Time Series HW # 4

Liam Flaherty

Professor Martin

NCSU: ST534-001

October 7, 2024

1) Analyze the third data set on Moodle.

a. Determine possible models for the data set using diagnostics such as the ACF, PACF, and white noise test. Include a unit root test and discuss those results as well. Include relevant plots and tables with your submission.

Our first step is to plot the data. We show it in Figure 0.1 below.

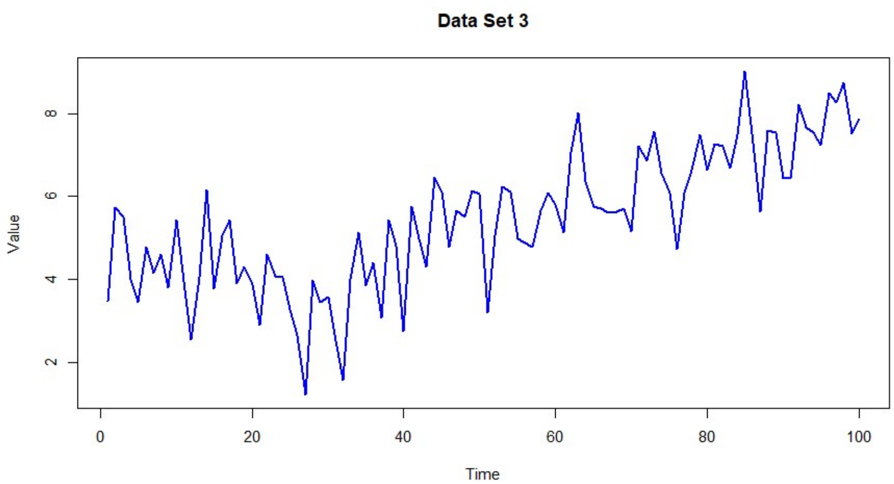

Figure 0.1: Posted Time Series Data

There is no obvious change in variance. The data may have some trending (e.g. from time point 30 to 100), but it is not abundantly obvious either way. Additionally, there might be some seasonality (the gaps between local peaks and valleys is approximately equal).

Before proceeding further, we should test if a model is needed in the first place (i.e. if the series is just a random walk). We use the Ljung-Box Q Test in Figure 0.2 below at lags of 6 and 12 to determine if a fit is needed. With p-values near machine-epsilon, we can comfortably reject white-noise at any reasonable significance level α.

```
> whitenoise6=Box.test(ts_data3,                #Do we need to fit model?#
+                      lag=6,
+                      type="Ljung-Box")
> whitenoise6                                   #p small \implies yes#

        Box-Ljung test

data:  ts_data3
X-squared = 254.02, df = 6, p-value < 2.2e-16

> whitenoise12=Box.test(ts_data3,
+                       lag=12,
+                       type="Ljung-Box")
> whitenoise12

        Box-Ljung test

data:  ts_data3
X-squared = 422.81, df = 12, p-value < 2.2e-16
```

Figure 0.2: Ljung-Box Q Test For White Noise

Now that we know we need to fit a model, we use R's built in ACF() and PACF() functions to get an idea of which models we want to try fitting. The code is shown in Figure 0.3 below.

```
###1b. Plot ACF and PACF###
ts_data3=ts(data3$val)              #Convert from DF to Time Series Object#
length2=nrow(data3)

par(mfrow=c(1,2))
data3_acf=acf(ts_data3, lag.max=40,
              main=paste0("Time Series Data Of Length ", length2, "\n", "Estimated ACF"),
              ci.col="blue",
              col="red",
              lwd=4)
data3_acf


data3_pacf=pacf(ts_data3, lag.max=40,
                main=paste0("Time Series Data Of Length ", length2, "\n", "Estimated PACF"),
                ci.col="blue",
                col="red",
                lwd=4)

data3_pacf
```

Figure 0.3: R Code For ACF And PACF Functions

The plots in Figure 0.4 are the result. Notice how the ACF dies out slowly, while the PACF seems to significantly cut off at the second lag; a natural choice for our model is an AR(2). While that second lag has the largest partial autocorrelation, the PACF does not completely die out. To account for the PACF's reluctance to cut off, some type of ARIMA model may be necessary.

