ARIMA Models

Handout

Stationarity

Strictly stationary time series

Properties

- $E(y_s) = E(y_t) = \mu$ for all s, t

Weakly stationary time series

Finite Variance Process

Properties

- μ_t - Constant (does not depend on t)
- Constant variance (does not depend on t)
- $\gamma_{s,t} = \gamma_{s+h,t+h}$
  - If they have the same time lag, the covariance should be the same. (i.e. $Cov(y_1, y_2) = Cov(y_2, y_3) = Cov(y_3, y_4)$. This is pretty useful in checking if the series is stationary
  - Note that if it is a function of h, it can still be stationary so long as it is not a function of t

Identification of Stationary Time Series

Time Series Plot

1. **Stationary time series**
   a. Roughly horizontal
   b. Has constant variance
   c. No predictable patterns in the long term (i.e. no trend)

2. **Non-stationary time series**
    a. Has trend or seasonality. They are not stationary since the mean is changing

ACF

1. Stationary series drops to 0 relatively quickly
2. Non-stationary series decreases slowly
3. Non-stationary series for r_1 is often large and positive

- Note that for ACF plots, it does not need to be all within the blue lines since we are not looking at white noise.

Differencing

$$y'_t = y_t - y_{t-1}$$

💡 Note that we will only have T-1 data since we can't compute the difference for the first observation y'_1

Purpose: To get the ACF to be well behaved. But note that we don't need to get it to white noise (but this will be ideally)

Unit Root Test

Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

$$y_t = \mu + \rho t + \mu_t + e_t$$

H_0 - Data are stationary and non seasonal

H_1 - Data are not stationary

Idea: We will fit a linear regression line on the data. If the test statistic is large it means that it is on one side of the line and therefore it could be non stationary.

- We can keep applying the KPSS test repeatedly to successive differencing to determine the number of differencing that should be carried out.

p-value: If the p-value is significant then we can reject the null hypothesis and conclude that we need to difference the data

Backshift Notation

$$By_t = y_{t-1}$$

The backshift operator B operates on the y_t and it will shift it back in time by 1

d-th order difference:

$$(1-B)^d y_t$$

Seasonal Difference:

First seasonal difference is given by

$$(1-B^m)y_t$$

Autoregressive Model

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t$$

💡 Note that c is not the mean of the time series. It is a function of the mean and the ϕ coefficients. $c = \mu - \phi_1\mu - \phi_2\mu - \cdots - \phi_p\mu$

- $\phi_1, \cdots, \phi_p$ - All constants
  - Changing these values will result in different time series patterns
  - $\phi_p = 0$
- e_t - Variance of the e_t will only change the scale of the series, it doesn't change the pattern

Roots determine if it is a casual process or not

▼ Linear Processes

> 💡 **Theorem**: A linear process y_t is defined to be a linear combination of white noise e_t and is given by $y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j e_{t-j}$ where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

Autocovariance Function:

Note that this is infinite-MA representation

$$\gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi j$$

- This will show that the autocovariance is a function of h and therefore is stationary
- Note that once we have the ϕ weights then we can work out $\gamma(h)$

Purpose

- This gives a way "in" to work out $\gamma(h)$ for $AR(p)$ processes. Because once we can determine that it is a linear process, we can use the expression for $\gamma(h)$
- Linear Processes are also **stationary**

▼ AR(1) Process is stationary

Proof is within OneNote

Is it possible to have a stationary AR(1) process when $\phi_1 > 1$

$y_t = -\sum_{j=1}^{\infty} \phi^{-j} e_{t+j}$

Stationary but it is not casual since it depends on future values

- Non casual time series is not really useful for us!
- We want to only look at casual processes instead

Casual AR Process

An AR Process is casual if it can be expressed by the previous values and not in terms of future values .

