STAT460-First Major, Term 222, Due Date: 14-Feb-2022

Name: ID:

1 (a) Write each of the models

(i) $Y_t = 0.3Y_{t-1} + e_t$

(ii) $Y_t = e_t - 1.3e_{t-1} + 0.4e_{t-2}$

(iii) $Y_t = 0.5Y_{t-1} + e_t - 0.3e_{t-1} + 1.2e_{t-2}$

(iv) $Y_t = 0.4Y_{t-1} + 0.45Y_{t-2} + e_t + e_{t-1} + 0.25e_{t-2}$

using backshift notation and determine whether the model is stationary and/or invertible.

(b) Write the following equations in lag operator form

(i) $X_t - \alpha X_{t-1} - \beta X_{t-2} = \epsilon_t$

(ii) $X_t = \alpha X_{t-1} + \epsilon_t$

(iii) $X_t = \alpha X_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$

2 Total annual precipitation is recorded yearly for 19 years. This record is examined to see if the amount of precipitation is tending to increase or decrease. The precipitation in inches was

```
45.25,   45.83,   41.77,   36.26,   45.37,   52.25,   35.37,   57.16,
35.37,   58.32,   41.05,   33.72,   45.73,   37.90,   41.72,   36.07,
49.83,   36.24,   39.90.
```

Does Cox-Stuart test find evidence at $\alpha = 0.05$ of a trend? You are expected to clearly state all the steps, including: (i) Hypotheses, (ii) compute the test statistic, (iii) p-value, (iv) decision rule and (v) conclusion

3 The Table below shows the actual daily occurrence of sunshine in Atlanta during November 1974, as a percentage of possible time the sun could have shone if it had not been for cloudy skies. The data are from the U.S. Department of Commerce. Dichotomize the observations according to whether the amount of sunshine was more than 50% of possible or 50% or less, and test the null hypothesis that the pattern of occurrence of the two types of day is random.

Day	Percentage	Day	Percentage	Day	Percentage
1	85	11	31	21	87
2	86	12	86	22	100
3	99	13	100	23	100
4	70	14	0	24	88
5	17	15	100	25	50
6	74	16	100	26	100
7	100	17	46	27	100
8	28	18	7	28	100
9	100	19	12	29	48
10	100	20	54	30	0

You are expected to clearly state all the steps, including: (i) Hypotheses, (ii) compute the test statistic, (iii) p-value, (iv) decision rule and (v) conclusion

4 (a) Suppose $E(X) = 2, Var(X) = 9, E(Y) = 0, Var(Y) = 4, and \quad Corr(X, Y) = 0.25$. Find:

(i) $Var(X + Y)$.

(ii) $Cov(X, X + Y)$.

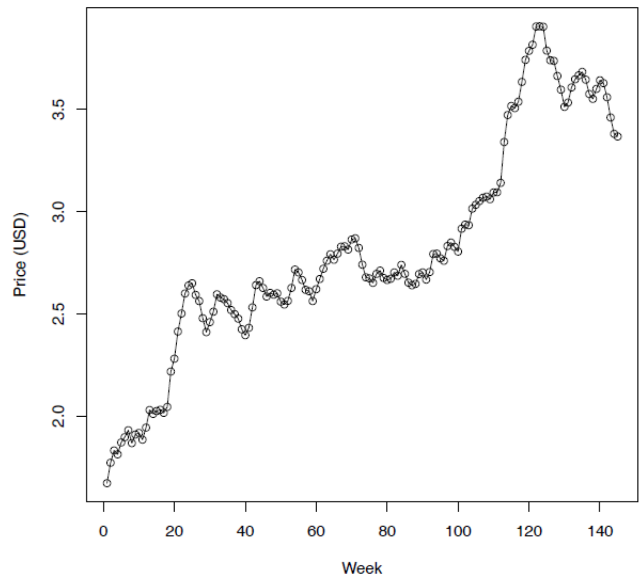
(iii) $Corr(X + Y, X - Y)$.

(b) Suppose that Z_1 and Z_2 are uncorrelated random variables with $E(Z_1) = E(Z_2) = 0$ and $var(Z_1) = var(Z_2) = 1$. Consider the process defined by

$$Y_t = Z_1 \cos(\omega t) + Z_2 \sin(\omega t) + e_t$$

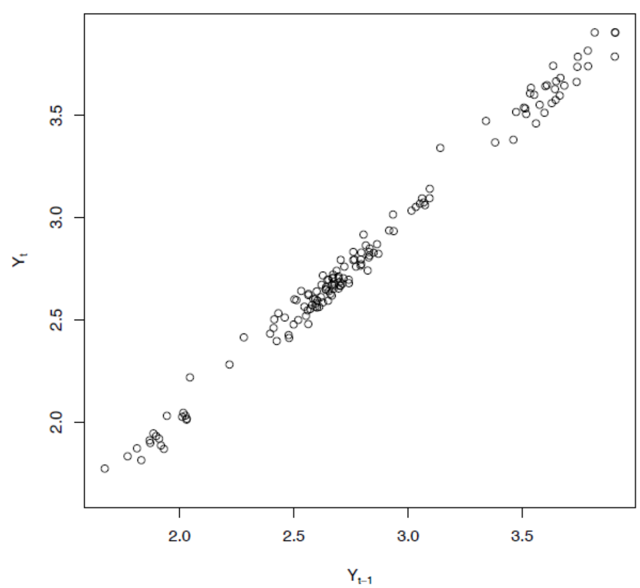
where $e_t \sim$ iid $N(0, \sigma_e^2)$ and $\{e_t\}$ is independent of both Z_1 and Z_2. Prove that $\{Y_t\}$ is stationary

