# lab7
PSTAT W 174/274

Sandy Gao

2023-11-17

We will continue to analyze adjusted monthly milk production measured in pounds per from Jan. 1962 to Dec. 1975. And we can import the dataset from tsdl package as milk in R, and denote the milk time series as Xt. For comparison, we split the dataset into training set train and testing set test. The training set is used for model building, and the testing set is used for prediction verification and comparison.

```
library(tsdl)
milk <- subset(tsdl, 12, "Agriculture")[[3]]
train <- milk[1:150]
test <- milk[151:156]
ts.plot(train, ylab = "Monthly milk production")
```
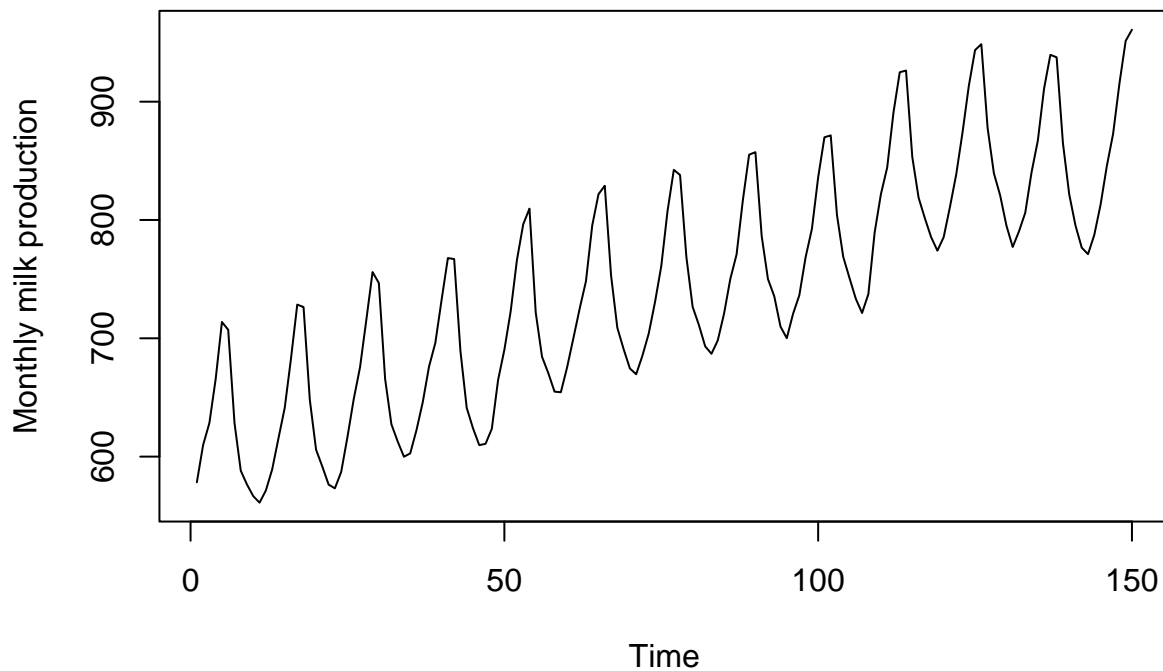
From the above graph, we can conclude that Xt is non-stationary because of the upward trend and seasonality. (You can think about whether we need to transform the series or not.) To make it more stationary, we need to remove trend and seasonality with the following code:

```
dmilk <- diff(train, 12)
ddmilk <- diff(dmilk, 1)
```

Let Yt denote the series ddmilk. Then, $Yt = (1 - B)(1 - B12)Xt$. As Lab Assignment 5 and 6, we choose SARIMA(0, 1, 0)x(0, 1, 1)12 as our final model. Now, we will conduct model diagnostic analysis.

```
fit.i <- arima(train, order=c(0,1,0), seasonal = list(order = c(0,1,1)
                                        , period = 12), method="ML")
```

(a). Perform diagnostics on the chosen model fit. Do the residuals appear to be white noise? Are they normally distributed? You should conduct hypothesis testing and plot some graphs to answer this questions.
1. Check normality assumptions:

```
# Calculate residuals from the model
resi = residuals(fit.i)

par(mfrow=c(2,2))

# Histogram of residuals
hist(resi,density=20,breaks=20, col="blue", xlab="", prob=TRUE,main="Histogram of res
iduals of model B")
m <- mean(resi)
std <- sqrt(var(resi))
curve( dnorm(x,m,std), add=TRUE )

# Time series plot of residuals
plot.ts(resi,ylab= "residuals of model B",main="Residuals plot of model B")
fitt <- lm(resi ~ as.numeric(1:length(resi)))
abline(fitt, col="red")
abline(h=mean(resi), col="blue")

# Normal Q-Q Plot
qqnorm(resi,main= "Normal Q-Q Plot for Model B")
qqline(resi,col="blue")
```

