ECON 504 Discussion Note

TA information:

Name: Xu Wang, PhD student in Economics

Email: xwang222@binghamton.edu

OH: M, noon – 1 pm, Wed, 6-7 pm

Discussion I: 10:50 – 11:50 AM

Discussion II: 12:00 – 1:00 PM

Three components:

1. Concepts review
2. Demonstrate codes from lecture notes
3. Talk about the homework (hint, demonstrate)

Discussion session I (Feb 19, 2021)

Concepts review:

1. Stochastic Process

2. Moments of probability distribution (univariate distribution)
   a. First moment: mean
   b. Second moment: variance
   c. Skewness

3. Moments of a joint probability distribution
   a. Covariance
      $cov(X, Y) = \sigma_{YX} = E\,(X - \mu_X)\,(Y - \mu_Y)$
   b. Correlation
      $corr(X, Y) = \rho_{YX} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$

4. In the context of time series
   a. Autocovariance
      $cov(Y_t, Y_{t-k}) = \sigma_{Y_t Y_{t-k}}$
   b. Autocorrelation
      $corr(Y_t, Y_{t-k}) = \rho_{Y_t Y_{t-k}} = \dfrac{\sigma_{Y_t Y_{t-k}}}{\sigma_{Y_t} \sigma_{Y_t - k}}$

5. Weakly/covariance-stationary vs Strict Stationary

$$\mu_{Y_1} = \mu_{Y_2} = \ldots = \mu_{Y_T}$$

$$\sigma^2_{Y_1} = \sigma^2_{Y_2} = \ldots = \sigma^2_{Y_T}$$

$$\rho_{Y_t Y_{t-k}} = \rho_{Y_t Y_{t+k}} = \rho_{|k|}$$

6. Non-stationary process – first difference

7. Basics of R
   # starts from a clean slate
   # sets the default directory
   # load library
   # read in data
   # use command from library, read documentation when needed
   # save plots
   # useful commands:
   - Cbind() ## combine column
   - par(mfrow=c(2,2)) ## set figure layout

Lab code: HW1.R

Discussion session II

Concepts review:

1. Concepts regarding forecasting
   a) Minimize the expected loss (different types of loss functions)
   b) If Y is i.i.d, then unconditional mean is the best forecast (Y_bar: mean)
   c) If Y is correlated to Y_t-1 (or X), then conditional mean is optimal (Y_t-1 and x should provide additional information)
   d) Use knowledge of regression to generate optimal forecasts (conditional mean of Y) (when we do regression models, we are implicitly model the conditional mean of Y as a function of X)

2. Difference between mean and conditional mean
   1) P(A|B) = P(AB)/P(B)
   2) If A and B are independent, then P(AB)=P(A)*P(B)

   Plug 2) into 1), we have:

   P(A|B) = P(AB)/P(B)= P(A)*P(B)/P(B)=P(A)

   If P(A|B) = P(A), A and B are independent.

3. ACF vs PACF

4. White noise
   A time series process with zero mean and a constant finite variance whose autocorrelations and partial autocorrelations are all zero:
   - Independent White Noise
   - Gaussian/Normal White Noise

5. White noise test
   a. Bartlett bands
   b. LBQ test

6. MA(1) process
   - Mean
   - Variance
   - Covariance
   - Correlation
   - ACF (for identification)
   - PACF (for identification)
   - Stationary
   - Invertible:
     - y_t has an autoregressive representation.
     - → can be written as a weighted sum of past Y + current shock
     - Model is identified (|theta|<1)

7. MA(2) process (see slides)

8. Calculation of autocorrelation

$$Y_t = 8 - 0.6\varepsilon_{t-1} + \varepsilon_t \text{ , } \varepsilon_t \sim iidN(0{,}0.25)\ \forall t \qquad (0)$$

To calculation autocorrelation $\rho_1 = \frac{\gamma_1}{\gamma_0}$, you need: 1) variance γ_0; 2) covariance γ_1

For variance,

$$\gamma_0 = Var(Y_t)$$

We know that

$$Var(X) = E\big(X - E(X)\big)^2$$

Therefore,

$$\gamma_0 = Var(Y_t) = E\big(Y_t - E(Y_t)\big)^2$$

We know that $E(Y_t) = E(8 - 0.6\varepsilon_{t-1} + \varepsilon_t) = 8$, therefore,

$$\gamma_0 = Var(Y_t) = E(8 - 0.6\varepsilon_{t-1} + \varepsilon_t - 8)^2 = E(-0.6\varepsilon_{t-1} + \varepsilon_t)^2$$
$$= E(0.36{\varepsilon_{t-1}}^2 + {\varepsilon_t}^2 - 2 * 0.6\varepsilon_{t-1}\varepsilon_t\ )$$
$$= 0.36 * 0.25 + 1 * 0.25$$

For covariance,

$$\gamma_1 = Cov(Y_t, Y_{t-1}\ )$$

We know that

$$Cov(X,Y) = E\big(X - E(X)\big)\big(Y - E(Y)\big)$$
$$\gamma_1 = E\big(Y_t - E(Y_t)\big)\big(Y_{t-1} - E(Y_{t-1})\big)$$
$$= E(-0.6\varepsilon_{t-1} + \varepsilon_t)(-0.6\varepsilon_{t-2} + \varepsilon_{t-1})$$
$$= E(-0.6\varepsilon_{t-1} * \varepsilon_{t-1})$$
$$= E(-0.6 * {\varepsilon_{t-1}}^2)$$
$$= -0.6 * E({\varepsilon_{t-1}}^2) = -0.6 * 0.25$$

9. Forecasting (Lecture Slide 8)
   - Point estimate
   - Confidence interval

10. Hypothesis testing
    a. How to state the hypothesis
    b. What can we infer from the test?

    H_0: x=0

    H_1: x>1

    P-value<0.05

    Reject H_0

    P_value =0.65 > 0.05

    Can say: can’t reject H_0

    Can’t say: accept H_0

Discussion session 3 (Mar 12, 2021)

1. Forecasting with MA(1) process:

$$Y_t = 8 - 0.5\varepsilon_{t-1} + \varepsilon_t \text{ , } \varepsilon_t \sim iidN(0,0.25)\ \forall t \qquad (0)$$

Suppose $\varepsilon_t = 0.8$

0) What do we want to know at period t?

Name	Notation	K=1	K=2	K=3
True value of $\mathrm{Y_{t+k}}$:	$\mathrm{Y_{t+k}}$	(1.1)	(1.2)	(1.3)
Unconditional mean of $\mathrm{Y_{t+k}}$:	$E(\mathrm{Y_{t+k}})$	8	8	8
Unconditional variance of $\mathrm{Y_{t+k}}$	$Var(\mathrm{Y_{t+k}})$	0.3125	0.3125	0.3125
Conditional mean of $\mathrm{Y_{t+k}}$:	$E(\mathrm{Y_{t+k}}\vert I_t)$	7.6	8	8
Conditional variance of $\mathrm{Y_{t+k}}$:	$Var(\mathrm{Y_{t+k}}\vert I_t)$	0.25	0.3125	0.3125
Forecast interval of $\mathrm{Y_{t+k}}$	$FI(\mathrm{Y_{t+k}})$	(7.11, 8.09)		
Point forecast:	$\widehat{\mathrm{Y_{t+k}}}$			
Forecasting error:	$e_{t,k}$	(9.1)	(9.2)	(9.3)
Forecasting variance:	$Var\left(e_{t,k}\middle\vert I_t\right)$	0.25	0.3125	0.3125

1) Where do I get actual $\mathrm{Y_{t+k}}$?

