M7
Predictive Analytics

ENİS KAYIŞ

ARIMA Models

- Time series forecasting
  - Moving Average
  - Exponential Smoothing
    - Trend and seasonality
  - ARIMA models
    - Autocorrelations

Stationarity

- A stationary time series is one whose properties do not depend on the time at which the series is observed.

Definition

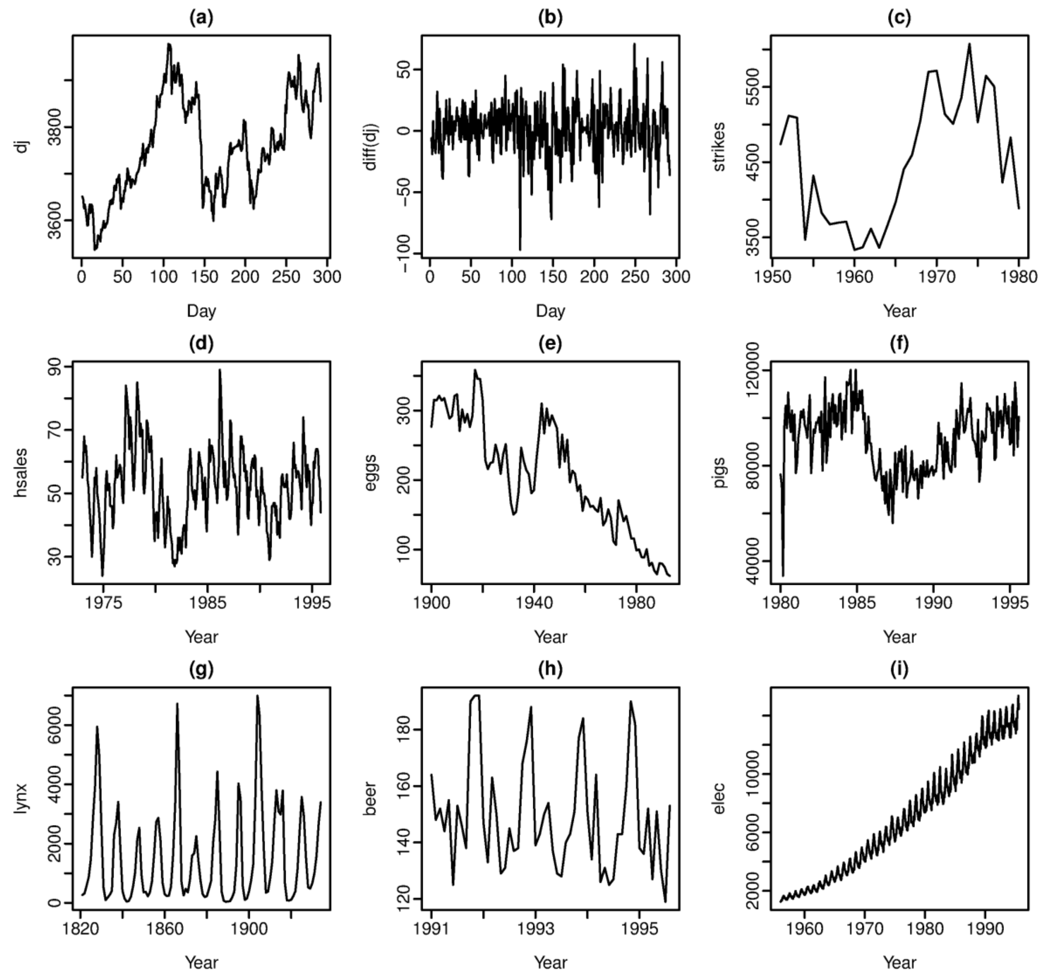
If $\{y_t\}$ is a stationary time series, then for all s, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on t.

- A **stationary series** is:
  - roughly horizontal
  - constant variance
  - no patterns predictable in the long-term
- What is NOT stationary?
  - Time series with trends
  - Time series with seasonality
- What is stationary?
  - Time series with cycles (cannot predict their time and duration)

ÖZYEĞİN UNIVERSITY

Stationarity

- Which of the following are stationary?
  - *(a) Dow Jones index on 292 consecutive days;*
  - *(b) Daily change in Dow Jones index on 292 consecutive days;*
  - *(c) Annual number of strikes in the US;*
  - *(d) Monthly sales of new one-family houses sold in the US;*
  - *(e) Price of a dozen eggs in the US (constant dollars);*
  - *(f) Monthly total of pigs slaughtered in Victoria, Australia;*
  - *(g) Annual total of lynx trapped in the McKenzie River district of north-west Canada;*
  - *(h) Monthly Australian beer production;*
  - *(i) Monthly Australian electricity production.*

