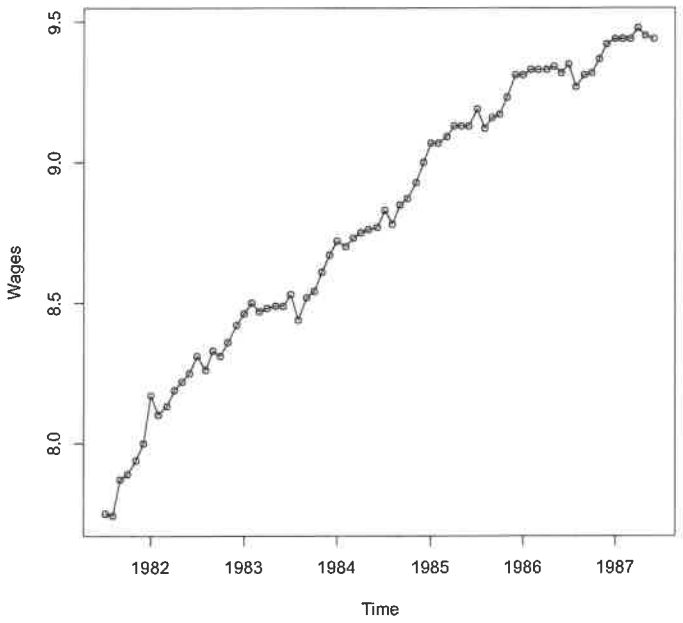
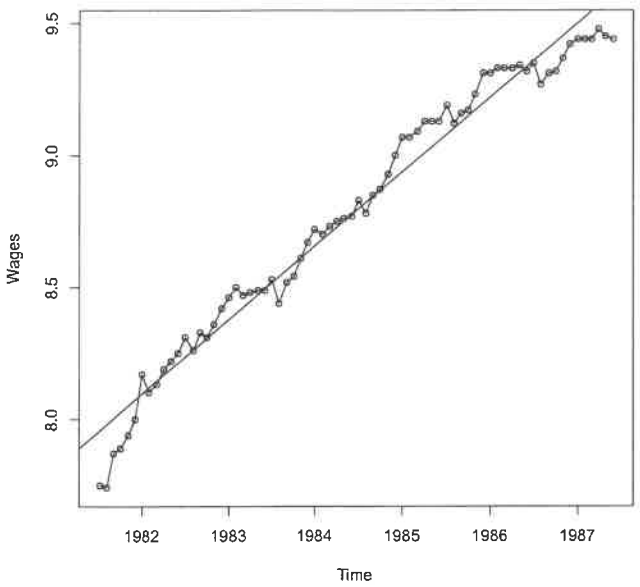
STAT 520 – Homework 2A – Fall 2023

[For this homework assignment, all students should do all the assigned problems.]

2) The monthly values of the average hourly wages for U.S. apparel and textile workers for July 1981 to June 1987 are in the `wages` object in the `TSA` package. Type `library(TSA); data(wages); print(wages)` in R to see the data set.
(a) Plot the time series. What basic pattern do you see from the plot?
From the plot below, we see that the mean function is increasing over time. The increase appears either linear or slightly nonlinear.

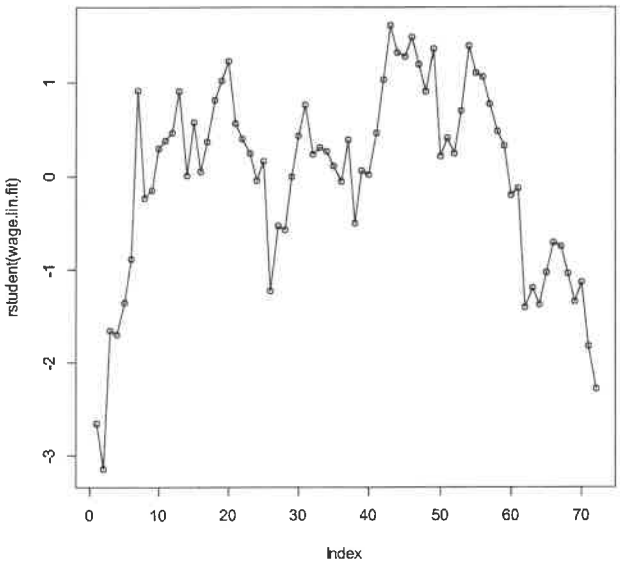



(b) Fit a linear time trend model using least squares. Give the plot of the linear trend overlain on the data, and give the estimated regression equation.
Fitted model equation:
$m_t = -549.0 + 0.2811\ t$




(c) Plot the standardized residuals from the linear regression over time. Comment on any notable pattern.

