STA457H1 S/ STA 2202 HS FINAL EXAM

1. **(25%, 5% each)** Definitions:

   a) Discuss a method taught in class for removing (or modeling) seasonality of time series data.

Ans: See course notes lecture one (eg. regression method) or three (differencing or small trend methods).

   b) Describe the Dickey Fuller unit root test.

Ans: See course note.

   c) Describe the Engle-Granger method to test cointegration.

Ans:

1. **Test if data of interests are I(1) or nonstationary using unit root tests**

2. **If the data of interest are indeed I(1) or nonstationary, run regression on data of interest using the least square method.**

3. **Collect the residuals of the aforesaid regression and test if the residuals are stationary using unit root tests. If the residuals pass the test, we say that the data of interest are cointegrated.**

Remarks: They may exist multiple cointegrated relationships among a set of variables.

   d) Describe a method to test (Granger) causality between two time series.

Ans: See Lecture note. Students can answer either the VAR approach or the univariate approach (i.e. Pierce and Haugh(1977)).

e) Describe the Granger-Newbold test for assessing forecast accuracy and the underlying assumptions.

Ans: Granger and Newbold (1976) proposed a statistical test for comparing the forecast accuracy of two time series. Suppose that the one-step-ahead forecast errors are e_{1t} and e_{2t} for the two models of interest, respectively. They defined that $x_t = e_{1t} + e_{2t}, z_t = e_{1t} - e_{2t}$. Given the assumptions that

- **The forecast errors have zero mean and are normally distributed.**
- **The forecast errors are serially uncorrelated.**

and under the assumption of equal forecast accuracy, x_t and z_t should be uncorrelated—i.e., $E(x_t z_t) = E\left(e_{1t}^2 - e_{2t}^2\right) = 0$. Let r_{XY} denote the sample correlation coefficient between x_t and z_t. Under the same assumptions, Granger and Newbold showed that

$$\frac{r_{XZ}}{\sqrt{\dfrac{1-r_{XZ}^2}{H-1}}} \sim t_{H-1},$$

where H is the series length of forecast errors, and t_{H-1} denotes the Student t distribution with H-1 degrees of freedom.

Thus, if r_{XY} is statistically different from zero, model 1 has larger mean prediction square errors if $r_{XY} > 0$; model 2 has larger mean prediction square errors if $r_{XY} < 0$.

2. **(10%)** Describe the additive Holt-Winters forecasting procedure taught in class and the corresponding h-step ahead forecast formula.

Hint: YAR & CHATFIELD (1990), "Prediction intervals for the Holt-Winters forecasting procedure", *International Journal of Forecasting* **6**, 127-137.

2. The Holt–Winters forecasting procedure

Exponential smoothing methods (see e.g., Gardner, 1985; Granger and Newbold, 1986) are popular, easy-to-use and generally work well in practice. Unfortunately, the literature is confused by many different notations. We adopt the notation of Chatfield and Yar (1988) for the reasons given there. Suppose we have an observed time series, denoted by $X_1, X_2, \ldots, X_n$, and wish to forecast X_{n+k}. The forecast made at time n for k steps ahead will be denoted by $\hat{X}_n(k)$. For a univariate forecast this depends only on X_n, $X_{n-1}, \ldots$. In simple exponential smoothing, the one-step-ahead predictor can be written in the recurrence form

$$\hat{X}_t(1) = \alpha X_t + (1-\alpha)\hat{X}_{t-1}(1), \tag{1}$$

where the smoothing parameter, α, is usually constrained so that $0 < \alpha < 1$. The Holt–Winters method (sometimes called the Winters method or seasonal exponential smoothing) generalises this approach to deal with trend and seasonality. Let α, γ, δ denote three smoothing parameters and let p denote the number of observations per seasonal cycle. If seasonality is additive (so that its size does not depend on the current mean level), then the error-correction recurrence formulae for updating the local mean level, L_t, the local trend, T_t, and the local seasonal index, I_t, when a new observation X_t becomes available, are

$$L_t = L_{t-1} + T_{t-1} + \alpha e_t, \tag{2}$$

$$T_t = T_{t-1} + \alpha\gamma e_t, \tag{3}$$

$$I_t = I_{t-p} + \delta(1-\alpha)e_t, \tag{4}$$

where $e_t = X_t - \hat{X}_{t-1}(1)$ is the one-step-ahead forecast error at time t. Then the new forecast made at time t of X_{t+k} is given by

$$\hat{X}_t(k) = L_t + kT_t + I_{t-p+k} \tag{5}$$

for $k = 1, 2, \ldots, p$. There are analogous formulae for the multiplicative-seasonality case.