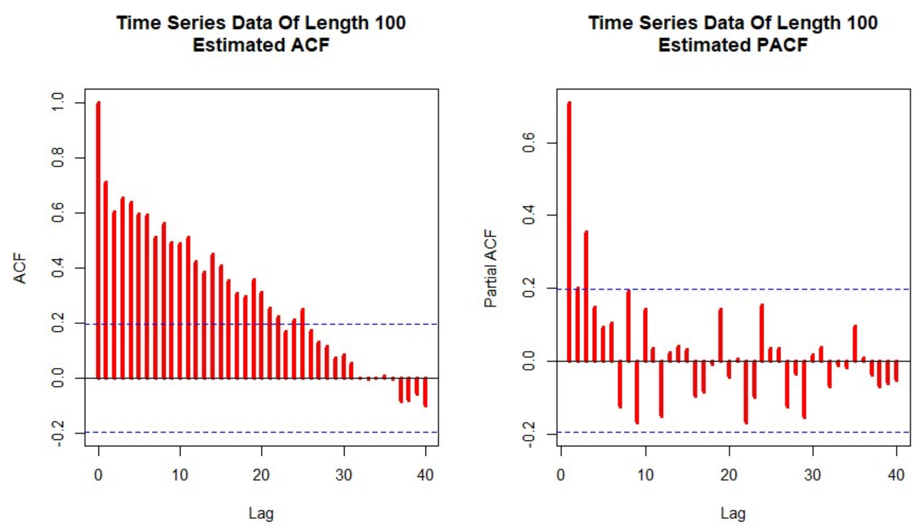

Figure 0.4: Autocorrelation And Partial Autocorrelation Of Data

To test if a difference is needed, we use the Augmented Dickey-Fuller Test in Figure 0.5 below. Under a significance level of $\alpha = 0.05$, we fail to reject the null hypothesis of "there is a unit root" ($p = 0.21$). As such, we will fit ARIMA models in addition to our AR(2).

```
> ###1c. Augmented Dickey-Fuller (Unit Root Test)###
> adf_result=adf.test(ts_data3)          #Null is that there is a root#
> adf_result                             #since p is 0.2, don't reject null; assume non-stationary#

        Augmented Dickey-Fuller Test

data:  ts_data3
Dickey-Fuller = -2.8748, Lag order = 4, p-value = 0.2142
alternative hypothesis: stationary
```

Figure 0.5: R Code For Augmented Dickey-Fuller Test

b. Fit the models that you identified as good possibilities and compare their fits using output diagnostics such as the residual test for white noise, AIC, SBC, etc.

After differencing the time series like indicated in the above, we see the following series.

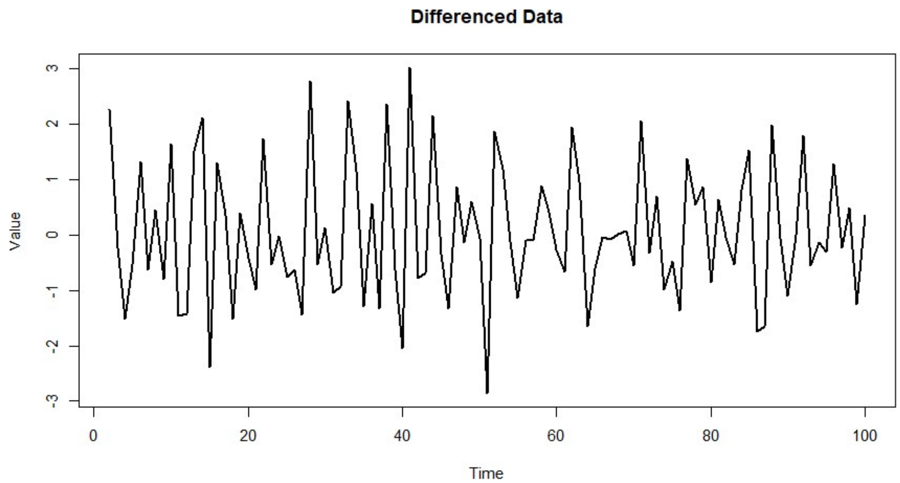

Figure 0.6: Plot Of Differenced Time Series

At least visually, this series shows better signs of weak stationarity than our first plot. Nevertheless we will proceed with our fitting of an AR(2) for comparison purposes.