Actual $\mathrm{Y_{t+1}}$ comes from the data generating process (estimated):

$Y_{t+1} = 8 - 0.5\varepsilon_t + \varepsilon_{t+1}$, $\varepsilon_t \sim iidN(0,0.25)\ \forall t$ (1.1)

$Y_{t+2} = 8 - 0.5\varepsilon_{t+1} + \varepsilon_{t+2}$, $\varepsilon_t \sim iidN(0,0.25)\ \forall t$ (1.2)

$Y_{t+3} = 8 - 0.5\varepsilon_{t+2} + \varepsilon_{t+3}$, $\varepsilon_t \sim iidN(0,0.25)\ \forall t$ (1.3)

Why is Y_{t+1} an equation instead of a concrete number?

Because now you are in period t, you can't observe ε_{t+1}. However, you do have all information up to period t (ie: Y_t, ε_t for period t, t-1, t-2, …)

2) Unconditional mean of $\mathrm{Y_{t+k}}$

$E(Y_{t+1}) = 8 - 0.5E(\varepsilon_t) + E(\varepsilon_{t+1})$, $\varepsilon_t \sim iidN(0,0.25)\ \forall t$ (2.1)

$E(Y_{t+2}) = 8 - 0.5E(\varepsilon_{t+1}) + E(\varepsilon_{t+2})$, $\varepsilon_t \sim iidN(0,0.25)\ \forall t$ (2.2)

$E(Y_{t+3}) = 8 - 0.5E(\varepsilon_{t+2}) + E(\varepsilon_{t+3})$, $\varepsilon_t \sim iidN(0,0.25)\ \forall t$ (2.3)

For (2.1), because we are finding unconditional mean, so we pretend that we don't observe ε_t. Therefore, we have:

$$E(Y_{t+k}) = E(Y_{t+1}) = 8 - 0.5E(\varepsilon_t) + E(\varepsilon_{t+1}) = 8$$

3) Unconditional variance of $\mathrm{Y_{t+k}}$

$$Y_{t+k} = 8 - 0.5\varepsilon_{t+k-1} + \varepsilon_{t+k} \text{ , } \varepsilon_t \sim iidN(0,0.25)\ \forall t$$

$$Var(\mathrm{Y_{t+k}}) = E\left[\left(\mathrm{Y_{t+k}} - E(\mathrm{Y_{t+k}})\right)^2\right] = E[(-0.5\varepsilon_{t+k-1} + \varepsilon_{t+k})^2]$$

$$= E(0.25{\varepsilon_{t+k-1}}^2 + {\varepsilon_{t+k}}^2 - 2 * 0.5 * \varepsilon_{t+k-1}\varepsilon_{t+k})$$

$$= 0.25 \underbrace{E(\varepsilon_{t+k-1}{}^2)}_{variance+0\ mean} + \underbrace{E(\varepsilon_{t+k}{}^2)}_{variance+0\ mean} - \underbrace{E(\varepsilon_{t+k-1}\varepsilon_{t+k})}_{covariance+0\ mean}$$
$$= 0.25 * var(\varepsilon_{t+k-1}) + var(\varepsilon_{t+k})$$
$$= 1.25 * 0.25$$
$$= 0.3125$$

4) Conditional mean of Y_{t+k}
When k=1:

$$E(Y_{t+1}|I_t) = 8 - 0.5 * E(\varepsilon_t|I_t) + \underbrace{E(\varepsilon_{t+1}|I_t)}_{=E(\varepsilon_{t+1})=0} = 8 - 0.5 * \varepsilon_t = 7.6$$

When k=2:

$$E(Y_{t+2}|I_t) = 8 - 0.5 * \underbrace{E(\varepsilon_{t+1}|I_t)}_{=E(\varepsilon_{t+1})=0} + \underbrace{E(\varepsilon_{t+2}|I_t)}_{=E(\varepsilon_{t+2})=0} = 8$$

When k=3:

$$E(Y_{t+3}|I_t) = 8 - 0.5 * \underbrace{E(\varepsilon_{t+2}|I_t)}_{=E(\varepsilon_{t+2})=0} + \underbrace{E(\varepsilon_{t+3}|I_t)}_{=E(\varepsilon_{t+3})=0} = 8$$

5) Conditional variance of Y_{t+k}
When k=1:

$$Var(Y_{t+1}|I_t) = E\left[\left(\underbrace{Y_{t+1}}_{8-0.5\varepsilon_t+\varepsilon_{t+1}} - \underbrace{E(Y_{t+1}|I_t)}_{8-0.5*\underbrace{E(\varepsilon_t|I_t)}_{\varepsilon_t}+\underbrace{E(\varepsilon_{t+1}|I_t)}_{=E(\varepsilon_{t+1})=0}}\right)^2 \middle| I_t\right]$$
$$= E[(\varepsilon_{t+1})^2|I_t]$$

Remember that:

$$Var(\varepsilon_{t+1}|I_t) = E[\varepsilon_{t+1}{}^2|I_t] - \underbrace{E[\varepsilon_{t+1}|I_t]^2}_{=E(\varepsilon_{t+1})^2}$$

Also remember that:

I. ε_t is white noise, so $E(\varepsilon_{t+1}|I_t) = E(\varepsilon_{t+1})$ [Note: if ε_t is not a white noise process, this equation doesn't hold.]
II. $E(\varepsilon_{t+1}) = 0$ [This comes from the assumption on distribution]

$$Var(Y_{t+1}|I_t) = Var(\varepsilon_{t+1}|I_t) = Var(\varepsilon_{t+1}) = 0.25$$

When k=2:

$$Var(Y_{t+2}|I_t) = E\left[\left(\underbrace{Y_{t+2}}_{8-0.5\varepsilon_{t+1}+\varepsilon_{t+2}} - \underbrace{E(Y_{t+2}|I_t)}_{8-0.5*\underbrace{E(\varepsilon_{t+1}|I_t)}_{=E(\varepsilon_{t+1})=0}+\underbrace{E(\varepsilon_{t+2}|I_t)}_{=E(\varepsilon_{t+2})=0}}\right)^2 \middle| I_t\right]$$

$$= E[(-0.5\varepsilon_{t+1} + \varepsilon_{t+2})^2 | I_t]$$

$$= E[0.25{\varepsilon_{t+1}}^2 + {\varepsilon_{t+2}}^2 - \varepsilon_{t+1}\varepsilon_{t+2} | I_t]$$

$$= 0.25 \underbrace{E({\varepsilon_{t+1}}^2 | I_t)}_{Var(\varepsilon_{t+1}|I_t)=Var(\varepsilon_{t+1})=0.25} + \underbrace{E({\varepsilon_{t+2}}^2 | I_t)}_{Var(\varepsilon_{t+2}|I_t)=Var(\varepsilon_{t+2})=0.25} - \underbrace{E(\varepsilon_{t+1}\varepsilon_{t+2} | I_t)}_{covariance+0\ mean=0}$$

$$= 0.25 * 0.25 + 1 * 0.25 = 0.3125$$

When k=3:

$$Var(Y_{t+3}|I_t) = E\left[\left(\underbrace{Y_{t+3}}_{8-0.5\varepsilon_{t+2}+\varepsilon_{t+3}} - \underbrace{E(Y_{t+3}|I_t)}_{8-0.5*\underbrace{E(\varepsilon_{t+2}|I_t)}_{=E(\varepsilon_{t+2})=0}+\underbrace{E(\varepsilon_{t+3}|I_t)}_{=E(\varepsilon_{t+3})=0}}\right)^2 \middle| I_t\right]$$

