Shpaner, Leon

Module 3
ARMA
-CH4
BACKSHIFT
Linear Regression

Linear Regression

$$x_t = \beta_0 + \underbrace{\beta_{z_{t_1}} + \beta_{2\,z_{t_2}}}_{z_i \text{ regressors}} + \cdots + \underbrace{w_t}_{\text{White noise}}$$

exogenous variables → coming from outside the model: some other information

i.e. trying to predict surface temperature of the ocean, or air temperature, past weather, or some other var. that aren't "x" itself.

The difference we're going to make in this chapter: we're going to replace these "z's" w/ past values of "x"

Autoregressive model

$$x_t = \alpha + \underbrace{\phi_1 x_{t-1} + \phi x_{t-2}}_{\text{past values of } x} + \cdots + w_t$$

→ one of the big differences in this chapter is the noise
· there may also be a moving average term of white noise.

$\phi$ ⇒ correspond directly to these regression coefficients.
telling us how much to add in from this part of the past and how much to add in from further back.

AR(1) $\quad x_t = \phi x_{t-1} + w_t$

→ if $|\phi| > 1$, magnitude of signal grows

this autoregressive model looks back one step into past, adds that in, plus adding new noise in @ each term in the cycle.

must be constrained to be $\leq 1$

$$x_t = \phi x_{t-1} + w_t$$
$$x_t = \phi\left(\phi_{t-2} + \cdots + w_t\right)$$

Shocks: some unexplained event causes this past values.

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}$$

mean = 0
→ b/c adding up a bunch of white noise terms and ultimately, these are all terms w/ mean 0.

$$x_t = \mu + \sum_{j=0}^{\infty} \phi^j w_{t-j}$$

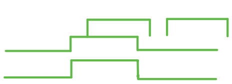
→ you can have a mean not dependent on t, and it's not saying that the "x" is getting recycled through or we wouldn't have a mean and wouldn't be stationary.

causal time series: if $\sum \phi^j < \infty$ → the future depends only on the past.

Moving Average Model

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots$$

Sum of multiple white noise models lagged in time w/ multiple coefficients.
- matches definition of a causal model
- no restriction on the parameters on the MA model.



moving average model
when you get out of the window, goes to zero.

one of the differences btwn this moving average model, which is that your autocorrelation is going to drop to zero once you're outside how far back you go.

$$\boxed{w_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}}$$ → condition of invertibility
→ one unique set of coefficients

$$\boxed{x_t = \alpha + \underbrace{\phi x_{t-1} + \phi_{x-t}}_{\substack{\text{AR} \\ p\text{-steps back}}} + \underbrace{w_t + \theta w_{t-1} + \theta w_{t-1}}_{\substack{\text{MA (invertible)} \\ q \text{ steps back}}}}$$

ARMA(p, q) → how far in the past to look at your own output, and how far back in the past to look at this correlated series of noise terms.

$ARMA = (1-0.9B)$ → where B = Backshift operator

$x_t = w_t$

$(1-0.9B)x_t = (1-0.9B)w_t$

$x_t - 0.9x_{t-1} = w_t - 0.9w_{t-1}$

$x_t = 0.9x_{t-1} - 0.9w_{t-1} + w_t$

→ this is a model that's overfit, but these are all equalities. These are the same model.

they fit identical, yet they are different

possible quiz question: what is the difference between the autocorrelation of the moving average model and the autocorrelation of the Autoregressive model? finite

Remember, the moving average model – you have the window and then you give up, so the autocorrelation is going to come back and then stop. But for this AR(1) model, $x_t$ depends on $x_t$ minus one, which depends on $x_t$ minus two, all the way back to the beginning of time so that autocorrelation for the Autoregressive model is never going to 0, no matter how far back you look.

$X, Y, Z$

$\hat{X} - B_z$ → partial correlation

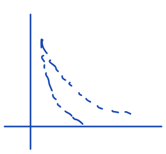
(or, $x-\hat{x}$, $y-\hat{y}$)

Partial Autocorrelation

$\phi_{11} = corr(x_1, x_0) = \rho(1)$

$\phi_{hh} = corr(x_h - \hat{x}_h, x_0 - \hat{x}_0)$

relationship of these two when you remove past window.



$x_h$ $x_1, x_0, \cdots x_{h-1}$

AR – partial ACF → x's depend on past values of x. You want to move that influence to look at the ACF.

MA – but if you do that on MA model, you start removing the influence of things that don't matter (introducing dependencies).

