Exam II: Information and Practice Problems with Solutions

Statistics 509 – Winter 2022

April 19, 2021

Information: The Second Exam will be Friday, April 22 from 1:30-3:30 in room **SEB1202** (in the School of Education Building). The exam will cover the material in Lectures 7-12 covered in class and corresponding relevant material in the book. The exam is open textbook (only the textbook) and open notes. Below is a set of practice Exam Problems. Additional practice problems are the exercise problems from the lectures and exercises in the book.

Practice Exam Problems

1. Problem 8 on page 514 in Ruppert/Matteson, but change part (d) to be a 4% expected return, and add the following part **(e)**.

   (e) Determine the β's of asset C and asset D relative to the tangency portfolio.

2. Problem 7 on page 513 of Ruppert/Matteson.

3. Suppose have two stationary AR time series

$$X_n = .5X_{n-1} + \epsilon_n$$

$$Y_n = \epsilon'_n + .6\epsilon'_{n-1}$$

   where ϵ and ϵ' are both white-noise processes with mean 0 and standard deviation of 0.5, and which are independent of each other.

   (a) Suppose $X'_n = X_n + X_{n+1}$. Show that $X' = \{X'_n\}$ is a weakly stationary process and derive it's auto-covariance function.

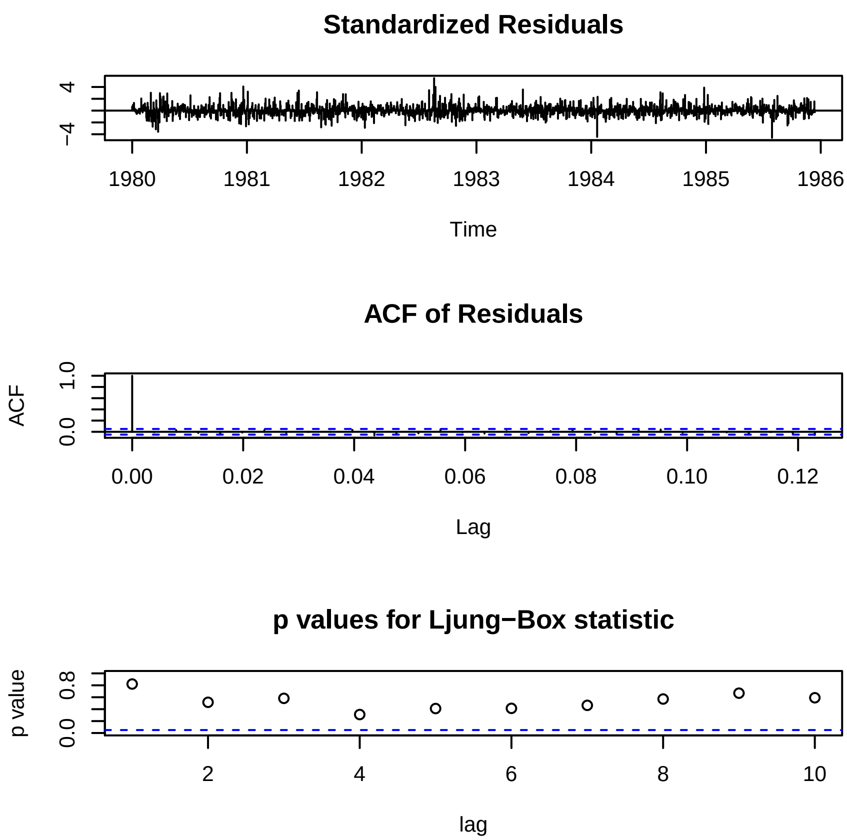
   (b) Suppose generate a new time series of $Z_n = X_n + Y_n$ – verify that this is a stationary process, and derive its mean and autocovariance function.

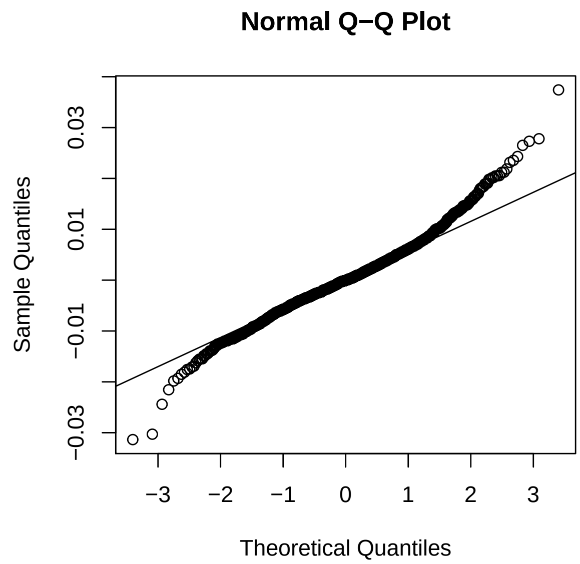
   (c) With Z as in **(b)**, find the optimal linear predictor of X_{25} based on Z_{24} and determine the mean-squared prediction error of this linear predictor.

4. Suppose had the following results a time series analysis in R applied to log of the daily price data from Dow Jones Utility Average Index.

```
> X = read.csv("Data//DJUtilityAv-daily.csv",header=TRUE)
> X_logprice = rev(log(X$Adj.Close))
> X_logprice.ts = ts(data=X_logprice,start=c(1980,1),frequency=252,
+ names=c("DJUtilityAv_DailyLogPrice"))
> auto.arima(X_logprice.ts,max.p=4,max.q=4,ic="bic")
Series: X_logprice.ts
ARIMA(1,1,0)
```

```
Coefficients:
         ar1
      0.2009
s.e.  0.0253

sigma^2 estimated as 4.633e-05:  log likelihood=5352.84
AIC=-10701.67   AICc=-10701.67   BIC=-10691.05
> DJUtility_ARIMA = arima(X_logprice.ts, order = c(1,1,0), method = "ML")
> tsdiag(DJUtility_ARIMA)
> qqnorm(DJUtility_ARIMA$residuals)
> qqline(DJUtility_ARIMA$residuals)
```





(a) What model was selected by auto.arima, and what criteria was used for this selection?

(b) From these results, what are the estimated coefficients and the estimated standard deviation for the white noise process?

(c) Based on the diagnostic plots, what can you say about the fit of the this ARIMA model to the daily log-price data, i.e., what is good and what is not?

(d) What alternative models/approaches might you consider based on these results?