$$= E[(-0.5\varepsilon_{t+2} + \varepsilon_{t+3})^2 | I_t]$$

$$= E[0.25{\varepsilon_{t+2}}^2 + {\varepsilon_{t+3}}^2 - \varepsilon_{t+2}\varepsilon_{t+3} | I_t]$$

$$= 0.25 \underbrace{E({\varepsilon_{t+2}}^2 | I_t)}_{Var(\varepsilon_{t+2}|I_t)=Var(\varepsilon_{t+2})=0.25} + \underbrace{E({\varepsilon_{t+3}}^2 | I_t)}_{Var(\varepsilon_{t+3}|I_t)=Var(\varepsilon_{t+3})=0.25} - \underbrace{E(\varepsilon_{t+2}\varepsilon_{t+3} | I_t)}_{covariance+0\ mean=0}$$

$$= 0.25 * 0.25 + 1 * 0.25 = 0.3125$$

6) Forecasting error:
the difference between actual Y_{t+1} and predicted (ie: conditional mean) of Y_{t+1}

$$e_{t,1} = Y_{t+1} - \mu_{t+1|t}$$

7) Find the forecast error $e_{t,1}$
When k=1

$$e_{t,1} = \overbrace{8 - 0.5\varepsilon_t + \varepsilon_{t+1}}^{Y_{t+1}} - \overbrace{[8 - 0.5E(\varepsilon_t|I_t) + E(\varepsilon_{t+1}|I_t)]}^{\mu_{t+1|t}}$$

$$e_{t,1} = 0.5 * E(\varepsilon_t|I_t) - 0.5\varepsilon_t + \varepsilon_{t+1} - E(\varepsilon_{t+1}|I_t)$$

Remember that you are now in period t, so you can observe ε_t (ie: ε_t is a constan, instead of a random variable). Therefore, $E(\varepsilon_t|I_t) = \varepsilon_t = C$, where C is a constant.

$$e_{t,1} = \varepsilon_{t+1} - E(\varepsilon_{t+1}|I_t)$$

What is $E(\varepsilon_{t+1}|I_t)$?
Remember $\varepsilon_t \sim iidN(0,0.25)\ \forall t$, this means:

a) ε_t is white noise, so $E(\varepsilon_{t+1}|I_t) = E(\varepsilon_{t+1})$ [Note: if ε_t is not a white noise process, this equation doesn't hold.]
b) $E(\varepsilon_{t+1}) = 0$ [This comes from the assumption on distribution]

$$e_{t,1} = \varepsilon_{t+1} \quad (9.1)$$

When k=2,

$$e_{t,2} = \overbrace{8 - 0.5\varepsilon_{t+1} + \varepsilon_{t+2}}^{Y_{t+2}} - \overbrace{\left[8 - 0.5\underbrace{E(\varepsilon_{t+1}|I_t)}_{0} + \underbrace{E(\varepsilon_{t+2}|I_t)}_{0}\right]}^{\mu_{t+2|t}}$$

$$= -0.5\varepsilon_{t+1} + \varepsilon_{t+2} \qquad (9.2)$$

When k=3,

$$e_{t,3} = \overbrace{8 - 0.5\varepsilon_{t+2} + \varepsilon_{t+3}}^{Y_{t+3}} - \overbrace{\left[8 - 0.5\underbrace{E(\varepsilon_{t+2}|I_t)}_{0} + \underbrace{E(\varepsilon_{t+3}|I_t)}_{0}\right]}^{\mu_{t+3|t}}$$

$$= -0.5\varepsilon_{t+2} + \varepsilon_{t+3} \qquad (9.2)$$

8） Now I know that $e_{t,1} = \varepsilon_{t+1}$, what else do I need to know? Isn't that enough?
Well, ε_{t+1} is itself a random variable. So, if you don't tell me the distribution of it, you basically didn't tell me anything useful. You need to at least tell me the conditional mean $E(e_{t,1}|I_t)$ and conditional variance $Var(e_{t,1}|I_t)$ of it.

9） Ok. Conditional mean and variance are easy.
10.1 Definition of optimal forecast (conditional mean=0)

$$E\left(e_{t,1}|I_t\right) = E(\varepsilon_{t+1}|I_t) = E(\varepsilon_{t+1}) = 0$$

$$E\left(e_{t,2}|I_t\right) = -0.5E(\varepsilon_{t+1}|I_t) + E(\varepsilon_{t+2}|I_t) = 0$$

$$E\left(e_{t,3}|I_t\right) = -0.5E(\varepsilon_{t+2}|I_t) + E(\varepsilon_{t+3}|I_t) = 0$$

10.2 Conditional variance

When k=1

$$Var\left(e_{t,1}\middle|I_t\right) = Var(\varepsilon_{t+1}|I_t)$$

ε_t is white noise, so $Var(\varepsilon_{t+1}|I_t) = Var(\varepsilon_{t+1})$. $\varepsilon_t \sim iidN(0,0.25)\ \forall t$, so:

$$Var\left(e_{t,1}\middle|I_t\right) = Var(\varepsilon_{t+1}|I_t) = Var(\varepsilon_{t+1}) = 0.25 \qquad (5)$$

When k=2

$$Var\left(e_{t,2}\middle|I_t\right) = Var(-0.5\varepsilon_{t+1} + \varepsilon_{t+2}|I_t) = 0.3125$$

When k=3

$$Var\left(e_{t,3}\middle|I_t\right) = Var(-0.5\varepsilon_{t+2} + \varepsilon_{t+3}|I_t) = 0.3125$$

10） Forecast interval of $\mathrm{Y_{t+k}}$

$$Y_{t+k}|I_t \sim N(\mu_{t+k,t}, {\sigma^2}_{t+k,t})$$

Forecast interval for $\mathrm{Y_{t+k}}$= $\mu_{t+k,t} \pm 1.96 * \sigma_{t+k,t}$

2. Derive 60th percentile of N(2,6)
   Let the 60th percentile be K*

$$p(K < K^*) = 0.6$$
$$p\left(\frac{K-2}{\sqrt{6}} < \frac{K^*-2}{\sqrt{6}}\right) = 0.6$$
$$p\left(Z < \frac{K^*-2}{\sqrt{6}}\right) = 0.6$$
$$\frac{K^*-2}{\sqrt{6}} = 0.25$$

K*=?

