ORIE5550_HW4_Markdown

Luis Alonso Cendra Villalobos (lc2234)

2024-02-28

Question 2

Consider the time series data gnp from the R package astsa.

(a) Take a suitable preliminary transformation of the series, and produce its time plot; In the following parts, work with the transformed series.

```
gnp_data <- ts(gnp, start = c(1947, 1), end = c(2002, 3) , frequency = 4)
gnp_data = log(gnp_data)
plot(gnp_data, type="l", main="Box-Cox Log-Transform of the GNP data",
     ylab = "log(gnp)", xlab = "Year")
```


Box–Cox Log–Transform of the GNP data
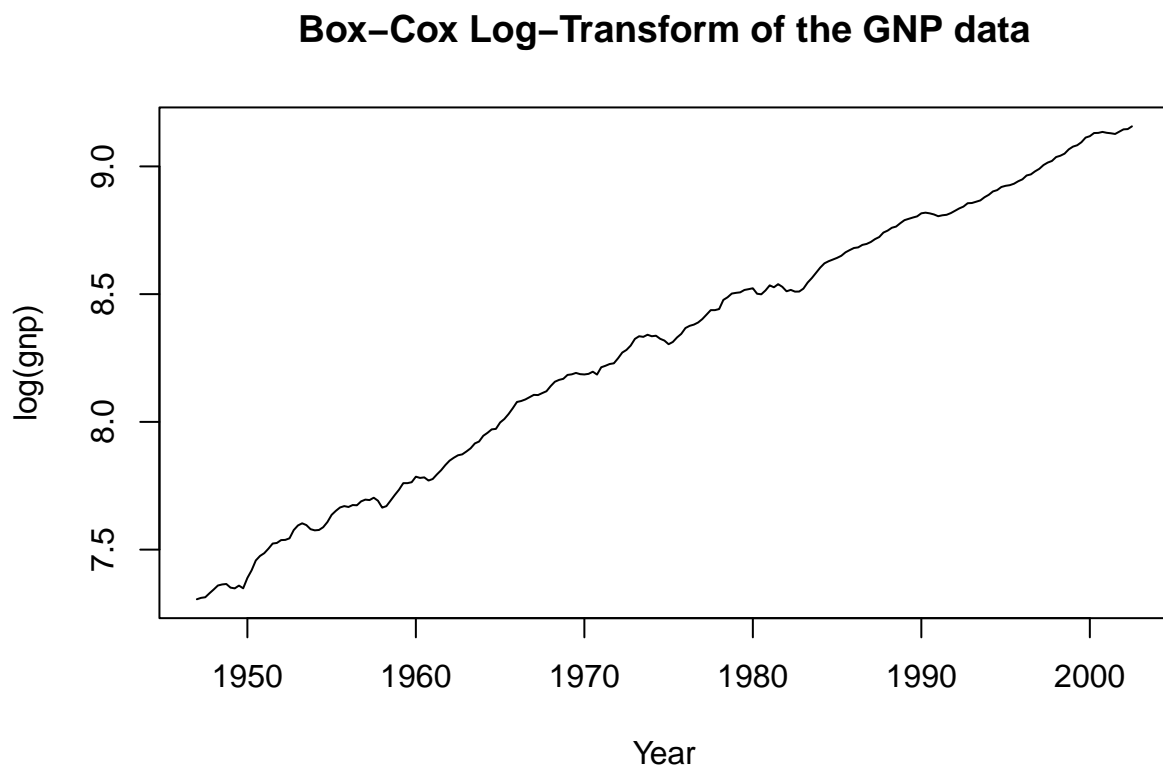
(b) Leave out the last 5 observations. Denote these samples as test data. The remaining observations are your training data. Fit a quadratic trend to the series using regression with the training data; Produce a time plot and a correlogram of the residuals obtained after removing the trend from the series

```r
# Leave out the last 5 observations as test data
training_data <- head(gnp_data, -5)
test_data <- tail(gnp_data, 5)

tt <- seq(1, length(training_data), by=1)
tt2 <- tt^2

fitModel <- lm(training_data ~ tt + tt2)
summary(fitModel)
```

```
##
## Call:
## lm(formula = training_data ~ tt + tt2)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.073915 -0.021840 -0.000729  0.024088  0.057608
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.289e+00  6.264e-03 1163.78   <2e-16 ***
## tt           1.025e-02  1.321e-04   77.63   <2e-16 ***
## tt2         -8.755e-06  5.840e-07  -14.99   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03054 on 215 degrees of freedom
## Multiple R-squared:  0.9967, Adjusted R-squared:  0.9966
## F-statistic: 3.225e+04 on 2 and 215 DF,  p-value: < 2.2e-16
```
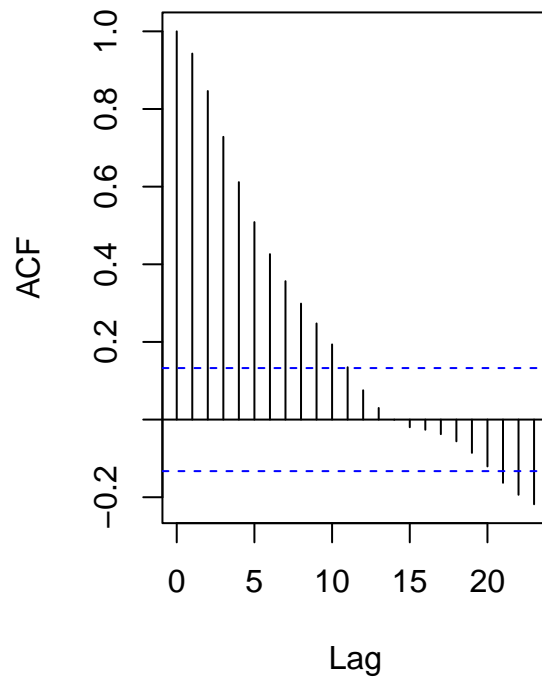
```r
residuals <- fitModel$residuals

par(mfrow = c(1, 2))
plot(residuals, type="l", main="Residuals of the fitted model")
acf(residuals, main="ACF of Residuals of the fitted model")
```