5 (a) The course web site contains the data set `gasprices`, which lists the average price (US dollars per gallon) for regular gasoline in the United States. There are $n = 145$ weekly observations collected from 1/5/2009 to 10/10/2011 (Source: Rajon Coles, Fall 2011). A time series plot for these data was constructed in R and is given as:



Describe all systematic patterns you see in the plot .

(b) A scatterplot with the observed series Y_t on the vertical axis and Y_{t-1} on the horizontal axis was created. This is called a **lag-1 scatterplot**. This plot displays the observed data plotted against the lag-1 series; i.e., the scatterplot of the 144 points $(Y_1, Y_2), (Y_2, Y_3), \cdots, (Y_{144}, Y_{145})$.

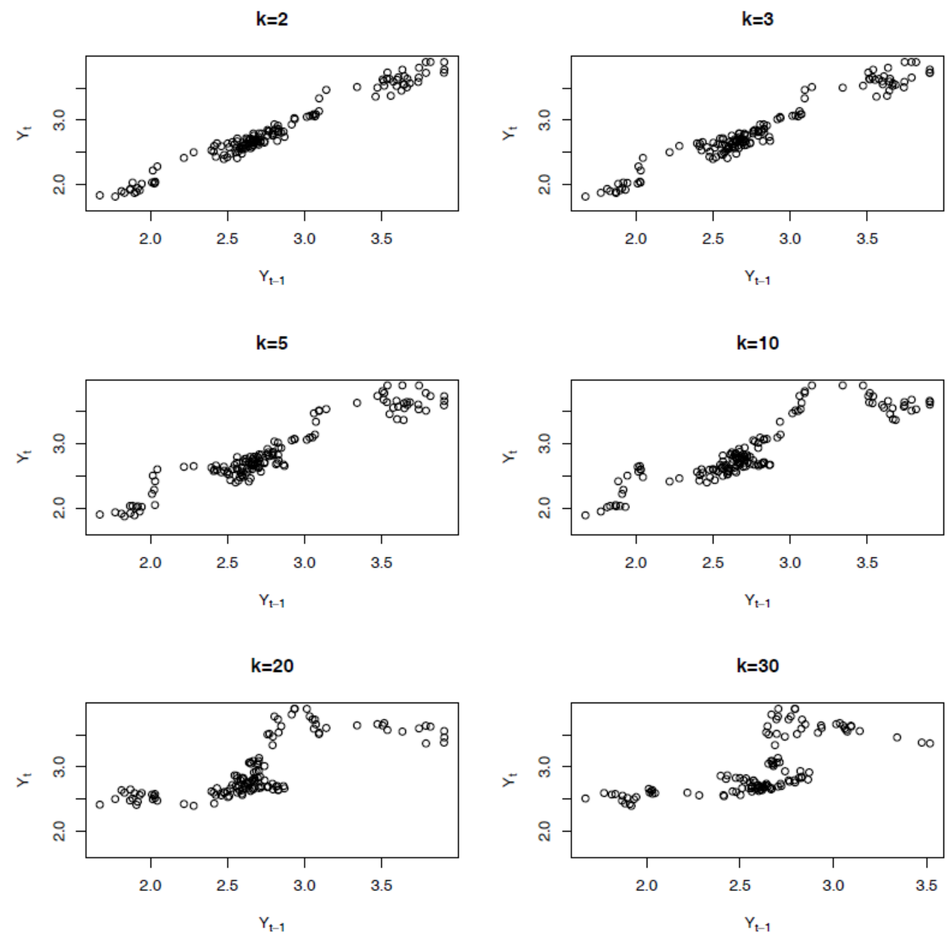


What does this plot suggests about the original series Y_t?

(c) To aid in your interpretation, you might also calculate the sample correlation for the data in the plot. You can do this using the following commands:

```
> cor(gasprices[2:145],zlag(gasprices,1)[2:145])
```

The `[2:145]` part in this code asks **R** to only consider the non-missing values because the 1st entry in `zlag(gasprices,1)` is vacuous. For larger lags (e.g., 2, 3, 5, 10, 20, etc.), the following plots are generated:



When compared to the lag-1 series, do the corresponding correlations between Y_t and Y_{t-k} increase or decrease as k increases? Interpret what is meant by this.