3. What is the probability of K>4.5? K~ N(2,6)

$$p(K > 4.5) = p\left(\frac{K-2}{\sqrt{6}} > \frac{4.5-2}{\sqrt{6}}\right) = p\left(Z > \frac{4.5-2}{\sqrt{6}}\right) = p(Z > 1.02)$$
$$= 1 - p(Z < 1.02) = 0.15$$

4. Forecasting with MA(2) process:

$$Y_t = 8 - 0.5\varepsilon_{t-1} + 0.1\varepsilon_{t-2} + \varepsilon_t \text{ , } \varepsilon_t \sim iidN(0,0.25)\ \forall t \qquad (0)$$

ε_t	1.4
ε_{t-1}	0.8
ε_{t-2}	-1
ε_{t-3}	0.6

1) Derive the point forecast and forecasting interval.
   Point forecast →conditional mean of y_{t+h}
   Forecasting interval → conditional mean of $y_{t+h} \pm Z * SEF$
   Standard error of forecasting (SEF)→sqrt of conditional variance of forecasting error
   Forecasting error→$y_{t+h} - E(y_{t+h}|I_t)$

2) Conditional mean of y_{t+h}
   h=1:

$$Y_{t+1} = 8 - 0.5\varepsilon_t + 0.1\varepsilon_{t-1} + \varepsilon_{t+1}$$
$$E(Y_{t+1}|I_t) = 8 - 0.5E(\varepsilon_t|I_t) + 0.1E(\varepsilon_{t-1}|I_t) + E(\varepsilon_{t+1}|I_t)$$
$$E(Y_{t+1}|I_t) = 8 - 0.5 * 1.4 + 0.1 * 0.8 = 7.38$$

h=2:

$$Y_{t+2} = 8 - 0.5\varepsilon_{t+1} + 0.1\varepsilon_t + \varepsilon_{t+2}$$
$$E(Y_{t+2}|I_t) = 8 - 0.5E(\varepsilon_{t+1}|I_t) + 0.1E(\varepsilon_t|I_t) + E(\varepsilon_{t+2}|I_t)$$
$$E(Y_{t+1}|I_t) = 8 - 0.5 * 0 + 0.1 * 1.4 = 8.14$$

h=3:

$$Y_{t+3} = 8 - 0.5\varepsilon_{t+2} + 0.1\varepsilon_{t+1} + \varepsilon_{t+3}$$
$$E(Y_{t+3}|I_t) = 8 - 0.5E(\varepsilon_{t+2}|I_t) + 0.1E(\varepsilon_{t+1}|I_t) + E(\varepsilon_{t+3}|I_t)$$
$$E(Y_{t+3}|I_t) = 8$$

3) Forecasting error

h=1:

$$e_{t,1} = Y_{t+1} - E(Y_{t+1}|I_t) = \varepsilon_{t+1}$$
$$E\left(e_{t,1}|I_t\right) = E(\varepsilon_{t+1}|I_t) = E(\varepsilon_{t+1}) = 0$$
$$Var\left(e_{t,1}\middle|I_t\right) = Var(\varepsilon_{t+1}|I_t) = Var(\varepsilon_{t+1}) = 0.25$$

h=2:

$$e_{t,2} = Y_{t+2} - E(Y_{t+2}|I_t) = -0.5\varepsilon_{t+1} + \varepsilon_{t+2}$$
$$E\left(e_{t,2}|I_t\right) = E(-0.5\varepsilon_{t+1} + \varepsilon_{t+2}|I_t) = 0$$
$$Var\left(e_{t,2}\middle|I_t\right) = Var(-0.5\varepsilon_{t+1} + \varepsilon_{t+2}|I_t)$$
$$= E\left[[(-0.5\varepsilon_{t+1} + \varepsilon_{t+2}) - E\left(e_{t,2}|I_t\right)]^2\right]$$
$$= E[0.25\varepsilon_{t+1}^2 + \varepsilon_{t+2}^2 - \varepsilon_{t+1}\varepsilon_{t+2}]$$
$$= 1.25Var(\varepsilon_{t+1}) = 1.25 * 0.25 = 0.3125$$

h=3:

$$e_{t,3} = Y_{t+3} - E(Y_{t+3}|I_t) = -0.5\varepsilon_{t+2} + 0.1\varepsilon_{t+1} + \varepsilon_{t+3}$$
$$E\left(e_{t,3}|I_t\right) = E(-0.5\varepsilon_{t+2} + 0.1\varepsilon_{t+1} + \varepsilon_{t+3}|I_t) = 0$$
$$Var\left(e_{t,3}\middle|I_t\right) = Var(-0.5\varepsilon_{t+2} + 0.1\varepsilon_{t+1} + \varepsilon_{t+3}|I_t)$$
$$= E[(-0.5\varepsilon_{t+2} + 0.1\varepsilon_{t+1} + \varepsilon_{t+3})^2]$$
$$= E[0.25\varepsilon_{t+2}^2 + 0.01\varepsilon_{t+1}^2 + \varepsilon_{t+3}^2 + cross - period\ terms]$$
$$= 1.26Var(\varepsilon_{t+1}) = 1.26 * 0.25 = 0.315$$

Discussion session 4 (Mar 19, 2021)

1. AR(1) process

$$y_t = c + \phi y_{t-1} + \epsilon_t, \quad \epsilon_t \sim iidN(0, \sigma^2) \quad (1)$$

   0) Stationary =➔ $|\phi| < 1$➔ y_t and future shocks are uncorrelated [15:54 of lecture recording]
   1) Mean

$$\underbrace{E(y_t)}_{\mu} = \underbrace{E(c)}_{c} + \underbrace{\phi}_{|\phi|<1} \underbrace{E(y_{t-1})}_{\mu} + \underbrace{E(\epsilon_t)}_{0}$$

$$\mu = c + \phi\mu \quad (2)$$

$$\mu = \frac{c}{1-\phi}$$

   2) ACF

$$\rho_1 = \frac{\gamma_1}{\gamma_0}$$

$$\gamma_1 = E(y_t - \mu)(y_{t-1} - \mu)$$

(1)-(2):

$$y_t - \mu = c - c + \phi y_{t-1} - \phi\mu + \epsilon_t$$

$$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t \quad (3)$$

(3)* $(y_{t-1} - \mu)$

$$(y_t - \mu)(y_{t-1} - \mu) = \phi(y_{t-1} - \mu)^2 + \epsilon_t(y_{t-1} - \mu)$$

$$\underbrace{E(y_t - \mu)(y_{t-1} - \mu)}_{\gamma_1} = \phi \underbrace{E(y_{t-1} - \mu)^2}_{\gamma_0} + E[\epsilon_t(y_{t-1} - \mu)]$$

$$\gamma_1 = \phi\gamma_0 + E[\epsilon_t(y_{t-1} - \mu)]$$

$$\gamma_1 = \phi\gamma_0 + E[\epsilon_t y_{t-1} - \epsilon_t\mu]$$

$$\gamma_1 = \phi\gamma_0 + E[\epsilon_t y_{t-1}] - E[-\epsilon_t\mu]$$

$$\gamma_1 = \phi\gamma_0$$

$$\frac{\gamma_1}{\gamma_0} = \phi = \rho_1$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0}$$

(3)* $(y_{t-2} - \mu)$

$$(y_t - \mu)(y_{t-2} - \mu) = \phi(y_{t-1} - \mu)(y_{t-2} - \mu) + \epsilon_t(y_{t-2} - \mu)$$

$$\underbrace{E(y_t - \mu)(y_{t-2} - \mu)}_{\gamma_2} = \phi \underbrace{E(y_{t-1} - \mu)(y_{t-2} - \mu)}_{\gamma_1} + \underbrace{E[\epsilon_t(y_{t-2} - \mu)]}_{0}$$

$$\gamma_2 = \phi\gamma_1$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \phi\frac{\gamma_1}{\gamma_0} = \phi\rho_1 = \phi^2$$

   3) γ_0: squaring eq(3) and take expectation [57:27]

$$(y_t - \mu)^2 = [\phi(y_{t-1} - \mu) + \epsilon_t]^2$$

$$E(y_t - \mu)^2 = E[\phi(y_{t-1} - \mu) + \epsilon_t]^2$$

$$\gamma_0 = \phi^2\gamma_0 + \sigma^2$$

$$\gamma_0 = \frac{\sigma^2}{1-\phi^2}$$

2. AR(2) process

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \quad \epsilon_t \sim iidN(0,\sigma^2) \quad (4)$$

