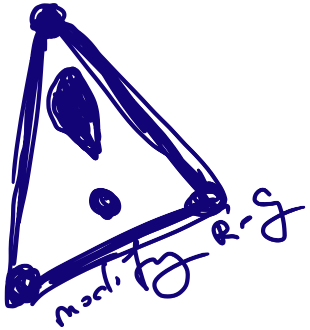


ARINA

Time Series

Peter Lukianchenko

31 January 2022

AR(p) MA(q)

Box-Jenkins procedure

Step 1
- Look at ACF and PACF
- Get an Idea about p and q (and d)

Step 2
- Estimate the candidate models

Step3
- Compute AIC or BIC, choose the best one
- Do diagnostics

Step 4
- Use the chosen model for forecasting

Choosing the best ARMA

- Hard to tell p and q from the picture
- In general, larger p and q ⇒ better fit. But we like models with smaller p and q
- **Akaike's Information Criterion (AIC):**

$$AIC = -\frac{2}{T}\log Likelihood + 2\frac{p+q+1}{T}$$

- **Schwarz's Bayesian Information Criterion (BIC, or SIC)**

$$BIC = -\frac{2}{T}\log Likelihood + \frac{p+q+1}{T}\log T$$

Choosing the best ARMA

- Choose the model with the smallest IC
- Akaike's Information Criterion (AIC):

$$AIC = -\frac{2}{T}\log Likelihood + 2\frac{p+q+1}{T}$$

- Schwarz's Bayesian Information Criterion (BIC, or SIC)

$$BIC = -\frac{2}{T}\log Likelihood + \frac{p+q+1}{T}\log T$$

- BIC is better at choosing the correct model asymptotically
- AIC might be better in small samples
- AIC, in general, aims at choosing a model with better forecasting power.

Direct forecasts

$\theta_2 \cdot y_{t-2}$

- Consider again AR(1): $Y_t = c + \theta_1 Y_{t-1} + \varepsilon_t$
- We **estimate the regression** of Y_{t+2} on Y_t
- $\{u_t\}$ is not a white noise anymore: it's MA(1)!
- The errors are serially correlated – we need the HAC variance estimator

Direct forecasts

- For the AR(1) example from the beginning:

$y_{t+1} = \alpha y_t$

$y_{100} \rightarrow \hat{y}_{101} \rightarrow \hat{y}_{102}$

- $\hat{\alpha} = 5.28,\ \hat{\beta} = 0.49$
- $\hat{y}^{df}_{102|100} = 5.28 + 0.494 \cdot 7.16 = 8.82$
- The iterated forecast was $\hat{y}_{102|100} = 3\,\frac{1-0.7^2}{1-0.7} + 0.7^2 \cdot 7.16 = 8.61$
- $Y_{102} = 8.31$

Forecast error

- We derived them *knowing the model parameters*
- But in real life we don't know them, we need to estimate them
- We use estimates of the coefficients to compute forecasts
- Now the forecast error also contains the error from the estimation of the coefficients
- Keep that in mind

Comparing models

- Several approaches to determine how good the model is
- Might care about how well the model fits the data
- Or, might care about how good is the predictive ability of the model
- Usually, there is a trade-off between the two

In sample fit

- **In-sample fit:** Estimate the model on existing data, look at an information criterion:

$$AIC = -\frac{2}{T}\log Likelihood + 2\frac{p+q+1}{T}$$

$$BIC = -\frac{2}{T}\log Likelihood + \frac{p+q+1}{T}\log T$$

- Show how well the model fits the existing sample

Comparing models

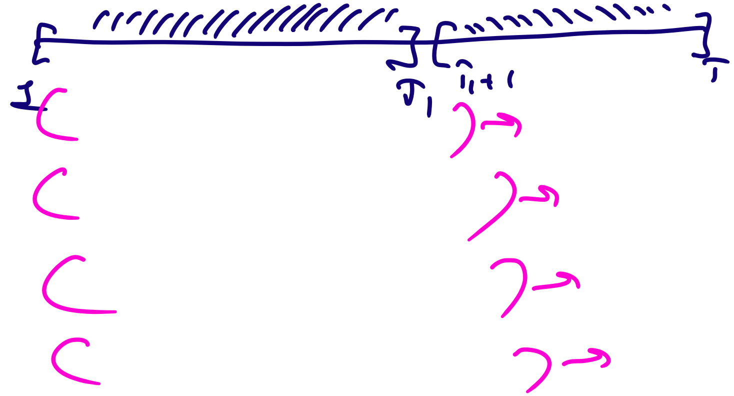
training $(y_0 \dots y_{100})$ test set $(y_{101} - y_{150})$