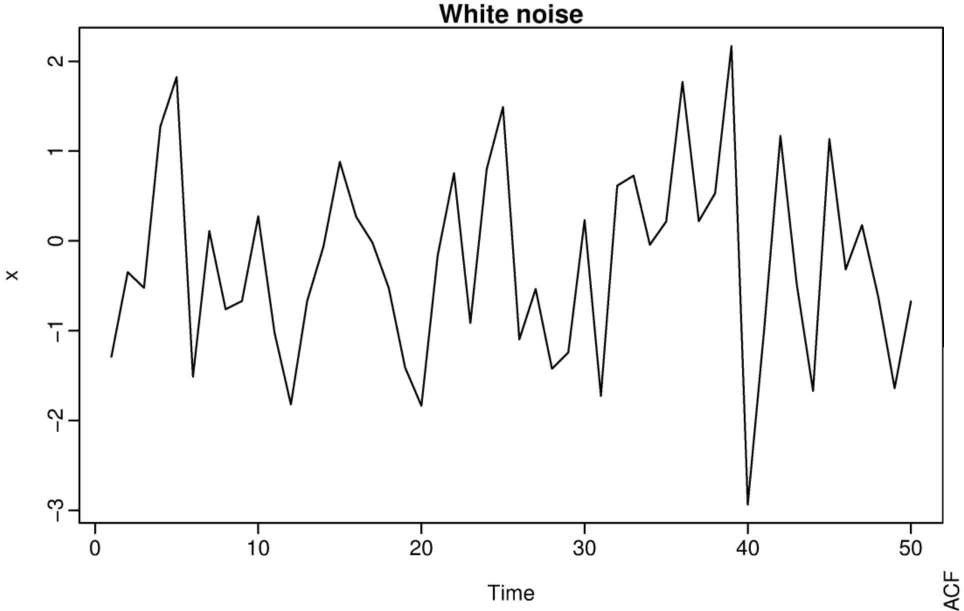


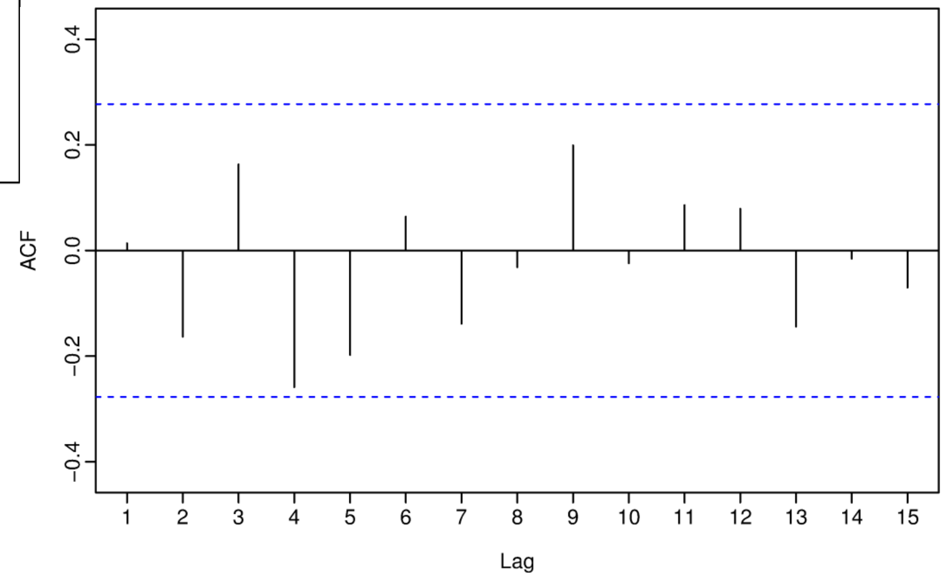
Stationarity

- Transformations help to **stabilize the variance**.
- For ARIMA modelling, we also need to **stabilize the mean**.
- Identifying non-stationary series:
  - Time plot.
  - The ACF of stationary data drops to zero relatively quickly.
  - The ACF of non-stationary data decreases slowly.
  - For non-stationary data, the value of r_1 is often large and positive.

Stationarity

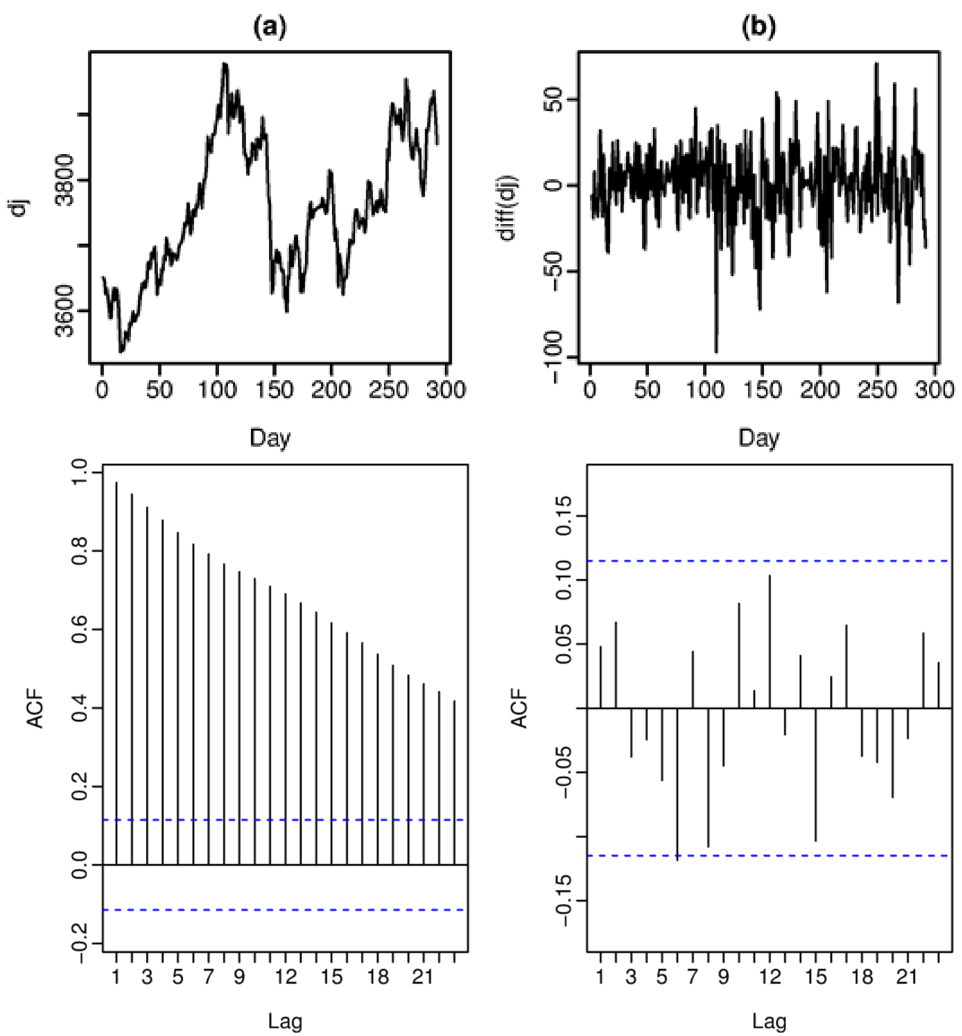
- A special time series:
  - White noise: no autocorrelations





Differencing

- Compare (a) and (b):
- One solution to make a series stationary: differencing:
  - Compute the differences between consecutive observations.
  - It can eliminate trend and seasonality
  - First, look at the ACF plot.