1) Stationarity condition
2) Mean

$$\underbrace{Ey_t}_{\mu} = c + \phi_1 \underbrace{Ey_{t-1}}_{\mu} + \phi_2 \underbrace{Ey_{t-2}}_{\mu} + E\epsilon_t$$

$$\mu = c + \phi_1\mu + \phi_2\mu \quad (5)$$

$$\mu = \frac{c}{1-\phi_1-\phi_2}$$

3) ACF

(4)-(5)

$$y_t - \mu = c - c + \phi_1 y_{t-1} - \phi_1\mu + \phi_2 y_{t-2} - \phi_2\mu + \epsilon_t$$

$$(y_t-\mu) = \phi_1(y_{t-1}-\mu) + \phi_2(y_{t-2}-\mu) + \epsilon_t \quad (6)$$

(6)* $(y_{t-1}-\mu)$

$$(y_t-\mu)(y_{t-1}-\mu) = \phi_1(y_{t-1}-\mu)(y_{t-1}-\mu) + \phi_2(y_{t-2}-\mu)(y_{t-1}-\mu) + \epsilon_t(y_{t-1}-\mu)$$

$$\mathrm{E}(y_t-\mu)(y_{t-1}-\mu) = \phi_1 E(y_{t-1}-\mu)(y_{t-1}-\mu) + \phi_2 E(y_{t-2}-\mu)(y_{t-1}-\mu) + E[\epsilon_t(y_{t-1}-\mu)]$$

$$\gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1$$

$$\frac{\gamma_1}{\gamma_0} = \phi_1 + \phi_2\frac{\gamma_1}{\gamma_0}$$

$$\rho_1 = \phi_1 + \phi_2\rho_1$$

$$\rho_1 = \frac{\phi_1}{1-\phi_2}$$

(6)* $(y_{t-2}-\mu)$

$$\mathrm{E}(y_t-\mu)(y_{t-2}-\mu) = \phi_1 E(y_{t-1}-\mu)(y_{t-2}-\mu) + \phi_2 E(y_{t-2}-\mu)(y_{t-2}-\mu) + E[\epsilon_t(y_{t-2}-\mu)]$$

$$\gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0$$

$$\frac{\gamma_2}{\gamma_0} = \phi_1\frac{\gamma_1}{\gamma_0} + \phi_2$$

$$\rho_2 = \phi_1\rho_1 + \phi_2$$

(6)* $(y_{t-3}-\mu)$

$$\mathrm{E}(y_t-\mu)(y_{t-3}-\mu) = \phi_1 E(y_{t-1}-\mu)(y_{t-3}-\mu) + \phi_2 E(y_{t-2}-\mu)(y_{t-3}-\mu) + E[\epsilon_t(y_{t-3}-\mu)]$$

$$\gamma_3 = \phi_1\gamma_2 + \phi_2\gamma_1$$

$$\rho_3 = \phi_1\rho_2 + \phi_2\rho_1$$

Discussion session 5 (Mar 26, 2021)

1. Forecasting with AR(1)

$$y_t = c + \phi y_{t-1} + \epsilon_t, \quad \epsilon_t \sim iidN(0,\sigma^2) \quad (1)$$

h=1

$$y_{t+1} = c + \phi y_t + \epsilon_{t+1}$$
$$E(y_{t+1}|I_t) = c + \phi E(y_t|I_t) + E(\epsilon_{t+1}|I_t)$$
$$\mu_{t+1|t} = E(y_{t+1}|I_t) = c + \phi y_t$$
$$Var(y_{t+1}|I_t) = E\left[\left(y_{t+1} - \mu_{t+1|t}\right)^2 \middle| I_t\right]$$
$$= E\left[{e_{t,1}}^2 \middle| I_t\right] = E[{\epsilon_{t+1}}^2|I_t]$$
$$= Var(\epsilon_{t+1}) = \sigma^2$$

h=2

$$y_{t+2} = c + \phi y_{t+1} + \epsilon_{t+2}$$
$$\mu_{t+2|t} = E(y_{t+2}|I_t) = c + \phi E(y_{t+1}|I_t) + E(\epsilon_{t+2}|I_t)$$
$$= c + \phi(c + \phi y_t)$$
$$Var(y_{t+2}|I_t) = E\left[\left(y_{t+2} - \mu_{t+2|t}\right)^2 \middle| I_t\right]$$
$$= E[(\phi\epsilon_{t+1} + \epsilon_{t+2})^2|I_t]$$
$$= (1+\phi^2)\sigma^2$$

2. Forecasting with AR(2)

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \quad \epsilon_t \sim iidN(0,\sigma^2) \quad (1)$$

h=1

$$y_{t+1} = c + \phi_1 y_t + \phi_2 y_{t-1} + \epsilon_{t+1}$$
$$E(y_{t+1}|I_t) = c + \phi_1 E(y_t|I_t) + \phi_2 E(y_{t-1}|I_t) + E(\epsilon_{t+1}|I_t)$$
$$\mu_{t+1|t} = E(y_{t+1}|I_t) = c + \phi_1 y_t + \phi_2 y_{t-1}$$
$$Var(y_{t+1}|I_t) = E\left[\left(y_{t+1} - \mu_{t+1|t}\right)^2 \middle| I_t\right]$$
$$= E\left[{e_{t,1}}^2 \middle| I_t\right] = E[{\epsilon_{t+1}}^2|I_t]$$
$$= Var(\epsilon_{t+1}) = \sigma^2$$

h=2

$$y_{t+2} = c + \phi_1 y_{t+1} + \phi_2 y_t + \epsilon_{t+2}$$
$$\mu_{t+2|t} = E(y_{t+2}|I_t) = c + \phi_1 E(y_{t+1}|I_t) + \phi_2 E(y_t|I_t) + E(\epsilon_{t+2}|I_t)$$
$$= c + \phi_1 \mu_{t+1|t} + \phi_2 y_t$$
$$= c + \phi_1(c + \phi_1 y_t + \phi_2 y_{t-1}) + \phi_2 y_t$$
$$Var(y_{t+2}|I_t) = E\left[\left(y_{t+2} - \mu_{t+2|t}\right)^2 \middle| I_t\right] = E\left[\left(e_{t+2,t}\right)^2 \middle| I_t\right]$$
$$e_{t+2,t} = y_{t+2} - \mu_{t+2|t} = c + \phi_1 y_{t+1} + \phi_2 y_t + \epsilon_{t+2} - \left(c + \phi_1 \mu_{t+1|t} + \phi_2 y_t\right)$$
$$= \phi_1\left(y_{t+1} - \mu_{t+1|t}\right) + \epsilon_{t+2}$$
$$= \phi_1(\epsilon_{t+1}) + \epsilon_{t+2}$$
$$Var(y_{t+2}|I_t) = E\left[\left(e_{t+2,t}\right)^2 \middle| I_t\right] = E[(\phi_1(\epsilon_{t+1}) + \epsilon_{t+2})^2|I_t]$$

h=3

$$y_{t+3} = c + \phi_1 y_{t+2} + \phi_2 y_{t+1} + \epsilon_{t+3}$$
$$\mu_{t+3|t} = E(y_{t+3}|I_t) = c + \phi_1 E(y_{t+2}|I_t) + \phi_2 E(y_{t+1}|I_t) + E(\epsilon_{t+3}|I_t)$$
$$= c + \phi_1 \mu_{t+2|t} + \phi_2 \mu_{t+1|t}$$
$$e_{t+3,t} = y_{t+3} - \mu_{t+3|t} = \phi_1\left(y_{t+2} - \mu_{t+2|t}\right) + \phi_2\left(y_{t+1} - \mu_{t+1|t}\right) + \epsilon_{t+3}$$

$$= \phi_2 e_{t,1} + \phi_1 e_{t+2,t} + \epsilon_{t+3}$$
$$= \phi_2 \epsilon_{t+1} + \phi_1(\phi_1(\epsilon_{t+1}) + \epsilon_{t+2}) + \epsilon_{t+3}$$
$$Var(y_{t+3}|I_t) = E\left[\left(e_{t+3,t}\right)^2\middle| I_t\right]$$