The ACF and PACF for the residuals of the AR(2) model, plotted in Figure 0.7 still show signs of a signal.

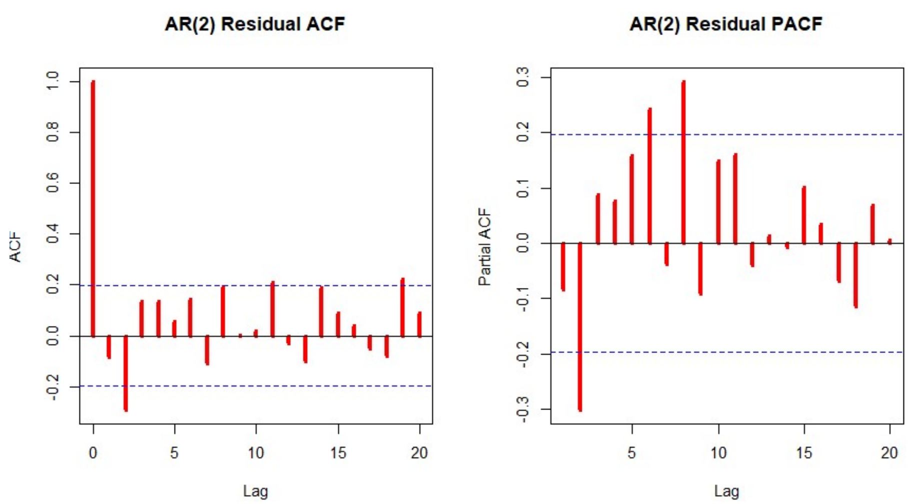

Figure 0.7: AR(2) Residual ACF and PACF

Our model diagnostics show that one cannot reject the presence of a signal at any significance level greater than $\alpha = 0.01$. The AIC for the model is about 314 while the BIC is about 325. These results are shown in Figure 0.8 below.

```
> AIC(ar2)
[1] 314.3033
> BIC(ar2)
[1] 324.7239
> result_ar2=Box.test(resid_ar2, lag=20, type="Ljung-Box")
> result_ar2                                #p-value still low \implies need to fit more#

        Box-Ljung test

data:  resid_ar2
X-squared = 41.797, df = 20, p-value = 0.002939
```

Figure 0.8: AR(2) Model Diagnostics

We can now try a variety of ARIMA models to see which one gives us the best fit. The model diagnostics are shown in Figure 0.9. The best model, in terms of all three of AIC (285.56), BIC (295.94), and Ljung-Box Q Test ($p = 0.48$), is an ARIMA(2,1,1) model.

```
> #ARIMA Model Diagnostics#
> ARIMA_model=vector()
> aic=vector()
> bic=vector()
> LBtest=vector()
>
> for (p in 1:3) {
+   for (q in 1:3) {
+     model=arima(ts_data3, order=c(p-1,1,q-1))
+     resid=residuals(model)
+
+     ARIMA_model[3*(p-1)+q]=paste0("ARIMA(", p-1, ",1,", q-1, ")")
+     aic[3*(p-1)+q]=round(AIC(model),2)
+     bic[3*(p-1)+q]=round(BIC(model),2)
+     LBtest[3*(p-1)+q]=round(Box.test(resid, lag=21, type="Ljung-Box")$p.value,2)
+   }
+ }
>
> df=data.frame(ARIMA_model, aic, bic, LBtest)
> df
  ARIMA_model    aic    bic LBtest
1 ARIMA(0,1,0) 322.44 325.03   0.00
2 ARIMA(0,1,1) 289.13 294.32   0.29
3 ARIMA(0,1,2) 288.72 296.50   0.24
4 ARIMA(1,1,0) 314.78 319.97   0.00
5 ARIMA(1,1,1) 289.99 297.78   0.27
6 ARIMA(1,1,2) 288.82 299.20   0.39
7 ARIMA(2,1,0) 294.31 302.10   0.02
8 ARIMA(2,1,1) 285.56 295.94   0.48
9 ARIMA(2,1,2) 287.51 300.49   0.48
```

Figure 0.9: ARIMA Model Diagnostics

We see the ACF and PACF of the residuals for our ARIMA(2,1,1) model in Figure 0.10.