5. Suppose have time series $X = \{X_n\}_{n=0}^{\infty}$ following a first-order AR model

$$X_n = \alpha X_{n-1} + \epsilon_n \qquad\qquad n = 1, 2, \ldots$$

where assume that $|\alpha| \leq 1$ and $\{\epsilon_n\}$ is a white noise process with mean zero and variance of σ^2. The least-squares estimates of α is given by

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} X_i X_{i-1}}{\sum_{i=1}^{n} X_i^2}.$$

(a) Assuming $|\alpha| < 1$, ergodicity, and $\{X_n\}$ is a stationary process, derive the limits of the normalized numerator and denominator of $\hat{\alpha}$, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_{i-1} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Justify your answer.

(b) Assuming $X_0 = 0$ and $\alpha = 1$, derive expressions for the expected value of the normalized numerator and denominator of $\hat{\alpha}$. *Hint:* Can use that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

(c) Based on results in **(a)** and **(b)**, would you say that $\hat{\alpha}$ is a reasonable estimate of α for all cases. Explain your answer.

(d) Based on results in **(a)** and **(b)** how might one test the hypothesis of $H_o : \alpha = 1$ vs. $H_1 : \alpha < 1$, i.e., what would be the general form of the test based on the test statistic $\hat{\alpha}$.

6. **(a)** Problem 12 on page 357 in Ruppert/Matteson.

   (b) Problem 16 on page 359 in Ruppert/Matteson, and add another part which is to compute the mean-squared prediction error of the 3 predictions, assuming perfectly accurate estimates on the coefficients.

7. Suppose that X_n is time series which corresponds to an ARCH(2) model of the form

$$X_n = \sigma_n \epsilon_n$$

   where

$$\sigma_n^2 = 1 + .4 \cdot X_{n-1}^2 + .2 \cdot X_{n-2}^2$$

   (a) Derive the conditional distribution of X_3 given $X_1 = 1, X_2 = 2$ assuming $\{\epsilon_n\}$ are iid $\mathcal{N}(0, 1)$.

   (b) Derive the conditional distribution of X_3 given $X_1 = 1, X_2 = 2$ assuming $\{\epsilon_n\}$ are iid DExp$(0, 1)$.

   (c) Answer **(a)** and **(b)** if the equations are

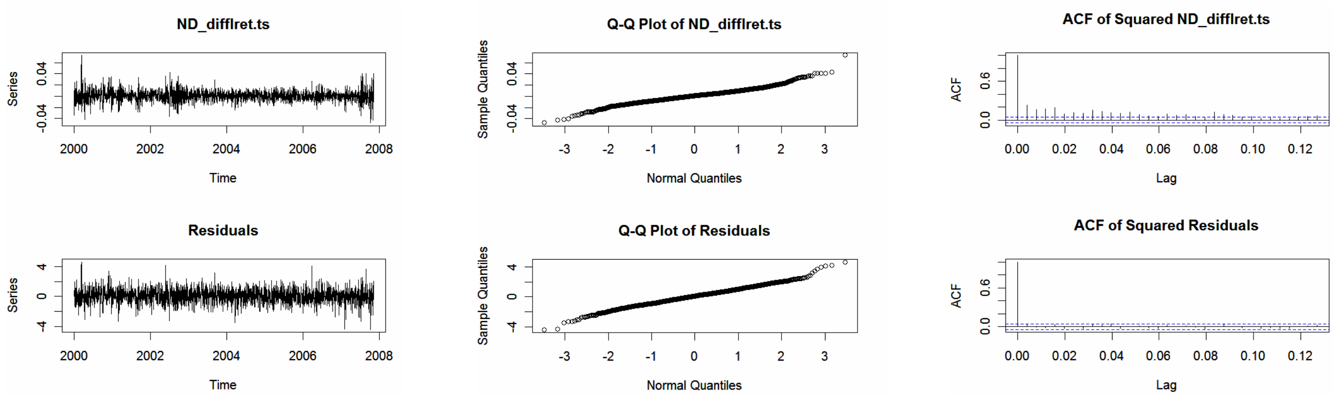
$$X_n = 0.5 + \sigma_n \epsilon_n$$

   where

$$\sigma_n^2 = 1 + .4 \cdot (X_{n-1} - 0.5)^2 + .2 \cdot (X_{n-2} - 0.5)^2$$

8. Suppose have the following results for a GARCH estimation on differenced log-returns of the Daily Adjusted Closing Prices of the NASDAQ from Jan-2000 to Nov-2007 – note we are assuming a mean of 0 for the process.

```
> summary(ND_garchest11)

Call:
garch(x = ND_difflret.ts, order = c(1, 1), itmax = 200, grad = c("analytic"))

Model:
GARCH(1,1)

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 2.743e-06   5.323e-07    5.153 2.57e-07 ***
a1 8.888e-02   9.438e-03    9.418  < 2e-16 ***
b1 8.848e-01   1.206e-02   73.375  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 75.8134, df = 2, p-value < 2.2e-16

        Box-Ljung test

data:  Squared.Residuals
X-squared = 2.1571, df = 1, p-value = 0.1419
```



(a) Based only on the plot of the differenced log-returns, does it appear that GARCH(1,1) model is better than a stationary white-noise for modeling this time series of the differenced log-returns. Explain your answer.

(b) Interpret the results of the GARCH analysis and say whether the estimates are appropriate/valid. Justify your answer and explain what are the deficiencies in this analysis.

(c) This data goes through the end of November 2007 – based on the estimated model being valid and assuming that the difference in log returns on the last trading day in November was $x_n = .0140$ and the standard deviation was $\hat{\sigma}_n = 0.0214$, estimate the standard deviation of the difference of log-returns on the first trading day in December (σ_{n+1}) and also estimate the square of differenced log-return (X_{n+1}^2).

9. Suppose have a weakly stationary process

$$X_n = \epsilon_n + .5\epsilon_{n-1} + .5\epsilon_{n-2} \qquad\qquad n = 1, 2, \ldots$$

where $\{\epsilon_n\}$ is a white noise process with variance σ^2.

(a) Derive the mean and autocovariance of X_n - you may derive it from first principles or rigorously quote results from class.

(b) Suppose $Y_n = X_n + X_{n-1}$. Is this time series $\{Y_n\}$ ARMA(p, q)? Justify your answer and specify parameters p and q if your answer is yes.