3. ARMA(p,q)→ARMA(1,1)

$$y_t = c + \phi y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t, \epsilon_t \sim iidN(0, \sigma^2)$$

   1) Stationarity + invertibility
   2) Mean of y_t [10:49]
   3) Point forecast (with $y_t = a$ and $\varepsilon_t = b$ given)

      h=1

$$y_{t+1} = c + \phi y_t + \theta \varepsilon_t + \varepsilon_{t+1}$$
$$E(y_{t+1}|I_t) = c + \phi E(y_t|I_t) + \theta E(\varepsilon_t|I_t) + E(\varepsilon_{t+1}|I_t)$$
$$E(y_{t+1}|I_t) = c + \phi * a + \theta b$$

      h=2

$$E(y_{t+2}|I_t) = c + \phi E(y_{t+1}|I_t) + \theta E(\varepsilon_{t+1}|I_t) + E(\varepsilon_{t+2}|I_t)$$
$$E(y_{t+2}|I_t) = c + \phi * (c + \phi * a + \theta b)$$

   4) Forecast error

      h=1

$$e_{t+1,t} = y_{t+1} - \mu_{t+1|t} = \varepsilon_{t+1}$$
$$y_{t+1} = c + \phi y_t + \theta \varepsilon_t + \varepsilon_{t+1}, \epsilon_t \sim iidN(0, \sigma^2)$$
$$E(y_{t+1}|I_t) = c + \phi E(y_t|I_t) + \theta E(\varepsilon_t|I_t) + E(\varepsilon_{t+1}|I_t)$$
$$Var(\epsilon_{t+1}) = \sigma^2$$

      h=2

$$e_{t+2,t} = y_{t+2} - \mu_{t+2|t} = \phi e_{t+1,t} + \theta \varepsilon_{t+1} + \varepsilon_{t+2}$$
$$= \phi \varepsilon_{t+1} + \theta \varepsilon_{t+1} + \varepsilon_{t+2}$$
$$y_{t+2} = c + \phi y_{t+1} + \theta \varepsilon_{t+1} + \varepsilon_{t+2}$$
$$E(y_{t+2}|I_t) = c + \phi E(y_{t+1}|I_t) + \theta E(\varepsilon_{t+1}|I_t) + E(\varepsilon_{t+2}|I_t)$$
$$= c + \phi E(y_{t+1}|I_t)$$
$$Var(\epsilon_{t+2}) = Var(\phi \varepsilon_{t+1} + \theta \varepsilon_{t+1} + \varepsilon_{t+2})$$

4. Forecasting with R (with given model + R results)

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \quad \epsilon_t \sim iidN(0, \sigma^2) \quad (1)$$

Coefficients:

Ar1	Ar2	Mean
0.6	-0.1	5

Signma^2 estimated as 25

   1) What is your forecasting model?

$$y_t = c + 0.6 y_{t-1} - 0.1 y_{t-2} + \epsilon_t, \quad \epsilon_t \sim iidN(0, 5^2) \quad (1)$$
$$E(y_t) = c + 0.6E(y_{t-1}) - 0.1E(y_{t-2}) + E(\epsilon_t)$$
$$5 = c + 0.6 * 5 - 0.1 * 5$$
$$2.5 = c$$

$$y_t = 2.5 + 0.6 y_{t-1} - 0.1 y_{t-2} + \epsilon_t, \quad \epsilon_t \sim iidN(0, 5^2) \quad (1)$$

2) What is your point forecast?
3) What is your forecasting interval?

5. R application

1) Model identification
    a. Plot the series and ensure stationarity
    b. Tentative identification based on ACF and PACF

Terms	ACF	PACF
AR	Geometric	p significant lags
MA	q significant lags	Geometric
ARMA	Geometric	Geometric

Suppose you have 3 candidates:

1)ARMA(4,0);

2) ARMA(2:2);

3) ARMA(3,1)

    c. Estimate the model
    d. Diagnostic
        i. Whether residuals are white noise?
            - Bartlett band and Q stat. Good if can’t reject→ residual is WN → regression utilized all useful information.
            - Degree of freedom

        ii. AIC + BIC → smaller is better.
    e. Forecast with selected model

2) Application

Discussion session 6 (Apr 9, 2021)

1. Technic details regarding OLS
2. Evaluating forecast
   a. Optimality of forecast
      i. Zero mean under quadratic loss (Test1, MPE test)
      ii. 1-step forecast errors are WN (Test2)
      iii. H-step –Ma(h-1)
      iv. Variance converge
      v. Optimality w/r/t information set (Test3)
   b. Accuracy of forecast
      i. RMSE – root mean squared error
      ii. MAE – mean absolute error
      iii. MAPE – mean absolute percentage error

3. Test1(MPE test)

$$E\left(e_{t,1}\right) = 0$$

Problem 1: How to obtain $e_{t,1}$?

1990-2021, monthly

Train: 1990-2015

Test: 2016-2021

Predicted: 2016-2021

Forecast error=test – predicted value

Answer: split data into training and test; using training data to train the model and use the actual value in test set to calculate $e_{t,1}$

Problem 2: How to obtain multiple values of $e_{t,1}$?

Answer: 1) Fixed scheme; 2) rolling; 3) rolling origin

Fixed scheme: training data and test data are fixed; only train model once; use multiple points in test data to obtain multiple $e_{t,1}$.[80:20]

Total: 100 data points

Train: first 80

Test: last 20

$$y_{t+1} = 3 + 0.5 * y_t + \epsilon_{t+1}$$

$$\widehat{y_{81}} = 3 + 0.5 * y_{80}$$

$$e_{80,1} = y_{81} - \widehat{y_{81}}$$

$$\widehat{y_{82}} = 3 + 0.5 * y_{81}$$

$$e_{81,1} = y_{82} - \widehat{y_{82}}$$

Rolling scheme: training data and test data are re-determined in each iteration; use only one point in test data to obtain $e_{t,1}$.[80:20] [81:19] [82:18]…

Iteration1: 80: 20

Iteration2: 81: 19

Iteration3:82: 18

$$y_{t+1} = 3 + 0.5 * y_t + \epsilon_{t+1}$$

$$e_{80,1} = y_{81} - \widehat{y_{81}}$$

$$y_{t+1} = 3.1 + 0.6 * y_t + \epsilon_{t+1}$$

Rolling origin: training data and test data are re-determined in each iteration; use only one point in test data to obtain $e_{t,1}$.[80:20] [80:19] [80:18]…

Iteration1: 80: 20

Iteration2: 80, started with index 2: 19

Iteration3:80, started with index 3: 18

Index	$e_{t,1}$
1	1.5
2	0.3
3	0.4
…	…

Problem 3: Suppose we have 20 observations of $e_{t,1}$, why not simply calculating $E(e_t, 1)$ and check if it equals 0?

Answer: suppose $mean(e_{t,1}) = 0.2$, but we don't know if it is statistically different from 0.

Problem 4: How to do the test?

