**Dr. Randall R. Rojas**
**Department of Economics**
**UCLA**

**Economics 144**
**Economic Forecasting**
**Spring, 2024**

# Midterm Exam
## May 9, 2024 by 10:45AM

For full credit on an analytical problem, you need to show all your work and the formula(s) used. For programming/analysis related problems, please make sure to include the respective code you wrote. You are not allowed to work with anyone on the exam. Your submitted midterm exam solutions must reflect your own work, otherwise it will be considered academic dishonesty. Please make sure to upload only one document in either PDF or HTML format that includes your R code.

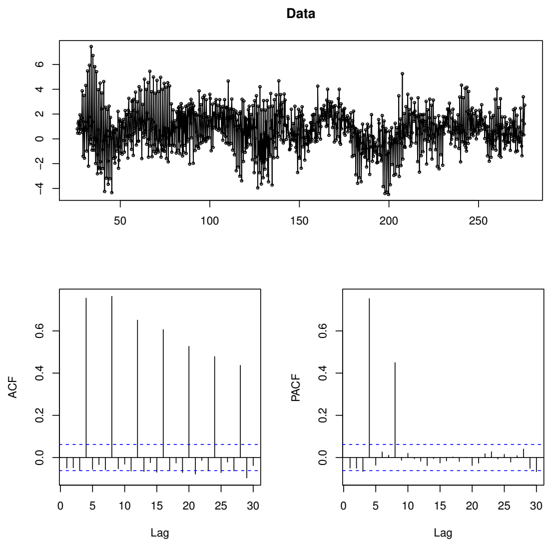
The exam time is from 9:30AM - 10:45AM. Please upload your exam solution through the upload link available on the course website. If there are any technical issues, you need to take a screen shot of your computer screen with the time stamp, and email me directly the screen shot along with your exam solution.

| | |
|---|---|
| **First Name** | |
| **Last Name** | |
| **UCLA ID #** | |

There are 7 questions in total.

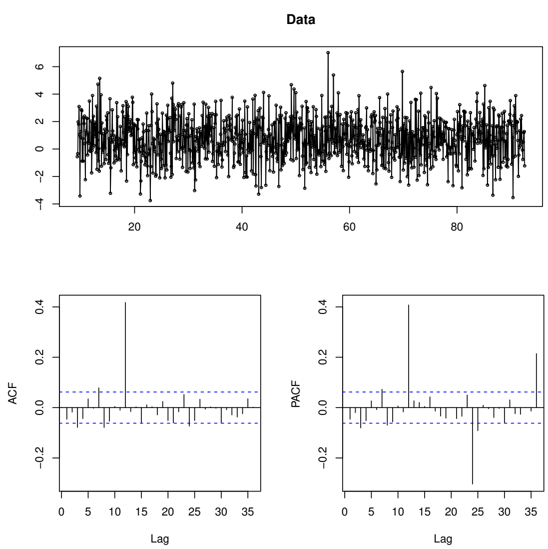
# Part I: Analytical/Conceptual Questions (50%)

1.(10%) (a) The `tsdisplay` plot below is based on quarterly observations. Which model fit would you recommend for these observations. Explain your answer in detail.



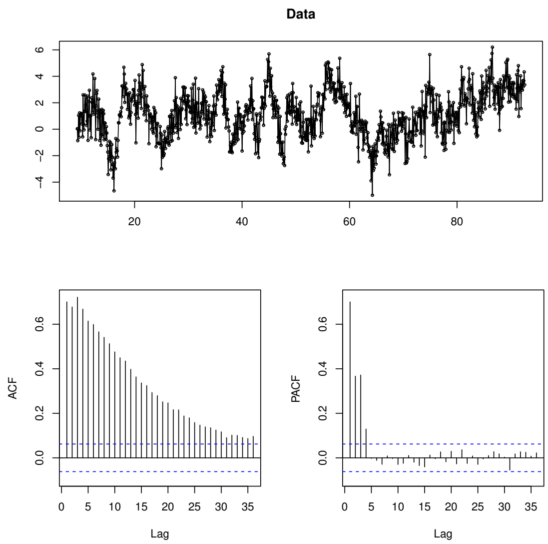

S-AR(2) with quarterly frequency. Spikes in the PACF at $k = 4, 8$, and decay in the ACF at quarterly spaced lags.

(b) The `tsdisplay` plot below is based on monthly observations. Which model fit would you recommend for these observations. Explain your answer in detail.



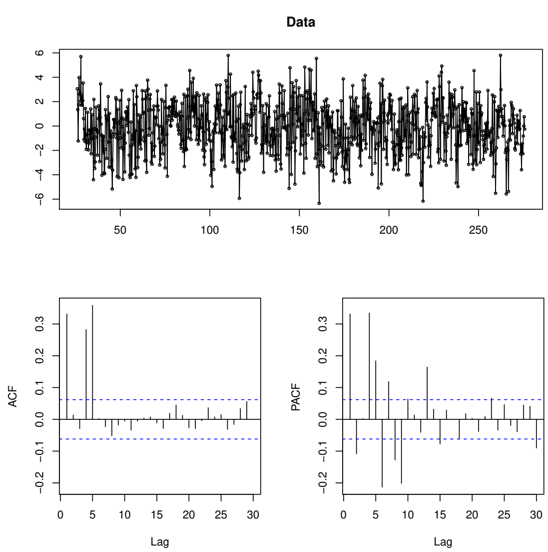

S-MA(1) with monthly frequency. Spike in the ACF at $k = 12$, and decay in the PACF at 12 month spaced lags.

(c) The `tsdisplay` plot below is based on monthly observations. Which model fit would you recommend for these observations. Explain your answer in detail.