Random Walk Model

- The differenced series is the *change* between consecutive observations in the original series:

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

- When the differenced series is white noise, the model for the original series can be written as:

$$y_t - y_{t-1} = e_t$$

- Widely used in finance and economics data. They have:
  - long periods of apparent trends up or down
  - sudden and unpredictable changes in direction.
- What is the best forecast?
- Random walk with drift:

$$y_t - y_{t-1} = c + e_t$$

Other types of differencing

- If the differenced data is not stationary, and it may be necessary to difference the data a second time to obtain a stationary series:

$$\begin{aligned} y''_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2}. \end{aligned}$$

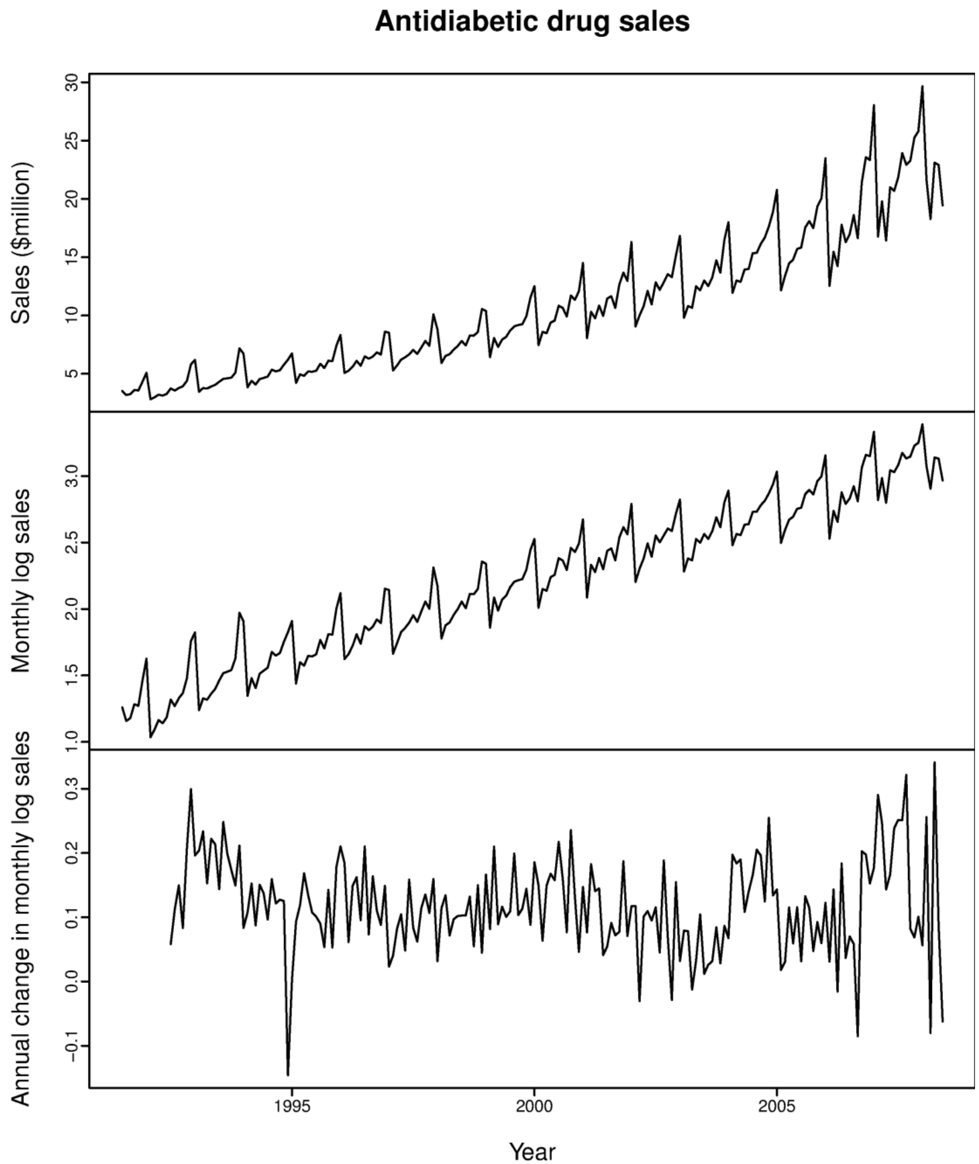
- In practice, it is almost never necessary to go beyond second-order dierences.
- Lag-m differencing for seasonality:

$$y'_t = y_t - y_{t-m} \qquad \text{where } m = \text{number of seasons.}$$

  - First differences: the change between **one observation and the next**
  - Seasonal differencing: the change between **one year to the next**