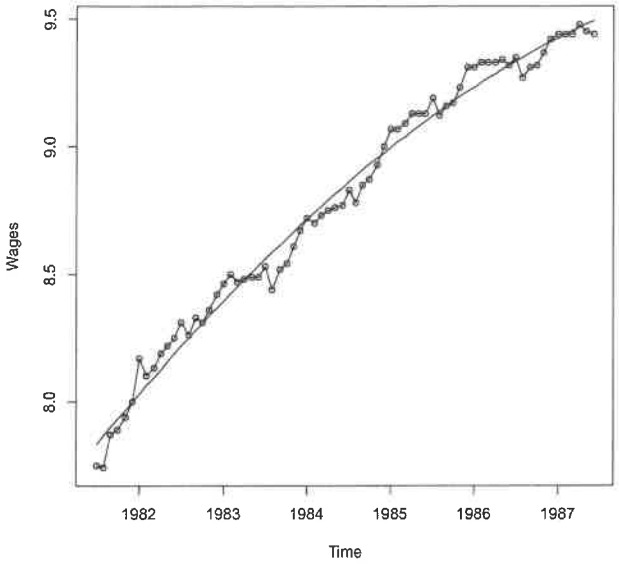
The residuals plotted over time show an "upside-down U" pattern in which they increase at the beginning and decrease at the end. This could indicate that the time trend is actually not linear.




(d) Fit a quadratic time trend model using least squares. Give the plot of the quadratic trend overlain on the data, and give the estimated regression equation.

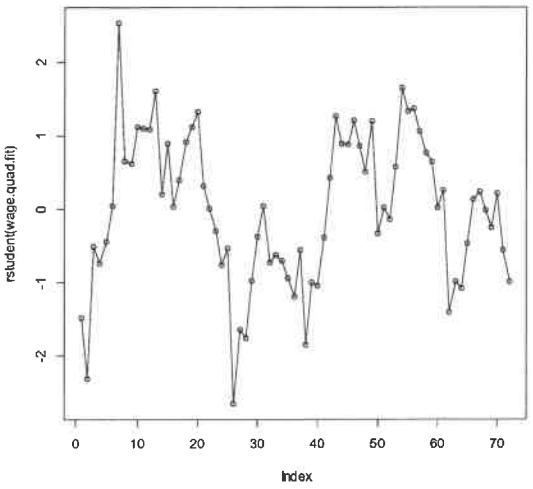
Fitted model equation:

$m_t = -84949.7 + 85.343\, t - 0.0214\, t^2$




(e) Plot the standardized residuals from the quadratic regression over time. Comment on any notable pattern.

These residuals look closer to white noise, fairly "stationary" in nature, without any strong pattern, although there is possibly some positive autocorrelation apparent.

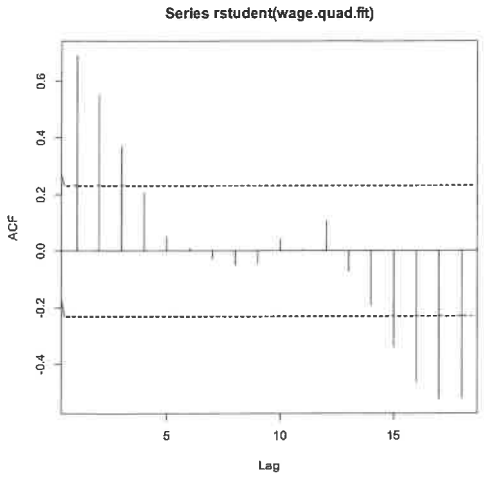



(f) Perform a runs test on the standardized residuals from the quadratic regression. What is your conclusion?