Residuals of the fitted model **ACF of Residuals of the fitted mod**
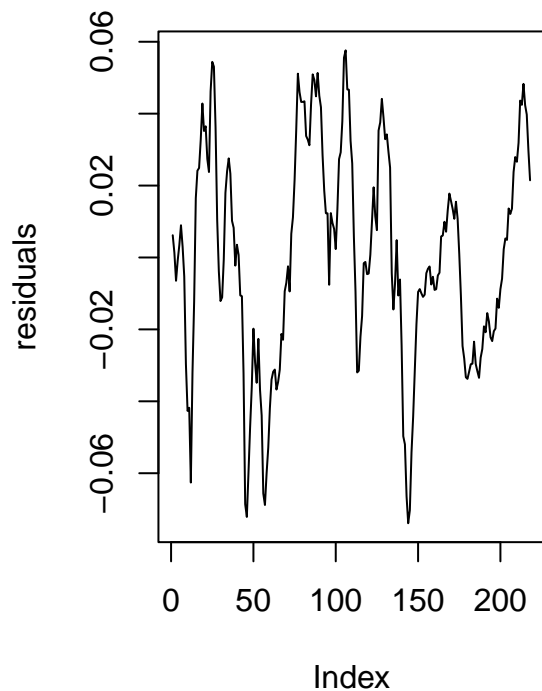
(c) Fit an ARMA(p,q) model to the residual of the regression with an order p,q determined by an information criterion; Include the output; Produce the sample ACF and PACF of the residuals of the ARMA(p,q) model; Check the assumptions of white noise and normality for the residuals

```r
auto.arima(residuals,max.p=10,max.q=10,ic="aic",allowmean = FALSE) # AIC
```

```
## Series: residuals
## ARIMA(1,0,3) with zero mean
##
## Coefficients:
##          ar1     ma1     ma2     ma3
##       0.8499  0.3880  0.3237  0.1553
## s.e.  0.0430  0.0744  0.0774  0.0733
##
## sigma^2 = 8.49e-05:  log likelihood = 713.16
## AIC=-1416.32   AICc=-1416.03   BIC=-1399.39
```

```r
auto.arima(residuals,max.p=10,max.q=10,ic="bic",allowmean = FALSE) # BIC
```

```
## Series: residuals
## ARIMA(2,0,0) with zero mean
##
```

3

```
## Coefficients:
##          ar1      ar2
##       1.3015  -0.3812
## s.e.  0.0623   0.0624
##
## sigma^2 = 8.638e-05:  log likelihood = 710.3
## AIC=-1414.61    AICc=-1414.5    BIC=-1404.46
```

```
arma.model <- arima(residuals, order=c(1,0,3), include.mean = FALSE, method = "ML")
par(mfrow = c(1, 2))
acf(arma.model$residuals, main="ACF of Residuals from the ARMA model")
pacf(arma.model$residuals, lag.max=20,
     ylim=c(-1, 1), main = "PACF of Residuals from the ARMA model") # PACF
```

