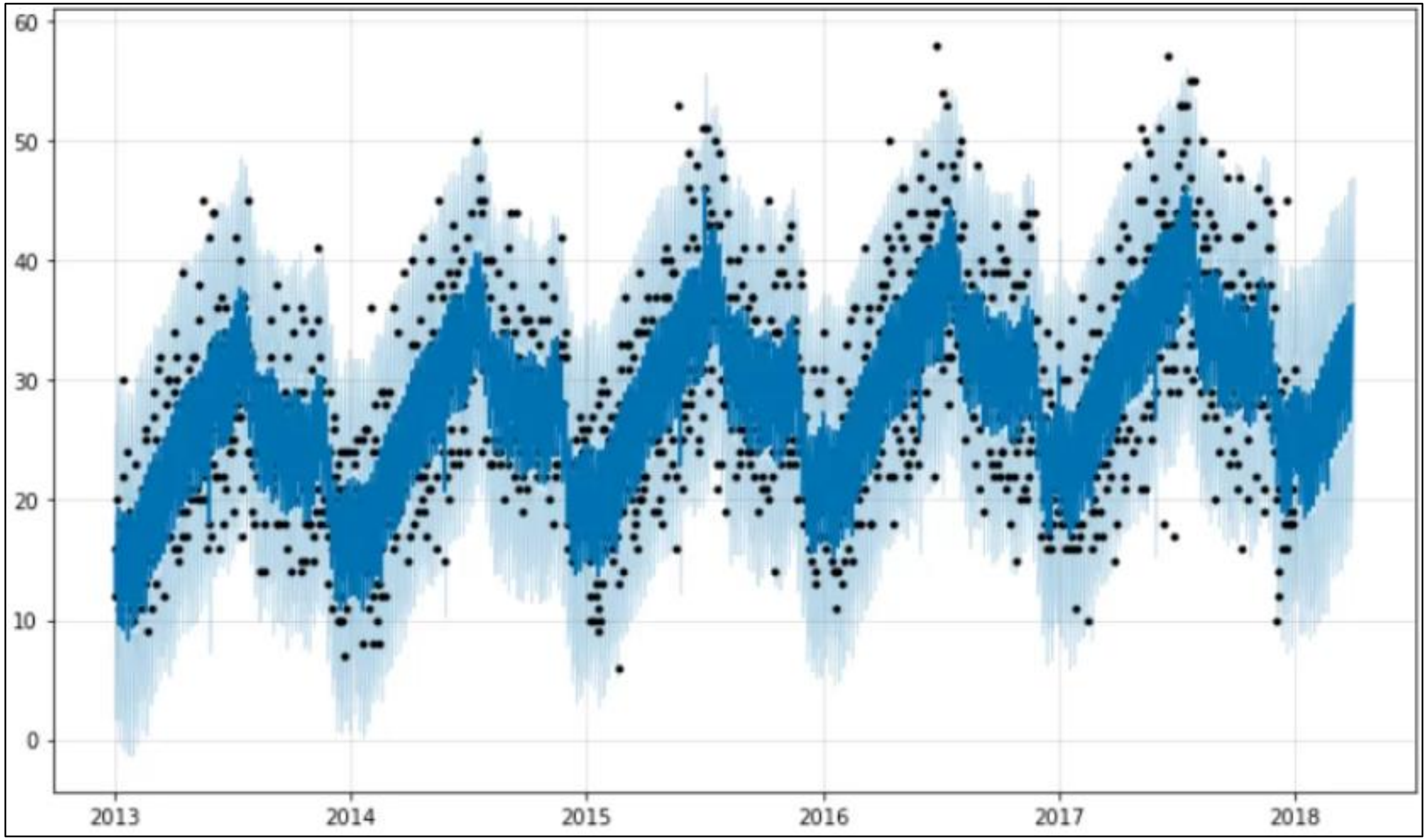
- There can be two types of trends:
  1. Deterministic Trend
  2. Stochastic Trend
- A deterministic trend is easy to observe. For instance upward or downward trend
- Stochastic Trend is difficult to observe. Not easy to catch with the bare eye. Need some tests to confirm.

Patterns in Time-Series Analysis

Seasonality:

- Ups and downs in the series at a regular interval.
- For example, commodity price hikes during the Ramadan season or the monthly average temperature of a city.
- There can be two types of seasonality:
  1. Deterministic/Predicted Trend
  2. Stochastic Trend/Random Walk

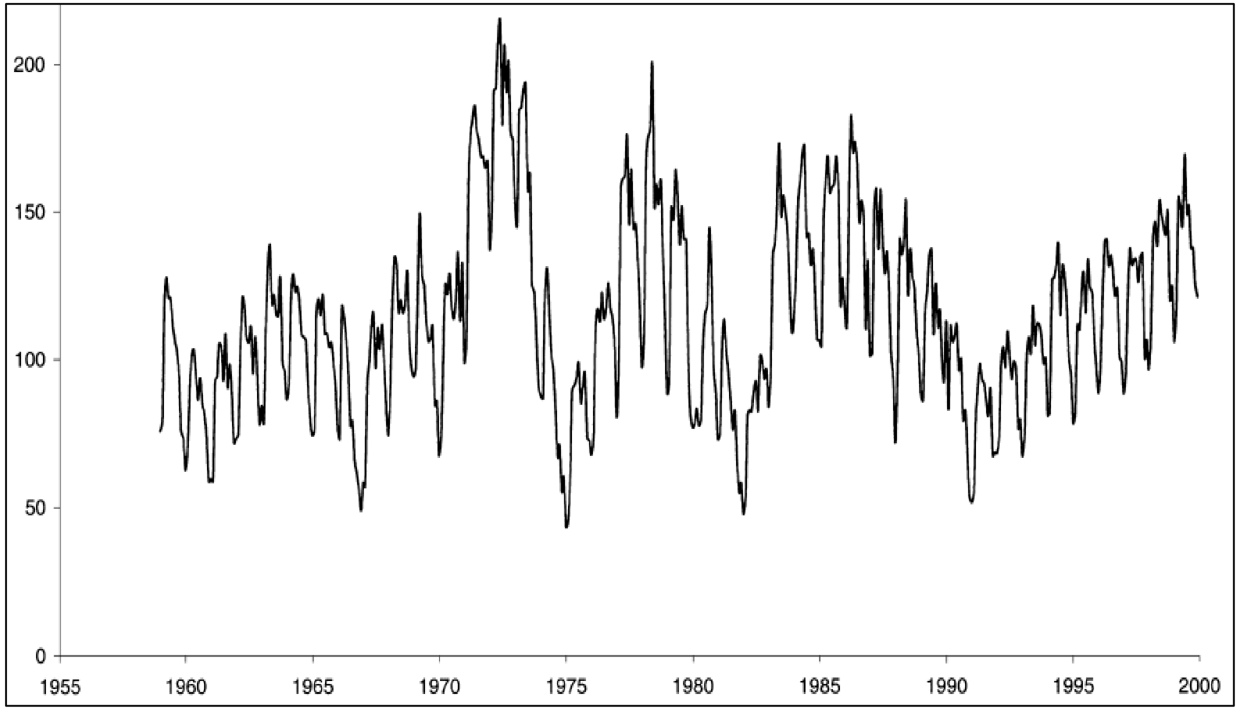
Patterns in Time-Series Analysis



Patterns in Time-Series Analysis

Cycle:

- Cycle is the most important component in time-series analysis. It captures of impact from shocks.
- It occurs in unpredicted intervals.
- By unpredictable, technically we mean there is no chance of getting probability of 1 at any data point of a particular series.



Patterns in Time-Series Analysis

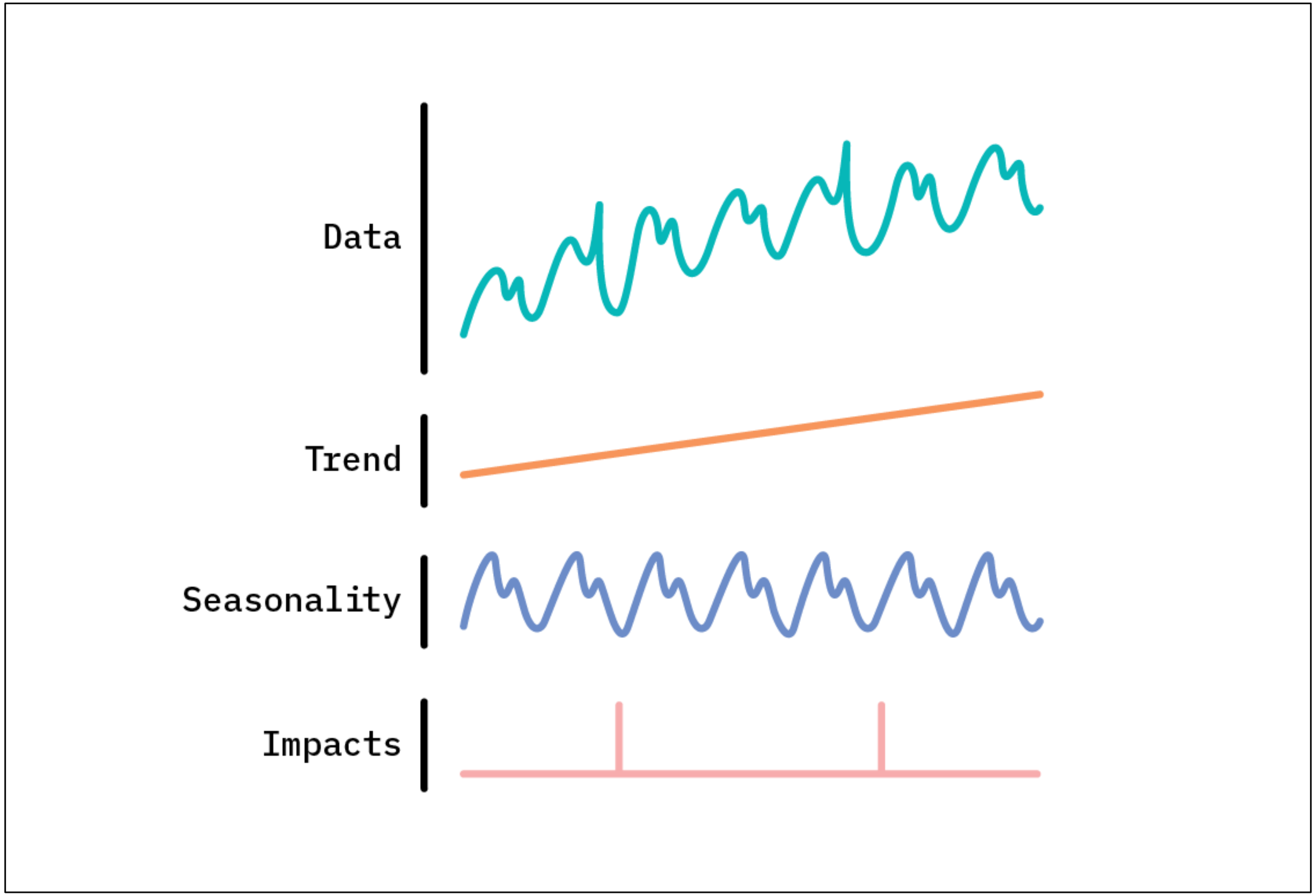
- Considering the effects of these patterns, we can define a time-series process in the following manner:

$$y_t = T_t + S_t + C_t + \varepsilon_t$$

$$y_t = T_t \times S_t \times C_t \times \varepsilon_t$$

- T_t= Trend, S_t= Seasonality, C_t= Cycle, ε_t= Irregularity (i.e. things that cannot be explained by T_t,S_t, and C_t).

Patterns in Time-Series Analysis



White Noise

- White Noise (W. N.) is the building block of time series models. White noise is a vital concept in the world of econometrics. It refers to a random variable that exhibits no specific pattern or correlation over time.
- This lack of correlation makes white noise an important tool in time series analysis and econometric modeling.
- **Properties of White Noise**:
  - **Zero means**: The average value of the white noise series is zero, which means there is no trending behavior.
  - **Constant variance**: The variance, or dispersion, of the white noise series remains constant over time.
  - **Independent distribution**: The values in a white noise series are not correlated and are independently distributed.
  - **Serial uncorrelatedness:** No value in a white noise series is autocorrelated, meaning there is no association between a value and any other value that occurs in the sequence before or after it.
  - **Stationary:** A white noise series is stationary, meaning that its statistical properties do not change over time.

White Noise

- $\{\varepsilon_t\}$ is a white noise process if in general we say the following:

$$\varepsilon_t \backsim \ N\,(0, \sigma_\varepsilon^2)$$

- Such expression depends on three distinct conditions:

1. $E(\varepsilon_t) = E(\varepsilon_t|\varepsilon_{t-1}, \dots, \varepsilon_{t-k}) = 0 \longrightarrow$ absence of predictability
2. $Var(\varepsilon_t) = Var(\varepsilon_t|\varepsilon_{t-1}, \dots, \varepsilon_{t-k}) = \sigma_\varepsilon^2$; constant $\forall_t = \{1,2, \dots, k\} \longrightarrow$ conditional homoskedasticity or constant variance at all the time
3. $Cov\,(\varepsilon_t, \varepsilon_{t-j}) = 0 \longrightarrow$ absence of any serial correlation within the process

- If you think carefully, these conditions tell you about randomness (at least closely).
- Also, if $\varepsilon_t \backsim \ N\,(0, 1)$, we define it as Gaussian White Noise.

Applications of White Noise in Econometrics

- **Model evaluation:** White noise is commonly used as a benchmark for evaluating econometric models. If the residuals or errors of a model exhibit white noise, it indicates that the model is capturing all relevant relationships in the data and performing well.
- **Model improvements:** In cases where residuals exhibit patterns or correlations instead of white noise, economists may need to refine their models to better account for these relationships or to adjust for seasonality.
- **Prediction accuracy:** The presence of white noise in a model's residuals can imply that the model is providing accurate predictions and that there is no information left in those residuals with which to improve future forecasts.
- Understanding white noise in econometrics is vital for economists, data analysts, and anyone involved in modelling, forecasting, or assessing complex economic relationships. By knowing this concept and its properties, one can ensure more accurate and robust econometric models for future developments in the economic landscape.

Testing for White Noise

- Testing whether a series is white noise is an essential part of model diagnostics. Various statistical tests can be used to check for white noise, including:
- **Ljung-Box Test:** This test checks for autocorrelation at different lag lengths. A failure to reject the null hypothesis suggests the residuals are white noise. An updated version of this test is the Portmanteau Test, which can be applied in STATA.
- **Runs Test:** The runs test assesses the randomness of data by analyzing the occurrence of runs above and below the median of the dataset.
- If a series is found not to be white noise, it may indicate that there are still patterns or structures in the data that a statistical model can exploit for prediction or explanation.

White Noise Correlogram

- Empirically, the correlation distribution of white noise is close to as:

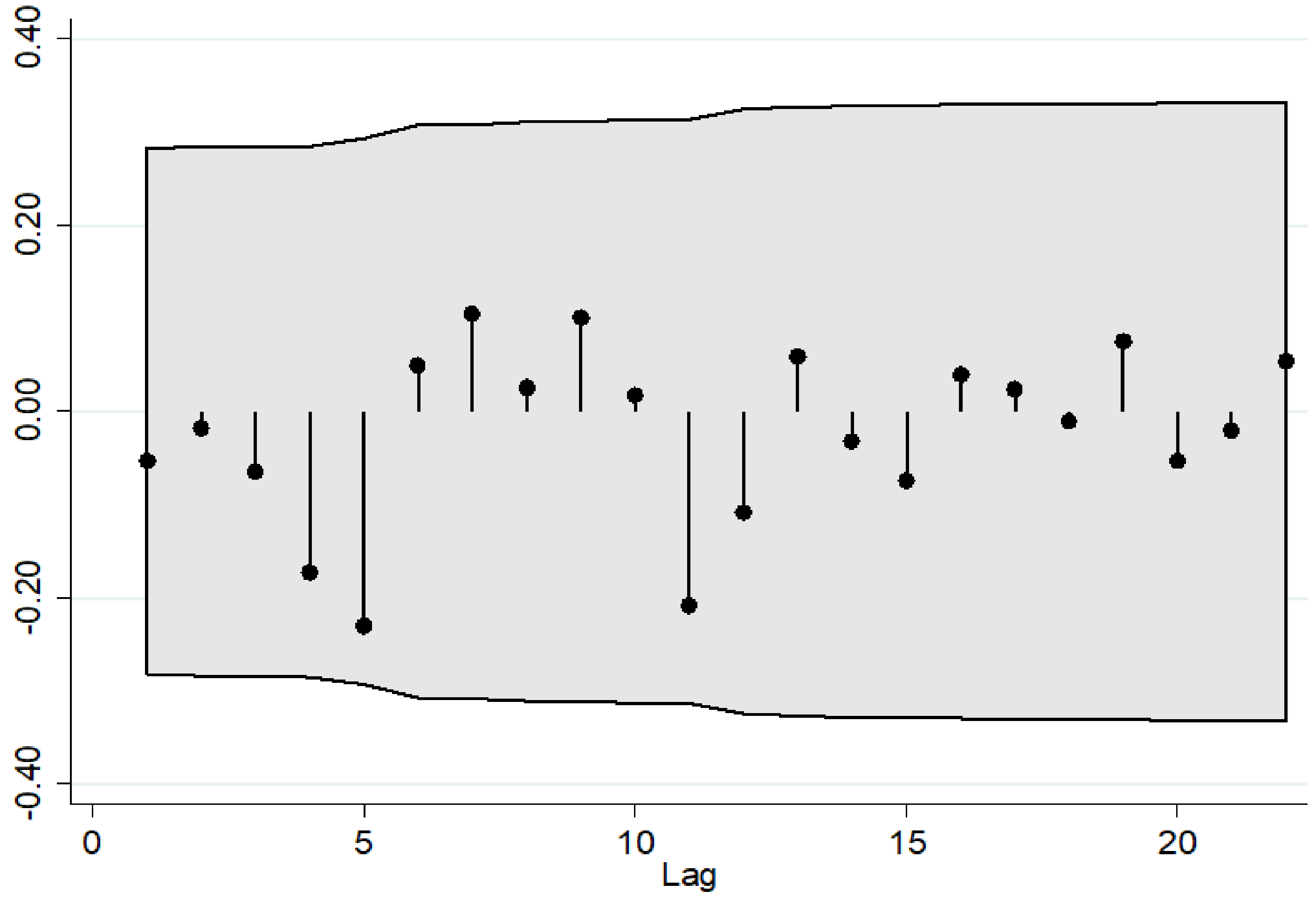
$$Corr\left(\varepsilon_t, \varepsilon_{t-j}\right) \backsim (0, \frac{1}{\sqrt{T}})$$

- from the above expression, we can develop a CI band that shows whether errors are serially uncorrelated.

$$CI = \hat{\varepsilon} \pm z\frac{1}{\sqrt{T}}$$

- The overall process is known as a correlogram.
- It is advisable that after finishing the estimation of any time-series model (i.e. considering all the patterns), one should check the correlogram since theoretically there should be no correlation in the residual.
- If no correlation is found, then the model outcome should be considered robust.

White Noise Correlogram



Stationary Process

- **Recall**: $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$
- A series $\{y_t\}$ is a stationary process if on average, there is no fluctuation in $\{y_t\}$.

