* Pick a project dataset
* No additional exams
* A homework due next Tuesday will be posted tonight
* Previous homework will be scanned and returned

Nonstationarity & ARIMA

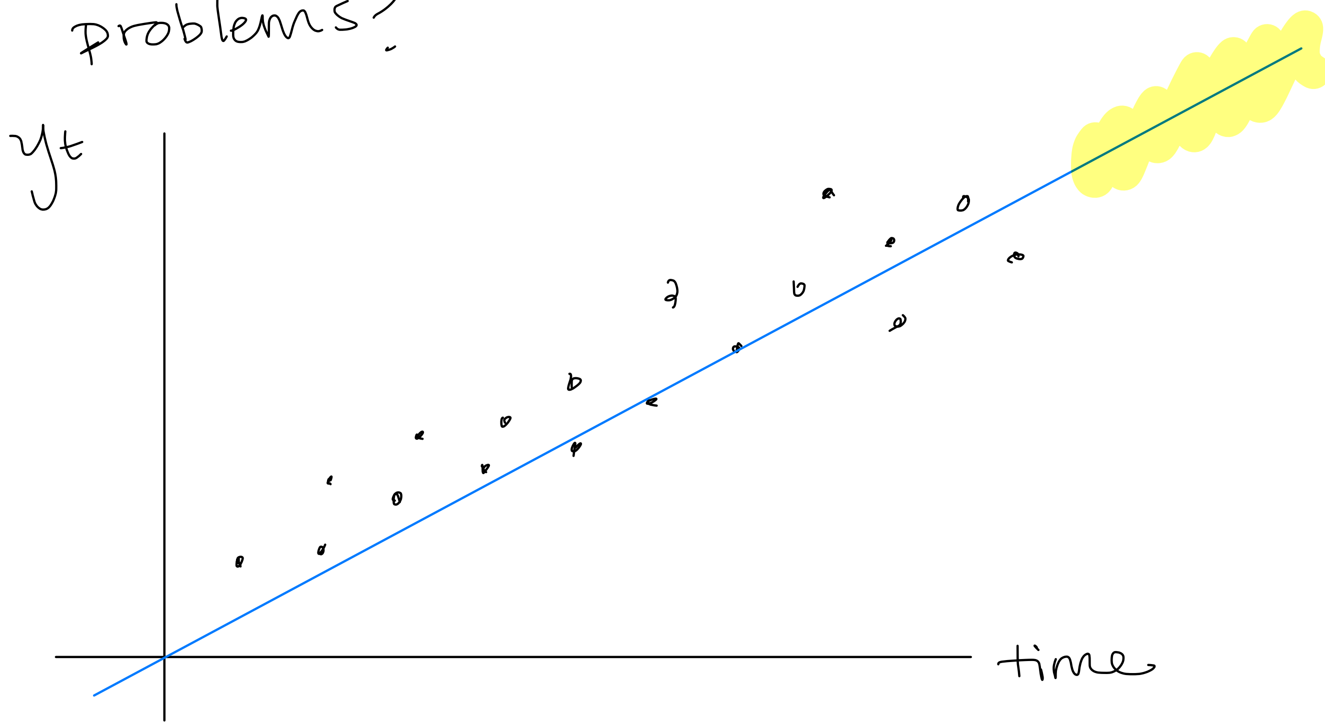
ARMA(p,q) models have the form

$$\Phi(B)(y_t - \mu_t) = \Theta(B) w_t,$$

* $y_t - \mu_t$ is stationary
* $w_t \overset{iid}{\sim} N(0, \sigma_w^2)$
* $\mu_t = \underset{\sim}{x}_t' \underset{\sim}{\beta}$

Before we discussed this, we would observe z_t, get y_t by regressing out sources of nonstationarity