ACF of Residuals from the ARMA mᵖACF of Residuals from the ARMA n
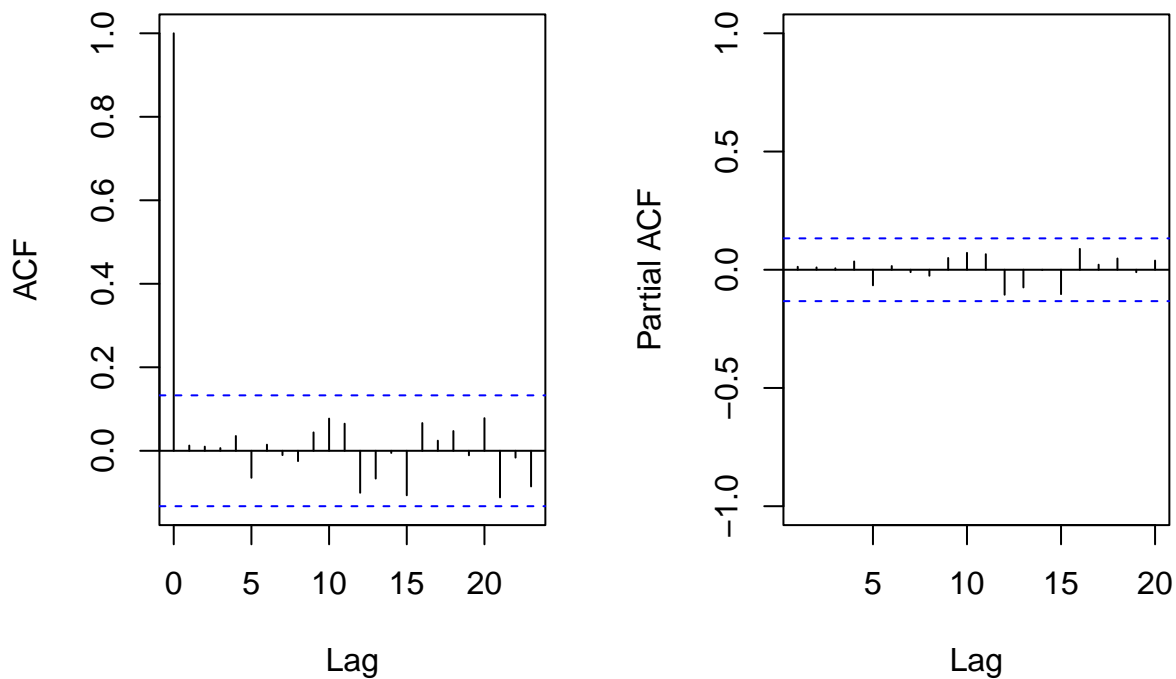

```
# QQ plot
qqnorm(arma.model$residuals, main = "QQ-plot of the regression residuals")
qqline(arma.model$residuals, col = "red", lwd = 2)

# Shapiro-Wilks test. Ho is normality
shapiro.test(arma.model$residuals)
```
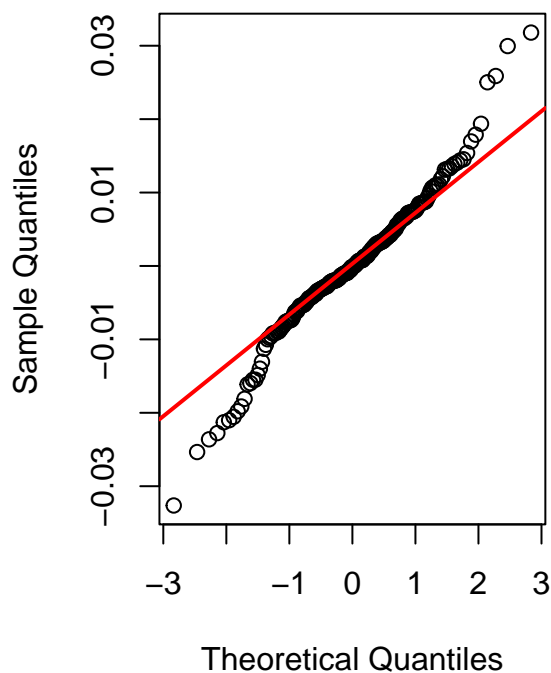
```
##
##  Shapiro-Wilk normality test
##
## data:  arma.model$residuals
## W = 0.97059, p-value = 0.0001647
```

```
#Box-Ljung test
Box.test(arma.model$residuals, lag=20, type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  arma.model$residuals
## X-squared = 13.588, df = 20, p-value = 0.8508
```

QQ–plot of the regression residua

We favor the ARIMA(1,0,3) due to its higher log-likelihood and lower information criteria.

From the Box-Ljung test, we cannot reject the no-autocorrelation among residuals up to lag 20, in accordance to the assumptions of white noise. The QQ-plot of the residuals exhibit heavy tails and we cannot reject the null hypothesis of normality for the Shapiro-Wilks test, thus, evidence suggests our residuals are not normally distributed; nevertheless, this is not a violation to white noise assumptions as there's no particular requirement that these are normally distributed.