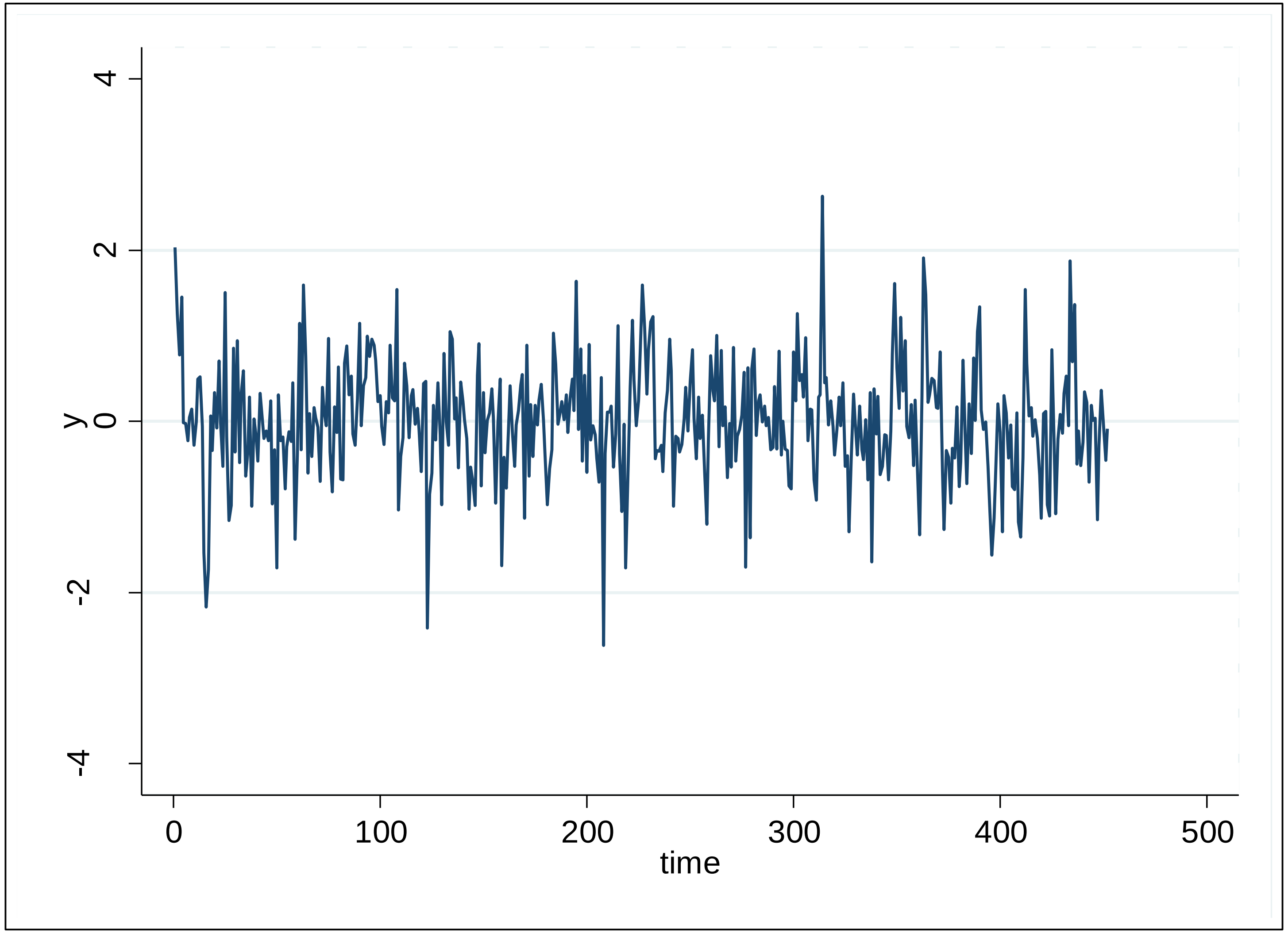
$$E(y_t) = \text{constant } \forall_t = \{1,2,\ldots,k\}$$

$$Var(y_t) = Var(y_t|y_{t-1},\ldots,y_{t-k}) = \sigma_\varepsilon^2; \text{ constant } \forall_t = \{1,2,\ldots,k\}$$

$$corr\ (y_t, y_{t-s}) = corr\ (y_t, y_{t+s}\ )$$

- Condition 3 is a bit tricky to understand.
- It indicates that the behaviour of $\{y_t\}$ between past and future (i.e. lags) should be the same.
- If say lag=5 then, $corr\ (y_t, y_{t-5})$ should be equal to $corr\ (y_t, y_{t+5})$.
- When a series satisfies these conditions, we call them weak stationary process/co-variance stationary process
- From the conditions of weak stationarity, we can further conceptualise the strong stationary process
  - In strong stationarity, we need conditions that go above the notion of second moment (say kurtosis = third moment)

Stationary Process



Stationary Process

Significance of Stationarity:

- To model the cycle, we use past/historical data. Only lag is used to get rid of trend and seasonality. So any pattern left in the data is due to the cycle.

- Statistical properties remain the same when data is stationary. Therefore, we can infer them in the future, leading to easy forecasting.
  - If data is non-stationary, then we cannot explain statistical properties unless there is confirmation of cointegration and other long-run properties.

ECO 372: Introduction to Econometrics

Spring 2024

Introduction to Time Series: Univariate Modelling

Sakib Bin Amin, Ph.D.

Associate Professor in Economics

Director, Accreditation Project Team (APT)

Outline

Our objectives for this lecture will be to learn:

- Moving Average (MA) Process
- Autoregressive (AR) Process
- Autocorrelation Function (ACF)
- Partial Autocorrelation Function (PACF)
- General Properties of AR (q) Process
- General Properties of MA (q) Process
- Goodness of Fit
- STATA Application
- STATA Codes
- Autoregressive Moving Average (ARMA) Process
- STATA Application

Moving Average (MA) Process

- This type of process is significantly different from Moving Average Smoother (MAS), which takes the average of the past values and attempts to predict the future value given a set of clearly defined boundaries.
- In MA, the future is predicted by talking about past shocks. The standard representation of an MA (q) process is as follows:

$$y_t = \beta_0 + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \cdots + \beta_q\varepsilon_{t-q} + \varepsilon_t;\ \varepsilon_t \sim N(0, \sigma^2)$$

- MA (1) and MA (2) processes thus can be written as:

$$y_t = \beta_0 + \beta_1\varepsilon_{t-1}; \qquad \varepsilon_t \sim N(0, \sigma^2)$$

$$y_t = \beta_0 + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2}; \qquad \varepsilon_t \sim N(0, \sigma^2)$$

- The β_q are the average weights of the shocks.
- Remember: if the dataset is not stationary we cannot use setup) MA under the Box-Jenkins setup (Box-Jenkins approach is discussed later).

Moving Average (MA) Process

- The MA model works as a finite impulse model, which means that the current noise value affects the present value of the model as well as "q" further values.
- One of the most important aspects of MA models is that the higher the value of the order of the MA model (q), the model will have longer memory and dependence on the past values.
- Fitting a MA model is generally more complicated. This is because the lagged error terms are not observable. This means that iterative non-linear fitting procedures (such as Maximum Likelihood) need to be used in place of linear least squares.
- Although $\{\varepsilon_t\}$ in the MA process is white noise, $\{y_t\}$ will not be a white noise.
- It can be shown y_t that will never be white noise even though it is modelled with a combination of white noise processes.

Autoregressive (AR) Process

- An AR expresses the conditional mean of a time series variable as a linear function of its own lagged values. Put simply, it's a regression of a variable with its past value.

- More specifically, an AR (p) model can be written as:

$$y_t = a_0 + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \varepsilon_t;\ \varepsilon_t \sim N(0, \sigma^2)$$

- AR (1) and AR (2) processes thus can be written as:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t; \qquad \varepsilon_t \sim N(0, \sigma^2)$$

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t; \qquad \varepsilon_t \sim N(0, \sigma^2)$$

- Similar to MA if the dataset is not stationary we cannot use AR under the Box-Jenkins setup.

- Autoregressive models are remarkably flexible at handling a wide range of different time series patterns.

Autocorrelation Function (ACF)

- Recall, correlation is a standardized measure of the relation between X and Y. It is bounded to be between -1 and +1.
- The autocorrelation function is a measure of the correlation between observations of a time series that are separated by s time units (y_t and y_{t-s}).

$$corr = \frac{cov(y,x)}{SE(y)SE(x)}$$

$$auto\ corr = \rho_s = \frac{cov(y_t, y_{t-s})}{SE(y_t)SE(y_{t-s})} = auto\ corr = \frac{cov(y_t, y_{t-s})}{Var(y_t)} = \frac{\gamma_s}{\gamma_0}$$

- When data is stationary, $SE(y_t)$ and $SE(y_{t-s})$ are same since the variability across time is constant.
- Squared term of SE is Var.
- ACF can be used for model identification as well as checking the correlation of the residuals.

Partial Autocorrelation Function (PACF)

- While the ACF shows all the relevant correlations across lags, the PACF shows the direct ones (like for specific lag).

- To understand it let's take an AR (2) process:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t; \qquad \varepsilon_t \sim N(0, \sigma^2)$$

$$\geqslant y_t - a_1 y_{t-1} - a_0 = a_2 y_{t-2} + \varepsilon_t$$

- Here, $\{y_t - a_1 y_{t-1} - a_0\}$ = in between lag effect is removed from $\{y_t\}$.

- Therefore, a_2 is the PACF of $\{y_t, y_{t-2}\}$.