3. **(15%)** Box, Jenkins, and Reinsel (1994) fit a transfer function model to data from a gas furnace. The input variable x_t is the volume of methane entering the chamber in cubic feet per minute and the output is the concentration of carbon dioxide emitted y_t. The transfer function model is

$$y_t = \frac{-(0.53 + 0.37B + 0.51B^2)}{1 - 0.57B} x_t + \frac{1}{1 - 0.53B + 0.63B^2} \varepsilon_t$$

where the input and output variables are measured every nine seconds.

a) What are the value of b, s, and r for this model?

Ans: b=0 (no lag), r=1 (one term in the denominator) and s=2 (two terms in the numerator)

b) What is the form of the *ARIMA* model for the errors?

Ans: There are no terms in the numerator and two in the denominator so the errors follow an AR(2) model.

c) If the methane input was increased, how long would it take before the carbon dioxide concentration in the output is impacted?

Ans: Since there is no lag, the carbon dioxide concentration in the output is impacted immediately.

4. **(10%)** $X_t = 0.5X_{t-1} + Z_t$, $Z_t \sim NID(0,1)$. What is the variance of $\dfrac{X_1 + X_2 + X_3}{3}$?

Hint: $\mathrm{var}(\dfrac{X_1 + X_2 + X_3}{3}) = \dfrac{1}{9}\mathrm{var}(X_1 + X_2 + X_3)$.

Ans:

$$
\begin{aligned}
\mathrm{var}(X_1 + X_2 + X_3) &= \mathrm{var}(X_1) + \mathrm{var}(X_2) + \mathrm{var}(X_3) \\
&\quad + 2\,\mathrm{cov}(X_1, X_2) + 2\,\mathrm{cov}(X_1, X_3) + 2\,\mathrm{cov}(X_2, X_3) \\
\because \text{ stationary} \quad &= 3\gamma_X(0) + 2\gamma_X(1) + 2\gamma_X(2) + 2\gamma_X(1)
\end{aligned}
$$

where $\gamma_X(0) = \dfrac{1}{1-0.5^2} = \dfrac{4}{3}$ **and** $\gamma_X(k) = 0.5^k \cdot \gamma_X(0)$ **.**

5. (25%) Forecast an $ARMA(1,1)$ model:

$$X_t - 0.5X_{t-1} = a_t + 0.25a_{t-1}, \quad a_t \sim NID(0,1) \qquad (*)$$

Suppose that $X_{97} = -0.7, X_{98} = -1, X_{99} = -0.8, X_{100} = -0.4$. Answer the following questions:

a) Write down the forecasting function for eqn. (*) (10%).

Ans:

- Lead time=1: $\hat{X}_t(1) = 0.5X_t + 0.25a_t = 0.5X_t + 0.25 \cdot (X_t - \hat{X}_{t-1}(1))$
- Lead time h, h>1: $\hat{X}_t(h) = 0.5\hat{X}_t(h-1)$

b) Calculate the best linear forecast of $X_{101} + X_{102} + X_{103}$. (5%).

Hint: $\hat{X}_{100}(1) + \hat{X}_{100}(2) + \hat{X}_{100}(3)$

Ans: Skip (For simplicity, students may assume a_t=0.

c) Calculate the 95% forecast (confidence) interval of the forecast in question 5b). For simplicity, use $Z_{0.975} \approx 2$ in your calculation. (10%)

Ans: Skip. Students need to calculate confidence level as var(et(1)+et(2)+et(3)), where et(h) is the forecast error at time t with lead time h.

(10%) Consider a two dimensional vector autoregressive process of order one

$$\begin{bmatrix} r_{2t} \\ r_{1t} \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 1.1 & -0.6 \\ 0.3 & 0.2 \end{bmatrix} \begin{bmatrix} r_{2,t-1} \\ r_{1,t-1} \end{bmatrix} + \begin{bmatrix} a_{2t} \\ a_{1t} \end{bmatrix},$$

where $\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ is the covariance matrix between a_{2t} and a_{1t}. Answer whether the above VAR(1) model is (weakly) stationary.

