MONASH University

MONASH BUSINESS SCHOOL

ETC3550/ETC5550 Applied forecasting

Week 8: ARIMA models

ARIMA models

AR: autoregressive (lagged observations as inputs)
I: integrated (differencing to make series stationary)
MA: moving average (lagged errors as inputs)

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

Stationarity

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.

Stationarity

Definition

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

Transformations help to **stabilize the variance**.

For ARIMA modelling, we also need to **stabilize the mean**.

Differencing

Differencing

- Differencing helps to **stabilize the mean**.
- First differencing: *change* between consecutive observations: $y'_t = y_t - y_{t-1}$.
- Seasonal differencing: *change* between years: $y'_t = y_t - y_{t-m}$.

Your turn

1. Does differencing make the Closing stock price series stationary for Amazon and Apple stocks?
2. What sorts of transformations and differencing are needed to make the `Gas` series from `aus_accommodation` stationary?

Automatic differencing

Statistical tests to determine the required order of differencing.

1. Augmented Dickey Fuller test: null hypothesis is that the data are **non-stationary** and non-seasonal.
2. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test: null hypothesis is that the data are **stationary** and non-seasonal.

Seasonal strength

STL decomposition: $y_t = T_t + S_t + R_t$

Seasonal strength $F_s = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t+R_t)}\right)$

If $F_s > 0.64$, do one seasonal difference.

Automatic differencing

Statistical tests to determine the required order of differencing.

1. Augmented Dickey Fuller test: null hypothesis is that the data are **non-stationary** and non-seasonal.
2. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test: null hypothesis is that the data are **stationary** and non-seasonal.

H_0: non-stationary

H_0: stationary

Seasonal strength

STL decomposition: $y_t = T_t + S_t + R_t$

Seasonal strength $F_s = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)}\right)$

If $F_s > 0.64$, do one seasonal difference.

Your turn

3. Do the unit root tests for the `Gas` series from `aus_accommodation`. Do they give the same numbers of difference as you chose?