TS Module 4 Regression methods

(The attached PDF file has better formatting.)

Time series practice problems variances and autocorrelations

** Exercise 4.1: σ^2_e, variance of Y_t, and variance of $\bar{y}$

A. An MA(1) process $Y_t = e_t - ½\ e_{t-1}$ has $\sigma^2_e = 4$. What is the variance of Y_t?
B. An MA(1) process $Y_t = e_t - ½\ e_{t-1}$ has ten observations, with $\sigma^2_e = 4$. What is the variance of $\bar{y}$, the average of the Y values?
C. An MA(1) process $Y_t = e_t - 0.6\ e_{t-1}$ has ten observations, with $\sigma^2_e = 4$. What is the variance of $\bar{y}$, the average of the Y values?

Part A: The variance $Y_t = (1 + \theta^2) \times \sigma^2_e = 1.25 \times 4 = 5$.

Y_t is the sum of independent random variables, so its variance is the sum of the variances of the random variables.

- The variance of e_t is σ^2_e.
- The variance of $-½\ e_{t-1}$ is $½ \times ½ \times \sigma^2_e = ¼\ \sigma^2_e$.

Jacob: Practice problems in module 2 determine variances of specific members of the time series, such as Y_2 or Y_3. This problem implies the variance does not depend on the subscript or the number of observations.

Rachel: Non-stationary time series do not have the same variance at each observation. In theory, even a non-stationary process may have a constant variance but it may have covariances that depend on t, but the non-stationary processes observed in practice have non-constant variances.) The random walks discussed in module 2 have linearly increasing variances. If the random walk has no beginning (t begins at $-\infty$) the variance of each observation is infinite. We assume instead that the time series starts at a certain point, such as t=1, with all previous values equal to zero (or any scalars). The variance differs for each observation.

Jacob: No time series is infinite; every time series starts at some point.

Rachel: For many time series (daily temperature, daily stock prices), the starting point is so long ago that it is not relevant. For a time series of daily temperature, we may have records of a few years or a few decades, but the time series itself is as old as the earth.

Part B: Adding the ten observations gives:

$$\sum Y_t = \sum(e_t - ½\ e_{t-1}) = e_{10} + ½\ e_9 + ½\ e_8 + ½\ e_7 \ldots + ½\ e_1 - ½\ e_0$$

The eleven random variables are independent, since we have netted terms with same subscript.

The variance of this sum is $\sigma^2 + (½)^2 \times 10\ \sigma^2 = 14$.

Dividing by 10^2 gives a variance of $14 / 10^2 = 0.14$.

The formula in the textbook is:

$$Var(\bar{Y}) \quad \frac{\gamma_0}{n}\left[1+2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\rho_k\right]$$

$\text{Var}(\overline{y}) = (\gamma_0 / n) \times [1 + 2 \times (1 - 1/n) \times (-0.4)] = (5/10) \times (1 + 2 \times (9/10) \times -0.4) = 0.14000$

(See Cryer and Chan page 28: equation 3.2.3)

Jacob: The variance of each term is 5. If we have a random sample of ten observations each of which has a variance of 5, the variance of the mean is $10 \times 5 / 10^2 = 0.5$. Why is the variance in this exercise smaller?

Rachel: The elements of this time series have a strong negative autocorrelation. If one observation is higher than expected, the next observation is expected to be lower. The expected value of the mean does not change but its variance is lower,

Part C: Adding the ten observations gives:

$$\sum Y_t = \sum(e_t - 0.6\, e_{t-1}) = e_{10} + 0.4\, e_9 + 0.4\, e_8 + 0.4\, e_7 \ldots + 0.4\, e_1 - 0.6\, e_0$$

The eleven random variables are independent, since we have netted terms with same subscript.

The variance of this sum is $\sigma^2 \times (1 + 0.4^2 \times 9 + 0.6^2) = 11.2$.

Dividing by 10^2 gives a variance of 0.112 for the mean.

** Exercise 4.2: MA(1) Process: Variance of mean

A. Two MA(1) processes with N observations each have $\sigma^2_\varepsilon = 1$.

$Y_t \ = \mu + e_t + e_{t-1}$
$Y_t' \ = \mu + e_t + \alpha \times e_{t-1}$, where $0 < \alpha < 1$

Which MA(1) has the greater variance of $\overline{y}$, the average of the N observations?

B. Two MA(1) processes with N observations each have $\sigma^2_\varepsilon = 1$.

$Y_t \ = \mu + e_t - \alpha \times e_{t-1}$, where $0 < \alpha < 1$
$Y_t' \ = \mu + e_t + \alpha \times e_{t-1}$, where $0 < \alpha < 1$

Which MA(1) has the greater variance of $\overline{y}$, the average of the N observations?

Part A: For Y_t, adding the ten observations gives $(\epsilon_t + \epsilon_{t-1}) + (\epsilon_{t-1} + \epsilon_{t-2}) + \ldots + (\epsilon_{t-9} + \epsilon_{t-10})$

$= \epsilon_t + 2\epsilon_{t-1} + 2\epsilon_{t-2} + \ldots + 2\epsilon_{t-9} + \epsilon_{t-10}$

The variance of this sum is $\sigma^2_\varepsilon + 2^2 \times 9 \times \sigma^2_\varepsilon + \sigma^2_\varepsilon$.

The variance of the mean is the variance of the sum divided by 10^2.

For Y_t', adding the ten observations gives $(\epsilon_t + \alpha \times \epsilon_{t-1}) + (\epsilon_{t-1} + \alpha \times \epsilon_{t-2}) + \ldots + (\epsilon_{t-9} + \alpha \times \epsilon_{t-10})$

$= \epsilon_t + (1 + \alpha) \times \epsilon_{t-1} + (1 + \alpha) \times \epsilon_{t-2} + \ldots + (1 + \alpha) \times \epsilon_{t-9} + \alpha \times \epsilon_{t-10}$

The variance of this sum is $\sigma^2_\varepsilon + (1 + \alpha)^2 \times 9 \times \sigma^2_\varepsilon + \alpha^2 \times \sigma^2_\varepsilon$

The variance of the mean is the variance of the sum divided by 10^2.

α is between 0 and 1, so $(1 + \alpha)^2 < 2^2$ and $\alpha^2 < 1$, so the variance of the mean of Y_t' is less than the variance of the mean of Y_t.

Part B: We computed the variance for Y_t' in Part A.

Y_t in Part B is like Y_t' except that $(1 + \alpha)$ is replaced by $(1 - \alpha)$.

α is between 0 and 1, so $(1 - \alpha)^2 < (1 + \alpha)^2$, and the variance of the mean of Y_t' is more than the variance of the mean of Y_t.