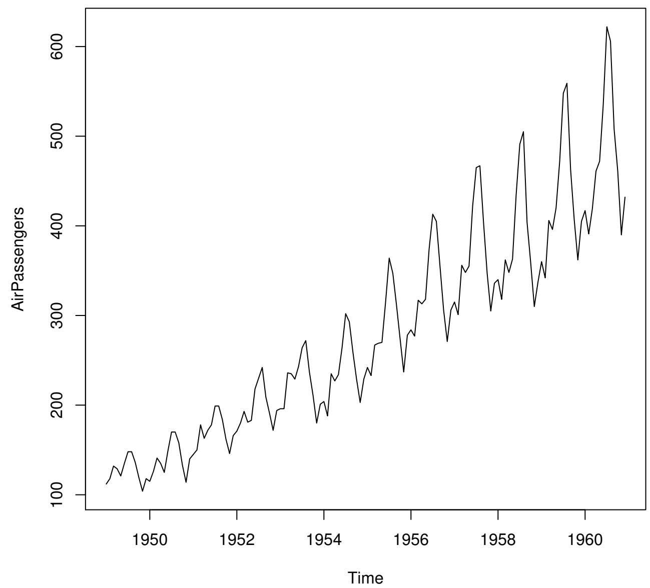

AR(4) with statistically significant lags $k = 1, 2, 3, 4$ given that they cross the Bartlett bands. The decay in the ACF confirms the AR dynamics.

(d) The `tsdisplay` plot below is based on quarterly observations. Which model fit would you recommend for these observations. Explain your answer in detail.




MA(5) with statistically significant lags $k = 1, 4, 5$ given that they cross the Bartlett bands. The decay in the PACF confirms the MA dynamics.

2.(10%) The plot below shows the number of international passengers per month on an airline in the United States for the period 1946-1960.



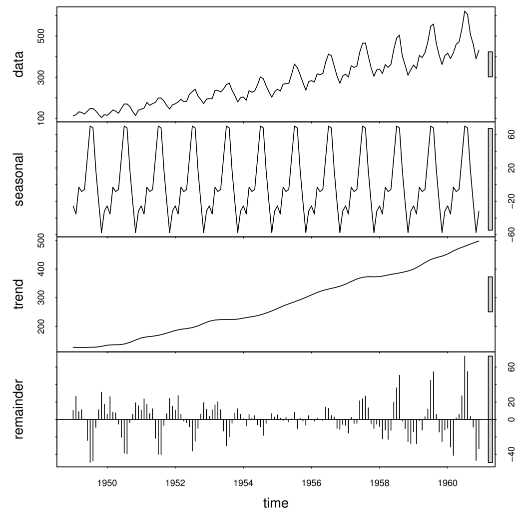

(a) If you had to choose between an AR(p) and an MA(q) model for this data set, which one would you choose and why?

I would choose an AR model given the strong persistence and lack of mean reversion

(b) Are the data covariance stationary? If so, explain why, if not, how would you transform it to a covariance stationary series?

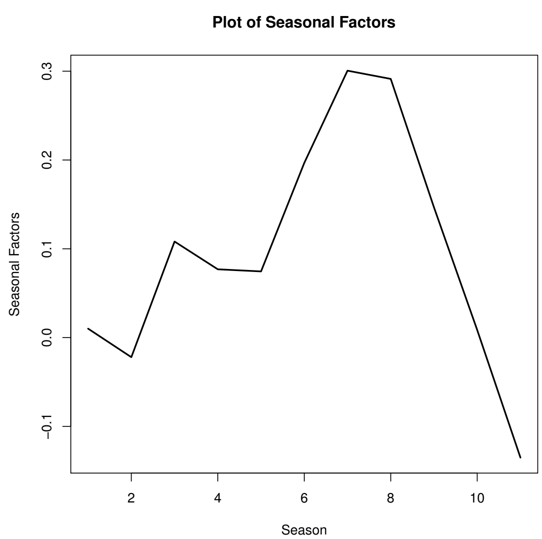
The series is not covariance stationary based on the same reasons as in part (a), and the fact that the variance is not constant (it seems to increase with time). I would consider taking the log to stabilize the volatility, and then taking the first difference to convert it to a mean reverting series with constant mean.

(c) Below we show the `stl` plot of the data. In terms of trend, seasonality, and cycles, briefly discuss which components you would include and why.



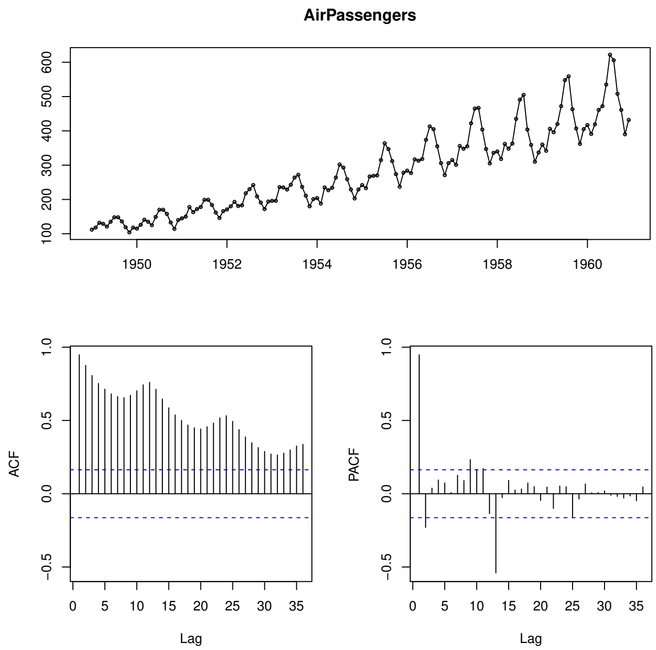

I would consider including all of them. Maybe a linear trend, monthly seasonal dummies and ARMA for the cycles. Notice that the remainder shows a strong pattern consistent with the presence of cycles.

(d) Comment on the seasonal factors plot below.