- **Out-of-sample fit:** See how good are the forecasts based on the model
- Forecasts for the dates we do not observe in the sample and didn't use to estimate our model
- Looks at **predictive ability** of a model
- Can be evaluated by computing the **Mean Squared Prediction Error**:

$$MPSE_h = \frac{1}{T_p} \sum_{j=1}^{T_p} (Y_j - \hat{Y}_{j|j-h})^2$$

Pseudo-out-of-sample fit

- Can check **pseudo-out-of-sample fit**:
- Take the sample of size T and split into two parts: periods $1, \ldots, T_1$ (for some T_1) and $T_1 + 1, \ldots, T$.
- Estimate candidate models:
  - Fixed scheme: on $1, \ldots, T_1$
  - Rolling scheme: on $i, \ldots, T_1 + i - 1$
  - Recursive scheme: on $1, \ldots, T_1 + i - 1$



- Use the estimated model to forecast for period $T_1 + i$, for $i = 1, \ldots, T - T_1$.
- Compute MSPE and compare models.

Diebold-Mariano test (matched pairs Test) (2-sample)

- Compares two sequences of forecasts: $\{\hat{Y}_{1t}\}$ and $\{\hat{Y}_{2t}\}$
- Forecasts are the primitives, not models
- Look at the loss differential:

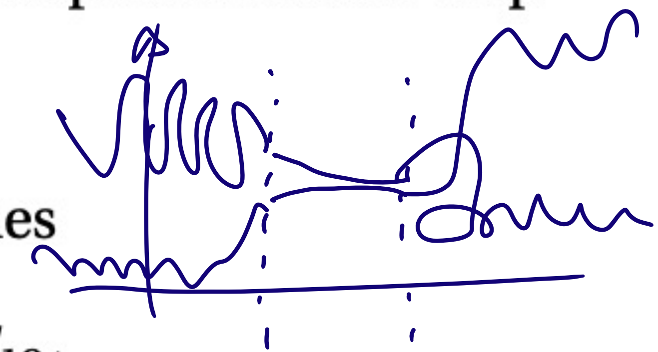
  $d_{12t} = L(e_{1t}) - L(e_{2t}) = (Y_t - \hat{Y}_{1t})^2 - (Y_t - \hat{Y}_{2t})^2$
- Assumption DM: $\{d_{12t}\}$ is covariance-stationary
- Two forecasts are equally good if $\mathrm{E}[d_{12t}] = 0$. That's H_0.
- Form the test statistic:

$$t = \frac{\frac{1}{T}\sum_{t=1}^{T} d_{12t}}{\sqrt{\hat{\sigma}_d / T}},$$

  where $\sigma_d = \sum_{j=-\infty}^{+\infty} \gamma_d(j)$
- $t \to^d \mathcal{N}(0,1)$
- If $t < -z_\alpha$, $\{\hat{Y}_{1t}\}$ is preferable; if $t > z_\alpha$, $\{\hat{Y}_{2t}\}$ is preferable.

West and Clark+McCracken

- Use DM test to investigate pseudo-out-of-sample fit for one-step ahead forecasts



- Estimate the model, using one of the schemes
- Be smart about estimating the variance of d_{12t}
- Be careful about whether the compared models are nested
- Be careful about the relative size of in-sample part and pseudo-out-of-sample parts

AR(p)

- $Y_t = \theta Y_{t-1} + \varepsilon_t$
- OLS:

$$\hat{\theta} = \frac{\sum_{t=2}^{T} Y_t Y_{t-1}}{\sum_{t=2}^{T} Y_{t-1}^2} = \theta + \frac{\sum_{t=2}^{T} \varepsilon_t Y_{t-1}}{\sum_{t=2}^{T} Y_{t-1}^2}$$

- $\sqrt{T}(\hat{\theta} - \theta) \rightarrow^d \mathcal{N}(0, V)$, where $V = \frac{\mathrm{Var}(Y_{t-1}\varepsilon_t)}{(\mathrm{E}[Y_t^2])^2}$ under no autocorrelation in $Y_{t-1}\varepsilon_t$