Ans: Let $\Phi = \begin{bmatrix} 1.1 & -0.6 \\ 0.3 & 0.2 \end{bmatrix}$. For the process to be staionary, the zeros of the determinantal equation $|I - \Phi B|$ must be outside the unit circle. Letting $\lambda = B^{-1}$, we have

$|I - \Phi B| = 0 \Leftrightarrow |\lambda I - \Phi_1| = 0$. Thus, the zeros of $|I - \Phi B|$ are related to the eigenvalues of Φ. Let λ_1 and λ_2 be the eigenvalues (assume that the eigenvectors are linear independent). Thus, we have $|I - \Phi B| = \prod_{i=1}^{2}(1 - \lambda_i B)$. Hence, the zeros of $|I - \Phi B|$ are outside the unit circle if and only if all the eigenvalues are inside the unit circle. Using R, we have

```
> m<-matrix(c(1.1,0.3,-0.6,0.2),2,2,F)

> eigen(m)$values

[1] 0.8 0.5
```

Therefore, the process is stationary.

6. **(15%)** Consider an ARCH(1) Process

$$X_t = \sigma_t Z_t,$$

where $Z_t \sim NID(0,1)$, $\sigma_t^2 = w_0 + w_1 X_{t-1}^2$, $w_0 > 0$, and $w_1 < 1$. Show that the above ARCH(1) process is "fat-tailed". Hint: Kurtosis greater than 3.

Ans: See next page (using different notation though).

To understand the ARCH models, it pays to carefully study the ARCH(1) model

$$a_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2,$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$. First, the unconditional mean of a_t remains zero because

$$E(a_t) = E[E(a_t | F_{t-1})] = E[\sigma_t E(\epsilon_t)] = 0.$$

Second, the unconditional variance of a_t can be obtained as

$$\begin{aligned}\mathrm{Var}(a_t) &= E(a_t^2) = E[E(a_t^2|F_{t-1})] \\ &= E(\alpha_0 + \alpha_1 a_{t-1}^2) = \alpha_0 + \alpha_1 E(a_{t-1}^2).\end{aligned}$$

Because a_t is a stationary process with $E(a_t) = 0$, $\mathrm{Var}(a_t) = \mathrm{Var}(a_{t-1}) = E(a_{t-1}^2)$. Therefore, we have $\mathrm{Var}(a_t) = \alpha_0 + \alpha_1 \mathrm{Var}(a_t)$ and $\mathrm{Var}(a_t) = \alpha_0/(1-\alpha_1)$. Since the variance of a_t must be positive, we require $0 \le \alpha_1 < 1$. Third, in some applications, we need higher order moments of a_t to exist and, hence, α_1 must also satisfy some additional constraints. For instance, to study its tail behavior, we require that the fourth moment of a_t is finite. Under the normality assumption of ϵ_t in Eq. (3.5), we have

$$E(a_t^4|F_{t-1}) = 3[E(a_t^2|F_{t-1})]^2 = 3(\alpha_0 + \alpha_1 a_{t-1}^2)^2.$$

Therefore,

$$E(a_t^4) = E[E(a_t^4|F_{t-1})] = 3E(\alpha_0 + \alpha_1 a_{t-1}^2)^2 = 3E(\alpha_0^2 + 2\alpha_0\alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4).$$

If a_t is fourth-order stationary with $m_4 = E(a_t^4)$, then we have

$$\begin{aligned}m_4 &= 3[\alpha_0^2 + 2\alpha_0\alpha_1 \mathrm{Var}(a_t) + \alpha_1^2 m_4] \\ &= 3\alpha_0^2\left(1 + 2\frac{\alpha_1}{1-\alpha_1}\right) + 3\alpha_1^2 m_4.\end{aligned}$$

Consequently,

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}.$$

This result has two important implications: (a) since the fourth moment of a_t is positive, we see that α_1 must also satisfy the condition $1 - 3\alpha_1^2 > 0$; that is, $0 \le \alpha_1^2 < \frac{1}{3}$; and (b) the unconditional kurtosis of a_t is

$$\frac{E(a_t^4)}{[\mathrm{Var}(a_t)]^2} = 3\frac{\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\alpha_0^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3.$$

Thus, the excess kurtosis of a_t is positive and the tail distribution of a_t is heavier than that of a normal distribution. In other words, the shock a_t of a conditional Gaussian ARCH(1) model is more likely than a Gaussian white noise series to produce "outliers." This is in agreement with the empirical finding that "outliers" appear more often in asset returns than that implied by an iid sequence of normal random variates.