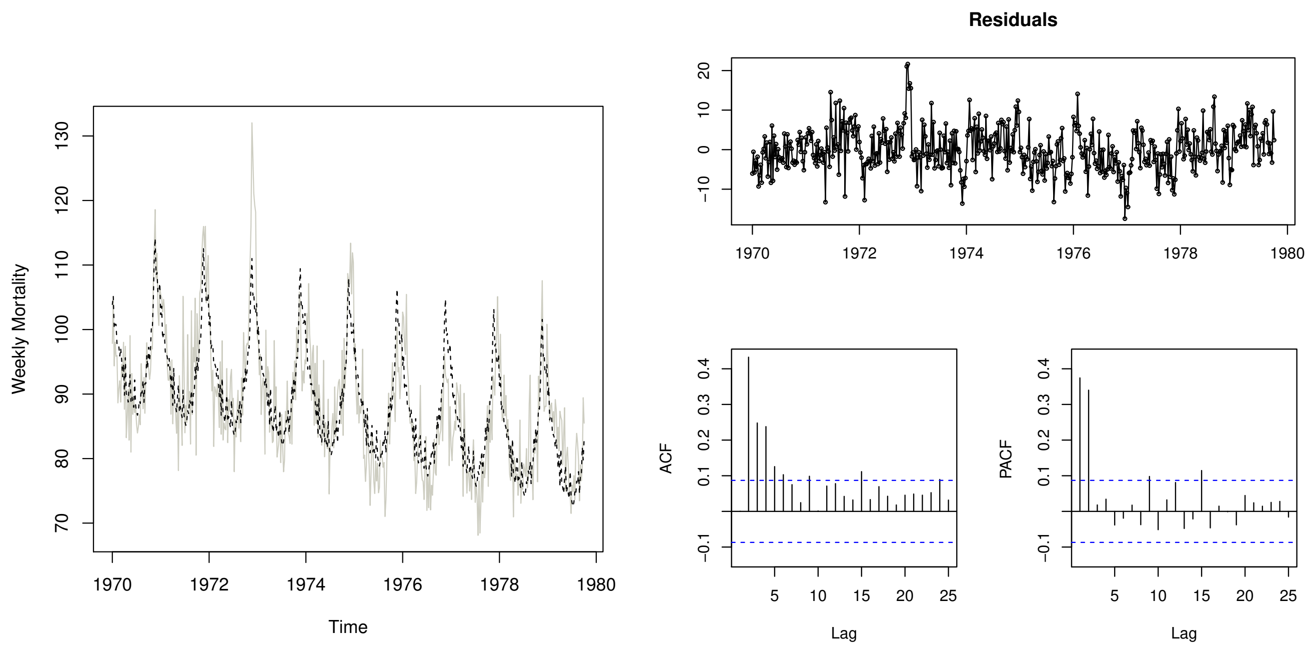

The plot suggests an increase the number of monthly international passengers from spring to summer, peaking around July, and then decreasing going into the fall season.

(e) Based on the `tsdisplay` plot below for the data, what model would you propose?




Based on the spikes in the PACF at lags $k = 1$ and $k = 2$, the decay in the ACF, and the strong spike at $k = 12$ I would suggest and AR(2) with possibly an S-AR(1) with monthly frequency.

3.(10%) For this problem the data consist of the average weekly cardiovascular mortality in Los Angeles County from 1970-1979 as shown in the left-figure below. We then fit a model with trend and seasonality components to the data (left-figure, the dashed line), and plot the respective residuals (right-figure), and the ACF and PACF of the residuals (two bottom-right plots).




(a) Based on the plots above, how would you improve the model fit?

I would consider either an ARMA or ARIMA model for characterizing the serial correlations (cycles) observed in the ACF and PACF plots of the residuals.

(b) Does the original data series seem to be covariance stationary? If so, how can you tell? If not, why not and what would you recommend to make it covariance stationary.

The downward trend in the series suggests that it might not be covariance stationary. This is also observed in the ACF and PACF plots. Therefore, I would try taking the first difference of the original series.

(c) Does the residuals plot show any serial correlation? Explain your answer in detail.

Yes, the spikes in both the ACF and PACF plots seem to suggest strong serial correlation.

(d) Suppose you use `auto.arima` to fit a model to the original data series, and you get the following output:

| ARIMA(2,1,1)(2,0,2)[52] | | | | | | | |
|---|---|---|---|---|---|---|---|
| Coefficients: | | | | | | | |
| | ar1 | ar2 | ma1 | sar1 | sar2 | sma1 | sma2 |
| | 0.3285 | 0.2939 | -0.9822 | 0.2424 | 0.4687 | -0.2648 | -0.2758 |
| s.e. | 0.0480 | 0.0422 | 0.0119 | 0.0632 | 0.0628 | 0.0753 | 0.0882 |

Provide a detailed interpretation of the ARIMA model suggested by R. Make sure to comment on any trend, seasonality, and/or cycles if appropriate.

Model: ARIMA (2,1,1) + Seasonal-ARIMA(2,0,2)
Since it has an I(1) component, this confirms that the data are not covariance stationary.

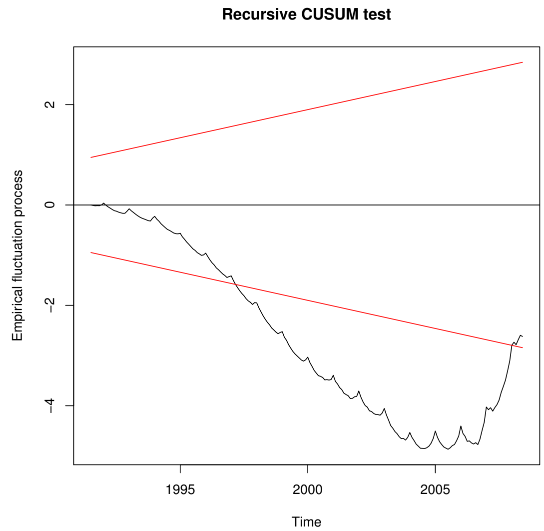
The fact that `auto.arima` estimated an additional seasonal component, suggest that not all the seasonality had been correctly adjusted for in the original model.

The trend is taken care of by the I(1) component of ARIMA.

The seasonality is taken care of by the S-ARMA(2,0,2) component.

The cycles are taken care of by the ARMA(2,2) component.

4.(10%) (a) Based on the CUSUM plot below, what can you conclude about the model used to fit the data?