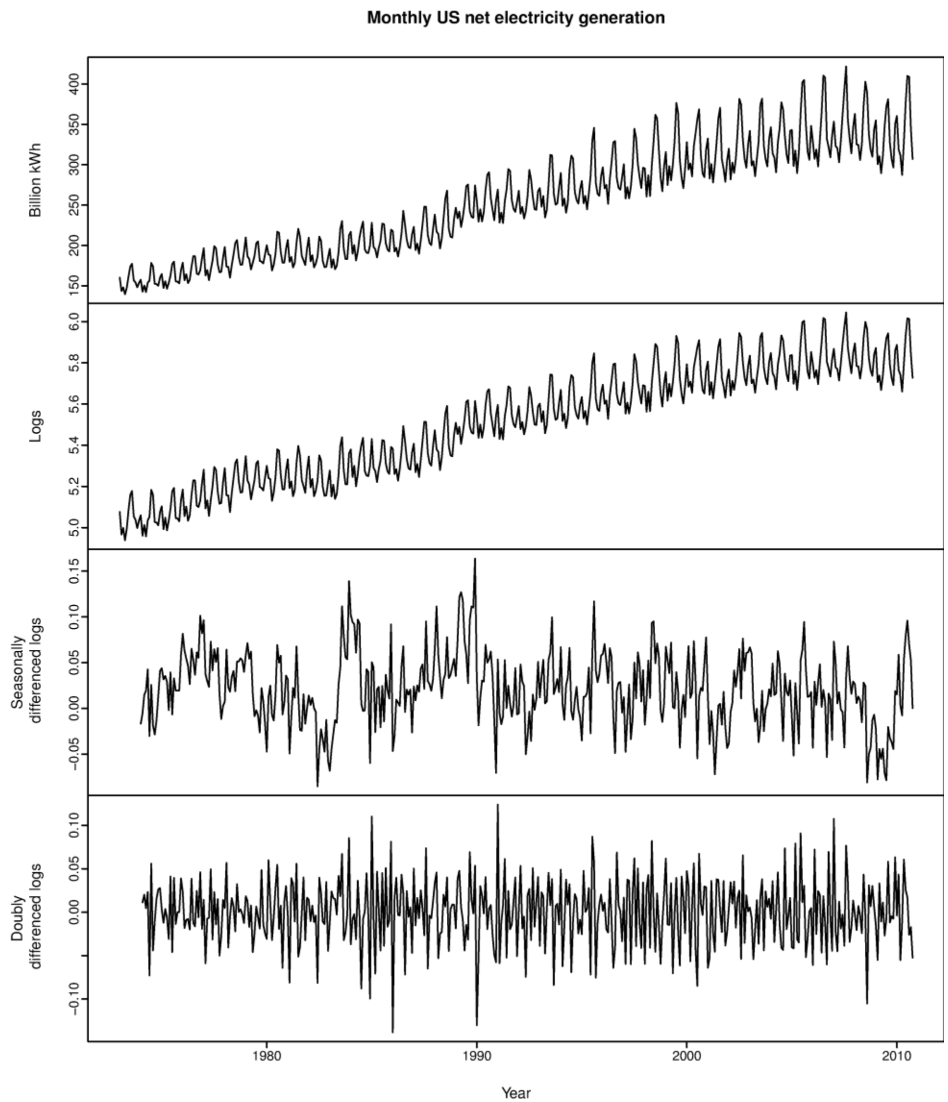
Differencing

- Example: Antidiabetic sales data (monthly)
- Observe each step:
  1. Take logarithm of the original series
  2. Compute lag-12 differences of (1)



Differencing

- Example 2: Electricity generation data (monthly)
- Observe each step:
  1. Take logarithm of the original series
  2. Compute lag-12 differences of (1)
  3. Compute first differences of (2)



Differencing

- When both seasonal and first differences are applied. . .
  - it makes no difference which is done first—the result will be the same.
  - If seasonality is strong, it is recommended that seasonal differencing be done first because sometimes the resulting series will be stationary and there will be no need for further first difference.
- Unit root tests:
  - Statistical tests to determine the required order of differencing.
  - Many are available, with possibly conflicting answers
    - Augmented Dickey Fuller test: null hypothesis is that the data are non-stationary and non-seasonal.
    - Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test: null hypothesis is that the data are stationary and non-seasonal.\
    - Other tests available for seasonal data.

Some notation

- Backshift Notation

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

- B operating on, y_t , has the effect of **shifting the data back one period**.

$$B(By_t) = B^2 y_t = y_{t-2}$$

- Seasonal lags: $B^{12} y_t = y_{t-12}$
- First differencing: $y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t$
- Second order differencing:

$$y''_t = y_t - 2y_{t-1} + y_{t-2} = (1 - 2B + B^2)y_t = (1 - B)^2 y_t \, .$$

- D^{th} order differencing: $(1 - B)^d y_t$

Some notation

- Second-order *difference* is denoted by $(1 - B)^2$
- *Second-order difference* is not the same as a *second difference*, which would be denoted by $1 - B^2$
- In general, a d^{th}-order *difference* can be written as $(1 - B)^d y_t$
- A seasonal *difference* followed by a first *difference* can be written as $(1 - B) \quad (1 - B^m) \quad y_t$

Autoregressive models

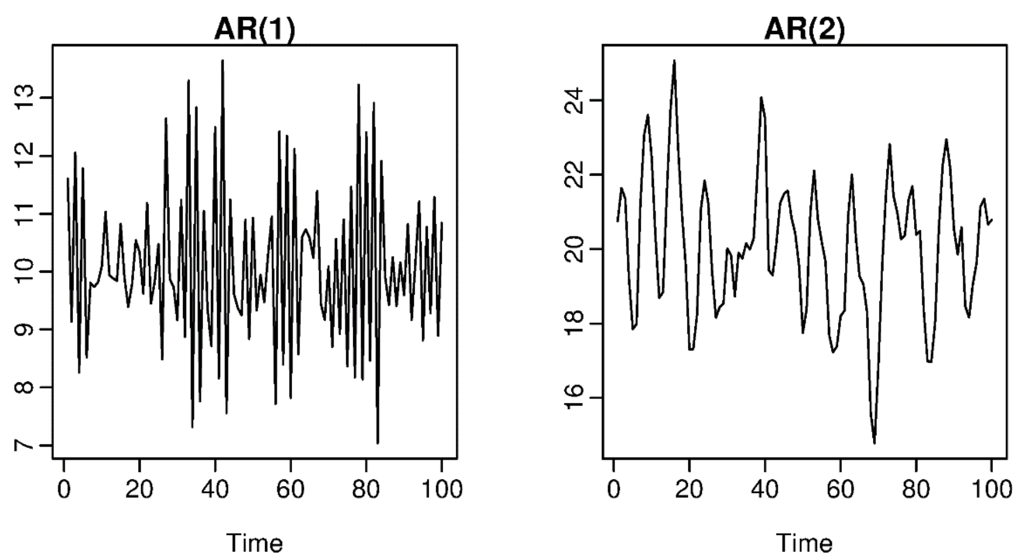
- Multiple Regression: forecast the variable of interest using a linear combination of predictors
- Autoregression: forecast the variable of interest using a linear combination of *past values of the variable*
  - An autoregressive model of order p, AR(p):