(d) Forecast the transformed time series for 5 steps into the future; Compute the mean squared forecast error (MSFE) by using the test data; Provide the outputs.

```
arma.model <- arima(residuals, order=c(1,0,3), include.mean = FALSE, method = "ML")

h = 5
arma.forecast <- predict(arma.model, h)
round(arma.forecast$pred, 3)
```

```
## Time Series:
## Start = 219
## End = 223
## Frequency = 1
## [1] 0.016 0.012 0.010 0.008 0.007
```

```r
round(arma.forecast$se, 3)
```

```
## Time Series:
## Start = 219
## End = 223
## Frequency = 1
## [1] 0.009 0.015 0.019 0.023 0.025
```

```r
forecasted_residuals <- matrix(0, nrow = 1, ncol = 5)
forecasted_residuals[1,1] <- as.numeric(arma.forecast$pred)[1]
forecasted_residuals[1,2] <- as.numeric(arma.forecast$pred)[2]
forecasted_residuals[1,3] <- as.numeric(arma.forecast$pred)[3]
forecasted_residuals[1,4] <- as.numeric(arma.forecast$pred)[4]
forecasted_residuals[1,5] <- as.numeric(arma.forecast$pred)[5]

estimated_coeffs <- matrix(0, nrow= 1, ncol = 3)
estimated_coeffs[1,1] <- fitModel$coefficients[1]
estimated_coeffs[1,2] <- fitModel$coefficients[2]
estimated_coeffs[1,3] <- fitModel$coefficients[3]

time_matrix <- matrix(0, nrow = 5, ncol = 3)
time_matrix[, 1] <- 1
time_matrix[, 2] <- 219:223
time_matrix[, 3] <- time_matrix[, 2]^2
time_matrix <-t(time_matrix)

trend.forecast = estimated_coeffs %*% time_matrix
ts.forecast.pred <- trend.forecast + forecasted_residuals
ts.forecast.pred
```

```
##           [,1]     [,2]     [,3]     [,4]     [,5]
## [1,] 9.130672 9.133126 9.137347 9.142227 9.147312
```

```r
lower <- ts.forecast.pred - (qnorm(0.975) * forecasted_residuals)
upper <- ts.forecast.pred + (qnorm(0.975) * forecasted_residuals)

msfe <- mean( (ts.forecast.pred - as.numeric(test_data))^2 )
msfe
```

```
## [1] 3.667099e-05
```

(e) Now, fit an ARIMA(p,d,q) model to the training data; Forecast the transformed time series for 5 steps into the future by using this model; Compute the MSFE again; Which model is preferred in terms of minimal MSFE? Include the output

```r
auto.arima(training_data,max.p=10,max.q=10, max.d=2, ic="aic",allowmean = FALSE) # AIC
```

```
## Series: training_data
## ARIMA(1,1,0) with drift
##
## Coefficients:
##          ar1    drift
##       0.3473  0.0084
## s.e.  0.0636  0.0010
##
## sigma^2 = 9.275e-05:  log likelihood = 700.81
## AIC=-1395.63   AICc=-1395.51   BIC=-1385.49
```

```r
auto.arima(training_data,max.p=10,max.q=10, max.d=2, ic="bic",allowmean = FALSE) # BIC
```

```
## Series: training_data
## ARIMA(1,1,0) with drift
##
## Coefficients:
##          ar1    drift
##       0.3473  0.0084
## s.e.  0.0636  0.0010
##
## sigma^2 = 9.275e-05:  log likelihood = 700.81
## AIC=-1395.63   AICc=-1395.51   BIC=-1385.49
```

```r
arma.model.2e <- arima(training_data, order=c(1,1,0), include.mean = FALSE, method = "ML")

h = 5
arma.forecast.2e <- predict(arma.model.2e, h)
round(arma.forecast.2e$pred, 3)
```

```
##        Qtr1  Qtr2  Qtr3  Qtr4
## 2001               9.128 9.128
## 2002 9.127 9.127 9.127
```

```r
msfe <- mean( (arma.forecast.2e$pred - as.numeric(test_data))^2 )
msfe
```

```
## [1] 0.0003259941
```

We can see that the ARMA(p,q) model minimizes the MSFE better than ARMA(p,d,q), for a forecasting horizon of 5 periods. This can be an unexpected result, but we can re conciliate these results by recognizing the fact that the test data has a somewhat of a linear monotonically increasing trend: $\{9.12, 9.13, 9.14, 9.15, 9.16\}$. This behavior is well captured by a trend model, as opposed to data with fluctuations which we believe the ARIMA model should be better at capturing. Additionally, the test data sample is small with only 5 observations, thus, making the measurement of the models performance challenging.