→ partial ACF not good - Fine w/ regular ACF

Forecasting

$x_t = \phi_{t-1} + w_1$

$x^n_{t+1} = \phi x_t + w^n_{t+1}$ → $E(w_t) = 0$

→ the value of $x_{t+1}$ given the history of $x^n$ (superscript, not power)

$$x^n_{n+m} = \phi^m x_n$$ } no more noise coming into the model.

ARIMA → integrated

→ differencing gives you an AR model.

AR

$\underbrace{x_t}_{\text{output}} = \alpha + \underbrace{\phi x_{t-1}}_{\text{past values}} \cdots$

MA

$\underbrace{x_t}_{\text{output}} = w_t + \theta w_{t-1}$ (important thing is that there's no x here).

ARMA (p, q)

$x_t = \alpha + \underbrace{\phi x_{t-1}}_{AR} \cdots + \overbrace{w_t + \theta_1 w_{t-1}}^{MA}$

p = past

q = past values of w

$x_t = \mu_t + y_t$ → this part is stationary

$\mu_t = \beta_0 + \beta_1 t$

$\nabla x_t = \beta_1 + \nabla y_t$

ARIMA → the idea here is that we might get something that's not (or might not be) stationary, and we're going to keep differencing it until we get something that we can work w/.

ARIMA $(p, d, q)$

$d$: how much differencing

forecasting

$$x^{n}_{n+1} = \phi x_n$$

$n$ → history

$n+1$ ← next step

$$y_t = \nabla^d x_t$$

⇒ we're going to forecast $y$, this difference signal, and then we can plug it in to get a prediction of $x$, our original output.

$$x^{n}_{n+m} = y^{n}_{n+m} + x^{n}_{n+m-1}$$

$$y^{n}_{n+m} = x^{n}_{n+m} - x^{n}_{n+m-1}$$

ARIMA $(0,1,1)$ → no past, 1 differencing, 1 noise term)

$$x_t = x_{t-1} + w_t - \lambda w_{t-1}$$

$\lambda$ → MA parameter

$$x^{n}_{n+1} = (1-\lambda)x_n + \lambda x^{n-1}_{n}$$

where $|\lambda| \le 1$

→ special case: exponentially MA

this model looks at the whole history of the past.



if there's some shock to the system, it will last forever

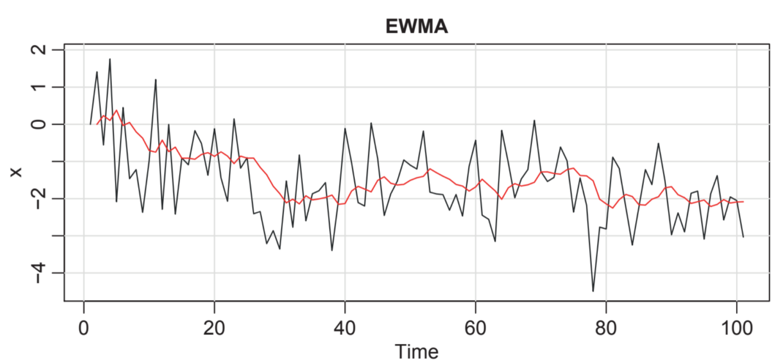
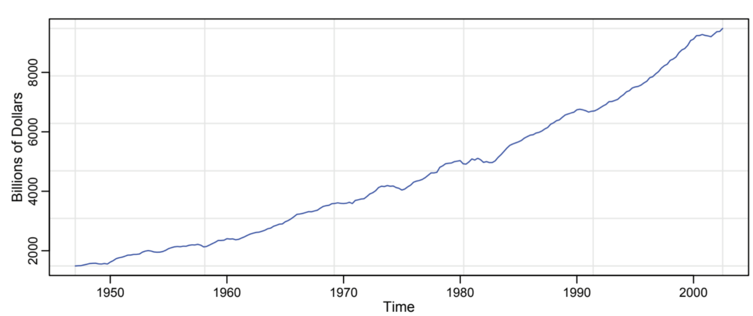
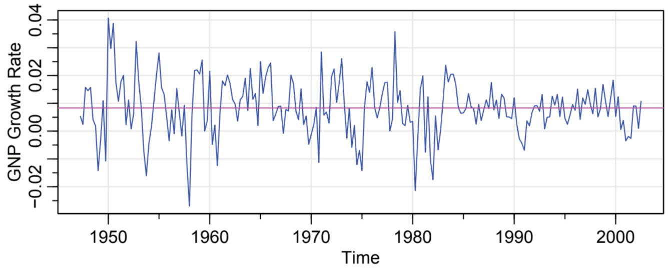



Figure 5.2 *Output for Example 5.5: Simulated data with an EWMA superimposed.*








→ differences { $x_t$ = GDP

$y_t = \nabla' x_t$

ARIMA (0, 1, 2)