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

  - where e_t is white noise.
- Examples:

a) $y_t = 18 - 0.8y_{t-1} + e_t$

b) $y_t = 8 + 1.3y_{t-1} - 0.7y_{t-2} + e_t$



Autoregressive models

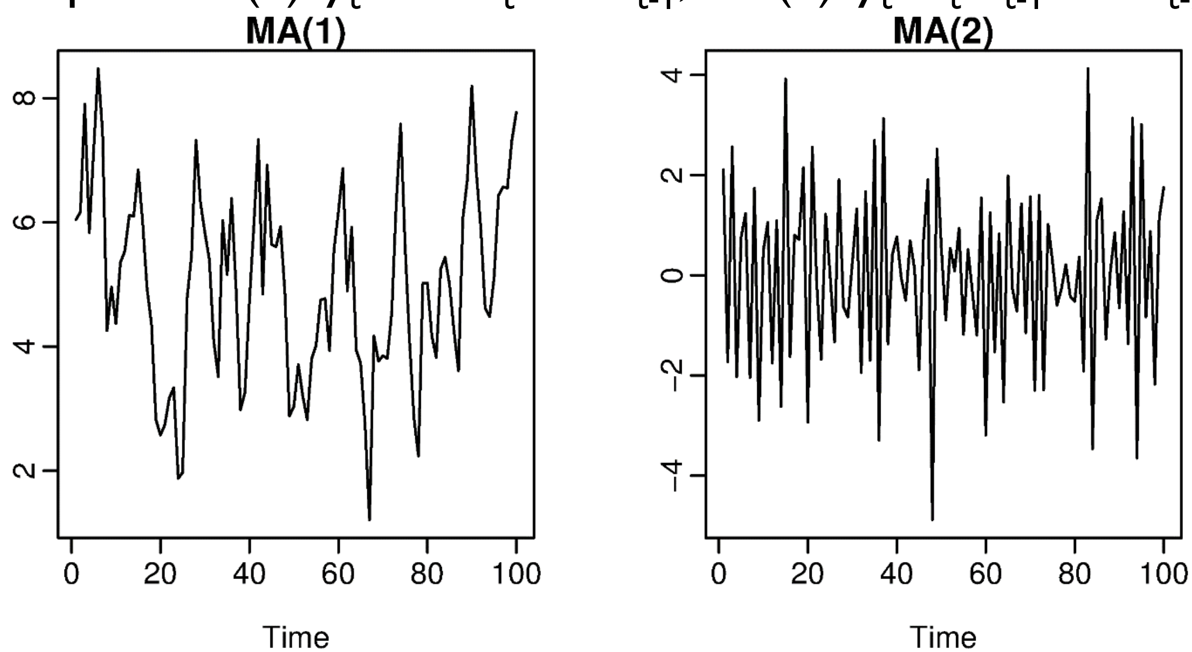
- For an AR(1) model: $y_t = c + \phi_1 y_{t-1} + e_t$
  - $\phi_1 = 0$, then the series is WN
  - $\phi_1 = 1$ and $c = 0$, then the series is RW
  - $\phi_1 = 1$ and $c \neq 0$, then the series is RW with drift
  - $\phi_1 < 0$, then the series oscillate between + and – values.
- Constraints in an AR model to ensure stationarity:
  - For an AR(1) model: $-1 < \phi_1 < 1$.
  - For an AR(2) model: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.
  - More complicated with higher orders, but R and Python takes care of them.

Moving Average Models

- Use past *forecast errors*, instead of past values.
  - An MA(q) model:

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

  - Interpretation: y_t is the weighted moving average of the past few forecast errors.
    - Example: MA(1): $y_t=20+e_t+0.8e_{t-1}$, MA(2): $y_t=e_t-e_{t-1}+0.8e_{t-2}$



AR and MA models

- Any AR(p) model could be written as an MA(∞) model.
  - Example: for an AR(1) model (assume c=0 for simplicity):

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

  - Provided $-1<\phi_1<1$, the value of ϕ_1^k will get smaller as k gets larger. So eventually we obtain

$$y_t = e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \cdots$$

  - Which is MA(∞) process.
- Similarly, an MA(q) model could be inverted to an AR(∞) model if:
  - For an MA(1) model: $-1 < \theta_1 < 1$.
  - For an MA(2) model: $-1 < \theta_2 < 1, \;\; \theta_2 + \theta_1 > -1, \;\; \theta_1 - \theta_2 < 1$.

ARIMA models

- Combination of differencing with AR and MA models:
  - AutoRegressive Integrated Moving Average.

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t,$$

    - Where y_t' is the differenced series (any number of times)
  - An ARIMA(p,d,q) model:
    - p: order of the autoregressive part
    - d: degree of first differencing involved
    - q: order of the moving average part
  - Alternative formulation using backshift notation:

$$\underset{\substack{\uparrow \\ \text{AR}(p)}}{(1 - \phi_1 B - \cdots - \phi_p B^p)} \; \underset{\substack{\uparrow \\ d \text{ differences}}}{(1 - B)^d} y_t = c + \underset{\substack{\uparrow \\ \text{MA}(q)}}{(1 + \theta_1 B + \cdots + \theta_q B^q)} e_t$$