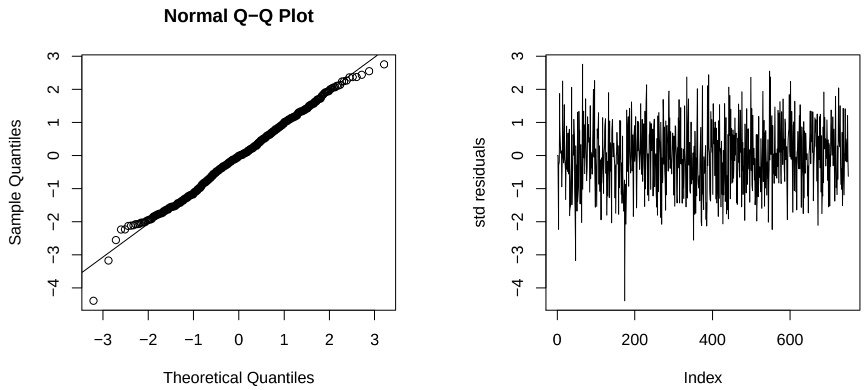
(c) What is your best linear prediction of X_{n+2} based on knowing only X_n? What is the the mean-squared prediction error of this linear predictor?

10. Suppose have daily price data for SP Midcap 400 index (for 2001-2003) and the following is the R-output of an analysis (commands and plots).

```
> X = read.csv("Data//SPMidcap400.csv",header=TRUE)
> X_logret = diff(log(X$AdjClose))
> X_logret.ts = ts(data=X_logret,start=c(2001,1),frequency=252,names=c("SPMidcap400"))
```

```
>
> SP400_garch <- garchFit(~arma(2,0)+garch(1,1), data=X_logret.ts[1:750], cond.dist=c("snorm
> summary(SP400_garch)

Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(2, 0) + garch(1, 1), data = X_logret.ts[1:750],
    cond.dist = c("snorm"), include.mean = T, include.skew = F,
    trace = F)

Mean and Variance Equation:
 data ~ arma(2, 0) + garch(1, 1)
<environment: 0x0000000004a0c358>
 [data = X_logret.ts[1:750]]

Conditional Distribution:  snorm

Std. Errors:  based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      6.879e-04   4.299e-04    1.600   0.1096
ar1     2.861e-02   3.788e-02    0.755   0.4501
ar2    -8.086e-02   3.757e-02   -2.152   0.0314 *
omega   5.324e-06   2.651e-06    2.008   0.0446 *
alpha1  8.996e-02   2.206e-02    4.078 4.53e-05 ***
beta1   8.792e-01   3.024e-02   29.071  < 2e-16 ***
---

Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  0.6848627 0.7100419
 Shapiro-Wilk Test  R    W      0.9953828 0.02412903
 Ljung-Box Test     R    Q(10)  10.54462  0.3940778
 Ljung-Box Test     R    Q(15)  17.11203  0.3122116
 Ljung-Box Test     R    Q(20)  20.2074   0.445025
 Ljung-Box Test     R^2  Q(10)  13.88344  0.1783752
 Ljung-Box Test     R^2  Q(15)  14.37448  0.4973381
 Ljung-Box Test     R^2  Q(20)  18.14707  0.5777204
 LM Arch Test       R    TR^2   16.49336  0.1696692

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-5.880602 -5.843641 -5.880729 -5.866360

> qqnorm(SP400_garch@residuals/SP400_garch@sigma.t)
> qqline(SP400_garch@residuals/SP400_garch@sigma.t)
```

```
> plot(SP400_garch@residuals/SP400_garch@sigma.t,type='l',ylab='std residuals')
```



(a) What model is being fit to this time series, and state clearly the final mathematical form of the model as estimated?

(b) Based on these estimates, what is the model for the conditional variance of the the white noise process?

(c) Based on diagnostics as shown here, how satisfied are you with this model? What alternative models might you investigate to get an even better fit? How would you select between this model and your proposed alternatives?

11. **(a)** Suppose X and σ^2 are two random variables and that X conditional on σ^2 is $\mathcal{N}(0, \sigma^2)$, i.e. that conditioned on the value of σ^2, X is mean-zero normal with variance σ^2. Show that

$$\frac{E(X^4)}{[\mathrm{Var}(X)]^2} \quad = \quad 3\left[1 + \frac{\mathrm{Var}(\sigma^2)}{[E(\sigma^2)]^2}\right].$$

What does this say about the excess kurtosis of X relative to that of a normal random variable?

Hint: For the above problem, recall that from your basic probability/statistics class the results that

$$E(X) = E(E(X|\sigma)),\ E(X^2) = E(E(X^2|\sigma)),\ E(X^4) = E(E(X^4|\sigma)).$$

(b) Suppose that X_n is a GARCH(1,1) process with parameters of α_o, α_1 and β_1, and the innovations $\{\epsilon_n\}$ are iid $\mathcal{N}(0, 1)$. Show that the kurtosis of X_n (assuming stationarity) is given by

$$\kappa = \frac{E(X_n^4)}{[\mathrm{Var}(X_n)]^2} = \frac{3\,[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3$$

when $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 < 1$.

(c) What are the implications for the GARCH(1,1) process if $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 \geq 1$?

12. Exercise 6 on page 448 in Ruppert/Matteson.

13. Exercise 5 on page 448 in Ruppert/Matteson.

14. Suppose $Y_{1,t}, Y_{2,t}$ are log-prices satisfying the following model for log-returns,

$$
\begin{aligned}
\triangle Y_{1,t} \equiv Y_{1,t} - Y_{1,t-1} &= \phi_1(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{1,t} \\
\triangle Y_{2,t} \equiv Y_{2,t} - Y_{2,t-1} &= \phi_2(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{2,t}
\end{aligned}
$$

where $\epsilon_{1,t}, \epsilon_{2,t}$ are white noise processes.

(a) Show that the above implies that $Y_{1,t} - \lambda Y_{2,t}$ satisfies an AR(1) model and identify the coefficients (subtract λ times second equation from the first equation).

(b) Specify sufficient conditions on ϕ_1, ϕ_2, λ implying that the AR(1) process is stationary.

(c) Problem 2 on page 463 in Ruppert/Matteson – basically adding μ_1 and μ_2 to the righthand side of the two equations above.