To check if a process is casual

The roots of the $AR(p)$ operator $\tau(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ determine if the $AR(p)$ process is casual, and hence stationary

Need to check if the complex roots lie outside the unit square, it is then it is a casual process

Example

For AR(1),

$$\tau(z) = 1 - \phi_1 z = 0$$
$$z = \frac{1}{\phi_1}$$

▼ MA Process

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

- ϵ_t - White Noise
- This is a multiple regression with **past errors** as predictors.
- **Note: This is not moving average smoothing!**

▼ Writing $AR(p)$ process as $MA(\infty)$

Note that it is possible to write any stationary $AR(p)$ process as an $MA(\infty)$ process

Example: **AR(1) Process**

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \epsilon_t \\ &= \phi_1(\phi_1 y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi_1^2 y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \\ &= \phi_1^3 y_{t-3} + \phi_1^2 e_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \\ &\dots \end{aligned}$$

Provided $-1 < \phi_1 < 1$

▼ Invertibility

General conditions for invertibility

> 💡 Complex roots of $1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ lie outside the unit circle on the complex plane

Essentially, solve for the roots of the moving average polynomial and check that the roots are outside the unit circle

- To check for complex roots, $z = x + yi$
  - Check the magnitude of $|z| = \sqrt{x^2 + y^2}$

- If $|z| > 1$ then it lies outside the unit circle

▼ **Conditions for different models**

- For $q = 1$: $-1 < \theta_1 < 1$
- For $q = 2$: $-1 < \theta_2 < 1 \qquad \theta_2 + \theta_1 > -1 \qquad \theta_1 - \theta_2 < 1$
- More complicated conditions hold for $q \geq 3$

Properties

1. If the **MA** model is invertible, any **MA(q)** process can be written as an **AR(∞)** process

Invertibility of an ARIMA model is equivalent to forecastability of an ETS model

▼ ARIMA

▼ **Components**

- **AR**: Auto Regressive (Lagged Observations as inputs)
- **I**: Integrated (Differencing to make series stationary)
- **MA**: Moving Average (Lagged errors as inputs)

Interpretability: It is rarely interpretable in terms of visible data structures like trend and seasonality. But it can capture a huge range of time series patterns.

ARMA Model Equation

$$\begin{aligned} y_t = c &+ \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} \\ &+ \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t \end{aligned}$$

- Note that this is when we have not included the **I** part which represents the differencing component of it.

ARIMA Model Equation

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_p e_{t-p} + e_t$$
$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B) y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q) e_t$$

Writing in a more succinct manner

$$\tau(B)(1-B)^d y_t = c + \kappa(B)e_t$$

$\tau(B)$ - AR Component

$(1-B)^d$ - Number of Difference

$\kappa(B)$ - MA Component

$c = \mu(1-\phi_1 - \cdots - \phi_p)$

- **Telling different processes apart**

  Sometimes some processes may look like others if we add in some redundant terms.

  - Note that if we add terms and reorder it, it does not change the ARIMA order

  What we can do is to factorize the y_t side and e_t side to see if there are any terms that are similar for both the autoregressive operator and moving average operator

  - We try to get the equation to the base form and that will be the underlying ARIMA model. We want the Parsimonious Model (model with the least parameters)

  - **Example**

    Looking at the following equation, we may think that it is a **ARIMA(2, 0, 1)** process

    $$y_t - 0.8y_{t-1} + 0.15y_{t-2} = e_t - 0.3e_{t-1}$$

    However, we can do the following

    $$\begin{aligned}(1-0.8B+0.15B^2)y_t &= (1-0.3B)e_t \\ (1-0.3B)(1-0.5B)y_t &= (1-0.3B)e_t \\ (1-0.5B)y_t &= e_t\end{aligned}$$

    We can see that since there is a common term between them, we can cancel it and this is just an **ARIMA(1, 0, 0)** model

    - Note that if we fit the 2 models, they will still produce similar forecasts

- **Predictors**: **Lagged values** of y_t and **lagged errors**
- Conditions on AR coefficients ensure stationarity
- Conditions on MA coefficients ensure invertibility

Link between ARIMA and ARMA

If y_t follows ARIMA(p, d, q) $\Leftrightarrow$ $(1-B)^d y_t$ follows ARMA(p, q)

Special Cases of ARIMA Models

Model Name	Model Parameters
White Noise	**ARIMA(0, 0, 0)** with no constant
Random Walk	**ARIMA(0, 1, 0)** with no constant
Random Walk with Drift	**ARIMA(0, 1, 0)** with a **constant**
Autoregression	**ARIMA(p, 0, 0)**
Moving Average	**ARIMA(0, 0, q)**

Relationship of c, d with long term forecasts

- Note that the p, q part affects the short term look of the data and the c, d affects how the long term forecasts look like

c	d	Long Term Forecast Pattern
0	0	Long-term forecasts goes to 0
0	1	Long-term forecasts will go to a non-zero constant
0	2	Long-term forecasts will follow a straight line
$\neq 0$	0	Long-term forecasts will go to the mean of the data
$\neq 0$	1	Long-term forecasts will follow a straight line
$\neq 0$	2	Long-term forecasts will follow a quadratic trend

