ARIMA(p, d, q) Model for y_t

$$\phi(B)\left(\nabla^d y_t - \mu\right) = \Theta(B) w_t,$$

$$w_t \overset{iid}{\sim} N(0, \sigma_w^2)$$

* we'll assume $\nabla^d y_t$ is stationary
* $\mu = E[\nabla^d y_t]$

In practice, need to choose d, how much to difference

way we often do this ... come up with a test that tests the null hypothesis that a process z_t is non-stationary

Start with $k=0$.

Test if $\nabla^k y_t$ is nonstationary

If we accept set $k=k+1$, repeat test

If we reject, set $d=k$

Tests of Non-Stationarity

Dickey-Fuller Test

- without trend, $\nabla y_t = K y_{t-1} + w_t$
  Null: $K = 0$, Alt: $K \neq 0$
- with trend, $\nabla y_t = a + K y_{t-1} + w_t$
  Null: $K = 0$, Alt: $K \neq 0$

Augmented Dickey-Fuller

- without trend
  $\phi(B)\nabla y_t = K y_{t-1} + w_t$
  Null: $K = 0$, Alt: $K \neq 0$
- with trend $\phi(B)\nabla y_t = a + K y_{t-1} + w_t$
  Null: $K = 0$, Alt: $K \neq 0$

Phillips-Perron Test

- without trend
  $\phi(B)\nabla y_t = K y_{t-1} + \Theta(B) w_t$
  Null: $K = 0$, Alt: $K \neq 0$
- with trend → $\phi(B)\nabla y_t = a + K y_{t-1} + \Theta(B) w_t$
  Null: $K = 0$, Alt: $K \neq 0$

↳ Based on a fact that tell's us that <u>any</u> ARIMA(p, 0, q) process can be arbitrarily well approximated by an ARIMA(K, 0, 0) process

Tests of Nonstationarity

Test	Null	Alternative
Dickey Fuller	ARIMA(0,1,0) (stationary)	ARIMA(1,0,0) (stationary)
Augmented Dickey Fuller	ARIMA(p, 1, 0) (stationary)	ARIMA(p+1, 0,0) (stationary)
Phillips Perron	ARIMA(p, 1, q) (stationary)	ARIMA(p+1, 0, q) (stationary)

fitting null model requires choice of p → use AIC, AICc, BIC to choose p

fitting null requires choice of k, approx AR model → choose k in a deterministic way as a function of time series length

* All of these tests work by fitting the exact or approximate null AR-type model to the differenced data, and estimating k via regression, and using $\frac{\hat{k}}{se\{\hat{k}\}}$ as test statistic

Forecasting for ARIMA(p, d, q) models (given a specific choice of d)

$$\min_{c_1,\dots,c_m} \mathbb{E}\left[\left(y_{m+1} - \left(\sum_{j=1}^{m} c_{mj} y_{m+1-j}\right)\right)^2\right] =$$

$$\mathbb{E}[y_{m+1}^2] + \underset{\sim}{c_m}' A_m \underset{\sim}{c_m} - 2\underset{\sim}{b_m}' \underset{\sim}{c_m},$$

where

$$A_{mij} = \mathbb{E}[y_{m+1-i}\, y_{m+1-j}]$$
$$b_{mj} = \mathbb{E}[y_{m+1}\, y_{m+1-j}]$$

previously these were autocovar. of a stationary process

when $d > 0$, y_t isn't stationary anymore, A_n, $\underset{\sim}{b_n}$ difficult to evaluate and work with

How we'll fix this... write

$$y_t = y_0 + \sum_{i=1}^{t} \nabla^d y_t$$

$$\min_{c_1, \dots, c_m} \mathbb{E}\left[\left(y_0 + \sum_{i=1}^{m+1} \nabla^d y_i - \sum_{j=1}^{m} c_{m,j}\left(y_0 + \sum_{k=1}^{m+1-j} \nabla^d y_k\right)\right)^2\right]$$

* This depends on

$\mathbb{E}[y_0 \nabla^d y_i]$ for $i > 0$ — it's not a big deal to assume these are zero

$\boxed{\mathbb{E}[\nabla^d y_i \nabla^d y_j]}$ autocovariances of a stationary process! ☺

Tedious to work out by hand, very similar to ARMA setting, so we'll just use R for this