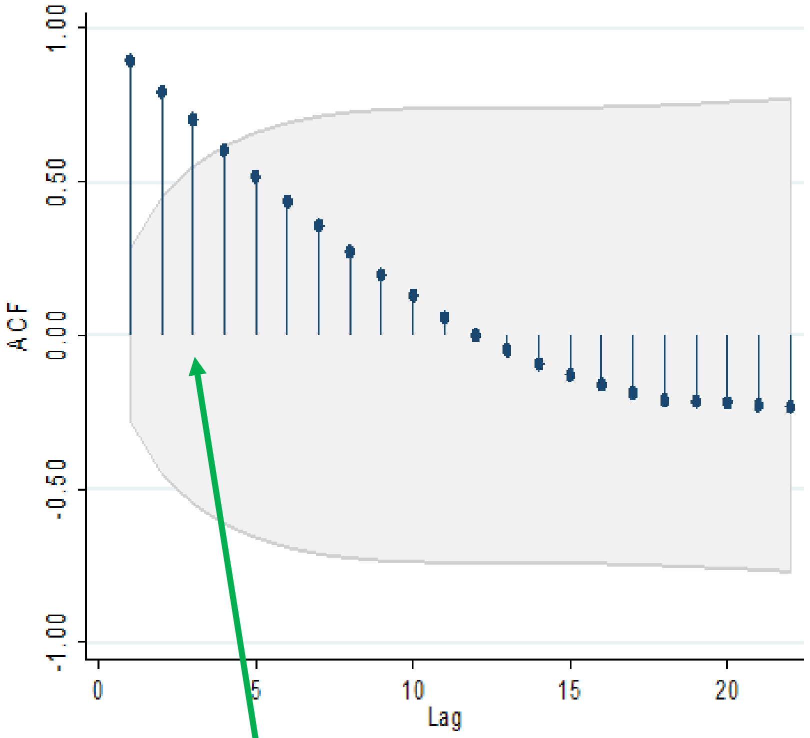
- PACF is used for model selection along with ACF.

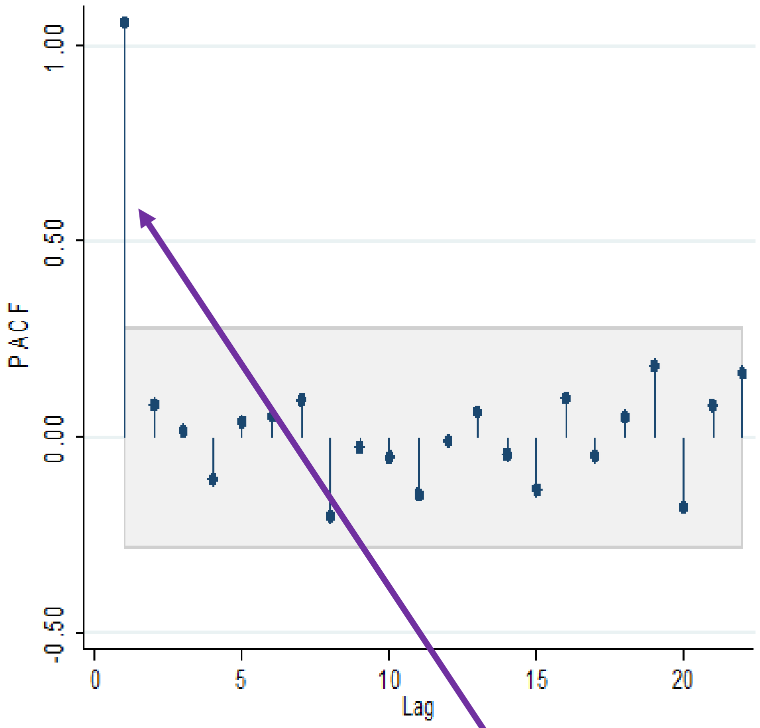
General Properties of AR (q) Process

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \varepsilon_t; \qquad \varepsilon_t \sim N(0, \sigma^2)$$

- $E(y_t)$ is constant across the time.
- $Var(y_t)$ is constant and depends on error variance (σ^2) and coefficient of regressor (i.e., $a_1, \ldots a_p$).
- The coefficient of the regressors determines whether AR process and be utilised.
- If $a_1 > 1$ AR(1) model is not valid because when the coefficient is greater than 1, it indicates non-stationarity.
- If $a_2 > 1$ and $(a_1 + a_2) > 1$ AR(2) model is not valid.
- If $a_3 > 1$ and $(a_1 + a_2 + a_3) > 1$ AR(3) model is not valid.
- ACF of the AR (p) process converges with a geometric decay with or without oscillation.
- PACF of AR (p) process cuts after the p^{th} lag. However, in real-life data, you will find zero and near zero. You may also find smelling spikes at higher lags, resulting from sudden shocks.

ACF and PACF of AR Process





Geometric decay of ACF, which converges to zero over time. This means the AR process needs to be used

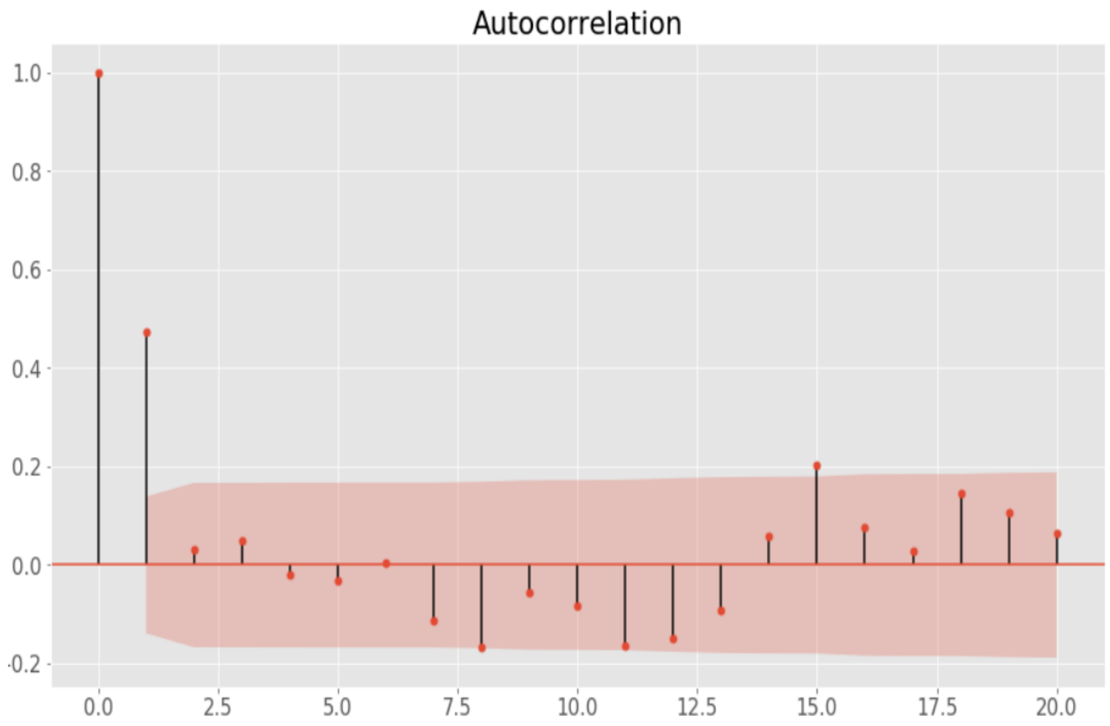
PACF cuts off after 1 spike, meaning AR(1) needs to be used

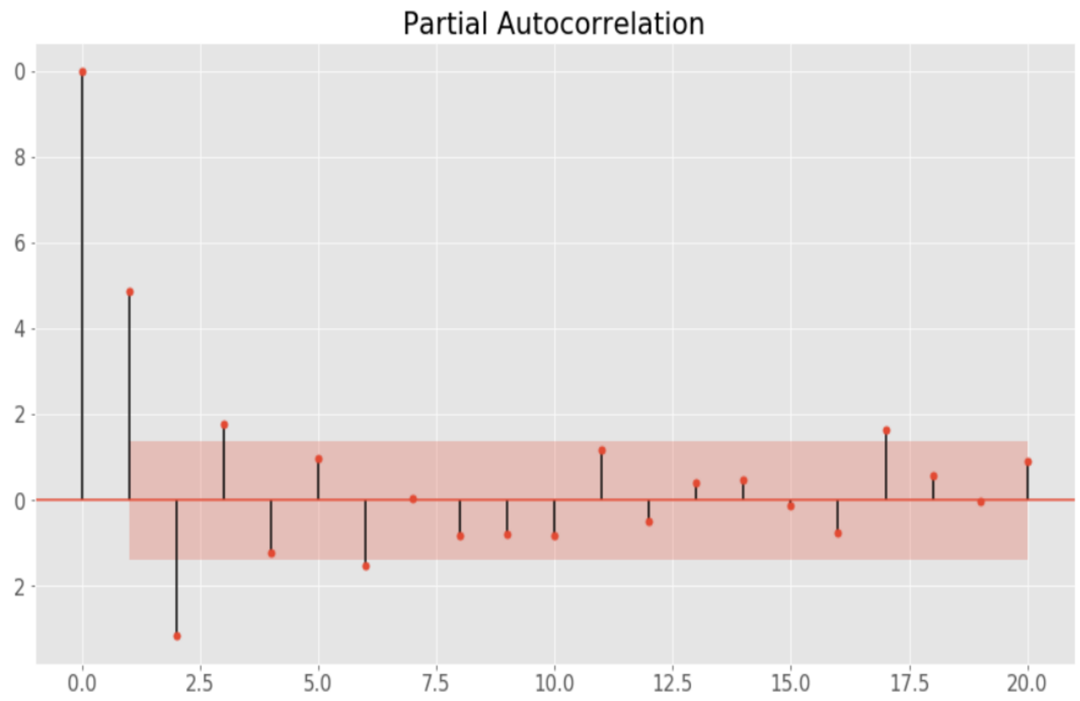
General Properties of MA (q) Process

$$y_t = \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q} + \varepsilon_t; \qquad \varepsilon_t \sim N(0, \sigma^2)$$

- $E(y_t)$ is constant across the time.
- $Var(y_t)$ is constant and depends on error variance (σ^2) and coefficient of regressor (i.e., $\beta_1, \ldots \beta_q$).
- For an MA(1) model $\beta_1 < 1$
- For an MA(2) model $\beta_1 < 1$ and $(\beta_1 + \beta_2) < 1$
- ACF of MA(q) jumps to zero after q^{th} lag . You may also find smelling spikes at higher lags, resulting from sudden shocks or seasonalities.
- However, PACF of MR (1) process converges with a geometric decay with or without oscillation.
- The change in pattern mainly happens due to the utilisation of past shocks and their underlying characteristics (i.e., white noise).