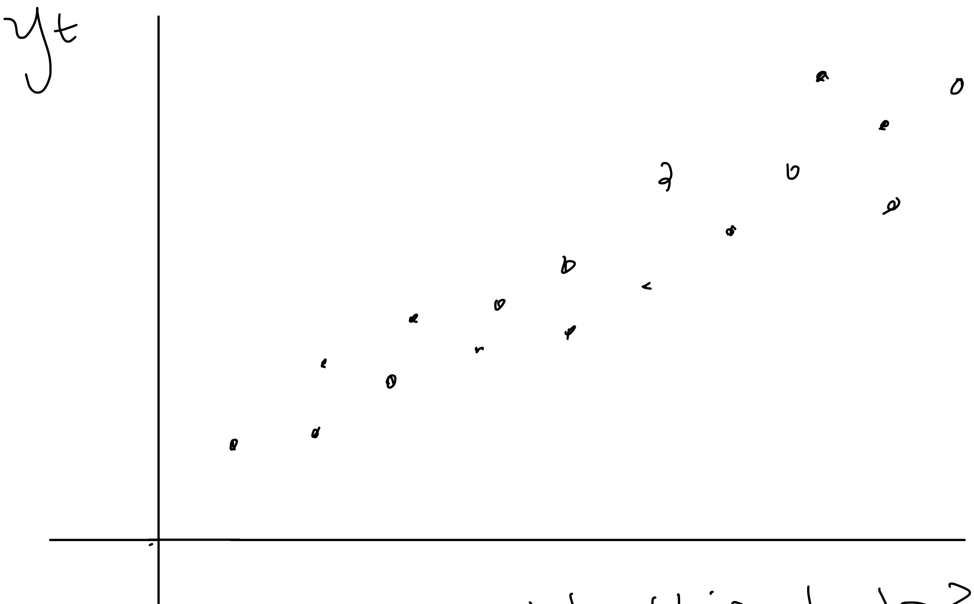
- That was ad-hoc, but using the framework above, our ARMA models <u>can</u> have time varying means why isn't incorporating $\mu_t = \underset{\sim}{x}_t' \underset{\sim}{\beta}$, a time trend enough to solve our problems?



$\underset{\sim}{x}_t = (1, t)$

Another approach to non-stationarity –

Differencing



time

Maybe

$$\underbrace{y_t - y_{t-1}}_{\text{increments}}$$

is stationary

Why should this help? Suppose y_t is AR(1) and $\mu = 0$, then $y_t = \phi y_{t-1} + w_t$

* Nonstationary if $\phi > 1$

$$y_t - y_{t-1} = (\phi - 1) y_{t-1} + w_t$$

Notation for Differencing

∇^d (\nabla^d)] Differencing operator

$$\nabla^1 y_t = \nabla y_t = y_t - y_{t-1}$$

$$\nabla^2 y_t = \nabla y_t - \nabla y_{t-1} = y_t - 2y_{t-1} + y_{t-2}$$

$$\nabla^3 y_t = \nabla^2 y_t - \nabla^2 y_{t-1} = \cdots$$

$$\vdots$$

$$\nabla^k y_t = \nabla^{k-1} y_t - \nabla^{k-1} y_{t-1}$$

Motivation for Differencing

Suppose $y_t = a + bt + w_t$, $a, b \neq 0$ are fixed unknown parameters

$$\nabla y_t = y_t - y_{t-1}$$

$$= a + bt + w_t - a - b(t-1) - w_{t-1}$$

$$= b + w_t - w_{t-1}$$

This is an MA(1) process with $\theta_1 = -1$

$$\mu = E[\nabla y_t] = b$$

Suppose we had

$$y_t = a + bt + ct^2 + w_t,$$

$a, b, c \neq 0$ are fixed, unknown parameters

$$y_t - y_{t-1} = bt + ct^2 + w_t - b(t-1) - c(t-1)^2 - w_{t-1}$$

$$= b + 2ct - c + w_t - w_{t-1}$$

$$\nabla^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

$$= b - c + 2ct + w_t - w_{t-1} - (b - c + 2c(t-1) + w_{t-1} - w_{t-2})$$

$$= 2c + w_t - 2w_{t-1} + w_{t-2},$$

MA(2) process

In general,

$\nabla^k y_t$ will be stationary if

$$y_t = a_0 + a_1 t + a_2 t^2 + \dots + a_k t^k + w_t,$$

where $a_0, a_1, \dots, a_k$ are fixed but unknown coefficients of a polynomial time trend