ARIMA models

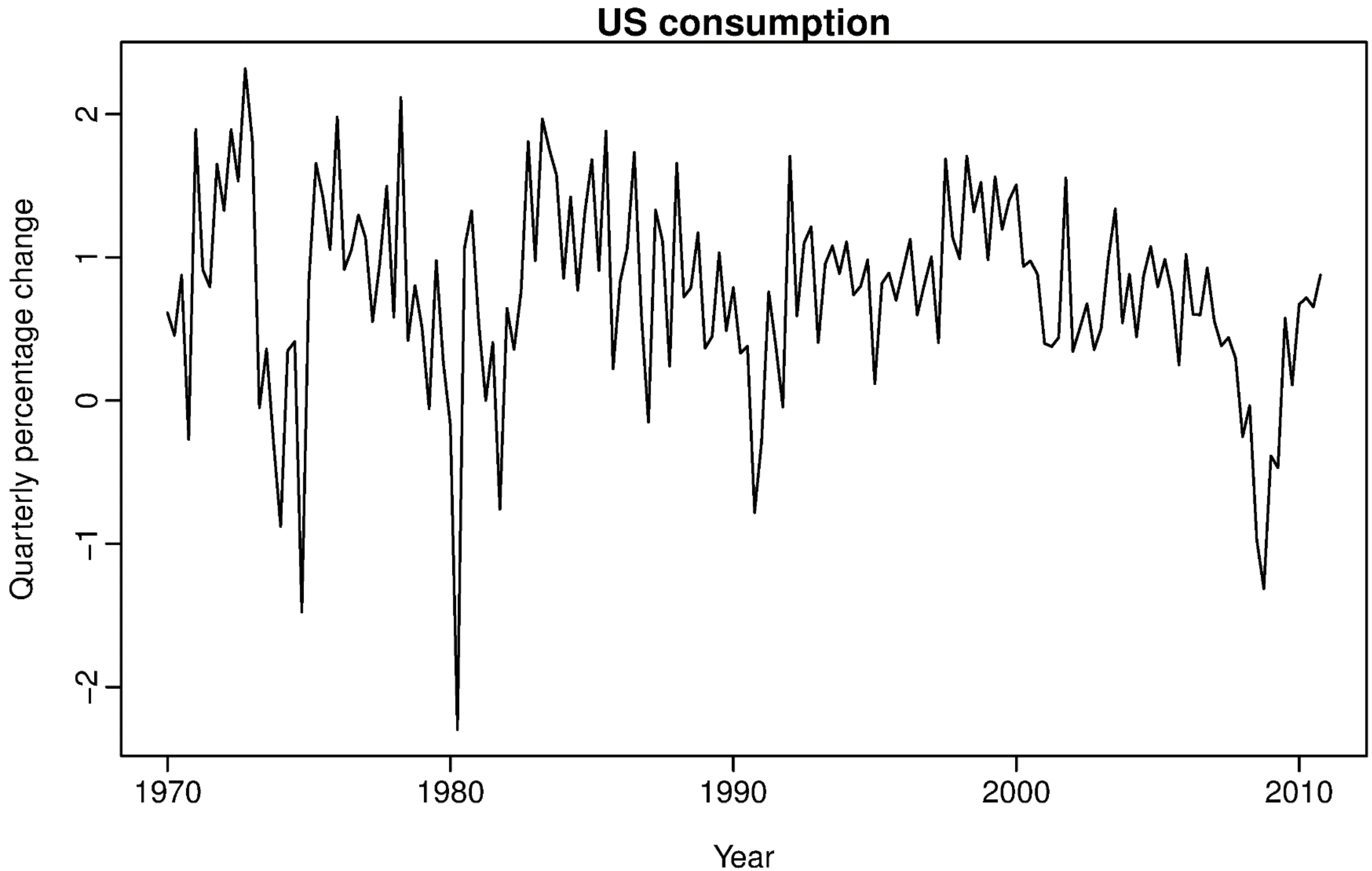
- Some special cases:

White noise	ARIMA(0,0,0)
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)

- How to select calues for p, d, and q?
  - auto.arima() function in R or Python.

ARIMA models

- Example: US personal consumption (quarterly, percentage change)



ARIMA models

- Mathematical (our) formulation:

Intercept form

$$(1 - \phi_1 B - \cdots - \phi_p B^p) y'_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

- R or Python (ARIMA of statsmodels, **NOT** auto_arima of pmdarima) formulation:

Mean form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(y'_t - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

- $y'_t = (1 - B)^d y_t$
- μ is the mean of y'_t.
- $c = \mu(1 - \phi_1 - \cdots - \phi_p)$.

ARIMA models

```
In [74]: Uschange.tail()
         Uschange_arima = sm.tsa.ARIMA(Uschange, order=(2, 0, 2)).fit(trend='c')
         print(Uschange_arima.summary())
```

```
                              ARMA Model Results
==============================================================================
Dep. Variable:            Consumption   No. Observations:                  187
Model:                     ARMA(2, 2)   Log Likelihood                -165.142
Method:                       css-mle   S.D. of innovations              0.585
Date:                Fri, 03 May 2019   AIC                            342.284
Time:                        06:51:04   BIC                            361.671
Sample:                    03-31-1970   HQIC                           350.140
                         - 09-30-2016
=======================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                   0.7463      0.085      8.830      0.000       0.581       0.912
ar.L1.Consumption       1.3908      0.255      5.444      0.000       0.890       1.891
ar.L2.Consumption      -0.5811      0.208     -2.796      0.006      -0.989      -0.174
ma.L1.Consumption      -1.1799      0.238     -4.951      0.000      -1.647      -0.713
ma.L2.Consumption       0.5582      0.140      3.978      0.000       0.283       0.833
```

ARIMA(2,0,2) model:

$$y_t = c + 1.391y_{t-1} - 0.581y_{t-2} - 1.180\varepsilon_{t-1} + 0.558\varepsilon_{t-2} + \varepsilon_t$$

where c=0.746*(1-1.391+0.581)=0.142 and ε_t is white noise with a standard deviation of 0.585.