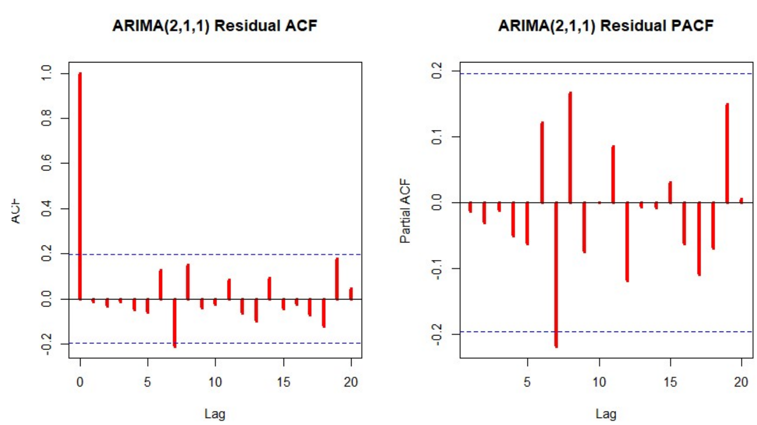

Figure 0.10: ARIMA(2,1,1) Residual ACF

c. Use your model to forecast the series 12 time units into the future.

In totality, our model is $(1 - 0.214B + 0.3056B^2)(1 - B)Z_t = (1 + 0.6546)a_t$. The results come from the code in Figure 0.11 below.

```
> #ARIMA(2,1,1)#
> arima211=arima(ts_data3, order=c(2,1,1))
> summary(arima211)

Call:
arima(x = ts_data3, order = c(2, 1, 1))

Coefficients:
         ar1      ar2      ma1
      0.0214  -0.3056  -0.6546
s.e.  0.1338   0.1138   0.1153

sigma^2 estimated as 0.9568:  log likelihood = -138.78,  aic = 285.56

Training set error measures:
                    ME      RMSE       MAE       MPE    MAPE      MASE        ACF1
Training set 0.1414653 0.9732733 0.7752467 -2.498994 17.7722 0.7913862 -0.01409562
```

Figure 0.11: Coefficients Of ARIMA(2,1,1) Model

We can forecast the next 12 time units into the future with R's predict() function. The 101st value is predicted to be about 8.15, the 102nd value is predicted to be about 8.05, and so on until the 112th value is predicted to be about 8.00.

```
> pred=predict(arima211, n.ahead = 12)
> pred
$pred
Time Series:
Start = 101
End = 112
Frequency = 1
 [1] 8.147981 8.045737 7.954332 7.983623 8.012179 8.003839 7.994934 7.997293 8.000064 7.999403 7.998542 7.998725

$se
Time Series:
Start = 101
End = 112
Frequency = 1
 [1] 0.9781764 1.0418812 1.0429226 1.0678142 1.1172006 1.1505038 1.1759539 1.2041510 1.2338354 1.2618514 1.2886182 1.3151066
```

Figure 0.12: Forecast Next 12 Values From ARIMA(2,1,1)

2) Consider the AR(2) model $(1 - 1.2B + 0.6B^2)(Z_t - 65) = a_t$ where $\sigma_a^2 = 1$ and we have the observations are $Z_{76} = 60.4, Z_{77} = 58.9, Z_{78} = 64.7, Z_{79} = 70.4,$ and $Z_{80} = 62.6$.

a. Forecast Z_{81}, Z_{82}, Z_{83}, and Z_{84}.