Cyclic Behaviours

- Note that this is something unique to ARIMA models that ETS models can't do. ETS models can only handle seasonality and not cyclic patterns
- **Conditions**:
  - $p \geq 2$ and some restrictions on coefficients are required
  - If $p = 2$, we need $\phi_1^2 + 4\phi_2 < 0$. Then average cycle of length will be

$$(2\pi)/[\arccos\left(-\phi_1(1-\phi_2)/4\phi_2\right)]$$

▼ **Notations**

p - Refers to the AR part

d - Refers to the degree of differencing involved

q - Refers to the MA part

▼ **Determining p, d, q**

- d - We can repeatedly use KPSS test to decide on d
- p - Finding the last significant lag for **PACF** plot (Note that this is only for **pure AR models**)
- q - Finding the last significant lag for **ACF** plot (Note that this is only for **pure MA models**)

▼ **Code Output**

```
fit <- global_economy |>
  filter(Code == "EGY") |>
  model(ARIMA(Exports))
report(fit)
```

```
## Series: Exports
## Model: ARIMA(2,0,1) w/ mean
##
## Coefficients:
##          ar1      ar2     ma1  constant
##        1.676  -0.8034  -0.690     2.562
## s.e.   0.111   0.0928   0.149     0.116
##
## sigma^2 estimated as 8.046:  log likelihood=-142
```

ARIMA(2,0,1) model:

$$y_t = 2.56 + 1.68y_{t-1} - 0.80y_{t-2} - 0.69\varepsilon_{t-1} + \varepsilon_t,$$

where ε_t is white noise with a standard deviation of 2.837 = $\sqrt{8.046}$.

Partial Autocorrelations

- Measures the relationship between y_t and y_{t-k}, when the effects of other time lags - $1, 2, 3, \cdots, k-1$ are removed

▼ **Definition**

Given a time series y_t, the partial autocorrelation of lag k is denoted, $\phi_{k,k}$, is the autocorrelation between y_t and y_{t+k} with the linear dependence of y_t on y_{t+1} to

y_{t+k-1} removed.

It is the autocorrelation between y_t and y_{t+k} that is not accounted for by lags 1 through $k-1$ inclusive

$$\phi_{1,1} = \rho(z_{t+1}, z_t), \qquad \text{for } k = 1$$
$$\phi_{k,k} = \rho(z_{t+k} - \hat{z}_{t+k}, z_t - \hat{z}_t), \qquad \text{for } k \geq 2$$

- where $\hat{z}_{t+k}, \hat{z}_t$ are linear combinations of $\{z_{t+1}, z_{t+2}, \cdots, z_{t+k-1}\}$ that minimises the mean squared error of z_{t+k} and z_t respectively.

Note: For stationary processes, the coefficients in $\hat{z}_{t+k}$ and $\hat{z}_t$ are the same but reversed

- $\hat{z}_{t+k} = \beta_1 z_{t+k-1} + \cdots + \beta_{k-1} z_{t+1}$
- $\hat{z}_t = \beta_1 z_{t+1} + \cdots + \beta_{k-1} z_{t+k-1}$

▼ **Reasoning for this**

- When we are looking at the correlation between y_t and y_{t-1}, it may induce k lags of autocorrelation since y_{t-1} may be autocorrelated with y_{t-2} and so on

$$\begin{aligned} \alpha_k &= \text{kth partial autocorrelation coefficient} \\ &= \text{estimate of } \phi_k \text{ in the following regression:} \\ y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_k y_{t-k} + e_t \end{aligned}$$

Idea behind this:

- Regression on the various lagged values and taking out the coefficient for the kth lag value since this tells us how much the kth lag explains the current time value

Features:

- $\alpha_1 = \rho_1$
- Same critical values of $\pm 1.96/\sqrt{T}$ as for ACF
- Last significant α_k indicates the order of an **AR Model**

▼ Forecasting

The point forecast $\hat{y}_{t+h|t}$ is chosen to be $g(y_1, \cdots, y_t)$ s.t.