Question 3

(a) Use set.seed(99) to generate the following random walk with drift: $X_t = -0.2t + 0.8\sum_{s=1}^{t} Z_s$ with $t = 1, ..., 100$ where Z_t is IID standard normal; Include a time plot of the series and sample ACF and PACF of the series.

```r
set.seed(99)
n_steps <- 100
Z_t <- rnorm(n_steps)
t <- seq_len(n_steps)
X_t <- -1 * 0.2 * t + 0.8 * cumsum(Z_t)

par(mfrow = c(2,2))
plot(t, X_t, type = 'l', col = 'blue', xlab = 'Time', ylab = 'X_t',
     main = 'Random Walk: X_t = -0.2*t + 0.8* sum(Z_t)')
acf(X_t, main="ACF of Residuals from the ARMA model")
pacf(X_t, lag.max=20, ylim=c(-1, 1), main = "PACF of Residuals from the ARMA model") # PACF
```
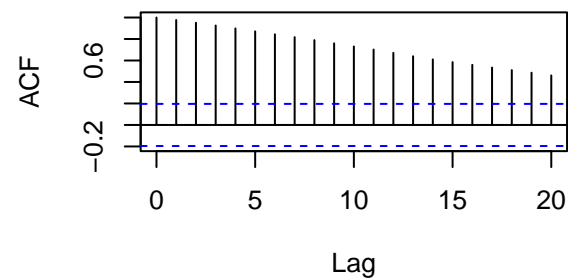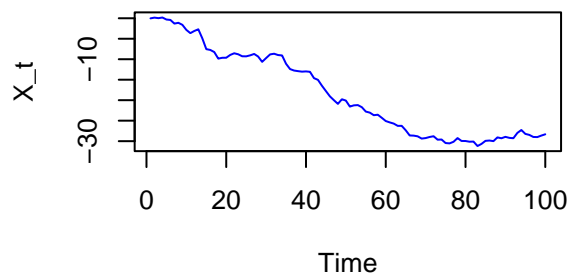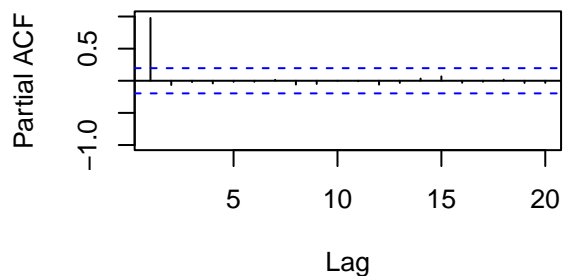

(b) Use the function auto.arima with suitable inputs to recover this model. That is, can you find an outcome of the model indicating the random walk with drift?

```
auto.arima(X_t,max.p=20,max.q=20, max.d=2, ic="aic", allowdrift=TRUE) # AIC
```

```
## Series: X_t
## ARIMA(0,1,1) with drift
##
## Coefficients:
##          ma1     drift
##       0.2996  -0.2830
## s.e.  0.1001   0.0905
##
## sigma^2 = 0.492:  log likelihood = -104.4
## AIC=214.8   AICc=215.06   BIC=222.59
```

```
auto.arima(X_t,max.p=20,max.q=20, max.d=2, ic="bic", allowdrift=TRUE) # BIC
```

```
## Series: X_t
## ARIMA(0,1,1) with drift
##
## Coefficients:
##          ma1     drift
##       0.2996  -0.2830
## s.e.  0.1001   0.0905
##
## sigma^2 = 0.492:  log likelihood = -104.4
## AIC=214.8   AICc=215.06   BIC=222.59
```

Thus, suggesting that the series is integrated of order 1 or random walk with a drift parameter. Given that we know the true model of the data, we can compare the estimations with the true parameters; for instance, the drift parameter is ≈ -0.28 close to the true value of -0.2 and the variance of the model is close to the true variance parameter of $(0.8)^2 \approx 0.64$.

(c) Go through the testing procedure for unit roots with significance level $\alpha = 0.05$ (for all steps). Indicate the conclusion at each step of the procedure; Check if the testing result corresponds to the used model

```
ur.gt <- ur.df(X_t, lags=0, type='trend')
summary(ur.gt)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt)
##
```

9

```
## Residuals:
##      Min      1Q   Median      3Q      Max
## -2.16040 -0.42329  0.03346  0.41717  1.42773
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.545465   0.146325  -3.728 0.000327 ***
## z.lag.1      0.001874   0.025610   0.073 0.941817
## tt           0.005879   0.009311   0.631 0.529242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7154 on 96 degrees of freedom
## Multiple R-squared:  0.04302,    Adjusted R-squared:  0.02308
## F-statistic: 2.158 on 2 and 96 DF,  p-value: 0.1212
##
##
## Value of test-statistic is: 0.0732 6.7048 2.1578
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

We cannot reject the null $\tau : \pi = 0$ at a significance level of 5%. Therefore, we proceed to the existence of a trend by testing if $\beta_2 = 0$ given $\pi = 0$ with the null $\phi_3 : (\pi, \widetilde{\beta_1}, \widetilde{\beta_2}) = (0, \widetilde{\beta_1}, 0)$. Similarly, we cannot reject this null at a significance level of 5%. Therefore, we take $\beta_2 = 0$ and proceed to fit another model without the trend term.

```
ur.gt <- ur.df(X_t, lags=0, type='drift')
summary(ur.gt)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1635 -0.4045  0.0486  0.4480  1.4270
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.536400   0.145167  -3.695 0.000364 ***
## z.lag.1     -0.013696   0.006899  -1.985 0.049940 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7132 on 97 degrees of freedom
## Multiple R-squared:  0.03904,    Adjusted R-squared:  0.02914
## F-statistic: 3.941 on 1 and 97 DF,  p-value: 0.04994
##
##
## Value of test-statistic is: -1.9853 9.9193
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

We cannot reject the null $\tau : \pi = 0$ at a significance level of 5%. Therefore, we proceed to the existence of a drift by testing if $\beta_1 = 0$ given $\pi = 0$ with the null $\phi_1 : (\pi, \widetilde{\beta_1}) = (0, 0)$. In this case, we reject this null at a significance level of 5% and conclude that a drift exists. We proceed to test whether $\pi = 0$, excluding the β's. By checking the t-statistic reported, we can state that π is significantly different then zero at a 5% confidence level, and conclude there is not a unit root.