We first write out the model as $Z_t = 65 + 1.2(Z_{t-1} - 65) - 0.6(Z_{t-2} - 65) + a_t$. Our forecast for l steps in the future is: $\widehat{Z_{80}}(l) = 65 + 1.2(\widehat{Z_{80+l-1}} - 65) - 0.6(\widehat{Z_{80+l-2}} - 65)$. Explicitly:

$$\begin{aligned}
\widehat{Z_{80}}(1) &= 65 + 1.2(62.6 - 65) - 0.6(70.4 - 65) = 58.88 \\
\widehat{Z_{80}}(2) &= 65 + 1.2(58.88 - 65) - 0.6(62.6 - 65) = 59.096 \\
\widehat{Z_{80}}(3) &= 65 + 1.2(59.096 - 65) - 0.6(58.88 - 65) = 61.5872 \\
\widehat{Z_{80}}(4) &= 65 + 1.2(61.5872 - 65) - 0.6(59.096 - 65) = 64.44704
\end{aligned}$$

b. Determine the 95% forecast limits for the forecasts in part a.

The standard error for our forecast is given by $\sqrt{\sum_{j=0}^{l-1} \psi_j^2}$. This follows from the fact that $\mathbb{V}(e_n(l)) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2$ and we are given $\sigma_a^2 = \sigma_a = 1$. The critical value is $z_{\alpha/2}$ which is the value with $\frac{\alpha}{2}$ of the mass of the standard normal distribution to it's right (a choice of $\alpha = 0.05$ yields about 1.96). It remains to be seen what our ψ weights are from the MA representation of the above model. We can write out our model as $(1 - 1.2B + 0.6B^2)(1 + \psi_1 B + \psi_2 B^2 + \cdots)a_t = a_t$. Equating the B coefficients, we find (where $n \geq 3$):

$$\begin{aligned}
&\psi_1 B - 1.2B = 0 \implies \psi_1 = 1.2 \\
&\psi_2 B^2 - 1.2\psi_1 B^2 + 0.6B^2 = 0 \implies \psi_2 = 1.2(1.2) - 0.6 = 0.86 \\
&\psi_n B^n - 1.2\psi_{n-1} B^n + 0.6\psi_{n-2} B^n \implies \psi_n = 1.2\psi_{n-1} - 0.6\psi_{n-2}
\end{aligned}$$

Then the forecast limits (point-estimate plus/minus margin of error) are approximately:

$$\widehat{Z_{80}}(1) : 58.88 \pm 1.96 \left(\sum_{j=0}^{0} \psi_j^2 \right)^{1/2} \approx 58.88 \pm 1.96\,(1)^{\frac{1}{2}} \approx (56.92, 60.84)$$

$$\widehat{Z_{80}}(2) : 59.10 \pm 1.96 \left(\sum_{j=0}^{1} \psi_j^2 \right)^{1/2} \approx 59.10 \pm 1.96 \left(1^2 + 1.2^2\right)^{\frac{1}{2}} \approx (56.03, 62.16)$$

$$\widehat{Z_{80}}(3) : 61.59 \pm 1.96 \left(\sum_{j=0}^{2} \psi_j^2 \right)^{1/2} \approx 61.59 \pm 1.96 \left(1^2 + 1.2^2 + 0.86^2\right)^{\frac{1}{2}} \approx (58.09, 65.08)$$

$$\widehat{Z_{80}}(4) : 64.45 \pm 1.96 \left(\sum_{j=0}^{3} \psi_j^2 \right)^{1/2} \approx 64.45 \pm 1.96 \left(1^2 + 1.2^2 + 0.86^2 + 0.312^2\right)^{\frac{1}{2}} \approx (60.90, 68.00)$$

c. Suppose that the observations at $t = 81$ turns out to be $Z_{81} = 62.2$. Determine the updated forecasts Z_{82} , Z_{83}, and Z_{84}.