SOLUTIONS

1. **(a)** The expected return of the tangency portfolio is

$$\mu_T = .6 * .04 + .4 * .06 = .048$$

(b) The standard deviation is given by

$$\sigma_T = \sqrt{.6^2 * .10^2 + .4^2 * .18^2 + 2 * .6 * .4 * \underline{.5 * .10 * .18}} = .114$$

(c) Now the efficient portfolio is a mixture of the tangency portfolio and the fixed-rate asset, with the weight w on the tangency portfolio being such that

$$w\sigma_T = .03$$

or

$$w = \frac{.03}{\sigma_T} = \frac{.03}{.114} = .263$$

so proportion in the risk-free asset should be .737, i.e., 73.7%. This is the only solution.

(d) To achieve an efficient portfolio with expected return being 4% (recall we changed problem to be 4% instead of the original 7%), want to again take a mixture of the tangency portfolio and the fixed-rate asset with the weight for the tangency portfolio being

$$w\mu_T + (1 - w)\mu_f = .04$$

or

$$w = \frac{.04 - .012}{.048 - .012} = .778$$

i.e., want 22.2% in the risk-free, want .778*60=46.68% in asset C, and .778*40=31.12% in asset D.

(e) Based on work in class, we know that

$$\beta_C = \frac{.04 - .012}{.048 - .012} = .778$$

$$\beta_D = \frac{.06 - .012}{.048 - .012} = 1.333$$

2. **(a)** If have a portfolio equally weighted between the 3 assets, then

$$R_p = \frac{1}{3}R_1 + \frac{1}{3}R_2 + \frac{1}{3}R_3$$

and so

$$\begin{aligned} R_p - \mu_f &= \frac{1}{3}(\beta_1 + \beta_2 + \beta_3) * (R_M - \mu_f) + \frac{1}{3}(\epsilon_1 + \epsilon_2 + \epsilon_3) \\ &= \tilde{\beta} * (R_M - \mu_f) + \tilde{\epsilon} \end{aligned}$$

where

$$\tilde{\beta} = \frac{1}{3}(\beta_1 + \beta_2 + \beta_3) = \frac{2.6}{3} = .867$$

and

$$\tilde{\epsilon} = \frac{1}{3}\left(\epsilon_1 + \epsilon_2 + \epsilon_3\right)$$

and so the beta is .867.

(b) The variance of the excess return is given by

$$\begin{aligned}
\mathrm{Var}(R_p - \mu_f) &= \mathrm{Var}(.867 * (R_M - \mu_f)) + \mathrm{Var}(\tilde{\epsilon}) \\
&= .867^2 * .014 + \left(\frac{1}{3}\right)^2 * (.010 + .015 + .011) = 0.0145
\end{aligned}$$

(c) The proportion of risk of asset 1 due to market risk is

$$\frac{.9^2 * .014}{.9^2 * .014 + .01} = .531$$

so just a little over 50%.

3. **(a)** Now X_n is an AR(1) process with parameter $\alpha_1 = .5$, so

$$\gamma_X(n) = \frac{(.5)^{|n|} \cdot (.5)^2}{1 - .5^2} = \frac{.5^{|n|+2}}{.75} \qquad \text{for any } n$$

For $n, n+m$,

$$\begin{aligned}
\mathrm{Cov}(X'_n, X'_{n+m}) &= \mathrm{Cov}(X_n + X_{n+1}, X_{n+m} + X_{n+m+1}) \\
&= \mathrm{Cov}(X_n, X_{n+m}) + \mathrm{Cov}(X_n, X_{n+m+1}) + \mathrm{Cov}(X_{n+1}, X_{n+m}) + \mathrm{Cov}(X_{n+1}, X_{n+m+1}) \\
&= \gamma_X(m) + \gamma_X(m+1) + \gamma_X(m-1) + \gamma_X(m) \\
&= \frac{2 \cdot (.5)^{|m|+2} + .5^{|m+1|+2} + (.5)^{|m-1|+2}}{.75}
\end{aligned}$$

and this is only a function of the difference of m, so $X' = \{X'_n\}$ is a stationary process and so the above is the autocovariance, i.e.,

$$\gamma_{X'}(m) = \frac{2 \cdot (.5)^{|m|+2} + .5^{|m+1|+2} + (.5)^{|m-1|+2}}{.75}$$

(b) Now Y_n is a MA(1) process with $\theta_1 = .6$, so by work in class it is stationary with autocovariance of

$$\gamma_Y(n) = \begin{cases} (.5)^2 \cdot (1 + .6^2) & n = 0 \\ (.5)^2 \cdot (1 \cdot .6) & n = \pm 1 \\ 0 & |n| \geq 2 \end{cases}$$

Now

$$E(Z_n) = E(X_n + Y_n) = E(X_n) + E(Y_n) = 0 + 0 = 0.$$

and

$$\begin{aligned}
\mathrm{Cov}(Z_n, Z_{n+m}) &= \mathrm{Cov}(X_n + Y_n, X_{n+m} + Y_{n+m}) \\
&= \mathrm{Cov}(X_n, X_{n+m}) + \mathrm{Cov}(Y_n, Y_{n+m}) \qquad \text{by independence of } X \text{ and } Y \\
&= \gamma_X(m) + \gamma_Y(m)
\end{aligned}$$

Since the mean function does not depend on n and auto-covariance only depends on the difference, we have Z is weakly stationary and

$$\mu_Z = 0$$

$$\gamma_Z(m) = \gamma_X(m) + \gamma_Y(m) = \frac{(.5)^{|m|+2}}{1-(.5)^2} + \gamma_Y(m).$$

(c) The optimal linear predictor of X_{25} based on Z_{24} is

$$\begin{aligned}
\hat{X}_{25} &= E(X_{25}) + \frac{\mathrm{Cov}(X_{25}, Z_{24})}{\mathrm{Var}(Z_{24})} \cdot (Z_{24} - E(Z_{24})) \\
&= \frac{\mathrm{Cov}(X_{25}, X_{24} + Y_{24})}{\gamma_Z(0)} \cdot Z_{24} \\
&= \frac{\gamma_X(1)}{\gamma_Z(0)} \cdot Z_{24} \\
&= \frac{\frac{(.5)^2(.5)}{1-(.5)^2}}{\frac{(.5)^2}{1-(.5)^2} + (.5)^2 \cdot (1 + .6^2)} \cdot Z_{24} \\
&= 0.248 \cdot Z_{24}
\end{aligned}$$

where $\gamma_X(1)$ and $\gamma_Z(0)$ are as above. The mean-squared prediction error is given by

$$\begin{aligned}
\text{MSPE} &= \gamma_X(0) - \frac{(\gamma_X(1))^2}{\gamma_Z(0)} \\
&= \frac{(.5)^2}{1-(.5)^2} - \frac{\left\{\frac{(.5)^2(.5)}{1-(.5)^2}\right\}^2}{\frac{(.5)^2}{1-(.5)^2} + (.5)^2 \cdot (1 + .6^2)} \\
&= 0.292
\end{aligned}$$

4. **(a)** The model selected was ARIMA(1,1,0), i.e., an AR(1) model on the difference process. The criteria used for this selection with the BIC information criteria, i.e., choosing the model that minimized that information criteria.