$$E[y_{t+h} - g(y_1, \cdots, y_t)]^2 \qquad \text{is minimised}$$

In general, the appropriate $g(.)$ is

$$E[y_{t+h}|y_1, \cdots, y_t]$$

Prediction Error

$$E[y_{t+h} - \hat{y}_{t+h|t}]^2$$

▼ Example: $AR(p)$

$$\hat{y}_{t+h|t} = \phi_1 \hat{y}_{t+h-1|t} + \phi_2 \hat{y}_{t+h-2|t} + \cdots + \phi_p \hat{y}_{t+h-p|t}$$

▼ Example: $ARMA(p, q)$

$$\hat{y}_{t+h|t} = \phi_1 \hat{y}_{t+h-1|t} + \phi_2 \hat{y}_{t+h-2|t} + \cdots + \phi_p \hat{y}_{t+h-p|t} + \theta_1 \hat{e}_{t+h-1|t} + \theta_q \hat{e}_{t+h-q|t}$$

$$\begin{gathered} \hat{y}_{i|t} = y_i \qquad \text{for } i \leq t \\ \hat{e}_{i|t} = 0 \qquad \text{for } i > t \\ \hat{e}_{i|t} = \tau(B)\hat{y}_{i|t} - \theta_1 \hat{e}_{i-1|t} - \cdots - \theta_q \hat{e}_{i-q|t} \end{gathered}$$

Forecast Variance Relationship with d

- Higher value of d → More rapidly the prediction intervals increase in size
- For $d = 0$, the long-term forecast standard deviation will go to the standard deviation of the historical data

▼ SARIMA

$$\text{ARIMA(p,d,q)(P, D, Q)}_m$$

Model Form

$$\tau(B)(1-B^{12})^D(1-B)^d y_t = c + \kappa(B)e_t$$

$(1-B^{12})$ - Extra differencing term where we are taking the difference between 12 months apart. Note that this is 12 because this is monthly data

$\tau(B)$ - It is a function of 2 polynomials

- $(1+\phi_1 B+\cdots+\phi_q B^q)(1+\Phi_1 B^{12}+\Phi_2 B^{24}+\cdots+\Phi_Q B^{12Q})$
- Note that there are 2 components where one is the non seasonal part and the other is the seasonal part

$\kappa(B)$ - It is a function of 2 polynomials

- $(1+\theta_1 B+\cdots+\theta_q B^q)(1+\theta_1 B^{12}+\theta_2 B^{24}+\cdots+\theta_Q B^{12Q})$
- Note that there are 2 components where one is the non seasonal part and the other is the seasonal part

▼ **Features**

- If it is a pure seasonal model, we can fall back on the same guidelines for identifying ARIMA model. But note that the spikes will be at the seasonal periods

	Pure Seasonal AR(P)	**Pure Seasonal MA(Q)**	**Pure Seasonal ARMA(P, Q)**
ACF	Tail off at lags $k \cdot s$	Cut off after $Q \cdot s$	Tails off
PACF	Cut off after lag $P \cdot s$	Tail off at $k \cdot s$	Tails off

▼ **Example 1**

$SARIMA(0,0,0)(1,0,1)_{12}$

$$\begin{aligned} y_t &= \Phi y_{t-12} + e_t + \Theta e_{t-12} \\ y_t - \Phi y_{t-12} &= e_t + \Theta e_{t-12} \\ (1-\Phi B^{12})y_t &= (1+\Theta B^{12})e_t \\ \tau(B)y_t &= \kappa(B)e_t \end{aligned}$$

Note that since we do note have any components that are non multiples of 12 (there are no non seasonal components)

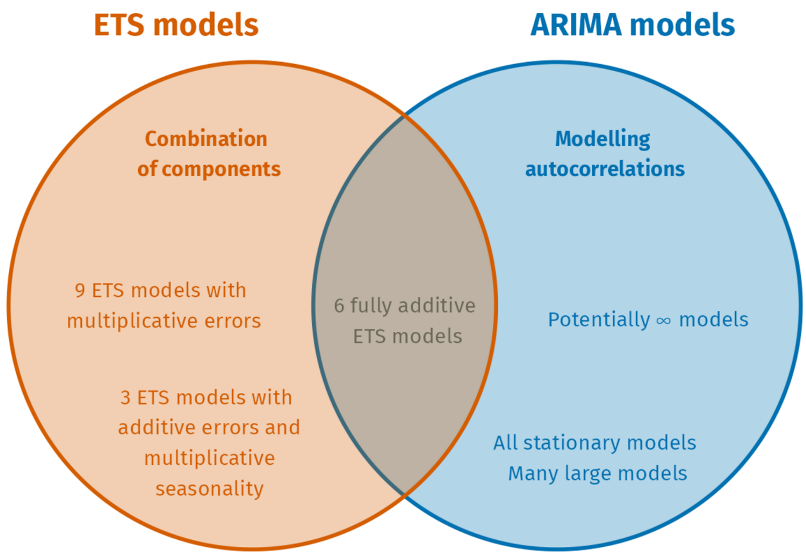
- This is a pure seasonal model, because there are no non-seasonal parameters