We have:

$$\begin{aligned}
\widehat{Z_{81}}(1) &= \widehat{Z_{80}}(2) + \psi_1 \left[Z_{81} - \widehat{Z_{80}}(1)\right] \\
&= 59.096 + 1.2\,[62.2 - 58.88] \\
&= 63.08 \\
\widehat{Z_{81}}(2) &= \widehat{Z_{80}}(3) + \psi_2 \left[Z_{81} - \widehat{Z_{80}}(1)\right] \\
&= 61.5872 + 0.86\,[62.2 - 58.88] \\
&= 64.4424 \\
\widehat{Z_{81}}(3) &= \widehat{Z_{80}}(4) + \psi_3 \left[Z_{81} - \widehat{Z_{80}}(1)\right] \\
&= 64.44704 + 0.312\,[62.2 - 58.88] \\
&= 65.48288
\end{aligned}$$

3) A sales series was fitted by the ARIMA(2,1,0) model $(1-0.14B+0.48B^2)(1-B)Z_t = a_t$ where $\sigma_a^2 = 58000$ and the last three observations are $Z_{n-2} = 640$, $Z_{n-1} = 770$, and $Z_n = 800$.

a. Calculate the forecast of the next three observations.

We can write the model as $(1 - 0.14B + 0.48B^2 - B + 0.14B^2 - 0.48B^3)Z_t = a_t$ or equivalently $Z_t = 1.14Z_{t-1} - 0.62Z_{t-2} + 0.48Z_{t-3} + a_t$. Our forecast for l steps in the future is: $\widehat{Z_n}(l) = 1.14\widehat{Z_{n+l-1}} - 0.62\widehat{Z_{n+l-2}} + 0.48\widehat{Z_{n+l-3}}$. Explicitly:

$$\widehat{Z_n}(1) = 1.14(800) - 0.62(770) + 0.48(640) = 741.8$$
$$\widehat{Z_n}(2) = 1.14(741.8) - 0.62(800) + 0.48(770) = 719.252$$
$$\widehat{Z_n}(3) = 1.14(719.252) - 0.62(741.8) + 0.48(800) = 744.0313$$

b. Calculate the 95% forecast limits for the forecasts in part a.

We know we can write $(1 - 1.14B + 0.62B^2 - 048B^3)Z_t = a_t$. Writing Z_t in terms of it's AR representation, we have $(1 - 1.14B + 0.62B^2 - 048B^3)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = a_t$. Equating coefficients of B, we arrive at our AR coefficients. We have:

$\psi_1 B - 1.14B = 0 \implies \psi_1 = 1.14$
$\psi_2 B^2 - 1.14\psi_1 B^2 + 0.62B^2 = 0 \implies \psi_2 = 1.14(1.14) - 0.62 = 0.6796$
$\psi_3 B^3 - 1.14\psi_2 B^3 + 0.62\psi_1 B^3 - 0.48B^3 = 0 \implies \psi_3 = 1.14(0.6796) + 0.62(1.14) + 0.48 = 1.961544$

Since $\mathbb{V}(e_n(l)) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2$ and we are given $\sigma_a^2 = 58000$, our standard error is $240.8319 \left(\sum_{j=0}^{l-1} \psi_j^2\right)^{\frac{1}{2}}$. As such, our forecast limits are approximately:

$$\widehat{Z_n}(1) : 741.8 \pm (1.96)240.8 \left(\sum_{j=0}^{0} \psi_j^2\right)^{\frac{1}{2}} \approx 741.8 \pm 471.97 \left(1^2\right)^{\frac{1}{2}} \approx (269.83, 1213.77)$$

$$\widehat{Z_n}(2) : 719.25 \pm (1.96)240.8 \left(\sum_{j=0}^{1} \psi_j^2\right)^{\frac{1}{2}} \approx 719.25 \pm 471.97 \left(1^2 + 1.14^2\right)^{\frac{1}{2}} \approx (3.54, 1434.96)$$

$$\widehat{Z_n}(3) : 744.03 \pm (1.96)240.8 \left(\sum_{j=0}^{2} \psi_j^2\right)^{\frac{1}{2}} \approx 744.03 \pm 471.97 \left(1^2 + 1.14^2 + 0.6796^2\right)^{\frac{1}{2}} \approx (-40.27, 1528.33)$$