General Properties of MA (q) Process





ACF cuts off after 2 spike but the first spike is at lag 0, meaning MA(1) needs to be used

Geometric decay of PACF, which converges to zero over time. This means MA process needs to used

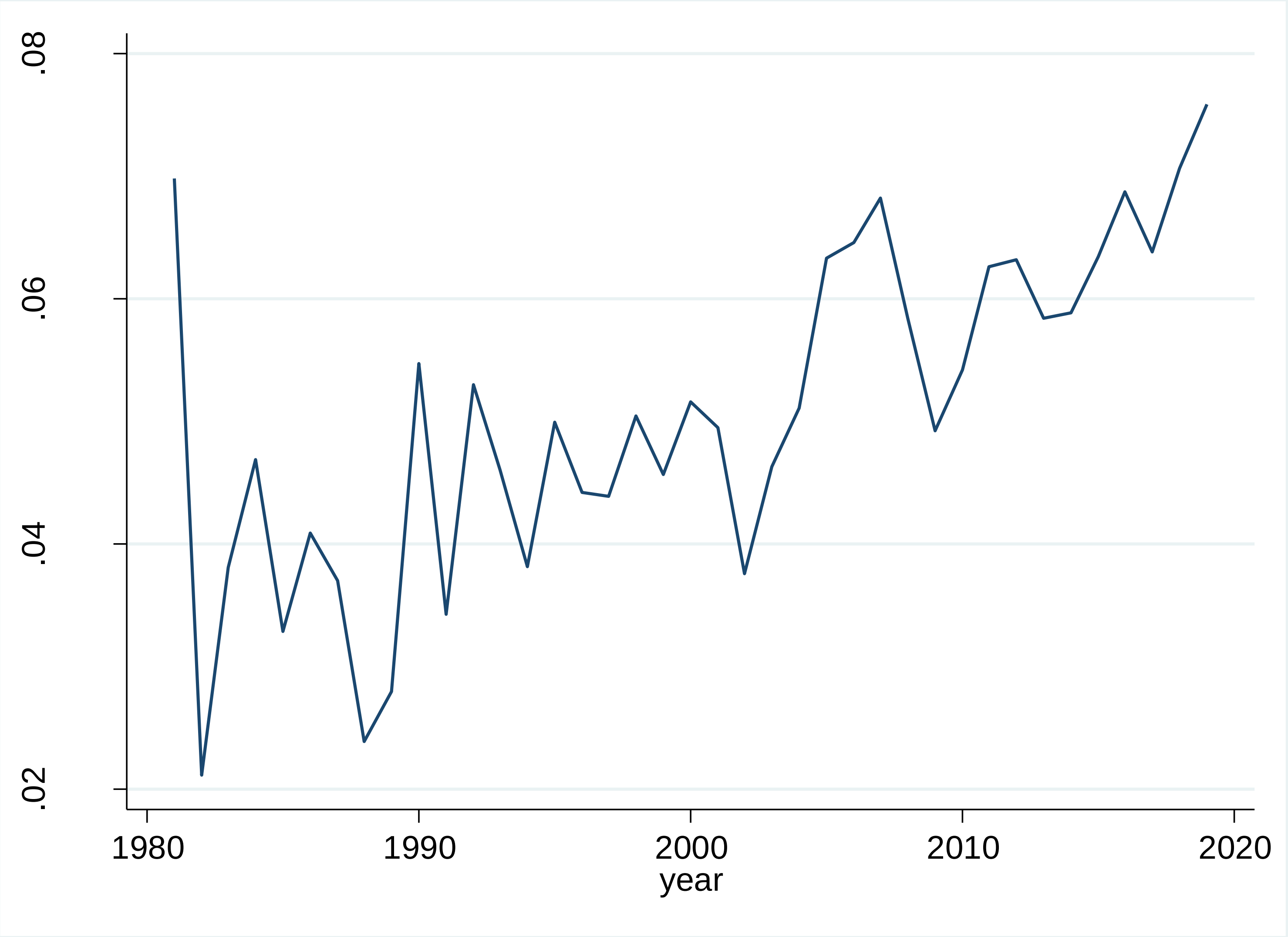
Goodness of Fit

- R^2 and Adjusted- R^2 can be used to determine the goodness of fit for the proposed model partially
  - When the maximum likelihood approach is used we don't need them. Then likelihood value closer to 0 indicates better model. Or, we can calculate a Pseudo-R^2 if needed.
- Additional tests need to be run to ensure overall goodness of fit.
- First, we need to check whether the error is residual is a process after estimation. The residual should be a white noise process.
  - If the residual is not a white noise process, not all the patterns are indeed addressed. Therefore, additional information is available in the residual.
  - The autocorrelation of residuals will be zero or close to zero.
- Second, we need to check the selected model is appropriate for inference. So we use the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).
  - When sample is less than 100, use AIC. When sample is more than 100, use BIC.
  - Both tests restricts use of unnecessary lags in the model. They always favour parsimonious models.
  - Lower the AIC/BIC value, better the model.

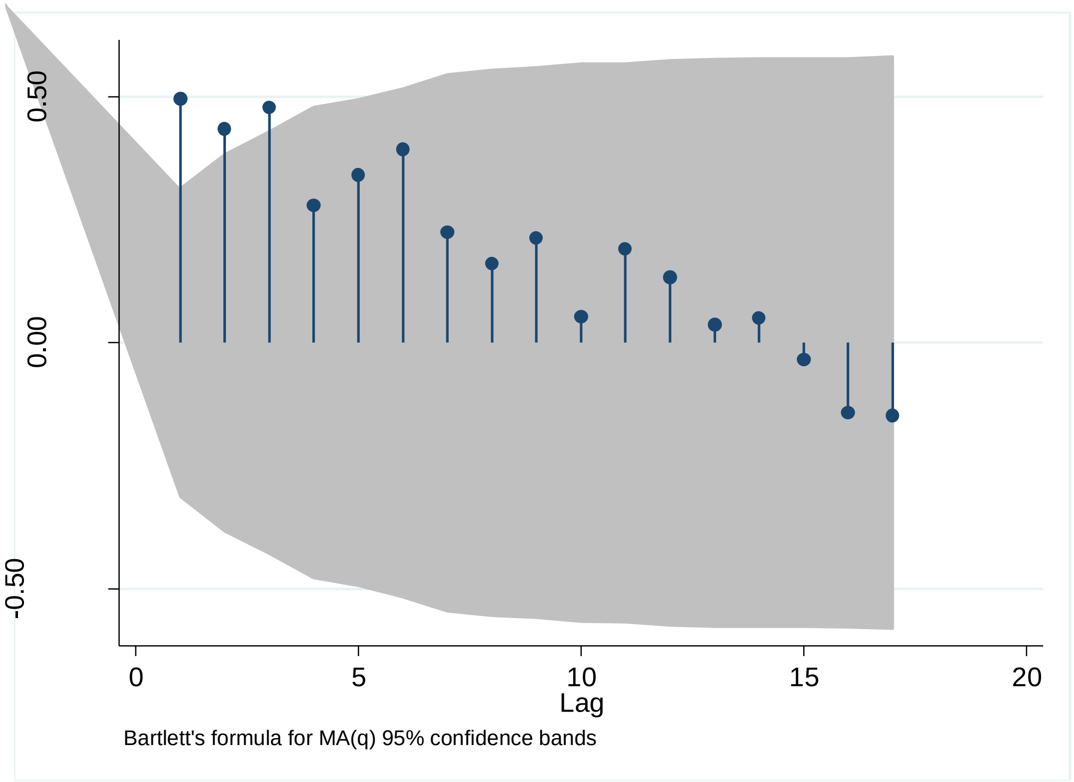
STATA Application

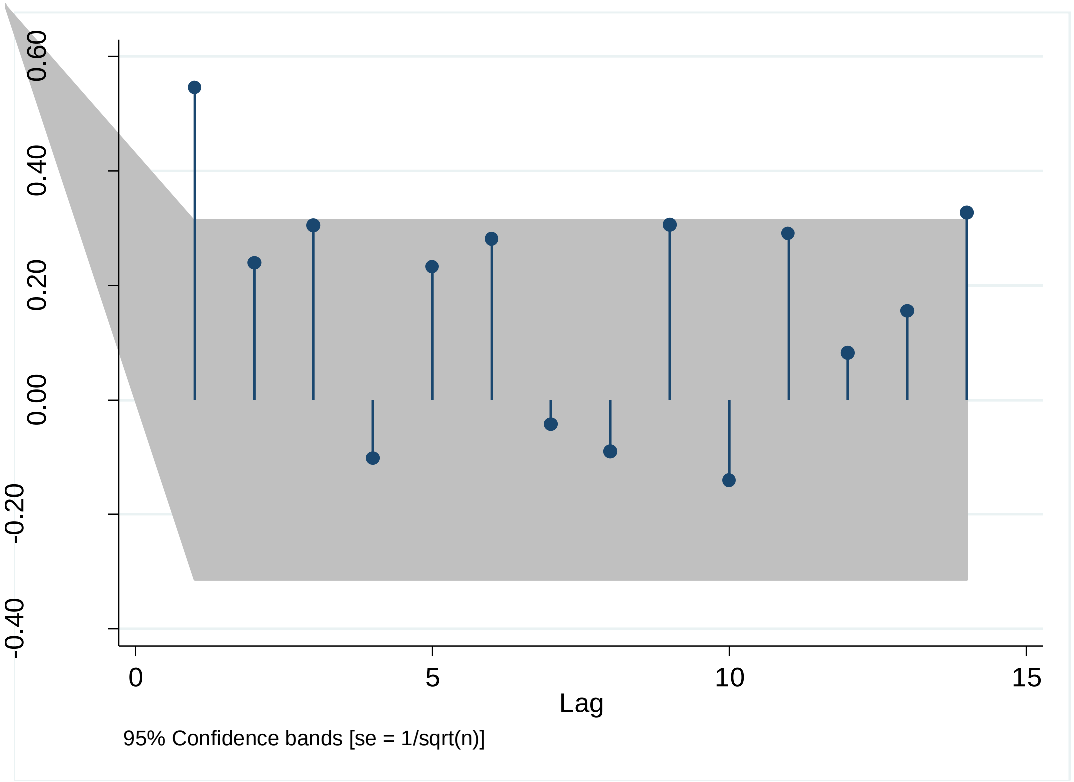
- Suppose we have data for the natural GDP growth of Bangladesh from 1980-2019
- We aim to model this series by following the flow known as the **Box-Jenkins** method:
  - First, we plot the data to visualise
  - Second, we check the ACF and PACF to determine whether to use AR or MA
  - Third, we specify/estimate the econometric form/function(s)
  - Fourth, we estimate the model(s)
  - Finally, we check the model selection criteria and diagnostics