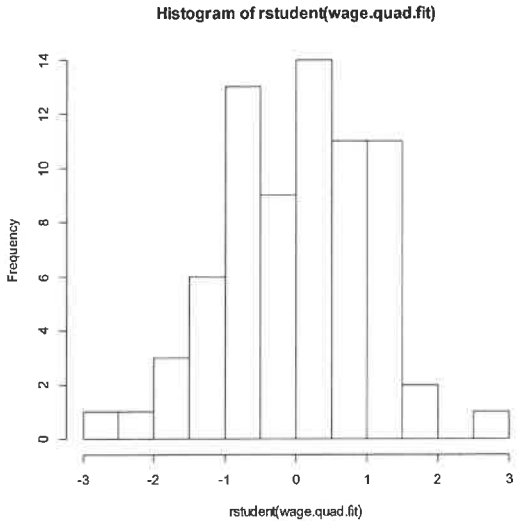
Based on the p-value near 0, we conclude that the residuals are not independent across time. In fact, since the observed number of runs is much lower than the expected number of runs under independence, we see evidence of positive autocorrelation.

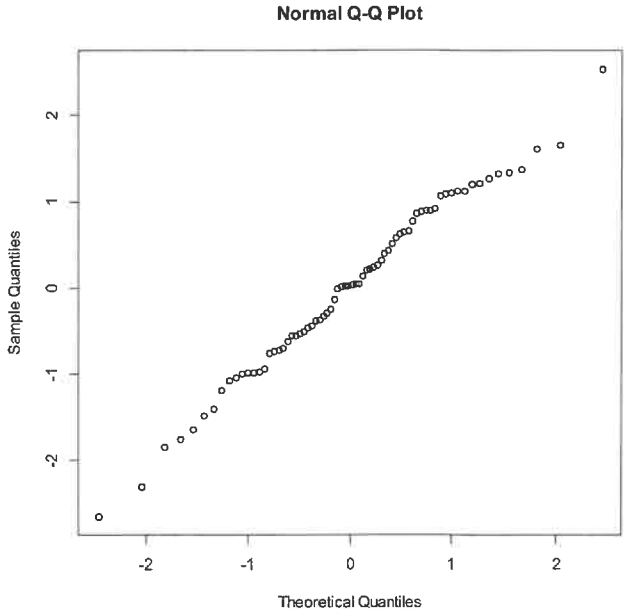
(g) Plot the autocorrelation function for the standardized residuals from the quadratic regression. What do you conclude about the standardized residuals?

Based on the ACF, we see there is notable positive autocorrelation at lags 1, 2, and 3, and negative autocorrelation at higher lags. We conclude the residuals do not look like white noise.




(h) Investigate the normality of the standardized residuals (error terms) from the quadratic regression. What is your conclusion?

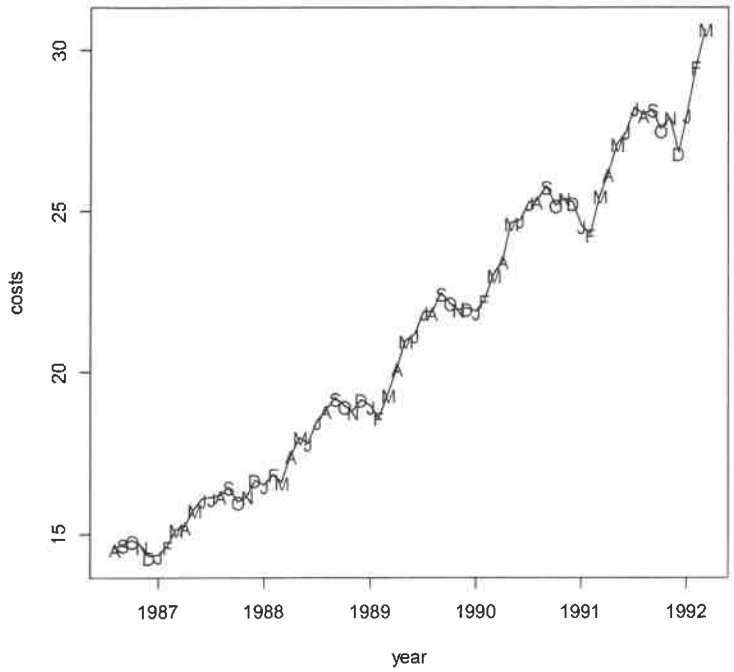



The histogram and normal Q-Q plot both indicate the errors may be normally distributed. The Shapiro-Wilk test does not reject the hypothesis of normality (p-value = 0.77).

3) The monthly U.S. prescription costs per claim for August 1986 to March 1992 are in the `prescrip` object in the `TSA` package. Type `library(TSA); data(prescrip); print(prescrip)` in R to see the data set.
(a) Plot the time series, using plotting symbols that allow you to check for seasonality. What basic pattern do you see from the plot?