Question 4

Raotbl3, description: This dataset contains the time series used by Darryl Holden and Roger Perman in their article: "Unit Roots and Cointegration for the Economist"

```r
data(Raotbl3)
attach(Raotbl3)
lc_TS <- ts(lc, start = c(1966, 4), end = c(1991, 2), frequency = 4)
plot(lc_TS, type="l", main="Real consumption exp., U.K., 1966-Q4 to 1991-Q2.",
     ylab = "Expenditure", xlab = "Quarter")
```

Real consumption exp., U.K., 1966–Q4 to 1991–Q2.
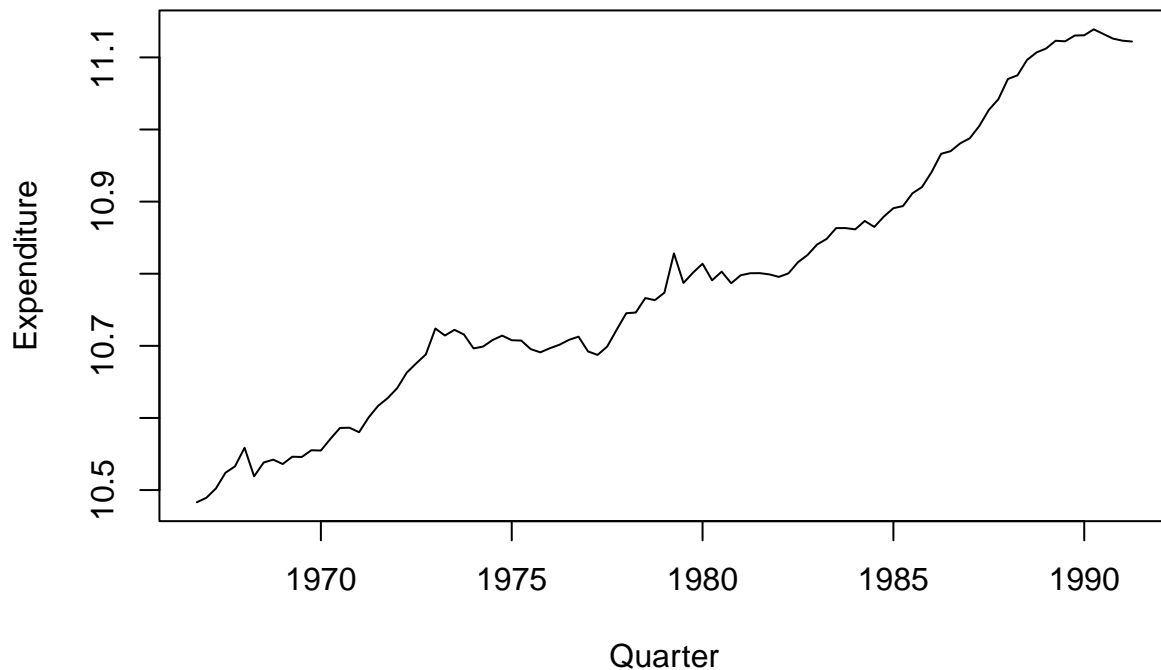
(a) Consider the time series Raotbl3$lc of real consumption expenditure from the United Kingdom starting in 1966:4 until 1991:2 in the R package urca. Produce a time plot of the series; Go through the testing procedure for unit roots discussed in class taking $k = 3$ for the number of lagged series differences to include in the regression; Indicate the conclusion at each step of the procedure.

```
ur.gt <- ur.df(lc_TS, lags=3, type='trend')
summary(ur.gt)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##       Min        1Q     Median        3Q       Max
## -0.044714 -0.006525  0.000129  0.006225  0.045353
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.7976591  0.3547775   2.248   0.0270 *
## z.lag.1     -0.0758706  0.0338880  -2.239   0.0277 *
## tt           0.0004915  0.0002159   2.277   0.0252 *
## z.diff.lag1 -0.1063957  0.1006744  -1.057   0.2934
## z.diff.lag2  0.2011373  0.1012373   1.987   0.0500 .
## z.diff.lag3  0.2998586  0.1020548   2.938   0.0042 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01307 on 89 degrees of freedom
## Multiple R-squared:  0.1472, Adjusted R-squared:  0.09924
## F-statistic: 3.071 on 5 and 89 DF,  p-value: 0.01325
##
##
## Value of test-statistic is: -2.2389 3.7382 2.5972
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

We cannot reject the null $\tau : \pi = 0$ at a significance level of 5%. Therefore, we proceed to the existence of a trend by testing if $\beta_2 = 0$ given $\pi = 0$ with the null $\phi_3 : (\pi, \widetilde{\beta_1}, \widetilde{\beta_2}) = (0, \widetilde{\beta_1}, 0)$. Similarly, we cannot reject this null at a significance level of 5%. Therefore, we take $\beta_2 = 0$ and proceed to fit another model without the trend term.