STATA Application



STATA Application





Both ACF and PACF are suggesting AR(1) model. $Growth_t = a_0 + a_1 Growth_{t-1} + \varepsilon_t$

STATA Application

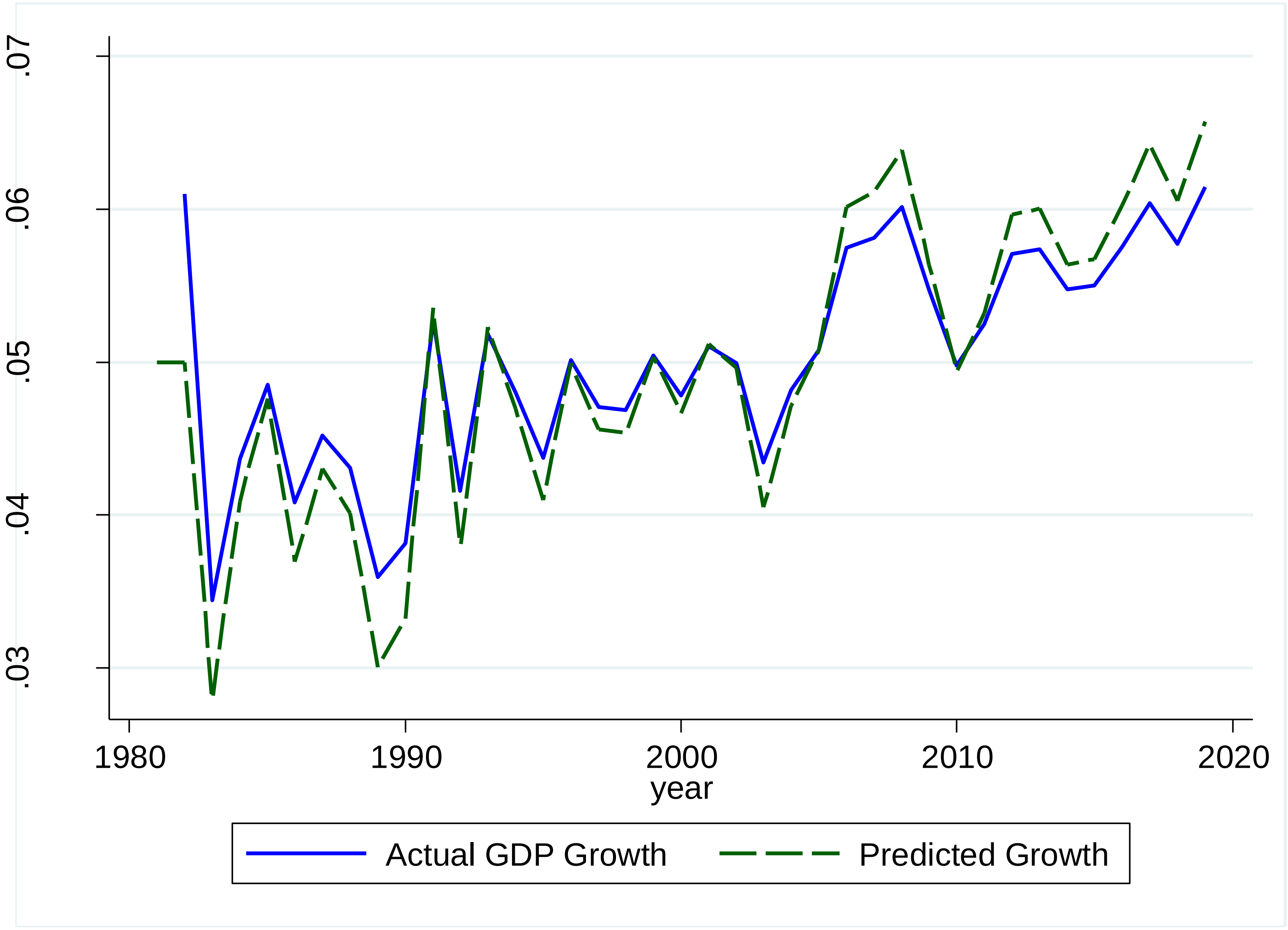
Source	SS	df	MS
Model	.001857627	1	.001857627
Residual	.004643959	36	.000128999
Total	.006501586	37	.000175719

Number of obs	=	38
F(1, 36)	=	14.40
Prob > F	=	0.0005
R-squared	=	0.2857
Adj R-squared	=	0.2659
Root MSE	=	.01136

growth	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
growth						
L1.	.546138	.1439183	3.79	0.001	.2542581	.838018
_cons	.022884	.007438	3.08	0.004	.0077991	.037969

- STATA expresses $a_1 Growth_{t-1}$ as growth.L1. L1 is lag of 1 year or (t-1)
- $a_1 < 1$ ACF and PACF indeed predicted modelling technique well.

STATA Application



STATA Application

Akaike's information criterion and Bayesian information criterion

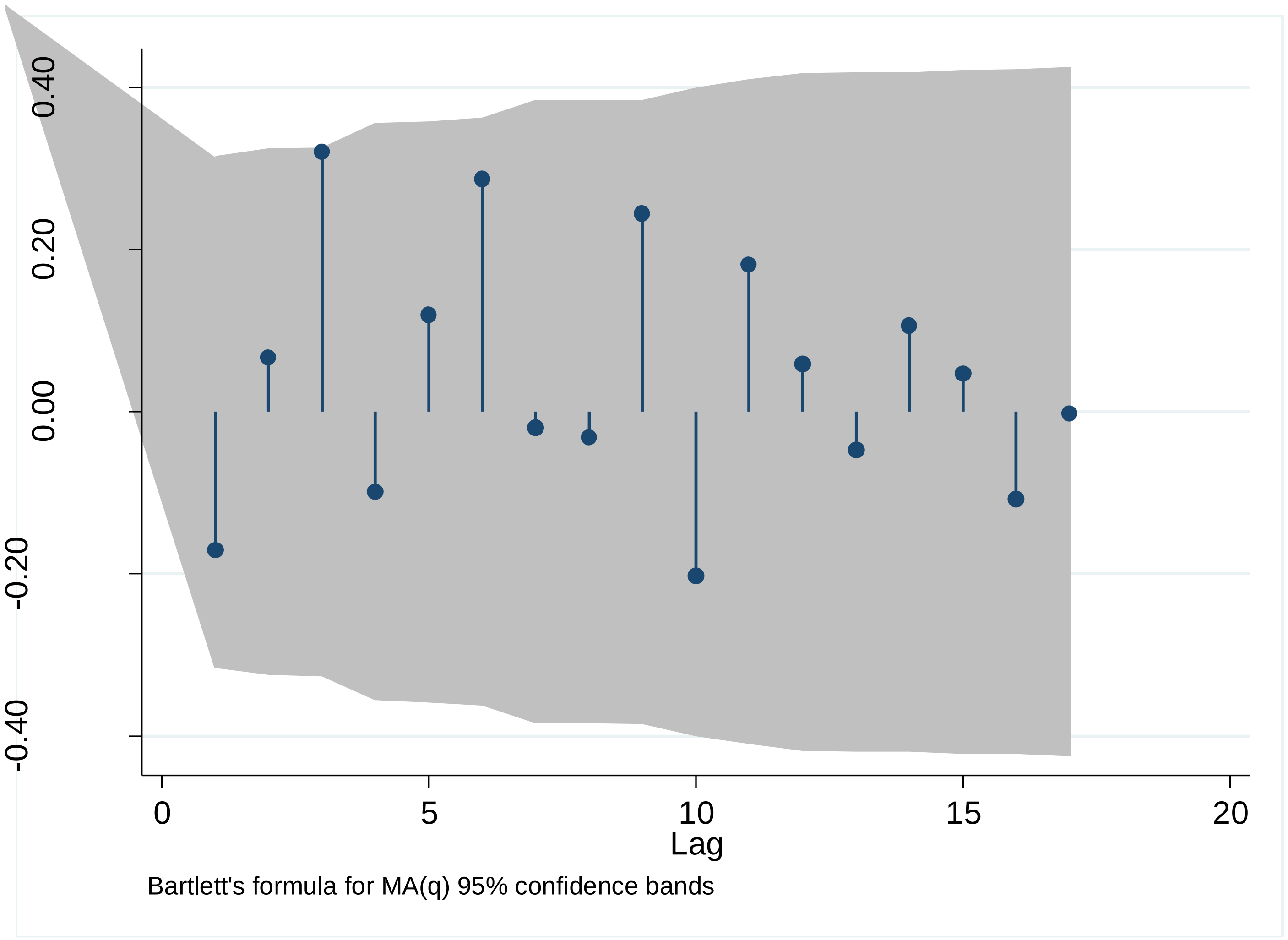
Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	38	110.8729	117.266	2	-230.5321	-227.2569

Portmanteau test for white noise

Portmanteau (Q) statistic	=	20.5384
Prob > chi2(17)	=	0.2476

Null hypothesis: series is white noise
Alternate hypothesis: series is not white noise

STATA Application



STATA Codes

```
** declaring as time series dataset
        tsset year
        tsline growth

** acf  and pacf
        ac growth
        pac growth

** estimation with reg and arima command both nearly provide same result.
        arima growth, ar(1)
        reg growth l.growth

**aic and bic check
        estat ic

** predict growth and residual
        predict predicted_growth
        predict residual, resid
        tsline growth predicted_growth

** white noise check
        wntestq residual
        ac residual
```

Autoregressive Moving Average (ARMA) Process

- It is possible that a model can be explained by both the AR and MA process.
- We can specify such a process with the ARMA (p, q) model. That is p lags of AR and q lags of MA.