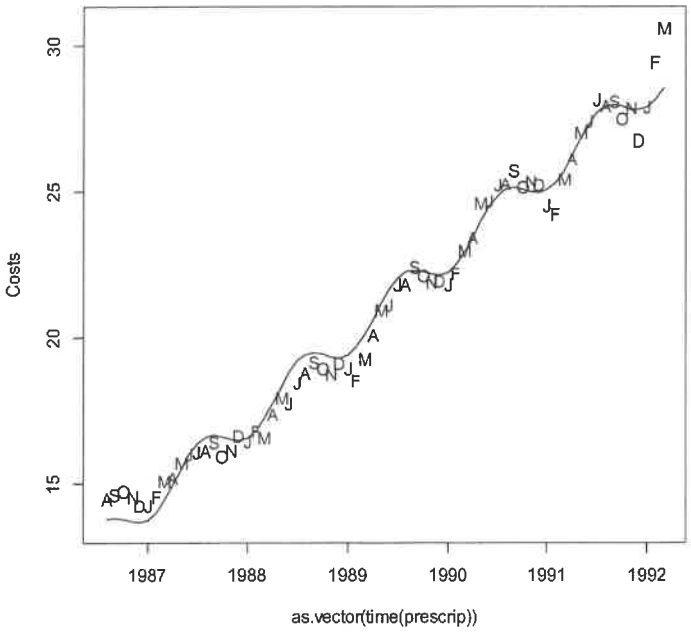
We see that the basic trend is for prescription costs to increase roughly linearly over time. There definitely seems to be some seasonality, as costs tend to be higher in late summer months and lower in cold-weather months, especially in the later timeframe of these data.




(b) Fit a harmonic regression model using least squares, including one pair of harmonic functions AND a linear time trend as predictors. Give the plot of the harmonic regression overlain on the data, and give the estimated regression equation.

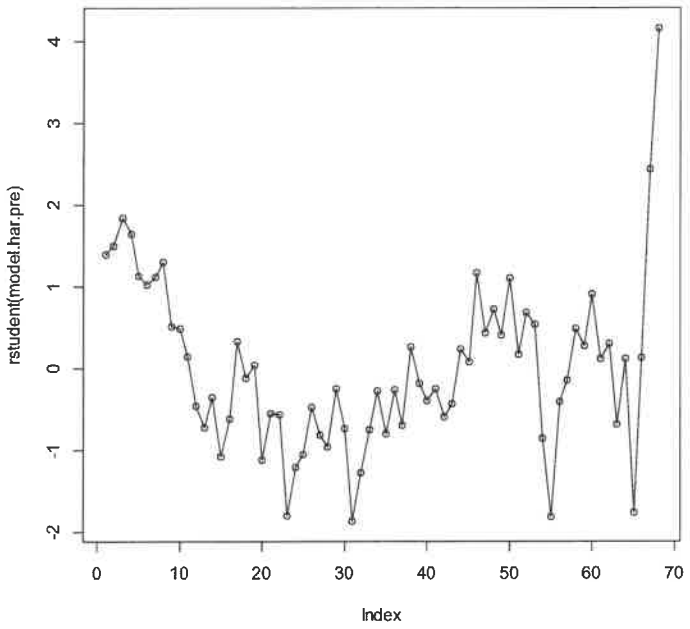
Fitted model equation:

$$m_t = -5611 + 2.831\,t - 0.6096\cos(2\pi t) - 0.1328\sin(2\pi t)$$




(c) Plot the standardized residuals from the harmonic regression over time. Comment on any notable pattern.

There looks like a bit of positive autocorrelation in places, and a couple of big outliers at the end of the series.