   (b) The estimated coefficients are that AR coefficient estimate is $\hat{\alpha}_1 = .2009$ and the variance estimate is $\hat{\sigma}_\epsilon^2 = .0000463$ or $\hat{\sigma}_\epsilon = 0.0068$.

   (c) The diagnostic plots for the plots of the residuals vs. time, the ACF, and the p-values of the Ljung-Box test are all pretty good in that they seem to suggests that the residual process is approximately uncorrelated. The QQ plot is not so good in that it shows a clear indication that the residuals are non-Gaussian and have signficantly heavier tails on both the upper and lower as compared to Gaussian.

   (d) Some alternative models are: (i) assume a t-distribution for the white noise error process and use this with the ARIMA model, and (ii) utilize a GARCH model on the errors, i.e., would be AR(1)-GARCH model on the difference process.

5. **(a)** Note that ergodicity implies that the limits of these sums is equal to the expected value. The expected value of the numerator is

$$\begin{aligned} E\left[\frac{1}{n}\sum_{i=1}^{n} X_i X_{i-1}\right] &= E(X_1 X_2) \\ &= \text{Cov}(X_1, X_2) \\ &= \gamma_X(1) = \frac{\sigma^2\alpha}{1-\alpha^2} \end{aligned}$$

The expected value of the denominator is

$$\begin{aligned} E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] &= \text{Var}(X_1) \\ &= \gamma_X(0) = \frac{\sigma^2}{1-\alpha^2} \end{aligned}$$

(b) In the case where $\alpha = 1$, X_n corresponds to a traditional random walk, so we have

$$\begin{aligned} E\left[\frac{1}{n}\sum_{i=1}^{n} X_i X_{i-1}\right] &= \frac{1}{n}\sum_{i=1}^{n} E(X_i X_{i-1}) \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left[(X_{i-1}+\epsilon_i)X_{i-1}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E(X_{i-1}^2) \\ &= \frac{1}{n}\sum_{i=1}^{n} (i-1)\sigma^2 \\ &= \frac{1}{n}\sum_{i=0}^{n-1} i\sigma^2 \\ &= \frac{1}{n}\frac{(n-1)n}{2}\sigma^2 = \frac{n-1}{2}\sigma^2 \end{aligned}$$

and

$$\begin{aligned} E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] &= \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) \\ &= \frac{1}{n}\sum_{i=1}^{n} i\sigma^2 \\ &= \frac{1}{n}\frac{n(n+1)}{2}\sigma^2 = \frac{n+1}{2}\sigma^2 \end{aligned}$$

(c) Based on results in **(a)** and **(b)**, it would seem that this estimate is good for all cases since the expected value of the numerator divided by the expected value of the denominator is exactly α when $|\alpha| < 1$ and in the case that $\alpha = 1$, we have that the ratio of expected values is $\frac{n-1}{n+1}$ which is very close to 1 when n is large.

(d) It is natural to reject H_o in favor of H_1 when the estimate $\hat{\alpha}$ is significantly less than 1, i.e., decision rule would be

$$\text{Reject } H_o \text{ if } \quad \hat{\alpha} < \kappa < 1$$

where κ is chosen so to achieve a fixed probability of error given that H_o is true. Could also restate this in terms of p-values.

6. **(a)** Note that

$$\begin{aligned} \hat{Y}_{n+1} &= \hat{\mu} + \hat{\theta}_1 \hat{\epsilon}_n + \hat{\theta}_2 \hat{\epsilon}_{n-1} \\ &= 45 + .3 * 1.5 - .15 * (-4.3) = 46.095 \end{aligned}$$

and then

$$\begin{aligned} \hat{Y}_{n+2} &= \hat{\mu} + \hat{\theta}_2 \hat{\epsilon}_n \\ &= 45 - .15 * 1.5 = 44.775 \end{aligned}$$

(b) Note that

$$\begin{aligned} \hat{Y}_{n+1} &= (1 - \hat{\phi}_1 - \hat{\phi}_2) * \hat{\mu} + \hat{\phi}_1 Y_n + \hat{\phi}_2 Y_{n-1} \\ &= (1 - .5 - .1) * 100.1 + .5 * 102.3 + .1 * 99.5 = 101.14 \end{aligned}$$

and then

$$\begin{aligned} \hat{Y}_{n+2} &= (1 - \hat{\phi}_1 - \hat{\phi}_2) * \hat{\mu} + \hat{\phi}_1 \hat{Y}_{n+1} + \hat{\phi}_2 Y_n \\ &= (1 - .5 - .1) * 100.1 + .5 * 101.14 + .1 * 102.3 = 100.84 \end{aligned}$$

and then

$$\begin{aligned} \hat{Y}_{n+3} &= (1 - \hat{\phi}_1 - \hat{\phi}_2) * \hat{\mu} + \hat{\phi}_1 \hat{Y}_{n+2} + \hat{\phi}_2 \hat{Y}_{n+1} \\ &= (1 - .5 - .1) * 100.1 + .5 * 100.84 + .1 * 101.14 = 100.574 \end{aligned}$$