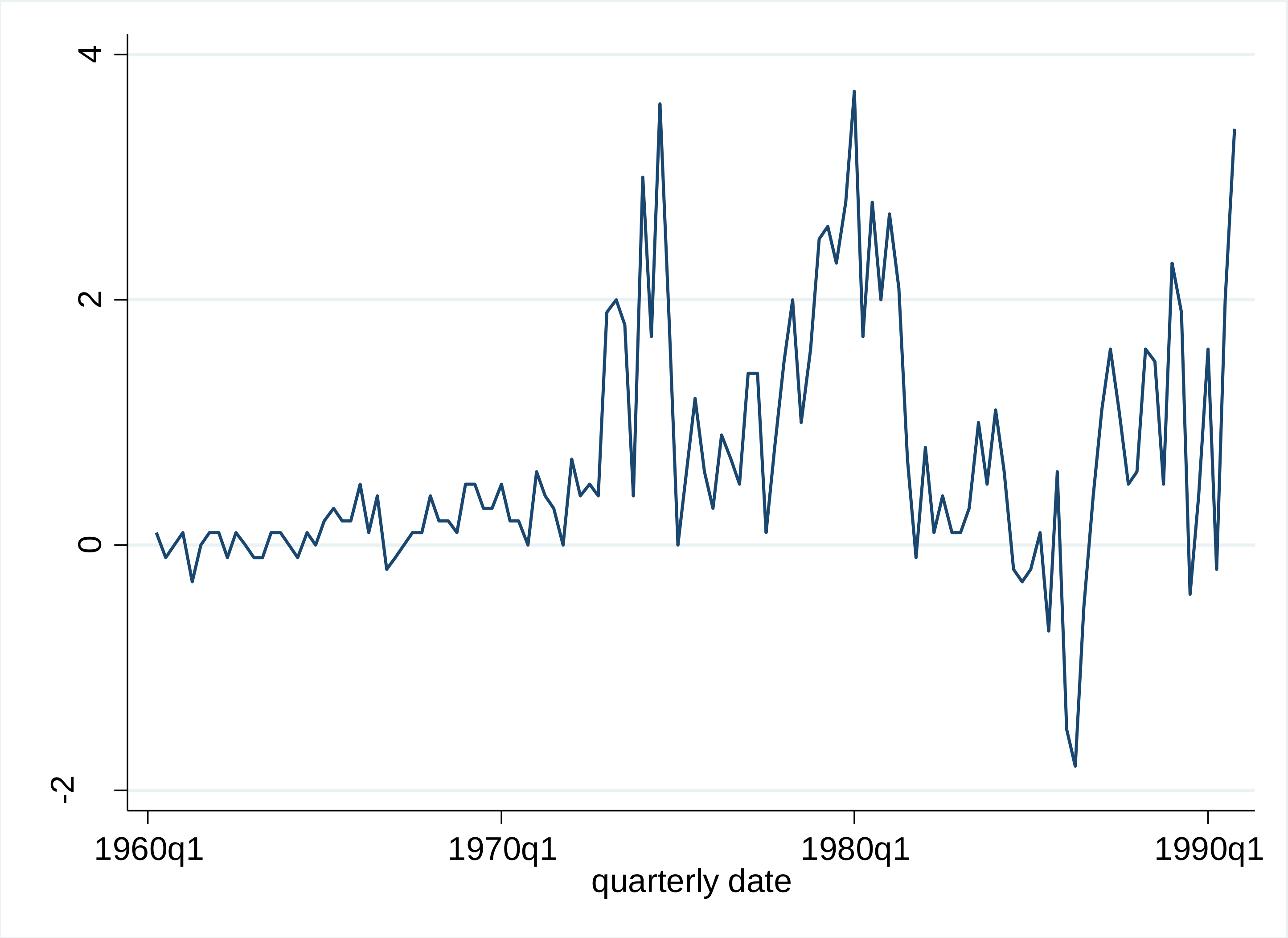
$$y_t = \mu_0 + a_1 y_{t-1} + \cdots + a_p y_{t-p} + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q} + \varepsilon_t; \varepsilon_t \sim N(0, \sigma^2)$$

- Since both AR and MA processes are present, the ACF and PACF will be different for the ARMA process.
- Both ACF and PACF will decay but with a jump after a specific lag.
- For example, **ARMA (2, 4)**: ACF will decay after lag 4. PACF will converge 2^{nd} lag.
- When data is not stationary ARMA Model can be run by taking the first difference of y_t (i.e., $\Delta y_t = y_t - y_{t-1}$). This model is known as ARIMA or Autoregressive Integrated Moving Average Model.