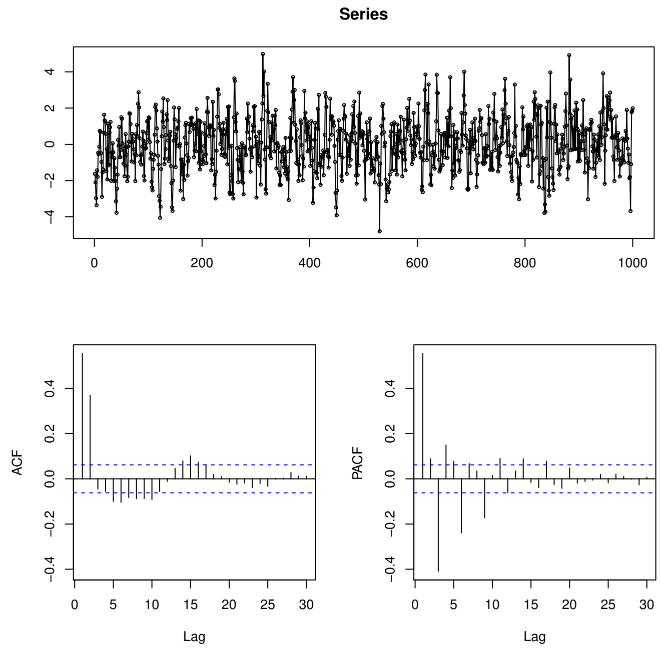

The model shows a breaking point around 1997-1998 and therefore, we would need to revise it since it would not be appropriate for the data

(b) From the list of four time-series processes below, which ones are expected to exhibit a deterministic cycle?

I. $Y_t = 20.5 \sin(t-1) + \varepsilon_t$
II. $Y_t = 100 Y_{t-10} + \varepsilon_t$
III. $Y_t = 0.25 Y_{t-1} + 0.25 Y_{t-2} + \varepsilon_t$
IV. $Y_t = t + 0.01 + \varepsilon_t$

Only I and IV exhibit a deterministic cycle

(c) What model would you fit to the series based on the figure below?




Based on the high degree of mean reversion, and the ACF and PACF plots, I would fit an MA(2).

5.(10%) For this problem we will look at monthly milk production in pounds per cow from January 1962 to December 1975. A 'linear trend + seasonality' model is fit to the data (the $y-$intercept is not included). The table below shows the summary statistics from the fit. In the figure below we show the relevant diagnostic plots.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| trend | 1.7278 | 0.0257 | 67.11 | 0.0000 |
| season1 | 590.5782 | 4.7698 | 123.82 | 0.0000 |
| season2 | 551.0647 | 4.7808 | 115.27 | 0.0000 |
| season3 | 643.5512 | 4.7920 | 134.30 | 0.0000 |
| season4 | 658.3949 | 4.8032 | 137.07 | 0.0000 |
| season5 | 719.3814 | 4.8146 | 149.42 | 0.0000 |
| season6 | 691.0108 | 4.8261 | 143.18 | 0.0000 |
| season7 | 641.2116 | 4.8377 | 132.55 | 0.0000 |
| season8 | 598.9124 | 4.8494 | 123.50 | 0.0000 |
| season9 | 556.3275 | 4.8612 | 114.44 | 0.0000 |
| season10 | 559.8140 | 4.8731 | 114.88 | 0.0000 |
| season11 | 528.8006 | 4.8851 | 108.25 | 0.0000 |
| season12 | 565.5014 | 4.8972 | 115.47 | 0.0000 |

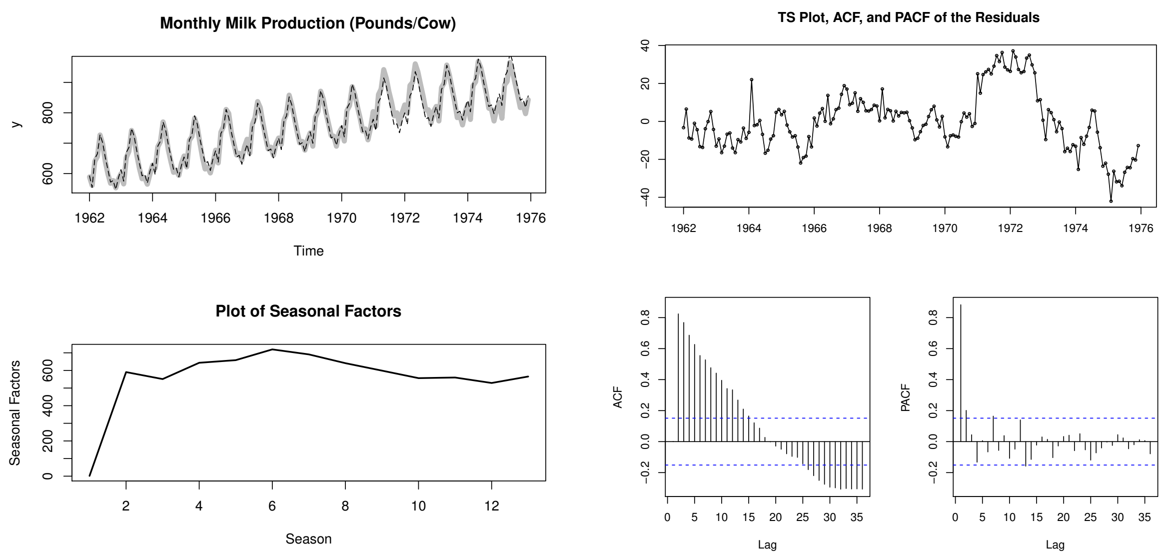



Fig. 1.— Data (top-left -gray line), fit (top-left -dashed line), seasonal factors (bottom-left), the residuals (top-right) and the ACF and PACF of the residuals (bottom-right).

(a) Discuss the results from the summary statistics table.
Based on the p-values, it appears that both the linear trend (excluding the y-intercept) and seasonal-monthly variables are all all statistically significant. In addition, as observed in figure above, the model seems like an excellent fir to the data.

(b) Interpret the seasonal factors plot.
From February to June/July there seems to be gradual increase in monthly milk production, followed by a decrease from July to December.

(c) After filtering the trend and seasonal components, do you believe that there might be cycles unaccounted for in your model? Explain your answer.
Yes. This is confirmed by looking the PACF of the residuals. According to this plot, an AR(2) model would seem like an appropriate model for characterizing the cycles unaccounted for in the original data.