```
ur.gt <- ur.df(lc_TS, lags=3, type='drift')
summary(ur.gt)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.047547 -0.007071  0.000265  0.007731  0.046880
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0123237  0.0851358   0.145   0.8852
## z.lag.1     -0.0007356  0.0079043  -0.093   0.9261
## z.diff.lag1 -0.1433015  0.1016454  -1.410   0.1620
## z.diff.lag2  0.1615256  0.1020242   1.583   0.1169
```

```
## z.diff.lag3  0.2585280  0.1027364    2.516    0.0136 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01337 on 90 degrees of freedom
## Multiple R-squared:  0.09747,    Adjusted R-squared:  0.05735
## F-statistic:  2.43 on 4 and 90 DF,  p-value: 0.05335
##
##
## Value of test-statistic is: -0.0931 2.8806
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

We cannot reject the null $\tau : \pi = 0$ at a significance level of 5%. Therefore, we proceed to the existence of a drift by testing if $\beta_1 = 0$ given $\pi = 0$ with the null $\phi_1 : (\pi, \widetilde{\beta_1}) = (0, 0)$. Similarly, we cannot reject this null at a significance level of 5%. Therefore, we take $\beta_1 = 0$ and proceed to fit another model without trend nor drift terms.

```r
ur.gt <- ur.df(lc_TS, lags=3, type='none')
summary(ur.gt)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.047220 -0.007276  0.000229  0.007674  0.046921
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      0.0004083  0.0001695   2.409   0.0180 *
## z.diff.lag1 -0.1444994  0.1007615  -1.434   0.1550
## z.diff.lag2  0.1599782  0.1009153   1.585   0.1164
## z.diff.lag3  0.2568572  0.1015353   2.530   0.0131 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0133 on 91 degrees of freedom
## Multiple R-squared:  0.2546, Adjusted R-squared:  0.2218
## F-statistic:  7.77 on 4 and 91 DF,  p-value: 1.967e-05
##
##
```

```
## Value of test-statistic is: 2.4089
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

We can reject the null $\tau : \pi = 0$ at a significance level of 5%. Therefore, we conclude there is no unit root.

(b) Repeat the testing procedure with the series in (a) with a smaller lag, k = 2; Check if a different choice of lag affects the conclusion.

```
ur.gt <- ur.df(lc_TS, lags=2, type='trend')
summary(ur.gt)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.042654 -0.006960  0.000573  0.007325  0.048921
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.5900161  0.3634079   1.624    0.108
## z.lag.1     -0.0558308  0.0347023  -1.609    0.111
## tt           0.0003664  0.0002218   1.652    0.102
## z.diff.lag1 -0.0755700  0.1047334  -0.722    0.472
## z.diff.lag2  0.1555229  0.1044548   1.489    0.140
##
## Residual standard error: 0.01366 on 91 degrees of freedom
## Multiple R-squared:  0.05993,    Adjusted R-squared:  0.01861
## F-statistic:  1.45 on 4 and 91 DF,  p-value: 0.2239
##
##
## Value of test-statistic is: -1.6089 5.3138 1.3643
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -4.04 -3.45 -3.15
## phi2  6.50  4.88  4.16
## phi3  8.73  6.49  5.47
```

We cannot reject the null $\tau : \pi = 0$ at a significance level of 5%. Therefore, we proceed to the existence of a trend by testing if $\beta_2 = 0$ given $\pi = 0$ with the null $\phi_3 : (\pi, \widetilde{\beta_1}, \widetilde{\beta_2}) = (0, \widetilde{\beta_1}, 0)$. Similarly, we cannot reject

this null at a significance level of 5%. Therefore, we take $\beta_2 = 0$ and proceed to fit another model without the trend term.

```
ur.gt <- ur.df(lc_TS, lags=2, type='drift')
summary(ur.gt)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##       Min       1Q    Median       3Q       Max
## -0.044303 -0.007044  0.001045  0.007454  0.049709
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.412e-03  8.590e-02   0.075    0.941
## z.lag.1     -1.105e-05  7.969e-03  -0.001    0.999
## z.diff.lag1 -1.069e-01  1.040e-01  -1.029    0.306
## z.diff.lag2  1.299e-01  1.043e-01   1.246    0.216
##
## Residual standard error: 0.01379 on 92 degrees of freedom
## Multiple R-squared:  0.03174,    Adjusted R-squared:  0.0001708
## F-statistic: 1.005 on 3 and 92 DF,  p-value: 0.3941
##
##
## Value of test-statistic is: -0.0014 6.4846
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.51 -2.89 -2.58
## phi1  6.70  4.71  3.86
```

We cannot reject the null $\tau : \pi = 0$ at a significance level of 5%. Therefore, we proceed to the existence of a drift by testing if $\beta_1 = 0$ given $\pi = 0$ with the null $\phi_1 : (\pi, \widetilde{\beta_1}) = (0, 0)$. We reject this null at a significance level of 5% and conclude that a drift exists. We proceed to test whether $\pi = 0$, excluding the β's. By checking the t-statistic reported, we cannot reject π is significantly different from zero, and conclude $\pi = 0$ and there is a unit root.

Therefore, there is a change in conclusions from a different lag choice.

(c) Go through the testing procedure with the transformed gnp used in Problem 1. Here, use the entire samples. Use the lag p determined for the ARMA model. If your ARMA model contains the MA part, use selectlags, contained to ur.df function; Check if the conclusion of this problem corresponds to the preference of the model (i.e. trend stationary or non-stationary) in Problem 1. (e);