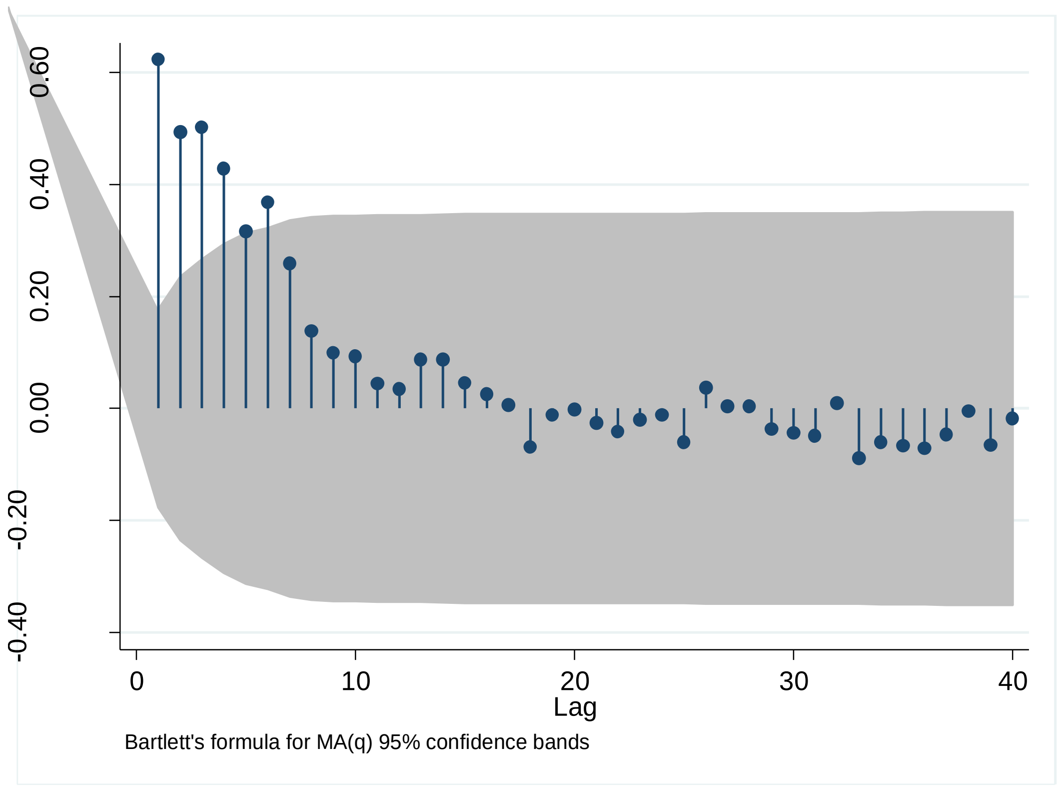
ARMA Application

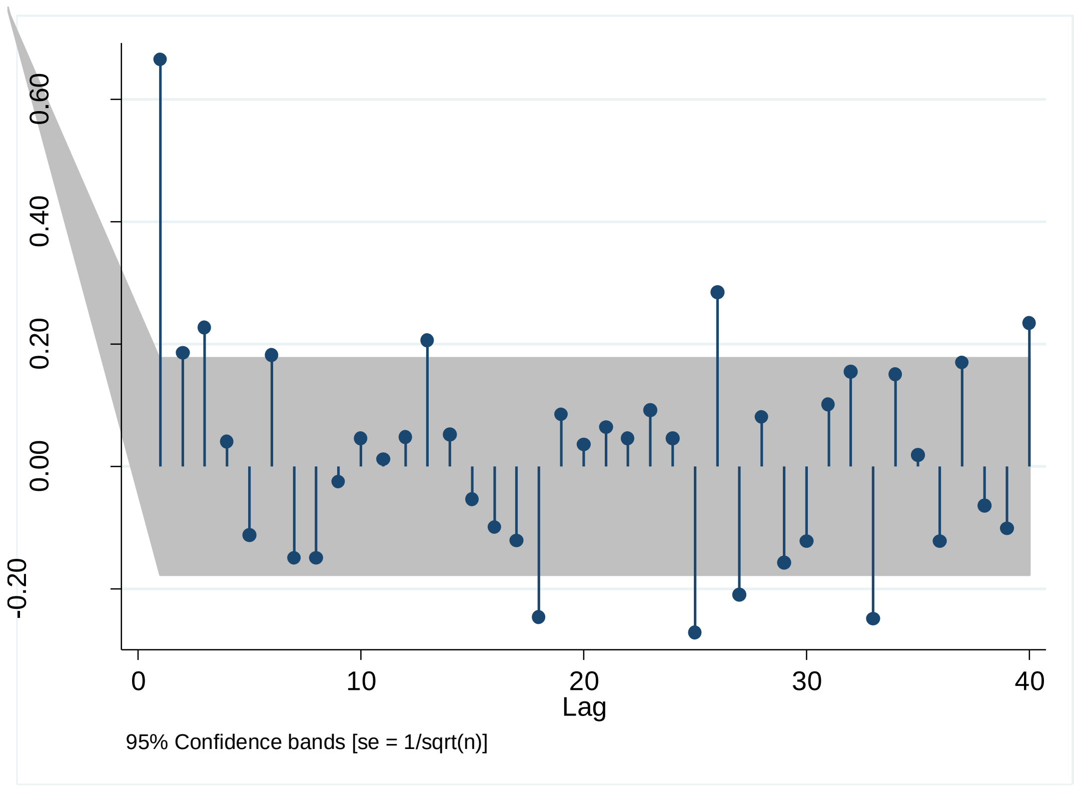
- Suppose we have data for wholesale price index from 1960-1990
- We want to model it
- We will use the Box-Jenkins approach
  - First, we plot the data to visualise
  - Second, we check the ACF and PACF to determine whether to use AR or MA
  - Third, we specify the econometric form(s)
  - Fourth, we Estimate the model(s)
  - Finally, we Check the model selection criteria and diagnostics

Autoregressive Moving Average (ARMA) Process



Autoregressive Moving Average (ARMA) Process



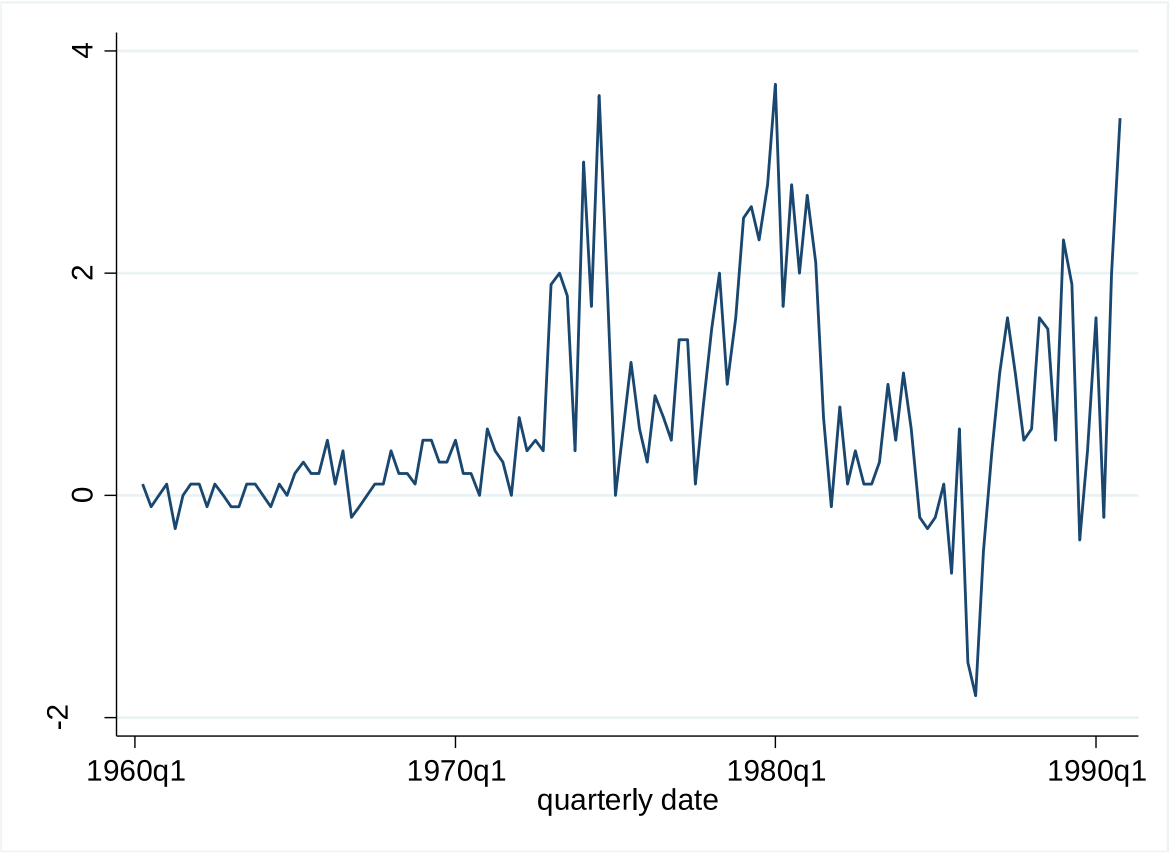


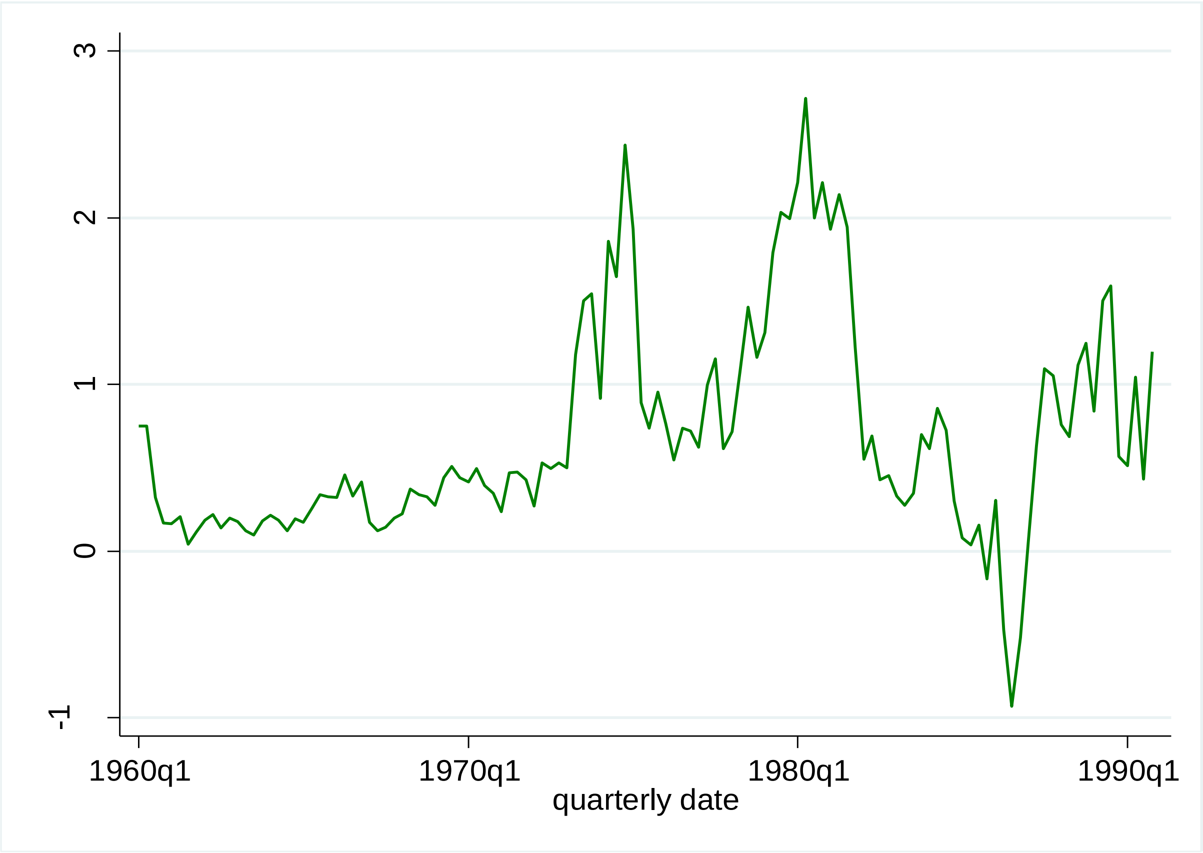
Both ACF and PACF are suggesting ARMA(1,1) model. Another model might be specified but let us stick to this model. $WPI_t = a_0 + a_1 WPI_{t-1} + \beta_1 \varepsilon_{t-1} + \varepsilon_t$

ARMA Application

D.wpi	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
wpi						
_cons	.7498197	.3340968	2.24	0.025	.0950019	1.404637
ARMA						
ar L1.	.8742288	.0545435	16.03	0.000	.7673256	.981132
ma L1.	-.4120458	.1000284	-4.12	0.000	-.6080979	-.2159938
/sigma	.7250436	.0368065	19.70	0.000	.6529042	.7971829

ARMA Application





ARMA Application

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	123	.	-135.3513	4	278.7026	289.9514

Portmanteau test for white noise

Portmanteau (Q) statistic	=	36.9722
Prob > chi2(40)	=	0.6073

Null hypothesis: series is white noise

Alternate hypothesis: series is not white noise

STATA Codes

```
** declaring as time series dataset
        webuse wpi1
        tsset year
        tsline d.wpi

** acf and pacf
        ac d.wpi
        pac d.wpi
** estimation with reg and arima command both nearly provide same result.
        arima d.wpi, arima(1,0,1)
        // arima (1,0,1) means ar(1) ma (1) and the series is stationarity (0)
**aic and bic check
        estat ic

** predict growth and residual
        predict predicted_wpi
        predict residual, resid
        tsline d.wpi
        tsline predicted_wpi

** white noise check
        wntestq residual
        ac residual
```