(d) Is the series (original data) covariance stationary? Explain your answer.
No. It appears that $\mu_Y$ does depend on time as observed in the upward drift of the original series, i.e., there is no evidence of 'reversion to the mean'.

(e) Based on the ACF and PACF pots, would you recommend improving the model? If so, how? If not, why not?
Yes. In addition to the trend with seasonality model, I would also include an AR(2) component to take care of the cycles.

# Part II: Data Analysis/Programming Questions (50%)

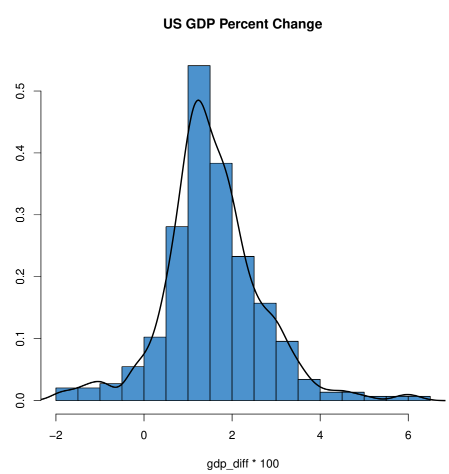
6.(25%) For this problem we will work with the US GDP percent change. To load the data, follow the two commands below:

```
library(Quandl)
gdp_diff <- Quandl("FRED/GDP", type="ts", transform="rdiff")
```

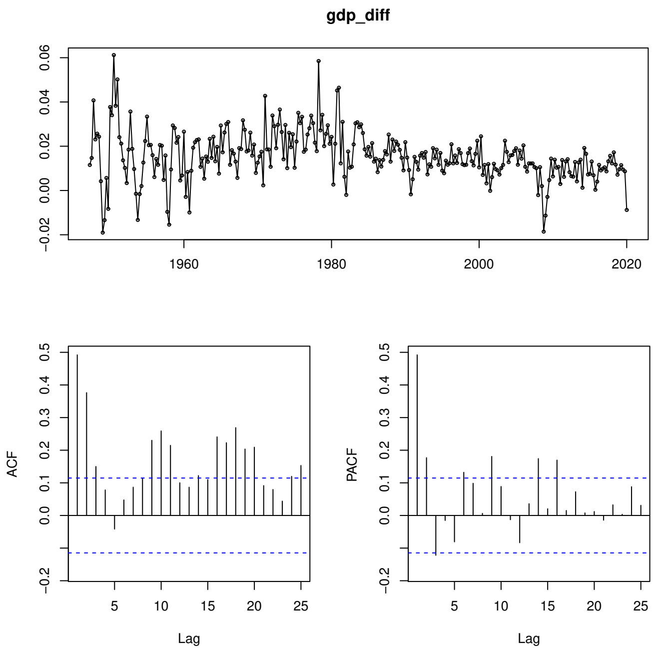
(a) Provide a descriptive analysis of your data. This includes the start and end dates, frequency, and relevant statistical summaries.

```
library(Quandl)
gdp_diff <- Quandl("FRED/GDP", type="ts", transform="rdiff")
head(gdp_diff)
           Qtr1       Qtr2       Qtr3       Qtr4
1947            0.01153131 0.01470516 0.04070757
1948 0.02308803 0.02568281 0.02432063
tail(gdp_diff)
             Qtr1         Qtr2         Qtr3         Qtr4
2018                                           0.007135121
2019  0.009619336 0.011443290 0.009478466 0.008661189
2020 -0.008798514
library(psych)
 describe(gdp_diff)
   vars   n mean   sd median trimmed mad  min  max range skew kurtosis se
X1    1 292 0.02 0.01  0.01    0.02 0.01 -0.02 0.06 0.08 0.34   2.15  0
truehist(gdp_diff*100,col='steelblue3',main="US GDP Percent Change")
lines(density(gdp_diff*100),lwd=2)
```

The data are quarterly observations from Q2/1947 to Q2/2020. The summary statistics are provided above. Below is the respective distribution.



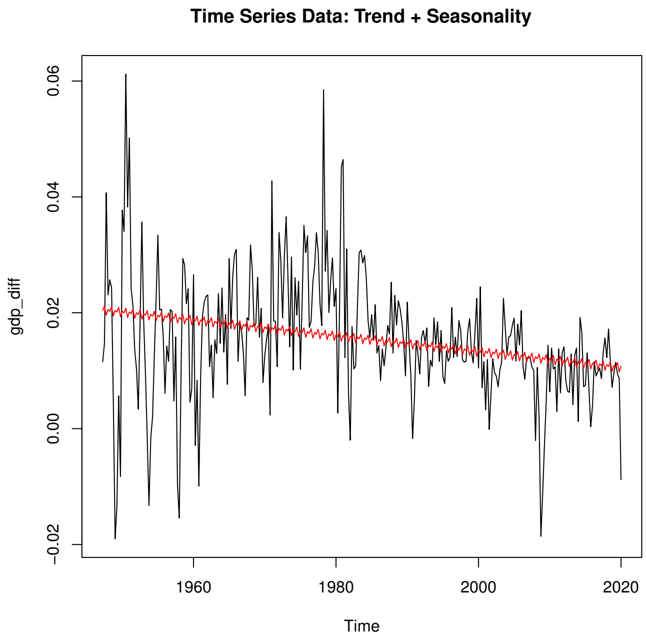
(b) Show the respective ACF and PACF plots and discuss what they suggest about the data.