Now we drop the hats on the parameter estimates as they are assumed to equal to the true, and we let $\phi_o = \mu(1 - \phi_1 - \phi_2)$ – then the MSPE for the 3 predictions are given by

$$
\begin{aligned}
\text{MSPE}(\hat{Y}_{n+1}) &= E\left[\left(\hat{Y}_{n+1} - Y_{n+1}\right)^2\right] \\
&= \underline{E(\epsilon_{n+1}^2) = \sigma_\epsilon^2} \\
\text{MSPE}(\hat{Y}_{n+2}) &= E\left[\left(\hat{Y}_{n+2} - Y_{n+2}\right)^2\right] \\
&= E\left[\left(\underline{\phi_o + \phi_1 \hat{Y}_{n+1} + \phi_2 Y_n} - Y_{n+2}\right)^2\right] \\
&= E\left[\left(\phi_1(\hat{Y}_{n+1} - Y_{n+1}) - \epsilon_{n+2}\right)^2\right] \\
&= E\left[(\phi_1 \epsilon_{n+1} - \epsilon_{n+2})^2\right] = (\phi_1^2 + 1)\sigma_\epsilon^2 = 1.25 * \sigma_\epsilon^2 \\
\text{MSPE}(\hat{Y}_{n+3}) &= E\left[\left(\hat{Y}_{n+3} - Y_{n+3}\right)^2\right] \\
&= E\left[\left(\phi_o + \phi_1 \hat{Y}_{n+2} + \phi_2 \hat{Y}_{n+1} - \underline{Y_{n+2}}\right)^2\right] \\
&= E\left[\left(\phi_o + \phi_1(\hat{Y}_{n+2} - Y_{n+2}) + \phi_2(\hat{Y}_{n+1} - Y_{n+1}) - \epsilon_{n+3}\right)^2\right] \\
&= E\left[(\phi_1(\phi_1 \epsilon_{n+1} - \epsilon_{n+2}) + \phi_2 \epsilon_{n+1} - \epsilon_{n+3})^2\right] \\
&= \left[(\phi_1^2 + \phi_2)^2 + \phi_1^2 + 1\right] \sigma_\epsilon^2 = 1.3725 * \sigma_\epsilon^2
\end{aligned}
$$

7. **(a)** In this case,

$$\sigma_3^2 = 1 + .4(2^2) + .2(1^2) = 2.8$$

so that the conditional distribution of X_3 is $\mathcal{N}(0, 2.8)$.

(b) In this case utilizing the same argument as in **(a)**, the conditional distribution of X_3 is double exponential with the distribution being that of $\sqrt{2.8} \cdot \epsilon$ where $\epsilon \sim \text{DExp}(0, 1)$. Based on the results in class, this conditional distribution is double exponential with parameter

$$\lambda = \frac{1}{\sqrt{2.8}} = 0.598.$$

Thus conditional distribution of X_3 is DExp$(0, .598)$.

(c) For part **(a)**, X_n is Garch(1,1) process with same parameters and mean of 0.5, and now

$$\sigma_3^2 = 1 + .4 * (2 - \underline{0.5})^2 + .2 * (1 - 0.5)^2 = 1.95$$

So the conditional of X_3 is $\mathcal{N}(.5, 1.95)$. For part **(b)**, going through the same logic, conditional distribution of $(X_3 - 0.5)$ is same as $\sqrt{1.95} \cdot \epsilon$ where $\epsilon \sim \text{DExp}(0, 1)$. And again, this conditional distribution is double exponential with parameter

$$\lambda = \frac{1}{\sqrt{1.95}} = 0.716.$$

Thus conditional distribution of X_3 is DExp$(0.5, 0.716)$.

8. **(a)** Based on the top left plot of difference log-returns, it appears visually that there are contiguous time periods with larger variances (volatility) and contiguous time periods of lower variance (volatility) – this would suggests the GARCH model (with non-zero coefficients) is a better approximation for this time series than the white-noise model with constant variance.

   (b) The significance values of the a1 and b1 estimates suggests that these are highly significant, thus suggesting that the GARCH model is a better model than the random white-noise model. Of course, we have to take this somewhat cautiously since the normal distribution QQ plot is not that great and the Jarqe-Bera test rejects that the residuals follow a normal distribution. However, it does seem that the ACF of the squared difference log-returns are quite significantly correlated (based on the top-right plot), and the plot of the ACF of the squared residuals appears to have all of the auto-correlations for non-zero lags being not sigificant (based on bottom-right plot). This latter plot along with the Box-Ljung test suggests that the GARCH model provides more relatively uncorrelated residuals (with non-zero lags) (than a simple constant white noise model) and hence suggests that this GARCH model could be used, but with some caveats based on the non-normality of the residuals.

   (c) In this case the estimate of the variance, $\hat{\sigma}^2_{n+1}$ (refer to this as "estimated" since relies upon estimated parameters for the GARCH model), is

$$
\begin{aligned}
\hat{\sigma}^2_{n+1} &= (2.743e-06) + (8.848e-01)\hat{\sigma}^2_n + (8.888e-02)x^2_n \\
&= (2.743e-06) + (8.848e-01)\cdot(.0214^2) + (8.888e-02)(.0140^2) \\
&= .000425
\end{aligned}
$$

   so that

$$\hat{\sigma}_{n+1} = 0.0206$$

   In terms of estimating X_{n+1}, note that

$$X^2_{n+1} = \sigma^2_{n+1}\epsilon^2_{n+1}$$

   Note that $E(\epsilon^2_{n+1}) = 1$, so it is natural to estimate

$$\widehat{\left[X^2_{n+1}\right]} = \hat{\sigma}^2_{n+1} = (.0206)^2.$$

9. **(a)** This process is a MA(2) process with parameters $\theta_1 = \theta_2 = .5$. The mean is

$$E(X_n) = E(\epsilon_n) + .5 * E(\epsilon_{n-1}) + .5 * E(\epsilon_{n-2}) = 0$$

   and based on work in class, the autocovariance is given by

$$
\begin{aligned}
\gamma_X(n) &= \begin{cases} (1 + .5^2 + .5^2)\cdot\sigma^2 & n = 0 \\ (1\cdot .5 + .5\cdot .5)\cdot\sigma^2 & |n| = 1 \\ (1\cdot .5)\cdot\sigma^2 & |n| = 2 \\ 0 & |n| \geq 3 \end{cases} \\
&= \begin{cases} 1.5\cdot\sigma^2 & n = 0 \\ .75\cdot\sigma^2 & |n| = 1 \\ .5\cdot\sigma^2 & |n| = 2 \\ 0 & |n| \geq 3 \end{cases}
\end{aligned}
$$

(b) Now

$$\begin{aligned} Y_n &= X_n + X_{n-1} \\ &= \epsilon_n + .5\epsilon_{n-1} + .5\epsilon_{n-2} + \epsilon_{n-1} + .5\epsilon_{n-2} + .5\epsilon_{n-3} \\ &= \epsilon_n + 1.5 \cdot \epsilon_{n-1} + \epsilon_{n-2} + .5\epsilon_{n-3} \end{aligned}$$

so this process is ARMA(0,3), with parameters of $\theta_1 = 1.5,\ \theta_2 = 1, \theta_3 = .5$.