Serial correlation

- Sometimes, there is autocorrelation
- Then $V \neq \frac{\mathrm{Var}(Y_{t-1}\,\varepsilon_t)}{(\mathrm{E}[Y_t^2])^2}$
- Instead, it is equal to

$$V = \left(\mathrm{E}[Y_t^2]\right)^{-2} \lim_{T\to\infty} \frac{1}{T-1} \mathrm{Var}\left(\sum_{t=2}^{T} Y_{t-1}\,\varepsilon_t\right)$$

- We need HAC variance estimator (was on the board)
- $\hat{V}^{HAC} = \hat{V}\hat{f}$, where

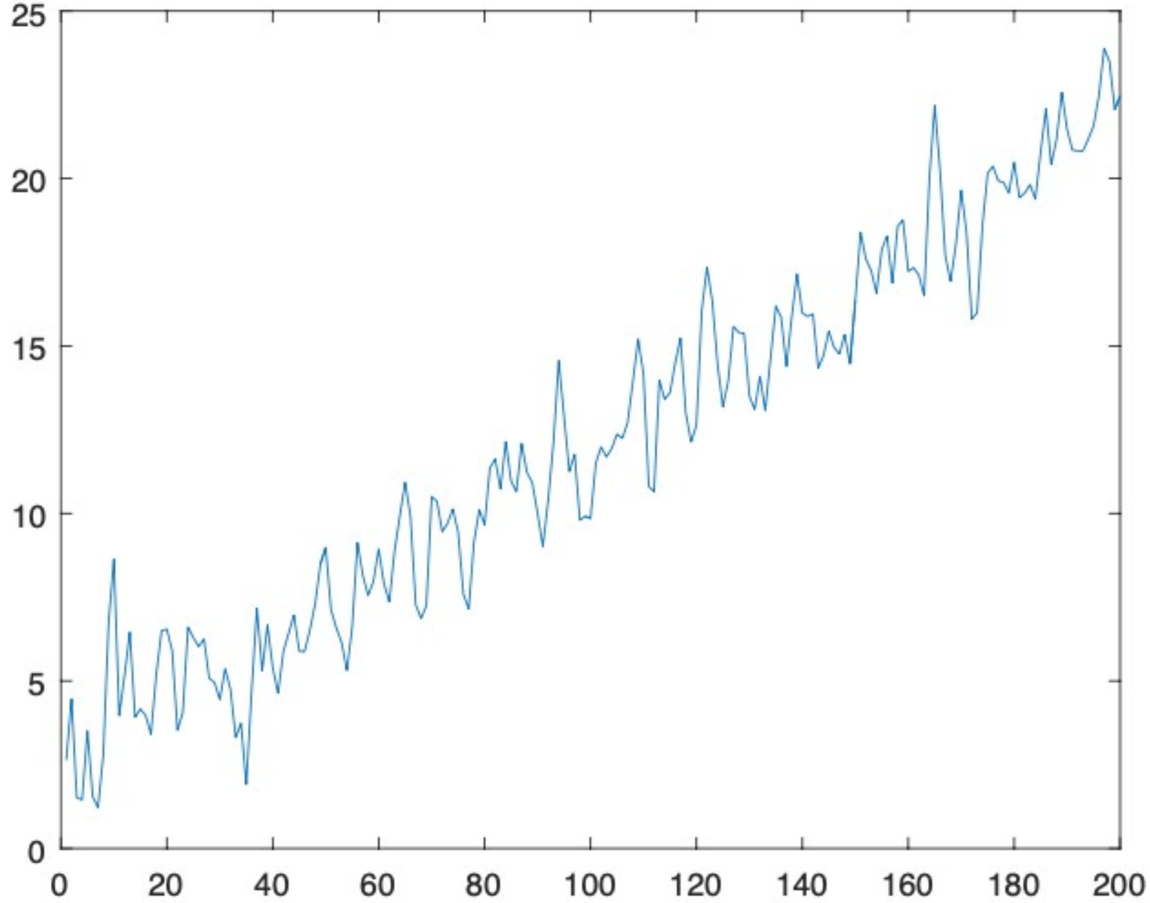
$$\hat{f} = 1 + 2\sum_{j=1}^{m} \frac{m-j}{m}\hat{\rho}(j),$$