Histogram of res
iduals of model B

Residuals plot of model B

Normal Q-Q Plot for Model B
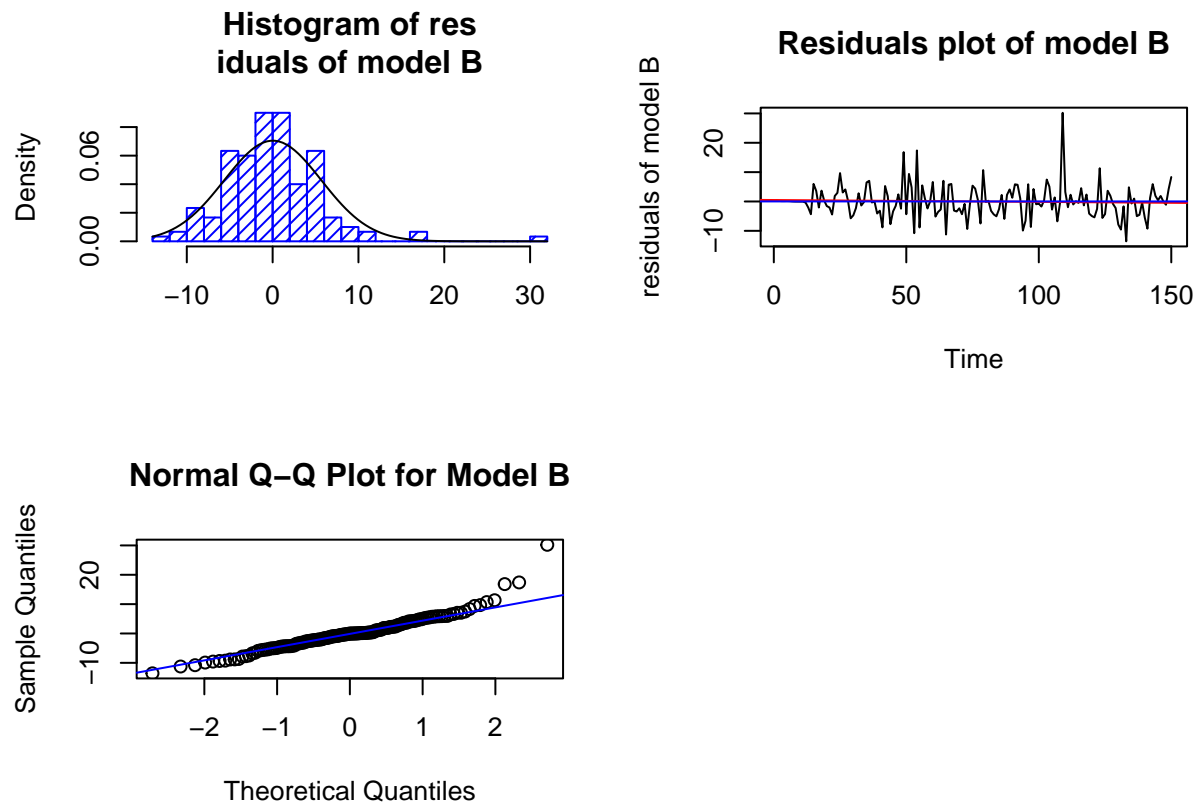
We can observe that it roughly follow a normal distribution from the q-q plot. Also, there is no trend or obvious seasonality from the time series plot of the residuals. The histogram indicates a right-skewed distribution, which might point to some model misspecification or data issues such as outliers or a non-linear relationship not captured by the model.