Answer: do the following linear regression and check if the intercept is significantly different from 0

$$e_{t,1} = \beta_0 * 1$$

If we can't reject the null: $\beta_0 = 0$, then the forecast is good.

4. Test3 (information efficiency test, MZ test)

$$e_{t,1} = \alpha_0 + \alpha_1 \mu_{t+1,t}$$

If $\alpha_1 = 0$, then your forecast error is not a function of past information, which is good.

In one word, all tests in the optimality check are to look for insignificance.

5. Compare one model with another model
   Get the difference of squared error between two models, then regress the difference on 1. If the coefficient is not significant, this means the difference is not significant, meaning one model is not significantly different/better than the other.

6. R application

   Step 1: import data and define as time series

   ```
   g <- ts(data$G,start=c(1990,2), end=c(2021,2), freq=4)
   ```

   Step 2: split into training and testing set

   ```
   train <- window(g, start=c(1990,2), end=c(2020,2))
   ```

   Step 3: Use training data to fit the model

   ```
   fit <- Arima(train, order=c(1,0,0))
   ```

   Step 4: Obtain $\mu_{t+1,t}$

   ```
   Forecast <- window(fitted(newfit), start=c(2020,3), end=c(2021,2))
   ```

   Step 5: Obtain $e_{t,1}$

   ```
   Forecast error <- actual - Forecast
   ```

   Step 6: Form the Table

   ```
   cbind(actual, Forecast, Forecast Error)
   ```

Actual	Forecast	Forecast Error	Constant
xxx	aaa	xxx-aaa	1
yyy	bbb	Yyy-bbb	1
…	…	…	..

   Step 7: Test1(MPE test)

   Regress Forecast Error on 1

   Step 8: Test3 (IE test)

   Regress Forecast Error on Forecast

Discussion session 7 (Apr 16, 2021)

1. Combined forecast
   100 data point
   Model A: high accuracy for the first 50 data points, low accuracy for last 50 data points
   Model B: high accuracy for 51-100, low accuracy for 1- 50.
   Ensemable model
   Model A:
   Model B:
   a. Equally weighted (CF1): 0.5*FA+0.5*FB=0.5*5+0.5*5.5=
   b. Inverse MSE (CF2): $\frac{\frac{1}{MSE_A}}{\frac{1}{MSE_A}+\frac{1}{MSE_B}} * FA + \frac{\frac{1}{MSE_B}}{\frac{1}{MSE_A}+\frac{1}{MSE_B}} * FB = a * 5 + b * 5.5 =$
   c. Regression method (CF3): $\beta_1 * FA + \beta_2 * FB$

Actual	Forecast_A	Forecast Error_A	Forecast_B	Forecast Error_B	CF_1	CF_2	CF3	Constant
xxx	aaa	xxx-aaa	ccc	xxx-ccc	0.5*aaa+0.5*ccc			1
yyy	bbb	Yyy-bbb	ddd	Yyy-ddd				1
…	…	…						..

2. Static model with time series data
   a. Residual analysis (see slides)
   b. Breusch-Godfrey test (H0: residual is WN) (see slides example)
      Objective: you want to choose a model where the residual is WN→you correctly specified the structure of autocorrelation in the error term
   c. Graphical test (plot residual against predictor/fitted value)
   d. Model selection --- predictor selection

3. Forecast with Static model (static in terms of the right-hand side variable)

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

One period forward

$$Y_{t+1} = \beta_0 + \beta_1 X_{t+1} + \epsilon_{t+1}$$

   a. Type of forecast
      i. Unconditional forecast (unconditional on X)

One period forecast

$$E(Y_{t+1}|I_t) = \beta_0 + \beta_1 E(X_{t+1}|I_t)$$

Unconditional means $E(X_{t+1}|I_t)$ needs to be forecasted. We don't have the true value to condition on.

      ii. Conditional forecast (Conditional on $X = \widetilde{X_{t+1}}$) (scenario analysis)

One period forecast

$$E(Y_{t+1}|I_t) = \beta_0 + \beta_1 \widetilde{X_{t+1}}$$

Here, the forecast means: if $X_{t+1} = x$, what is my Y?

iii. Lagged X

$$Y_t = \beta_0 + \beta_1 X_{t-4} + \epsilon_t$$

One period forward

$$Y_{t+1} = \beta_0 + \beta_1 X_{t-3} + \epsilon_{t+1}$$

One period forecast

$$E(Y_{t+1}|I_t) = \beta_0 + \beta_1 E(X_{t-3}|I_t) = \beta_0 + \beta_1 X_{t-3}$$

4. Forecast with Dynamic model
   a. Model specification
      i. Distributed Lag Model

$$Y_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \epsilon_t$$

   b. Regression with ARMA errors (Special case of ADLM)

$$Y_t = \beta_0 + \beta_1 X_t + \eta_t$$
$$\eta_t = \theta \eta_{t-1} + \epsilon_t + \phi \epsilon_{t-1}$$

   c. Regression with lags in both Y and X (Autoregressive Distributed Lag Model)

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \epsilon_t$$

Discussion session 8 (Apr 23, 2021)

1. Forecast with seasonality (deterministic)
    a. Model specification

$$y_t = c + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 + \epsilon_t$$

    b. Deterministic: same for each year

2. Forecast with seasonality (Stochastic)
    a. Model specification

$$y_t = c + \phi_{s1} y_{t-4} + \epsilon_t$$

    b. Forecast

$$E(y_{t+4}|I_t) = c + \phi_{s1} E(y_t|I_t)$$

3. Combine seasonality with ARMA model

4. Nonstationary

    a. Trend stationary – the process become stationary around its time trend
    b. Trend models
        i. Linear Trend
        ii. Log-linear Trend (adjustment)
        iii. Piece-wise Linear
        iv. Cubic Spline model
        v. Natural Spline model

Discussion session 9 (Apr 30, 2021)

1. Stochastic non-stationarity – Unit Root process

$$y_t = ay_{t-1} + by_{t-2} + \epsilon_t + c\epsilon_{t-1}$$

$$y_t = aLy_t + bL^2y_t + \epsilon_t + c\epsilon_{t-1}$$

$$\underbrace{(1 - aL - bL^2)}_{\phi(L)} y_t = \epsilon_t + c\epsilon_{t-1}$$

$\phi(L)$ is called the AR lag polynomial of the process and suppose there are two roots L_1 and L_2 to $\phi(L) = 0$. Suppose $L_1 = 1$, then the process is not covariance stationary because covariance stationary requires

$$\frac{1}{L} < 1$$

Because $L_1 = 1$, we can rewrite the above equation as:

$$(1 - dL)(1 - L)y_t = \epsilon_t + c\epsilon_{t-1}$$

$$(1 - dL)\Delta y_t = \epsilon_t + c\epsilon_{t-1}$$

$$\Delta y_t = d\Delta y_{t-1} + \epsilon_t + c\epsilon_{t-1}$$

Now, Δy_t is a stationary ARMA(1,1) process

2. Random work

$$y_t = y_{t-1} + \epsilon_t$$
$$y_{t-1} = y_{t-2} + \epsilon_{t-1}$$
$$y_{t-2} = y_{t-3} + \epsilon_{t-2}$$
$$y_{t-3} = y_{t-4} + \epsilon_{t-3}$$
$$\ldots$$
$$y_t = y_0 + \sum \epsilon_t$$

Mean of the process:

$$E(y_t) = y_0$$

Variance of the process:

$$Var(y_t) = E(y_t - y_0)^2 = E(\epsilon_1 + \epsilon_2 + \epsilon_3 + \cdots)^2 = t\sigma^2$$

Covariance of the process:

$$Cov(y_t, y_{t-k}) = E(y_t - y_0)(y_{t-k} - y_0) = (t - k)\sigma^2$$

Correlation of the process:

$$Corr(y_t, y_{t-k}) = \frac{Cov(y_t, y_{t-k})}{\sqrt{var(y_t)var(y_{t-k})}} = \sqrt{\frac{t-k}{t}} \approx 1$$

Therefore, it is not stationary.