and

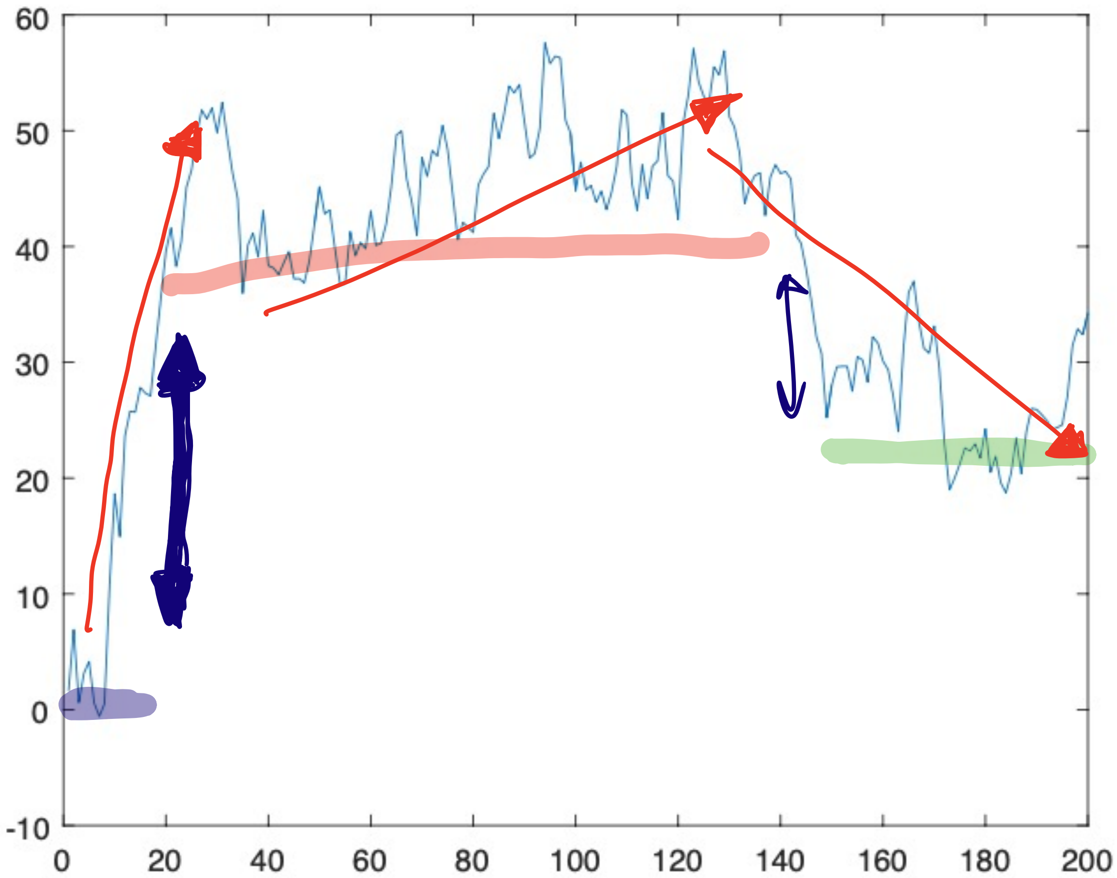
$$m = CT^{1/3}, \text{ where } C = \left(\frac{6\rho^2}{(1-\rho^2)^2}\right)^{1/3},$$

for the case when $\varepsilon_t\,Y_{t-1}$ is AR(1) with parameter ρ.