▼ Example 2

$SARIMA(0,1,1)(0,1,1)_{12}$

$$
\begin{gathered}
y_t - y_{t-1} - y_{t-12} - y_{t-13} = e_t + \theta e_{t-1} + \Theta e_{t-12} - \theta\Theta e_{t-13} \\
(1-B)y_t - (1-B)(B^{12})y_t = (1+\theta B)e_t + \Theta B^{12} B e_t + \theta\Theta B^{12} e_t \\
(1-B)(1-B^{12})y_t = (1+\theta B)e_t + \Theta B^{12}(1+\theta B)e_t \\
(1-B)(1-B^{12})y_t = (1+\theta B)(1+\Theta B^{12})e_t
\end{gathered}
$$

▼ ETS vs ARIMA



Equivalence relationship between ETS and ARIMA models

ETS Model	ARIMA Model	Parameters
ETS(A,N,N)	ARIMA(0,1,1)	$\theta_1 = \alpha - 1$
ETS(A,A,N)	ARIMA(0,2,2)	$\theta_1 = \alpha + \beta - 2$ $\theta_2 = 1 - \alpha$
ETS(A, A_d, N)	ARIMA(1,1,2)	$\phi_1 = \phi$ $\theta_1 = \alpha + \phi\beta - 1 - \phi$ $\theta_2 = (1-\alpha)\phi$
ETS(A,N,A)	ARIMA(0,1,m)$(0,1,0)_m$	

ETS Model	ARIMA Model	Parameters
ETS(A,A,A)	ARIMA(0,1,m+1) $(0,1,0)_m$	
ETS(A,A_d,A)	ARIMA(1,0,m+1) $(0,1,0)_m$	

- Note that once there are multiplicative terms, there will not be any equivalent ARIMA model (i.e. **Non-linear** exponential smoothing models do not have equivalent ARIMA counterparts)

Comparison between Models

- Note that since ETS and ARIMA are 2 different class of models, we can only compare their forecast accuracy. We **cannot** compare their AIC since they have different number of parameters and will not give a good enough estimate.

▼ **ETS Models**

- Combination of different components (**Error, Trend, Seasonality**)
- Flexible model since it allows for trend etc to change over time
- All **non-stationary**.
- Models with seasonality or non-damped trend (or both) have two unit roots; all other models have one unit root
    - Number of roots = Number of differencing required

▼ **ARIMA Models**

- Modelling autocorrelations after differencing
- Potentially ∞ number of models
- All stationary models & many large models

▼ **Example: Equivalence of $ETS(A,N,N)$ and $ARIMA(0,1,1)$**

$$y_t = l_{t-1} + e_t$$
$$l_t = l_{t-1} + \alpha e_t$$

Conversion

$$l_t - l_{t-1} = \alpha e_t$$

From $(*)$

$$\begin{aligned}(1-B)y_t &= (1-B)l_{t-1} + (1-B)e_t \\ &= \alpha e_{t-1} + e_t - e_{t-1} \\ (1-B)y_t &= e_t + (\alpha-1)e_{t-1}\end{aligned}$$

▼ **Invertibility Condition**

$$\begin{aligned}&|\alpha - 1| < 1 \\ &\Leftrightarrow 0 < \alpha < 2\end{aligned}$$

▼ **Example: Equivalence of $ETS(A,N,N)$ and $ARIMA(0,2,2)$**

$$\begin{aligned}y_t &= \ell_{t-1} + b_{t-1} + e_t \qquad \cdots(*) \\ \ell_t &= \ell_{t-1} + b_{t-1} + \alpha e_t \Rightarrow (1-B)\ell_t = b_{t-1} + \alpha e_t \\ b_t &= b_{t-1} + \beta e_t \Rightarrow (1-B)b_t = \beta e_t\end{aligned}$$