What to do? Take the difference. $\Delta y_t = \epsilon_t$ (y_t is difference stationary).

3. Random work with a drift

$$y_t = \delta + y_{t-1} + \epsilon_t$$
$$y_t = t\delta + y_0 + \sum \epsilon_t$$

Mean of the process:

$$E(y_t) = y_0 + t\delta$$

Variance of the process:

$$Var(y_t) = E\big(y_t - E(y_t)\big)^2 = E(\epsilon_1 + \epsilon_2 + \epsilon_3 + \cdots)^2 = t\sigma^2$$

Covariance of the process:

$$Cov(y_t, y_{t-k}) = E(y_t - y_0)(y_{t-k} - y_0) = (t-k)\sigma^2$$

Correlation of the process:

$$Corr(y_t, y_{t-k}) = \frac{Cov(y_t, y_{t-k})}{\sqrt{var(y_t)var(y_{t-k})}} = \sqrt{\frac{t-k}{k}} \approx 1$$

To summarize:

The same essential results hold:

– Shocks have a permanent effect.

– Variances grow without bound as t increases

– Highly persistent (i.e., not weakly dependent)

– Can be made stationary by differencing

4. Seasonal Unity Root Process

$$y_t = y_{t-s} + \epsilon_t$$

5. Diagnostics for Unity Root Process
   a. Plot the series to check for trend
   b. Check ACF and PACF (slowly die-down of ACF, very high PACF)
   c. KPSS Test
      i. Regular unit root process

         Step 1: check the graph to see if there is a linear trend. If yes, then specify “Tau”.
         Tau: null-→Trend stationary
         Blank: null →level stationary

         Step 2: check the t statistics. If very large, reject null. →there is unit root→need to know how many times of difference you need to take.

         Step 3: use ndiffs() to determine the number of differences you need to take.

         Step 4: generate the differenced process, then go back to step 1. Stop until you can’t reject your null.

      ii. Seasonal unit root process
         Step 1: check the graph to see if there is a linear trend. If yes, then specify “Tau”.
         Tau: null-→Trend stationary
         Blank: null →level stationary

Step 2: check the t statistics. If very large, reject null. →there is unit root→need to know how many times of difference you need to take.

Step 3: use nsdiffs() to determine the number of differences you need to take.

Step 4: generate the differenced process, do the kpss test again, check if there is still unit root.

Step 5: then use ndiffs() to determine the number of differences you need to take.

Step 6: generate the differenced process, then go back to step 1. Stop until you can't reject your null.

d. Goal is to make sure the resulting process is stationary on which our regression and forecast will be based.

6. Forecast with random work

$$y_t = y_{t-1} + \epsilon_t$$

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

$$y_{t+2} = y_{t+1} + \epsilon_{t+2}$$

$$E(y_{t+h}|I_t) = y_t$$

AR process: the current information will become less and less useful as time goes.
RW process: the current information is always as important and will not diminish with the time.

Forecast error:

$$e_{t,1} = Y_{t+1} - E(Y_{t+1}|I_t) = \varepsilon_{t+1}$$

$$e_{t,2} = \underbrace{Y_{t+2}}_{y_{t+1}+\epsilon_{t+2}} - \underbrace{E(Y_{t+2}|I_t)}_{y_t} = \varepsilon_{t+1} + \epsilon_{t+2}$$

Variance of Forecast error:

$$\sigma^2_{t+h|t} = h\sigma^2$$

7. Forecast with random work with drift

$$y_t = \delta + y_{t-1} + \epsilon_t$$
$$y_{t+1} = \delta + y_t + \epsilon_{t+1}$$
$$y_{t+2} = \delta + y_{t+1} + \epsilon_{t+2}$$

Forecast error:

$$e_{t,1} = Y_{t+1} - E(Y_{t+1}|I_t) = \varepsilon_{t+1}$$

$$e_{t,2} = \underbrace{Y_{t+2}}_{\delta + \underbrace{y_{t+1}}_{\delta + y_t + \epsilon_{t+1}} + \epsilon_{t+2}} - \underbrace{E(Y_{t+2}|I_t)}_{\delta+\delta+y_t} = \varepsilon_{t+1} + \epsilon_{t+2}$$

Variance of Forecast error:

$$\sigma^2_{t+h|t} = h\sigma^2$$

8. Cubic polynomial trend

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \epsilon_t$$

$$E(y_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \underbrace{E(\epsilon_t)}_{0}$$

$$Var(y_t) = E\big(y_t - E(y_t)\big)^2$$

$$Cov(y_t, y_{t-k}) = E\big(y_t - E(y_t)\big)\big(y_{t-k} - E(y_{t-k})\big)$$

9. Exponential trend

$$y_t = \beta_0 e^{\beta_1 t} + \epsilon_t$$

$$E(y_t) = E\big(\beta_0 e^{\beta_1 t} + \epsilon_t\big) =?$$

$$Var(y_t) = E\big(y_t - E(y_t)\big)^2 =$$

Discussion session 10 (May 13, 2021)

1. Vector Autoregression

$$Y_t = c_1 + \alpha_{11} Y_{t-1} + \alpha_{12} Y_{t-2} + \ldots \alpha_{1p} Y_{t-p} + \beta_{11} X_{t-1} + \beta_{12} X_{t-2} + \ldots + \beta_{1p} X_{t-p} + \varepsilon_{1t}$$

$$X_t = c_2 + \alpha_{21} Y_{t-1} + \alpha_{22} Y_{t-2} + \ldots \alpha_{2p} Y_{t-p} + \beta_{21} X_{t-1} + \beta_{22} X_{t-2} + \ldots + \beta_{2p} X_{t-p} + \varepsilon_{2t}$$

   a. Not coming from economic theory
   b. Treating all variable's endogenous (impossible to learn the contemporary causality)
   c. Mainly reveals correlation between variables
   d. Can contain arbitrary number of variables
   e. Only contain lagged values
   f. Order of lags are same for all variables
   g. Decision of order of lag (information criteria, library(vars), VARselect)
   h. Forecasting with VAR (need to use another equation for period beyond t+1)
   i. Granger causality test (after controlling the lags of itself, does the history of other variables help to forecast it?)
      i. Lmtest, grangertest (F-test, for multiple lags/constraints)
      ii. T-test (significance in the regression, 1-lag)

H_0: $\beta_{11}=0, \beta_{12}=0, \ldots, \beta_{1p}=0$ against
H_A: at least one of the β's $\neq 0$

2. Cointegration
   a. Test of long-run eqm
   b. Individual I(1) but linear combination I(0)
   c. $Z_t = Y_t - \beta_0 + \beta_1 X_t$
   d. Engel-Granger Method
      i. H0: no cointegration
      ii. Dickey-Fuller test on the residual (ur.df())
   e. Error Correction Model (link short-run adjustment with long-run eqm)

$$\frac{(R^2(u) - R^2(r))/\#\ of\ restriction}{(1 - R^2(u))/\ dof}$$