Series 1

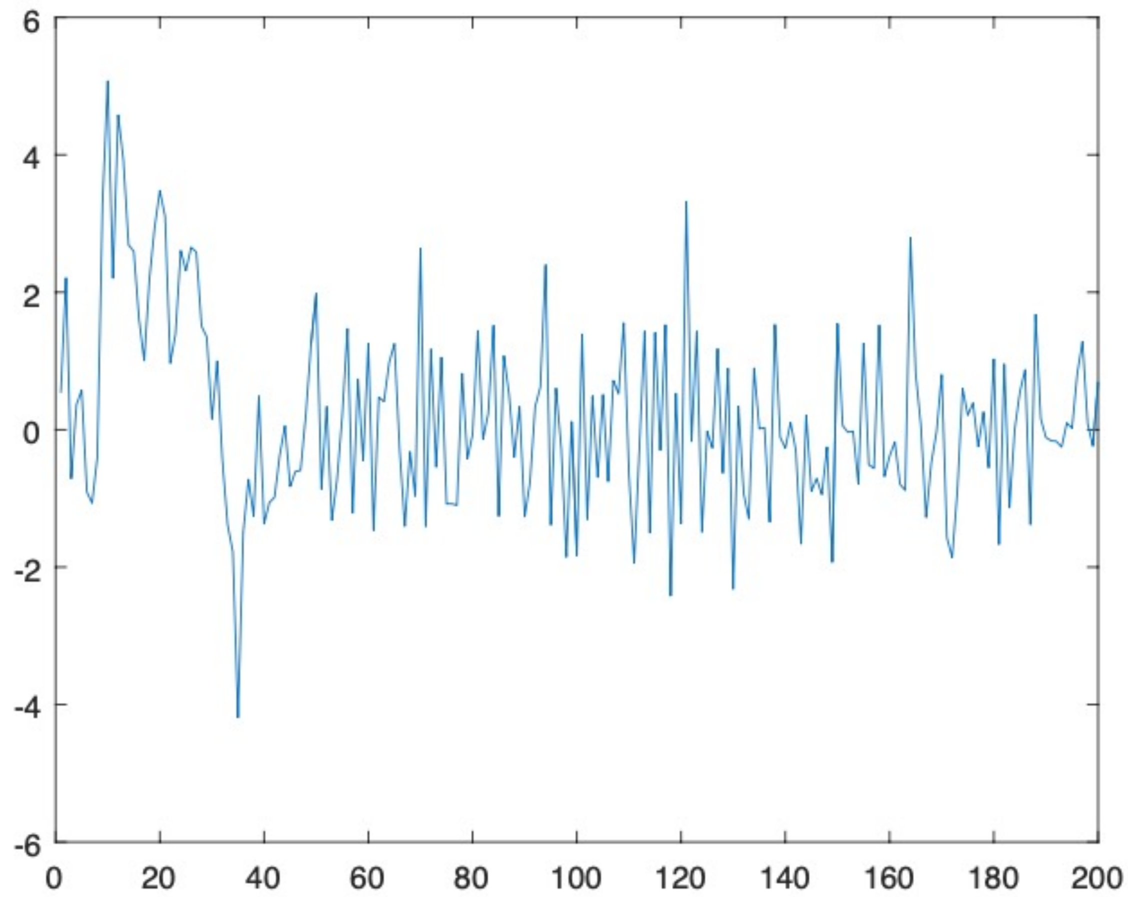


Series 2

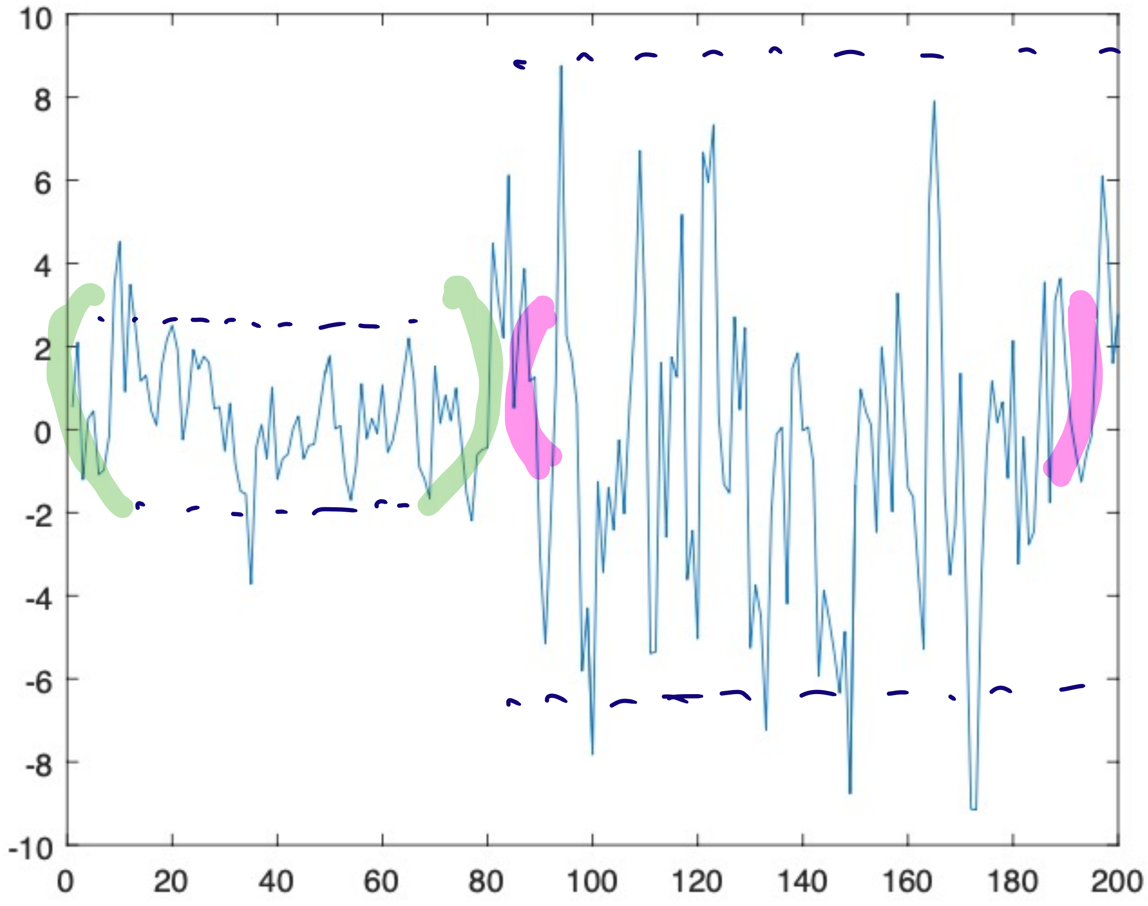


$\alpha \cdot t$

Series 3

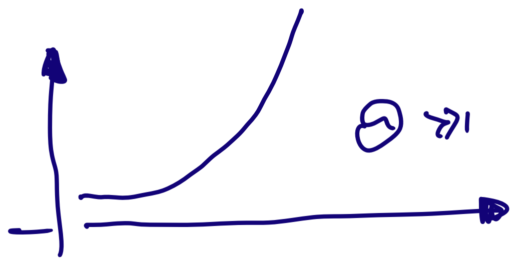


Series 4



Type of Non-Stationary TimeSeries

- Time trend
- Unit root
- Structural break in levels
- Structural break in variance



$y_t = \theta \cdot y_{t-1} + \varepsilon_t$

Trend-Stationary TimeSeries

$$Y_t = \mu + \delta t + \Psi(L)\varepsilon_t = \mu + \delta t + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\sum_{j=1}^{\infty} |\psi_j| < \infty$

- $Y_t - \delta t$ is stationary
- Forecasts:
  - $\hat{Y}_{t+h|t} = \mu + \delta(t+h) + \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + \ldots$
  - Forecast error: $e_{t+h|t} = \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \ldots + \psi_{h-1} \varepsilon_{t+1}$
  - Variance of the forecast error: $\mathrm{Var}(e_{t+h|t}) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 < \infty$
- Impulse response to a shock: $\frac{\partial Y_{t+h}}{\partial \varepsilon_t} = \psi_h \to 0$, as $h \to \infty$

Trend-Stationary TimeSeries

- Estimate the trend + arma
- If there is no trend, $\hat{\delta} \rightarrow^{p} 0$