(c) Since the mean is 0 and $\underline{\mathrm{Cov}(X_n, X_{n+2}) = \gamma_X(2)}$, the best linear predictor is given by

$$\begin{aligned} \hat{X}_{n+2} &= \frac{\gamma_X(2)}{\gamma_X(0)} X_n \\ &= \frac{.5 \cdot \sigma^2}{1.5 \cdot \sigma^2} X_n = \frac{1}{3} X_n \end{aligned}$$

and the MSPE of this best linear predictor is given by

$$\mathrm{MSPE} = \gamma_X(0) - \frac{(\gamma_X(2))^2}{\gamma_X(0)} = 1.5 \cdot \sigma^2 - \frac{(.5 \cdot \sigma^2)^2}{1.5 \cdot \sigma^2} = \left(1.5 - \frac{.25}{1.5}\right) \sigma^2 = 1.33\sigma^2$$

10. **(a)** This is an AR(2)-GARCH)(1,1) model, so if X_n is the log-return process, then the model for this process is an AR(2) process of the form

$$X_n = \alpha_o + \alpha_1 X_{n-1} + \alpha_2 X_{n-2} + \epsilon_n$$

where $\alpha_o = \mu(1 - \alpha_1 - \alpha_2)$, and the error process ϵ_n is a GARCH(1,1) process of the form

$$\epsilon_n = \sigma_n \delta_n$$

where

$$\sigma_n^2 = \alpha_{g,o} + \beta_{g,1}\sigma_{n-1}^2 + \alpha_{g,1}\epsilon_{n-1}^2$$

and δ_n is a white Gaussian noise process, i.e., iid mean 0 with variance of 1. The estimate for α_o in the AR(2) part of the model is

$$\hat{\alpha}_o = \hat{\mu}(1 - \hat{\alpha}_1 - \hat{\alpha}_2) = .00069 * (1 - .029 + .081) = 0.000726$$

so the final model with the estimated coefficients is

$$X_n = 0.00061 + .029 \cdot X_{n-1} - .081 \cdot X_{n-2} + \epsilon_n$$

and the error process ϵ_n is a GARCH(1,1) process of the form

$$\epsilon_n = \sigma_n \delta_n$$

where

$$\sigma_n^2 = .000005 + .878 \cdot \sigma_{n-1}^2 + .090 \cdot \epsilon_{n-1}^2$$

(b) The conditional variance of the error process ϵ_n in the AR process is

$$.000005 + .878 \cdot \sigma_{n-1}^2 + .090 \cdot \epsilon_{n-1}^2$$

(c) Overall, pretty satisfied with this model – the p-values for all of the correlation tests on the residuals and the squared residuals are non-significant. The only significant p-value was the Shapiro-Wilk test, and also there was some asymmetry on QQ-plot relative to the tails (below the line on the upper tail and above line on the lower tail). So might try a combination of APARCH to see if that captures the asymmetry or possibly the t-distribution, though that is not that likely to be better. For selecting these models, would use a combination of these same diagnostics and also the Information Criteria (AIC and BIC) for comparing, i.e., see if there is a significant decrease in the information criteria when going to these models.

11. **(a)** Note that

$$\begin{aligned}
E(X) &= E(E(X|\sigma)) = E(0) = 0, \\
\mathrm{Var}(X) &= E(X^2) && \text{by above} \\
&= E(E(X^2|\sigma)) = E(\sigma^2), \\
E(X^4) &= E((X^4|\sigma)) = 3E(\sigma^4).
\end{aligned}$$

Thus we have that

$$\frac{E(X^4)}{[\mathrm{Var}(X)]^2} = \frac{3E(\sigma^4)}{[E(\sigma^2)]^2} = \frac{3\left\{\mathrm{Var}(\sigma^2) + [E(\sigma^2)]^2\right\}}{[E(\sigma^2)]^2} = 3\left[1 + \frac{\mathrm{Var}(\sigma^2)}{[E(\sigma^2)]^2}\right].$$

(b) Note that in class we have shown that

$$E(\sigma_n^2) = \frac{\alpha_o}{1 - \alpha_1 - \beta_1}.$$

Thus

$$\begin{aligned}
E(\sigma_n^4) &= E\left[\left(\alpha_o + \alpha_1 X_{n-1}^2 + \beta_1 \sigma_{n-1}^2\right)^2\right] \\
&= E\left(E\left[\left(\alpha_o + \alpha_1 X_{n-1}^2 + \beta_1 \sigma_{n-1}^2\right)^2 \ |\sigma_{n-1}^2\right]\right) \\
&= E\left[\left(\alpha_o + \alpha_1 \sigma_{n-1}^2 \epsilon_{n-1}^2 + \beta_1 \sigma_{n-1}^2\right)^2\right] \\
&= \alpha_o^2 + \alpha_1^2 E(\sigma_{n-1}^4 \epsilon_{n-1}^4) + \beta_1^2 E(\sigma_{n-1}^4) + 2\alpha_o E(\alpha_1 \sigma_{n-1}^2 \epsilon_{n-1}^2 + \beta_1 \sigma_{n-1}^2) + 2\alpha_1\beta_1 E(\sigma_{n-1}^4 \epsilon_{n-1}^2) \\
&= \alpha_o^2 + \alpha_1^2 E\left[E(\sigma_{n-1}^4 \epsilon_{n-1}^4 | \sigma_{n-1}^2)\right] + \beta_1^2 E(\sigma_{n-1}^4) + 2\alpha_o E(\alpha_1 \sigma_{n-1}^2 \epsilon_{n-1}^2 + \beta_1 \sigma_{n-1}^2) \\
&\qquad + 2\alpha_1\beta_1 E\left[E(\sigma_{n-1}^4 \epsilon_{n-1}^2 | \sigma_{n-1}^2)\right] \\
&= \alpha_o^2 + 3\alpha_1^2 E(\sigma_{n-1}^4) + \beta_1^2 E(\sigma_{n-1}^4) + 2\alpha_o(\alpha_1 + \beta_1) E(\sigma_{n-1}^2) + 2\alpha_1\beta_1 E(\sigma_{n-1}^4) \\
&= \alpha_o^2 + 2\alpha_o(\alpha_1 + \beta_1) E(\sigma_{n-1}^2) + (3\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1) E(\sigma_{n-1}^4)
\end{aligned}$$