2. Portmanteau Statistics:

```
num_observations <- length(train)
num_observations
```

```
## [1] 150
```

```
# Shapiro test for normality
shapiro.test(resi)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resi
## W = 0.93102, p-value = 1.17e-06
```

```
# Box-Pierce test
Box.test(resi,lag = 12, type = c("Box-Pierce"), fitdf =  1)
```

```
##
##   Box-Pierce test
##
## data:  resi
## X-squared = 4.665, df = 11, p-value = 0.9463
```

```r
# Ljing-Box test
Box.test(resi, lag = 12, type = c("Ljung-Box"), fitdf = 1)
```

```
##
##   Box-Ljung test
##
## data:  resi
## X-squared = 4.8742, df = 11, p-value = 0.9371
```

```r
# Mclead-Li test
Box.test(resi^2, lag = 12, type = c("Ljung-Box"), fitdf = 0)
```

```
##
##   Box-Ljung test
##
## data:  resi^2
## X-squared = 1.8771, df = 12, p-value = 0.9996
```

lag: Here, we choose lag $= 12$ because we have n=150 observations and the square root is about 10. fitdf: Also, Model i has 1 estimated coefficients. Therefore, fitdf $= 1$. We can see that the p-values are larger than 0.05 for all the tests except shapiro test. It seems that the residuals of the model do not follow a normal distribution but do not show signs of autocorrelation or heteroscedasticity. This might indicate that while the independence assumption (no autocorrelation) for the residuals is met, the assumption of normality is violated.

3. Check if the residuals are AR(0):

```r
ar(resi, aic = TRUE, order.max = NULL, method = c("yule-walker"))
```

```
##
## Call:
## ar(x = resi, aic = TRUE, order.max = NULL, method = c("yule-walker"))
##
##
## Order selected 0  sigma^2 estimated as  32.02
```

We can see ar() function have recommend order 0 result for the residuals, which means that the residuals are WN.
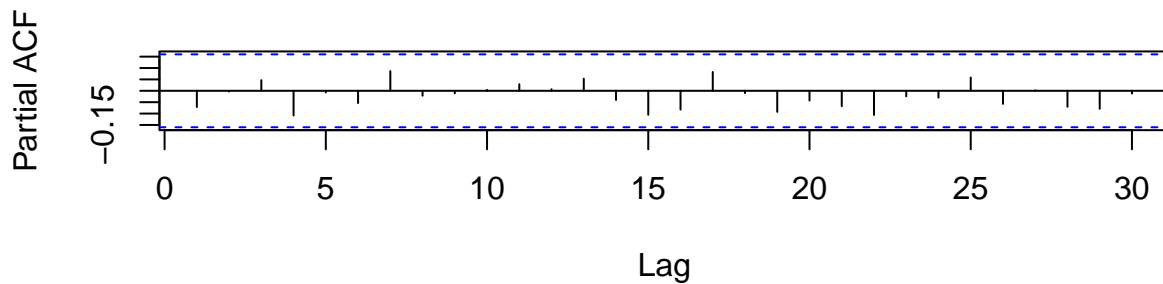
4. Check ACF and PACF

```r
par(mfrow=c(2,1))
acf(resi, lag.max=30,main="")
title("ACF of the residuals of Model B")
pacf(resi, lag.max=30,main="")
title("PACF of the residuals of Model B")
```

4

ACF of the residuals of Model B
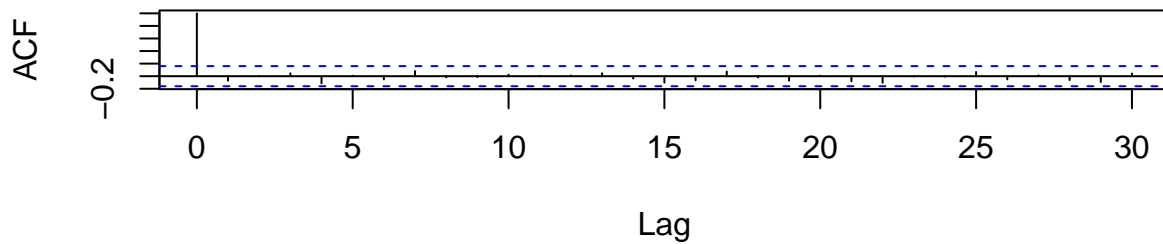
PACF of the residuals of Model B

ACF and PACF values at all lags are within the confidence interval excluding lag 0. So the residuals can be seen as White Noise.

(b). Forecast the next 6 observations using predict(), and plot your predictions. And you should also add true milk production points in test.

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##    method            from
##    as.zoo.data.frame zoo
```

```
pred.tr <- predict(fit.i, n.ahead = 6)
U.tr= pred.tr$pred + 2*pred.tr$se
L.tr= pred.tr$pred - 2*pred.tr$se
```

```
ts.plot(train, xlim=c(1,length(train)+6), ylim = c(min(train),max(U.tr)))
lines(U.tr, col="blue", lty="dashed")
lines(L.tr, col="blue", lty="dashed")
points((length(train)+1):(length(train)+6), pred.tr$pred, col="red")
```