```
auto.arima(gnp_data,max.p=10,max.q=10,ic="aic",allowmean = FALSE) # AIC
```

```
## Series: gnp_data
## ARIMA(1,1,0) with drift
##
## Coefficients:
##          ar1    drift
##       0.3467   0.0083
## s.e.  0.0627   0.0010
##
## sigma^2 = 9.136e-05:  log likelihood = 718.61
## AIC=-1431.22   AICc=-1431.11   BIC=-1421.01
```

```
auto.arima(gnp_data,max.p=10,max.q=10,ic="bic",allowmean = FALSE) # BIC
```

```
## Series: gnp_data
## ARIMA(1,1,0) with drift
##
## Coefficients:
##          ar1    drift
##       0.3467   0.0083
## s.e.  0.0627   0.0010
##
## sigma^2 = 9.136e-05:  log likelihood = 718.61
## AIC=-1431.22   AICc=-1431.11   BIC=-1421.01
```

For this result, we see that $p = 1$ and according to the lectures material, we select $k = p - 1 = 0$ for the ADF test.

```
ur.gt <- ur.df(gnp_data, lags=0, type='trend')
summary(ur.gt)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt)
##
## Residuals:
##        Min          1Q     Median          3Q        Max
```

```
## -0.037285 -0.004751  0.000541  0.005980  0.028789
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.2157840  0.1141539   1.890   0.0600 .
## z.lag.1     -0.0279548  0.0155053  -1.803   0.0728 .
## tt           0.0002174  0.0001292   1.682   0.0939 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01009 on 219 degrees of freedom
## Multiple R-squared:  0.02325,    Adjusted R-squared:  0.01433
## F-statistic: 2.607 on 2 and 219 DF,  p-value: 0.07606
##
##
## Value of test-statistic is: -1.8029 52.303 2.6068
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -3.99 -3.43 -3.13
## phi2  6.22  4.75  4.07
## phi3  8.43  6.49  5.47
```

We cannot reject the null $\tau : \pi = 0$ at a significance level of 5%. Therefore, we proceed to the existence of a trend by testing if $\beta_2 = 0$ given $\pi = 0$ with the null $\phi_3 : (\pi, \widetilde{\beta_1}, \widetilde{\beta_2}) = (0, \widetilde{\beta_1}, 0)$. Similarly, we cannot reject this null at a significance level of 5%. Therefore, we take $\beta_2 = 0$ and proceed to fit another model without the trend term.

```r
ur.gt <- ur.df(gnp_data, lags=0, type='drift')
summary(ur.gt)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1)
##
## Residuals:
##       Min       1Q   Median       3Q      Max
## -0.036450 -0.005358  0.000236  0.005875  0.030495
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.024555   0.010570   2.323   0.0211 *
## z.lag.1     -0.001957   0.001273  -1.537   0.1256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.01013 on 220 degrees of freedom
## Multiple R-squared:  0.01063,    Adjusted R-squared:  0.006132
## F-statistic: 2.364 on 1 and 220 DF,  p-value: 0.1256
##
##
## Value of test-statistic is: -1.5374 76.4037
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.46 -2.88 -2.57
## phi1  6.52  4.63  3.81
```

We cannot reject the null $\tau : \pi = 0$ at a significance level of 5%. Therefore, we proceed to the existence of a drift by testing if $\beta_1 = 0$ given $\pi = 0$ with the null $\phi_1 : (\pi, \widetilde{\beta_1}) = (0, 0)$. We reject this null at a significance level of 5% and conclude that a drift exists. We proceed to test whether $\pi = 0$, excluding the β's. By checking the t-statistic reported, we cannot reject π is significantly different from zero, and conclude $\pi = 0$ and there is a unit root. Same conclusion as in 4.*b*)

Our conclusion for problem 2.*e.* is that gnp_data looks like an ARIMA(1,1,0) with a drift. Which means that the series is integrated or order(1), thus, not stationary.

Our conclusion from 4.*c.* is that the $\beta_1 \neq 0$ and $pi = 0$, thus, there exists a unit root and there exists a drift.

So, we conclude that the models agree that the data is not stationary and has a drift.