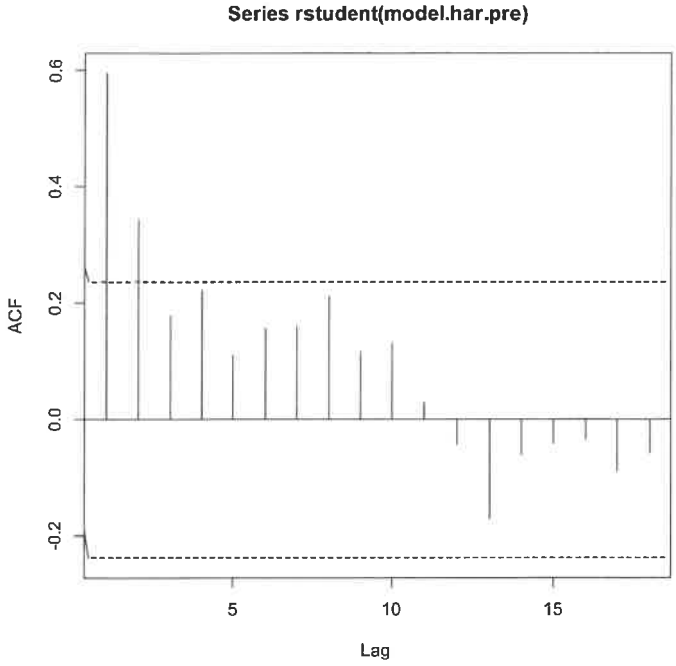



(d) Perform a runs test on the standardized residuals from the harmonic regression. What is your conclusion?

The p-value is near zero, so we reject the hypothesis that the errors are independent. There are fewer observed runs than would be expected under independence, which indicates positive autocorrelation.

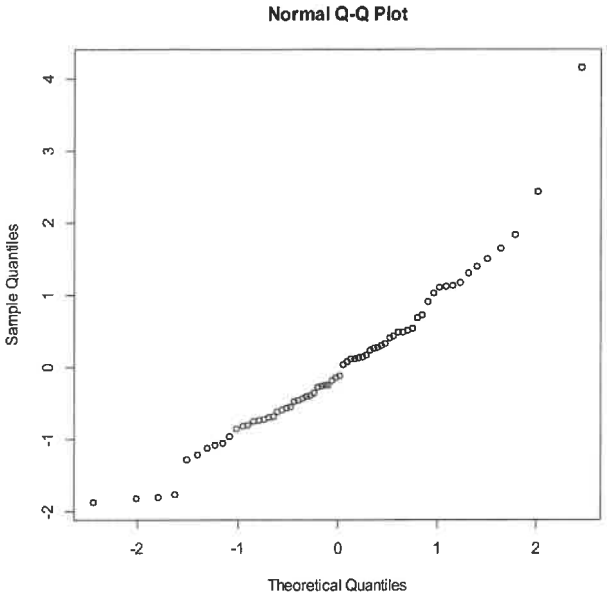
(e) Plot the autocorrelation function for the standardized residuals from the harmonic regression. What do you conclude about the standardized residuals?

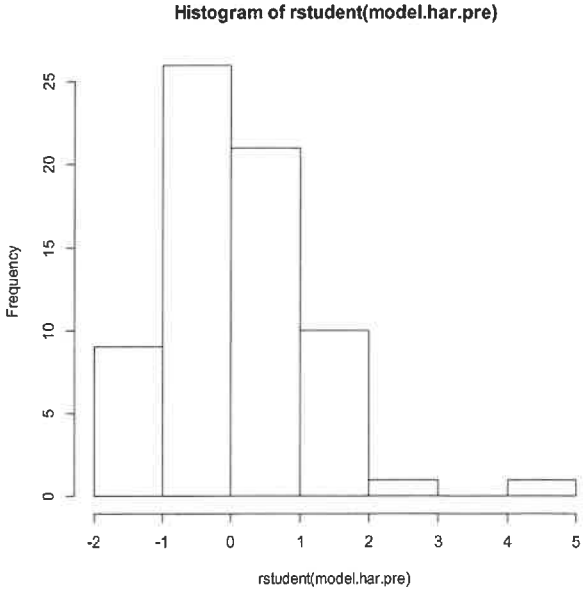
The autocorrelation function does not look like what we would expect if the errors were white noise. There looks like there is significant positive lag-1 and lag-2 autocorrelation.




(f) Investigate the normality of the standardized residuals (error terms) from the harmonic regression. What is your conclusion?

Based on the plots, the residuals appear mostly fairly normal, but with at least one notable outlier on the high end that casts doubt on the normality assumption. The Shapiro-Wilk test rejects the normality assumption (p-value = 0.0058).


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-2
-1
0
1
2
-2
-1
0
1
2
3
4


Histogram of rstudent(model.har.pre)
Frequency
rstudent(model.har.pre)
0
5
10
15
20
25
-2
-1
0
1
2
3
4
5