ACF versus PACF plots

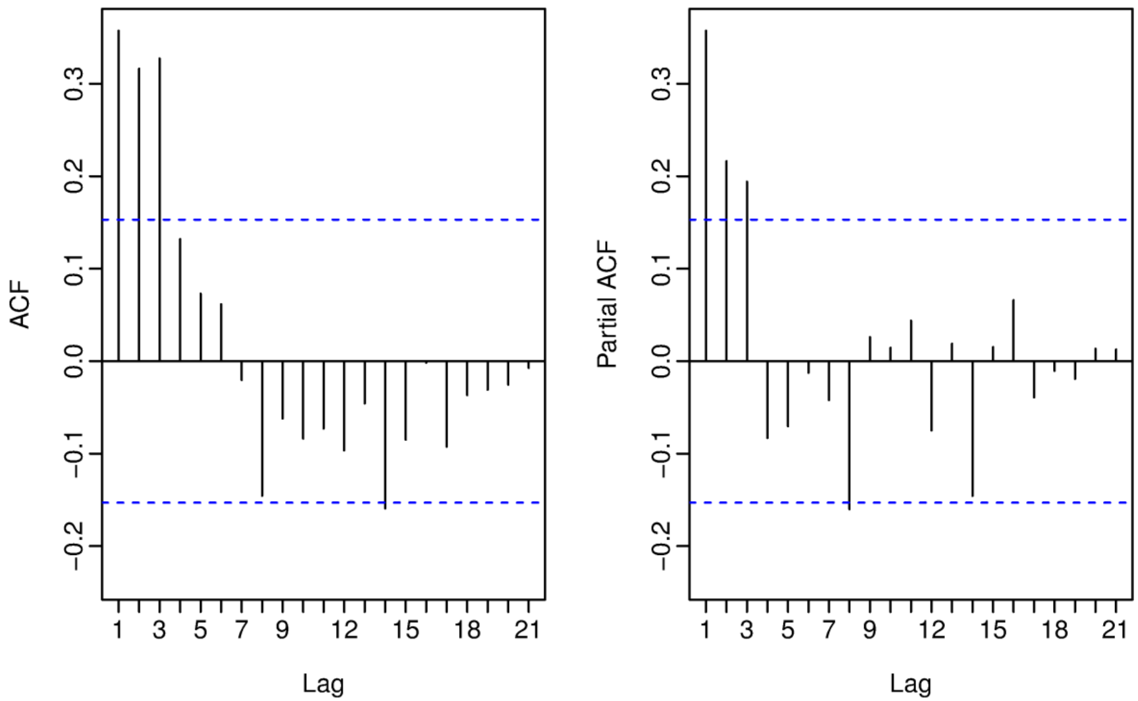
- It is usually not possible to tell, simply from a time plot, what values of p and q are appropriate for the data.
- Sometimes the ACF plot, and the closely related PACF plot, could help.
- Remember ACF, lag 1:
  - Correlation between y_t and y_{t-1}, and y_{t-1} and y_{t-2} ,…
  - So, y_{t-2} and y_t could be correlated somewhat as well.
  - Solution: Partial Autocorrelations:
    - Measure the relationship between y_t and y_{t-k} after removing other time lags, 1,2,…,k-1.
    - Note: ACF(1)=PACF(1)

ACF versus PACF plots

- How to compute PACF?

$$\alpha_k = k\text{th partial autocorrelation coefficient}$$
$$= \text{the estimate of } \phi_k \text{ in the autoregression model}$$
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_k y_{t-k} + e_t.$$

- Example: US consumption data

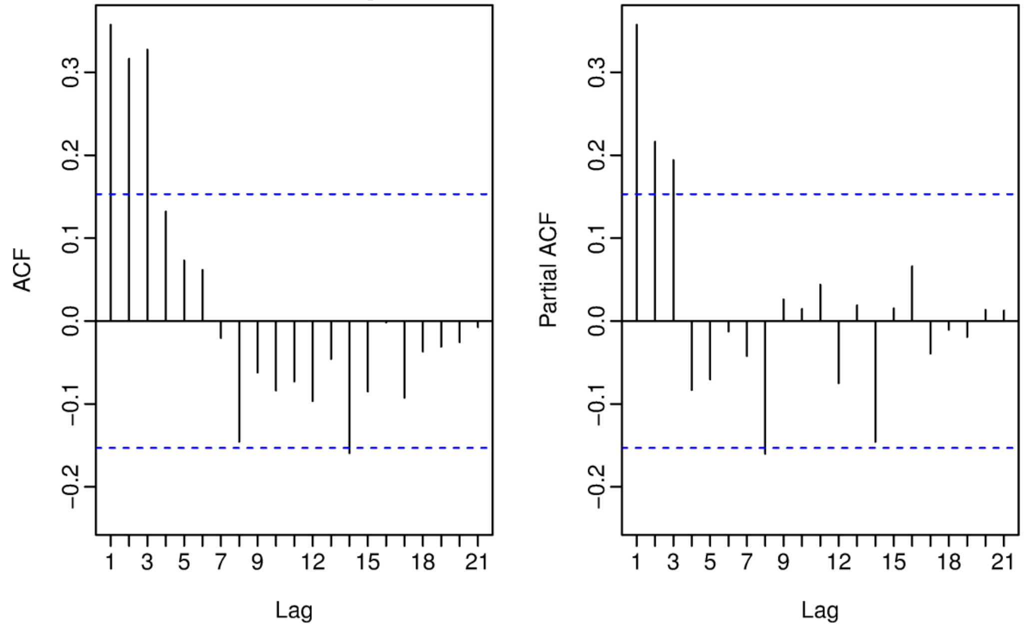


ACF versus PACF plots

- How to determine ARIMA(p,d,q) from ACF and PACF?
- The data may follow an ARIMA(p,d,0) model if the differenced data show the following patterns:
  - the ACF is exponentially decaying or sinusoidal;
  - there is a significant spike at lag p in PACF, but none beyond lag p.
- The data may follow an ARIMA(0,d,q) model if the differenced data show the following patterns:
  - the PACF is exponentially decaying or sinusoidal;
  - there is a significant spike at lag q in ACF, but none beyond lag q.
- Cannot easily determine ARIMA(p,d,q) if both p and q are positive.

ACF versus PACF plots

- Example: Can you determine the right ARIMA model based on the following ACF and PACF plots?



  - Three spikes in the ACF and then no significant spikes thereafter (apart from one just outside the bounds at lag 14)
  - Three spikes in the PACF decreasing with the lag, and then no significant spikes thereafter (apart from one just outside the bounds at lag 8)
  - We can ignore one significant spike in each plot if it is just outside the limits, and not in the first few lags.
    - The probability of a spike being significant by chance is about 1/20, and we are plotting 21 in each.

ARIMA Model Selection

- R and Python uses Maximum Likelihood Estimation (MLE)
  - Find the coefficients to maximize:
    - log likelihood: log(Prob(observed data is from the estimated model))
  - We have three criteria:

$$\mathrm{AIC} = -2\log(L) + 2(p+q+k+1)$$

$$\mathrm{AIC_c} = \mathrm{AIC} + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2}$$

$$\mathrm{BIC} = \mathrm{AIC} + \log(T)(p+q+k-1)$$

  - $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$.
  - Good models are obtained by minimizing either the AIC, $\mathrm{AIC_c}$, or BIC.
  - It is preferred to use the $\mathrm{AIC_c}$.

ARIMA Modeling

- Follow these steps to find the right model.
  - Caution: Do not simply use auto.arima to your raw data.