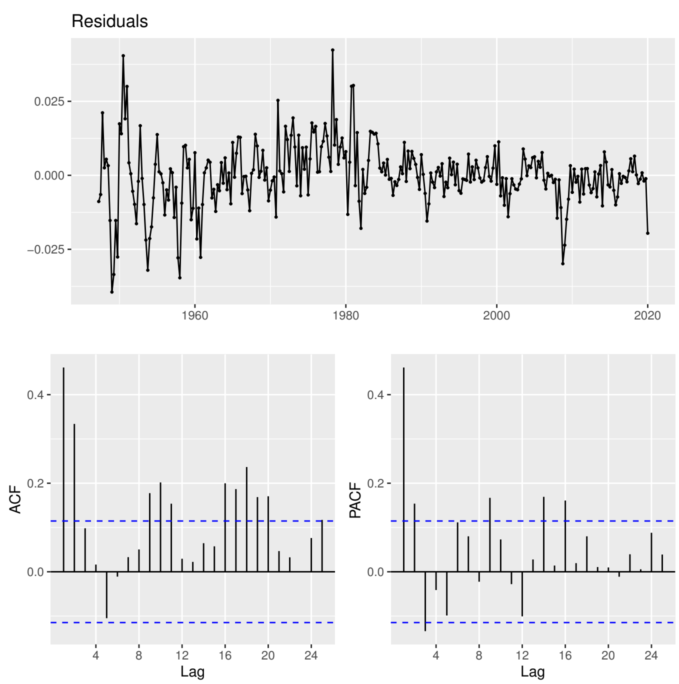



The strong spikes at lower lags in the PACF and ACF suggests possibly a low order ARMA, where the AR term would dominate, and possibly a seasonal AR as well given the strong spikes in the PACF at longer lags, and decay in the ACF.

(c) Fit a linear trend with seasonal (based on the frequency) dummy variables, and comment on the fit. Make sure to overlay the fit on the series.

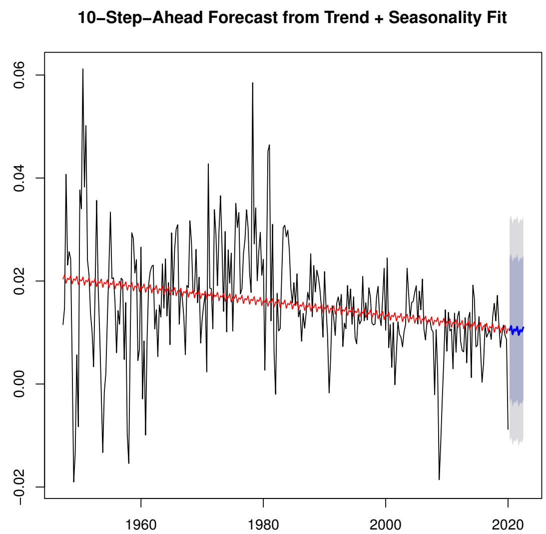
```
quartz()
par(mfrow=c(3,1))
fit=tslm(gdp_diff ~ trend + season)
plot(gdp_diff, main="Time Series Data: Trend + Seasonality")
lines(fit$fitted.values, col="red")
acf(fit$residuals)
pacf(fit$residuals)
```



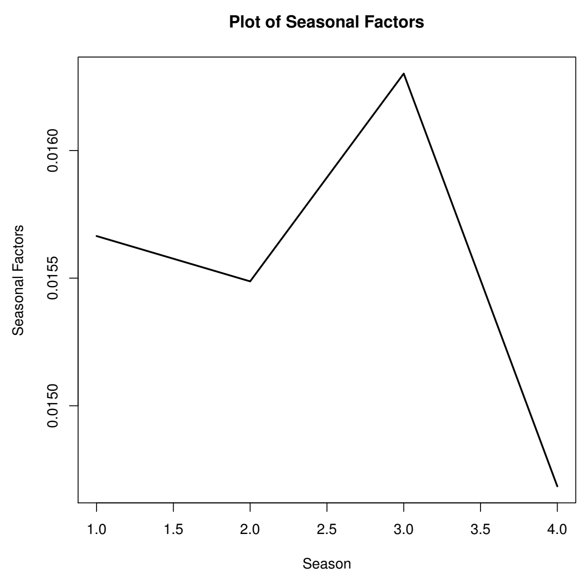




Visually, the fit does not look good at all. However, given that the seasonality does not appear constant, the seasonal dummies alone may not be enough to model correctly the seasonal variation. Based on the ACF and PACF, we can see that there are indeed seasonal and cyclical dynamics.

(d) Compute and plot a 10 period ahead forecast using your model from (c).




(e) Plot the seasonal factors from your model in (c) and comment on the plot.