Noting that

$$(3\alpha_1^2 + \beta_1^2 + 2\alpha_1\beta_1) = (\alpha_1 + \beta_1)^2 + 2\alpha_1^2,$$

recalling from lecture that

$$E(\sigma_{n-1}^2) = \frac{\alpha_o}{1 - \alpha_1 - \beta_1},$$

and utilizing that stationarity implies that $E(\sigma_n^2) = E(\sigma_{n-1}^2)$, we can solve for $E(\sigma_n^4)$ with the result that

$$E(\sigma_n^4) \quad = \quad \frac{\alpha_o^2 \left[\frac{1+\alpha_1+\beta_1}{1-\alpha_1-\beta_1}\right]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}$$

Thus we have

$$\begin{aligned}
\frac{E(X_n^4)}{[\mathrm{Var}(X_n)]^2} &= \frac{3E(\sigma_n^4)}{[E(\sigma_n^2)]^2} \\
&= \frac{3 \cdot \frac{\alpha_o^2 \left[\frac{1+\alpha_1+\beta_1}{1-\alpha_1-\beta_1}\right]}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}}{\frac{\alpha_o^2}{(1-\alpha_1-\beta)^2}} \\
&= \frac{3\left[1 - (\alpha_1 + \beta_1)^2\right]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}
\end{aligned}$$

(c) The implications are that this condition suggests that the kurtosis unbounded, i.e., the fourth moment is infinite. This can be true even under the conditions that the process is stationary with $(\alpha_1 + \beta_1) \leq 1$. As we have seen in other contexts, it is even more important to have a significant history (number of time samples) for doing estimation on these sorts of heavy tailed distributional processes. It is interesting that the additional term is α_1^2 (relative to $(\alpha_1 + \beta_1)^2$) for determining the kurtosis.

12. **(a)** For this, note that

$$\sigma_t = \sqrt{1 + .5 * a_{t-1}^2} = 1.086$$

as given in the problem, so that

$$E(Y_t | X_t = .01,\ a_{t-1}) = .06 + (.35) * (0.1) + .22 * \sigma_t = .334$$

(b)

$$\mathrm{Var}(Y_t | X_t) = \sigma_t^2 = (1.086)^2$$

(c) The conditional distribution of Y_t given X_t, a_{t-1} is normal, given that σ_t in the ARCH model is effectively known and the white noise process in the ARCH model is normal.

(d) The marginal distribution is virtually certain not to be normal, and that is due to the fact that random part of Y_t includes the product of the two independent random variables σ_t and normal random variable ϵ_t and this will only be normal if σ_t is a fixed constant, and that is not the case. In addition, Y_t includes X_t and $\delta\sigma_t$.

13. **(a)** This is an AR(1)-ARCH(1) process, and so the process Y_t is an AR(1) process in terms of its first and second order properties. Hence the mean is

$$\mu_Y = \frac{2}{1 - .67} = 6.06$$

(b) The ACF is given as

$$\rho_Y(k) = .67^{|k|}$$

(c) Since the ARCH(1) process is a white process as shown in class, $\rho_a(0) = 1$ and $\rho_a(k) = 0$ for all non-zero lags k.

(d) Since the square of an ARCH(1) process is an AR(1) process with the coefficient of .5, so that the autocorrelation is given as

$$\rho_{a^2}(k) = (.5)^{|k|}$$

14. We are given that the log-prices $Y_{1,t}, Y_{2,t}$ are log-prices satisfy the following model for log-returns,

$$\begin{aligned} \triangle Y_{1,t} \equiv Y_{1,t} - Y_{1,t-1} &= \phi_1(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{1,t} \\ \triangle Y_{2,t} \equiv Y_{2,t} - Y_{2,t-1} &= \phi_2(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{2,t} \end{aligned}$$

where $\epsilon_{1,t}, \epsilon_{2,t}$ are white noise processes.

(a) Note that by subtraction of the first equation by λ times the second (lefthand side from lefthand side and righthand side from righthand side), we have

$$Y_{1,t} - Y_{1,t-1} - \lambda(Y_{2,t} - Y_{2,t-1}) = \phi_1(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{1,t} - \lambda\left[\phi_2(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{2,t}\right]$$

or adding $(Y_{1,t-1} - \lambda Y_{2,t-1})$ to both sides, have

$$\begin{aligned} Y_{1,t} - \lambda Y_{2,t} &= Y_{1,t-1} - \lambda Y_{2,t-1} + \phi_1(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{1,t} - \lambda\left[\phi_2(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{2,t}\right] \\ &= (1 + \phi_1 - \lambda\phi_2)(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{1,t} - \lambda\epsilon_{2,t} \end{aligned}$$

So the above is an AR equation in the process $Y_{1,t} - \lambda Y_{2,t}$, with coefficient of $(1 + \phi_1 - \lambda\phi_2)$.

(b) Sufficient conditions are that

$$|1 + \phi_1 - \lambda\phi_2| < 1$$

(c) We suppose that there additive constants to righthand side of both of the original equations, i.e.,

$$\begin{aligned} \triangle Y_{1,t} \equiv Y_{1,t} - Y_{1,t-1} &= \mu_1 + \phi_1(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{1,t} \\ \triangle Y_{2,t} \equiv Y_{2,t} - Y_{2,t-1} &= \mu_2 + \phi_2(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{2,t} \end{aligned}$$

Carrying out the same subtraction as before, we have

$$Y_{1,t} - Y_{1,t-1} - \lambda(Y_{2,t} - Y_{2,t-1}) = \mu_1 - \lambda\mu_2 + \phi_1(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{1,t} - \lambda\left[\phi_2(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{2,t}\right]$$

or adding $(Y_{1,t-1} - \lambda Y_{2,t-1})$ to both sides, as above, we have

$$Y_{1,t} - \lambda Y_{2,t} \quad = \quad \mu_1 - \lambda\mu_2 + (1 + \phi_1 - \lambda\phi_2)\,(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{1,t} - \lambda\epsilon_{2,t}$$

Now is shown earlier in the Lecture on time series, the constant on the righthand side of the AR(1) model equation is related to the mean μ of the process $Y_{1,t} - \lambda Y_{2,t}$ as

$$\mu_1 - \lambda\mu_2 = [1 - (1 + \phi_1 - \lambda\phi_2)]\,\mu$$

or

$$\mu = \frac{\mu_1 - \lambda\mu_2}{1 - (1 + \phi_1 - \lambda\phi_2)} = \frac{\mu_1 - \lambda\mu_2}{-\phi_1 + \lambda\phi_2}$$