1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (e.g., log, polynomials) to stabilize the variance.
3. If the data are non-stationary: take first differences of the data until the data are stationary.
4. Examine the ACF/PACF: Is an AR(p) or MA(q) model appropriate?
5. Try your chosen model(s), and use the AICc to search for a better model.
6. Check the residuals from your chosen model by plotting the ACF of the residuals. If they do not look like white noise, try a modified model.
7. Once the residuals look like white noise, calculate forecasts.

ARIMA Modeling

- See the Jupyter Notebook example at LMS

How does auto.arima() work?

A non-seasonal ARIMA process

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t$$

Need to select appropriate orders: p, q, d

- **Hyndman and Khandakar (JSS, 2008) algorithm:**
  - Select no. differences d and D via KPSS test and seasonal strength measure.
  - Select p, q by minimizing AICc.
  - Use stepwise search to traverse model space.

How does auto.arima() work?

$$\text{AICc} = -2\log(L) + 2(p+q+k+1)\left[1 + \frac{(p+q+k+2)}{T-p-q-k-2}\right].$$

where L is the maximised likelihood fitted to the *differenced* data, $k = 1$ if $c \neq 0$ and $k = 0$ otherwise.

Step1: Select current model (with smallest AICc) from:

ARIMA$(2, d, 2)$
ARIMA$(0, d, 0)$
ARIMA$(1, d, 0)$
ARIMA$(0, d, 1)$

Step 2: Consider variations of current model:

- vary one of p, q, from current model by ± 1;
- p, q both vary from current model by ± 1;
- Include/exclude c from current model.

Model with lowest AICc becomes current model.

Repeat Step 2 until no lower AICc can be found.

Modeling with Arima

- Follow these steps to find the right model.
  - Caution: Do not simply use auto.arima to your raw data.

1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (e.g., log, polynomials) to stabilize the variance.
3. If the data are non-stationary: take first differences of the data until the data are stationary.
4. Examine the ACF/PACF: Is an AR(p) or MA(q) model appropriate?
5. Try your chosen model(s), and use the AICc to search for a better model.
6. Check the residuals from your chosen model by plotting the ACF of the residuals. If they do not look like white noise, try a modified model.
7. Once the residuals look like white noise, calculate forecasts.

Modeling with auto_arima()

- Follow these steps to find the right model.
  - Caution: Do not simply use auto.arima to your raw data.

1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (e.g., log, polynomials) to stabilize the variance.

3. Use auto.arima to select a model.

6. Check the residuals from your chosen model by plotting the ACF of the residuals. If they do not look like white noise, try a modified model.
7. Once the residuals look like white noise, calculate forecasts.

Forecasting with ARIMA

- Point Forecasts: Follow these steps:
  1. Expand the ARIMA equation so that y_t is on the left hand side and all other terms are on the right.
  2. Rewrite the equation by replacing t by T+h.
  3. On the right hand side of the equation, replace future observations by their forecasts, future errors by zero, and past errors by the corresponding residuals.
- Example: ARIMA(3,1,1) model with h=1 (one step ahead forecast)
  - Using backshift operator, we can write ARIMA(3,1,1) as:

$$(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)(1-B)y_t=(1+\theta_1 B)e_t$$

where the coefficients are available from model selection stage.

Forecasting with ARIMA

- Step 1a: Then expand the series:

$$\left[1-(1+\phi_1)B+(\phi_1-\phi_2)B^2+(\phi_2-\phi_3)B^3+\phi_3B^4\right]y_t=(1+\theta_1B)e_t,$$

- Step 1b: Now, apply backshift operator to get:

$$y_t-(1+\phi_1)y_{t-1}+(\phi_1-\phi_2)y_{t-2}+(\phi_2-\phi_3)y_{t-3}+\phi_3y_{t-4}=e_t+\theta_1e_{t-1}$$

- Step 1c: After rearranging:

$$y_t=(1+\phi_1)y_{t-1}-(\phi_1-\phi_2)y_{t-2}-(\phi_2-\phi_3)y_{t-3}-\phi_3y_{t-4}+e_t+\theta_1e_{t-1}$$

- Step 2: Now, we replace t by T+1

$$y_{T+1}=(1+\phi_1)y_T-(\phi_1-\phi_2)y_{T-1}-(\phi_2-\phi_3)y_{T-2}-\phi_3y_{T-3}+e_{T+1}+\theta_1e_T$$

- Step 3: Replace e_{t+1} by 0 and e_t by the last observed residual, $\hat{e}_t$

$$\hat{y}_{T+1|T}=(1+\phi_1)y_T-(\phi_1-\phi_2)y_{T-1}-(\phi_2-\phi_3)y_{T-2}-\phi_3y_{T-3}+\theta_1\hat{e}_T$$

Forecasting with ARIMA

- What if h=2 (two step ahead forecast)?
  - Follow step1 the same, in step 2, replace t by T+2, and in step 3 replace y_{T+1} by $\hat{y}_{T+1|T}$, e_{t+1} and e_{t+2} by 0:

$$\hat{y}_{T+2|T} = (1+\phi_1)\hat{y}_{T+1|T} - (\phi_1 - \phi_2)y_T - (\phi_2 - \phi_3)y_{T-1} - \phi_3 y_{T-2}.$$

- Find point forecasts for any h similarly.
- Range forecasts:
  - For h=1, the forecast range is $\hat{y}_{T+1|T} \pm 1.96\hat{\sigma}$, where $\hat{\sigma}$ is the standard deviation of the residuals
  - For h>1, it is more complicated, use function forecast(.).
  - BE CAREFUL:
    - If the residuals of ARIMA model is not WN, then forecast range from forecast() is incorrect.
    - The range will be wider for higher h, unless d=0 (in which case the range will be the same after some h).

Prediction intervals

95% Prediction interval

$$\hat{y}_{T+h|T} \pm 1.96\sqrt{v_{T+h|T}}$$

where $v_{T+h|T}$ is estimated forecast variance.

- Multi-step prediction intervals for ARIMA(0,0,q):

$$y_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

$$v_{T|T+h} = \hat{\sigma}^2 \left[1 + \sum_{i=1}^{h-1} \theta_i^2 \right], \qquad \text{for } h = 2, 3, \ldots.$$

- AR(1): Rewrite as MA(∞) and use above result.
- Other models beyond scope of this subject.