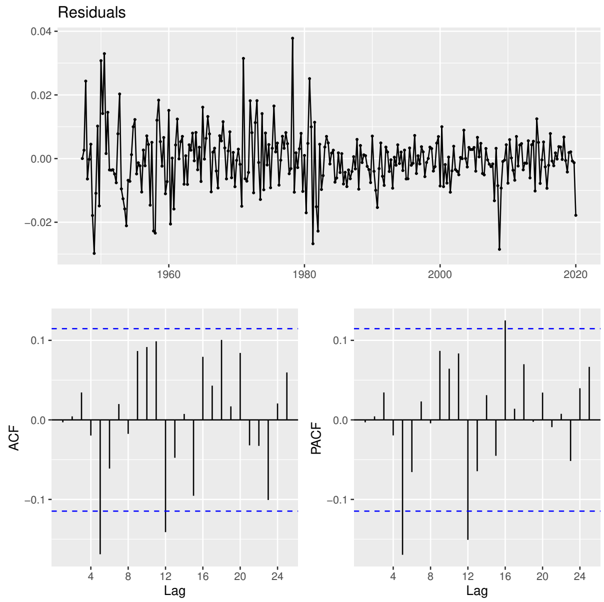


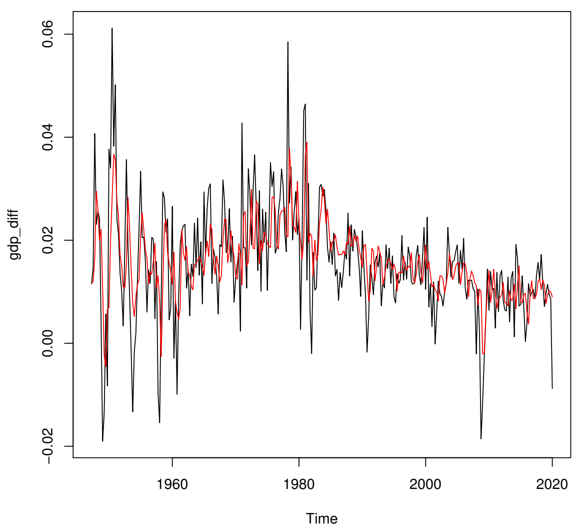

From the plot there appears to be an increase in US GDP change around Q3 followed by a strong decrease afterwards. However, based on the magnitude of the seasonal factors, I would suggest that they are not statistically significant. I Below is as summary of the model fit which confirms this as seen in the large p-values for the seasonal factors.

```
summary(fit)

Call:
tslm(formula = gdp_diff ~ season + trend)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.039442 -0.004886 -0.000059 0.005191 0.042338 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2.070e-02 1.682e-03 12.304 < 2e-16 ***
season2    -2.796e-04 1.788e-03 -0.156   0.876    
season3     5.687e-04 1.788e-03  0.318   0.751    
season4    -1.015e-03 1.788e-03 -0.568   0.571    
trend      -3.399e-05 7.499e-06 -4.533 8.54e-06 ***
---
Signif. codes: 0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 0.0108 on 287 degrees of freedom
Multiple R-squared: 0.06933, Adjusted R-squared: 0.05636 
F-statistic: 5.345 on 4 and 287 DF, p-value: 0.0003641
```

(f) Fit an ARMA + S-ARMA to the data and comment on the fit.

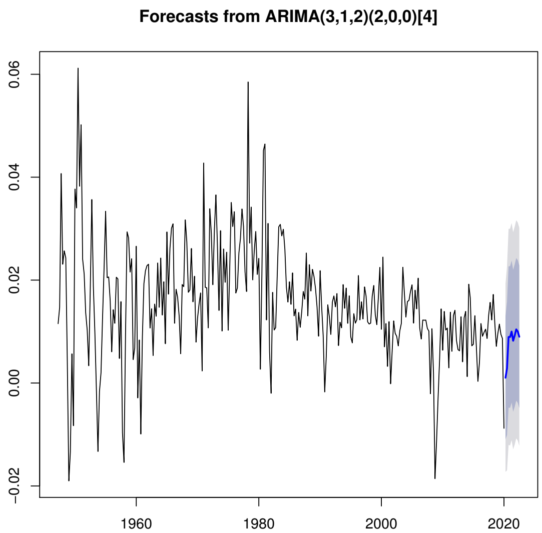
```
summary(auto.arima(gdp_diff))
Series: gdp_diff 
ARIMA(3,1,2)(2,0,0)[4] 

Coefficients:
          ar1    ar2      ar3      ma1      ma2     sar1     sar2
      -0.2763 0.4000 -0.0683 -0.3039 -0.6017 -0.1298 -0.1484
s.e.   0.1827 0.0788  0.0836  0.1742  0.1637  0.0670  0.0680

sigma^2 estimated as 8.701e-05: log likelihood=950
AIC=-1884   AIC=-1883.49  BIC=-1854.61

Training set error measures:
                      ME        RMSE         MAE     MPE     MAPE      MASE
                          ACF1
Training set -0.0003188589 0.009199076 0.006688316 16.2424 105.6547 0.6394436
    -0.003038839
```




Based on the training set errors, the model does not seems to be doing a good job at all, despite the overall fit (red curve) looking like a good one. The `tsdisplay` plot of the residuals shows that there are dynamics that have not been accounted for by the model.

(g) Compute and plot a 10 period ahead forecast using your model from (f).



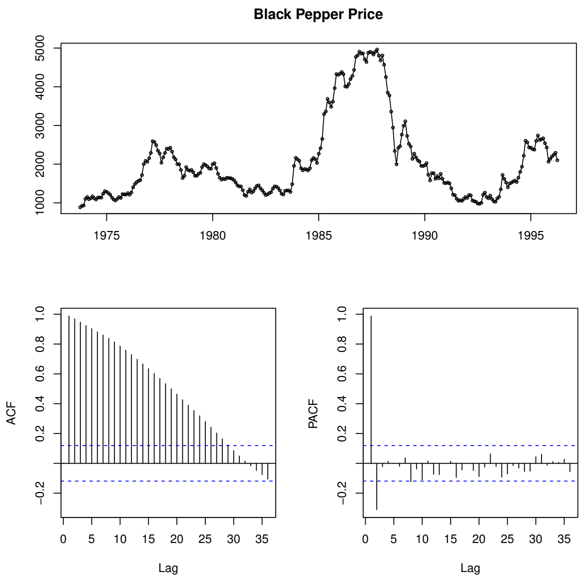

(h) Which model is better, (c) or (f), why?

The model from (f) seems like the best of the two models based on the overall fit, and ACF and PACF of the residuals.

7.(25%) For this problem we will work with the monthly spot price for black pepper from 10/1973 to 04/1996. To load and plot the data, follow the three commands below:

```
library(AER)
data("PepperPrice")
plot(PepperPrice[,1])
```

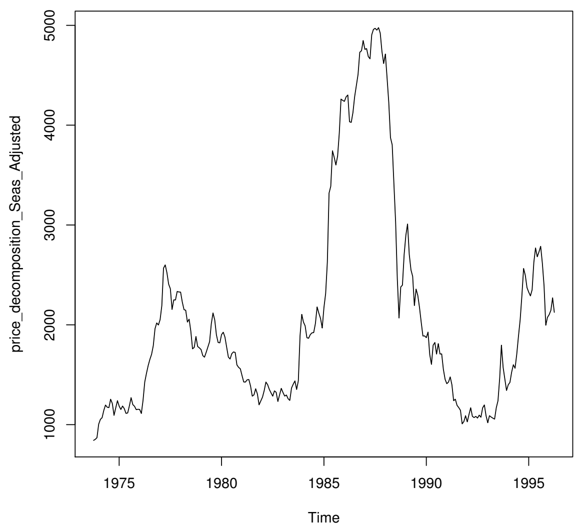
(a) Show and discuss (in terms of an MA and AR process) the respective `tsdisplay` plot.




Based on the spikes in the PACF and decay in the ACF, I would suggest modeling the series as an AR(2).