ARIMA(p, d, q) process is given by:

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

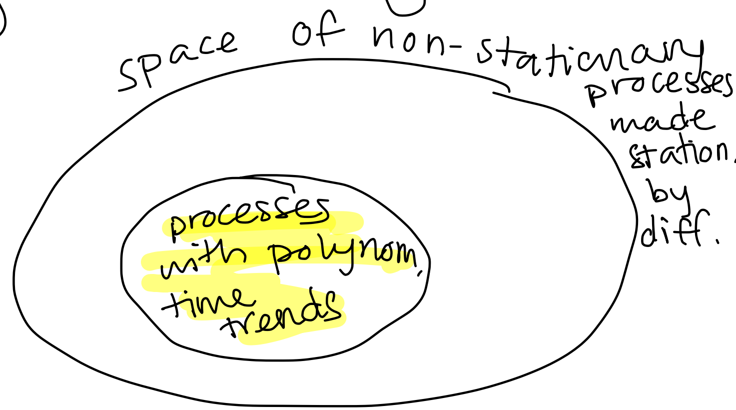
* $w_t \overset{iid}{\sim} N(0, \sigma_w^2)$
* $\nabla^d y_t$ is stationary, $\mu = \mathbb{E}[\nabla^d y_t]$
* If $\mu \neq 0$, then we are allowing for a degree d polynomial time trend

ARIMA(p, d, q) process is given by:

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

* $w_t \overset{iid}{\sim} N(0, \sigma_w^2)$
* $\nabla^d y_t$ is stationary, $\mu = E[\nabla^d y_t]$
* If $\mu \neq 0$, then we are allowing for a degree d polynomial time trend

* Even if $\mu \neq 0$, we are allowing for at least a degree d-1 polynomial time trend



Why difference?

Suppose $y_t = y_{t-1} + w_t$

* No explicit time trend
* Nonstationary

$\nabla y_t = y_t - y_{t-1}$ will give a stationary process

$= w_t$

Remaining Roadblock in Use of ARIMA(p, d, q) models is:

How do we choose d?

- Can't use AIC/AICc/BIC because our data is different for each d
- Cross-validation would be ok
- Implement a test, where we are testing the null hypothesis that our process is non-stationary

Tests of Non-Stationarity

* Dickey-Fuller test, tests the null that y_t is AR(1) with $\phi_1 = 1$

Assumes $\nabla y_t = a + k y_{t-1} + w_t$, $w_t \overset{iid}{\sim} N(0, \sigma_w^2)$

Null: $k = 0$, a could take on any value

Alternative: $k \neq 0$

Test statistic is $\hat{k}/se\{\hat{k}\}$,

under null, $\underline{\underline{\hat{k}/se\{\hat{k}\}}} \xrightarrow[n \to \infty]{d} DF$

Dickey Fuller distribution, quantiles are available

* Problem with this test is null is too restrictive - many nonstationary processes aren't AR(1) with $\phi_1 = 1$

Tests of Non-Stationarity Continued

Augmented Dickey Fuller Test

Tests null that y_t is an ARIMA(p,1,0) process

$$\nabla y_t = a + K y_{t-1} + \phi_1 \nabla y_{t-1} + \dots + \phi_p \nabla y_{t-p} + w_t$$

$w_t \overset{iid}{\sim} N(0, \sigma_w^2)$ and $\phi_1, \dots, \phi_p$ correspond to a AR parameters stationary process

If $K = 0 \implies y_t$ is ARIMA(p, 1, 0) process

If $K \neq 0 \implies y_t$ is ARIMA(p+1, 0, 0) and stationary

Compute $\hat{K} / se\{\hat{K}\} \xrightarrow[n \to \infty]{d} DF$