- If there is a trend, but just estimate arma, you'll get something close to a unit root (model is misspecified)
- Trends might be logarithmic or quadratic

Difference stationary TS

$$Y_t = \mu + Y_{t-1} + \Psi(L)\varepsilon_t = \mu + Y_{t-1} + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\sum_{j=1}^{\infty} |\psi_j| < \infty$

- $(1-L)Y_t = Y_t - Y_{t-1}$ is stationary
- Forecasts (for simplicity, let $\Psi(L) = I$):
  - $\hat{Y}_{t+h|t} = \mu h + Y_t$
  - Forecast error: $e_{t+h|t} = \sum_{j=1}^{h} \varepsilon_{t+j}$
  - Variance of the forecast error: $\mathrm{Var}(e_{t+h|t}) = \sigma^2 h \to \infty$, as $h \to \infty$
- Impulse response to a shock: $\frac{\partial Y_{t+h}}{\partial \varepsilon_t} = 1$

Difference Stationary TS

$y_t = \dots + \cancel{\delta t} + \dots$

$y_{t-1} = \dots + \delta(t-1) + \dots \rightarrow \delta t - \delta$

- Work with $Z_t = (1 - L)Y_t = Y_t - Y_{t-1}$, which is stationary

$\Delta y_t = y_t - y_{t-1} = + \dots + \delta + \dots$

- Need to determine if there is a unit root
- Look at ACF (but might confuse with just large $\theta < 1$)
- Do statistical testing