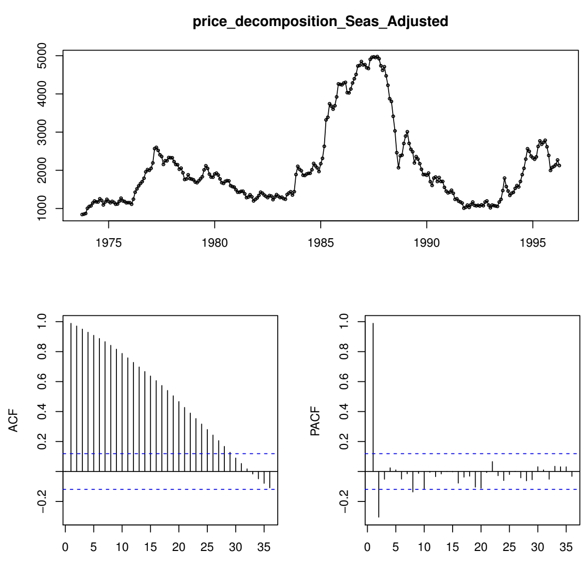
(b) Show a plot of your data after performing an additive seasonal adjustment to it. Discuss your results.

```
price_decomposition = decompose(PepperPrice[,1])
price_decomposition_Seas_Adjusted = PepperPrice[,1]-
    price_decomposition$seasonal
plot(price_decomposition_Seas_Adjusted)
```



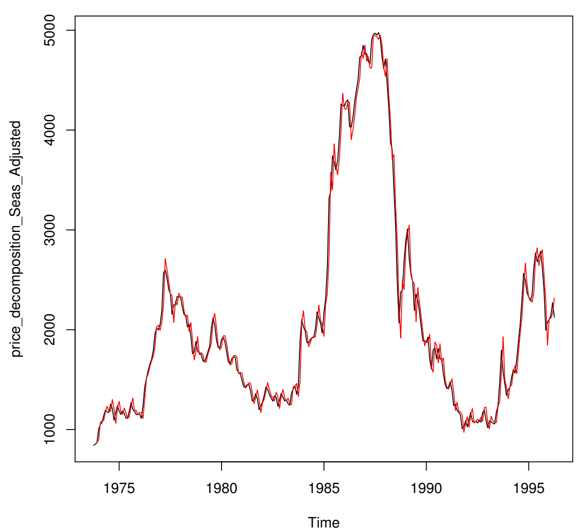

Based on the before and after seasonal adjustment, since the series does not appear to have changed much, I would infer that there is not much of a seasonal effect on the spot price of black pepper.

(c) For the seasonally adjusted series, using the tools discussed in class, identify and fit an appropriate model to the cycles. Discuss the overall fit.

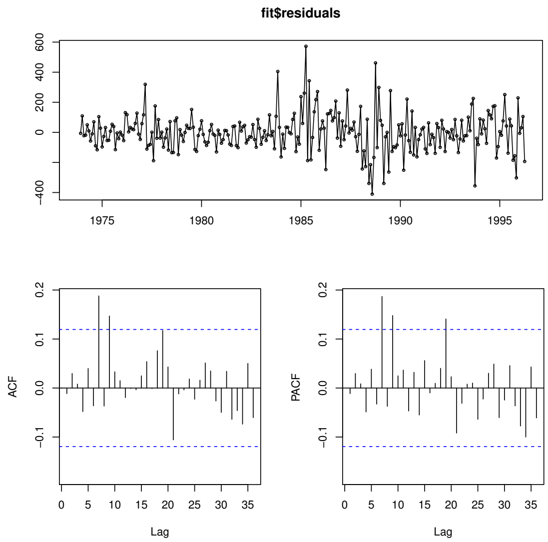



Based on the spikes in the PACF and decay in the ACF, as before, I would still consider an AR(2).

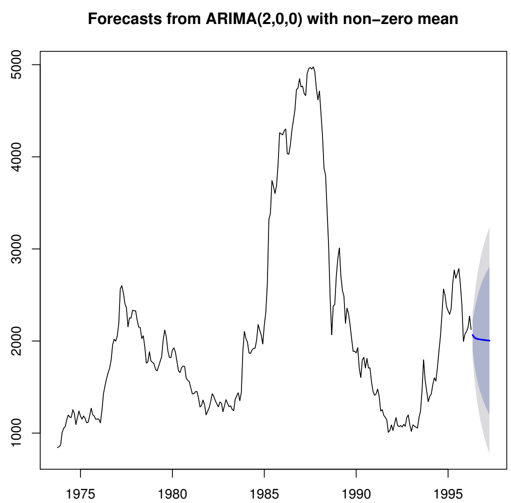
```
fit=arma(price_decomposition_Seas_Adjusted,order=c(2,0))
summary(fit)

Call:
arma(x = price_decomposition_Seas_Adjusted, order = c(2, 0))

Model:
ARMA(2,0)

Residuals:
     Min       1Q   Median       3Q      Max
-410.338 -78.176  -4.069  59.615  572.369

Coefficient(s):
          Estimate Std. Error t value Pr(>|t|)
ar1        1.37818    0.05600   24.608 < 2e-16 ***
ar2       -0.39182    0.05586   -7.014 2.32e-12 ***
intercept 31.01299   17.10609    1.813   0.0698 .
---
Signif. codes: 0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Fit:
sigma^2 estimated as 16163, Conditional Sum-of-Squares = 4331656, AIC = 3401.18
```




Based on the high statistical significance of the estimated parameters and overall looks of the fit (red curve) I would say the model fit did a very good job. However, we also have to look at the residuals to see if there are any dynamics unaccounted for. Based on the ACF and PACF plots below, we can see that we can still improve our model.




(d) Using your estimated model in part (c) provide a 12 step ahead forecast. Discuss your forecast.




The forecast seems to be heavily weighted towards the unconditional mean as expected, especially since we are now working with the seasonally adjusted series.

(e) In part (b) you applied an additive seasonal adjustment to your data, comment on whether a multiplicative seasonal adjustment would have been better.

Since the series did not have much of a seasonal component, and it did not vary with time as much, a multiplicative seasonal adjustment probably would not have been better.