# A Time Series Analysis on Lynx Trappings

Vivek Nimmagadda (110312285)

## Table of Contents



## 1. Introduction

A time series usually represented as $\{X_t: t = 1, 2, \ldots, n\}$ is a sequence of ordered observations that were collected over a specific period where 't' denotes the time of the recording, and 'n' denotes the number of observations recorded (Ranganathan & Dhanan 2012). For instance, recording the monthly sales of a product over a period of 5 years is an example of a time series.

Usually, three notable components can be observed in a time series data as follows: trend (indicates the long-term direction which can be either positive or negative), seasonality (systematic peaks and lows of the observed recordings), irregular component (short-term fluctuations). For example, let us assume that we tracked the daily total number of vehicles passing through a sensor located in the CBD of Adelaide over a period of 5 years from 2015 – 2020. Through basic intuition, it can be understood that the recordings would normally be a lot higher on the weekdays compared to the weekends which is the trend component. It is also expected that the sensor would record less traffic in the CBD during few of the public holidays such as Christmas which is the seasonality component. However, the records for 2020 would be completely different compared to the previous 4 years because of COVID and the accompanying lockdowns which is the irregular component associated with the data.

A time series analysis is the process of fitting a best possible model to the data that can accurately predict the long-term implications of the current observations. In simple terms, researchers were always keen on studying the patterns associated with systems over time, hoping to uncover the underlying principles associated with the data and build models to accurately predict the future events which led to the implementation of a time series analysis (Peixeiro 2019).

A Lynx refers to any one of the four species of medium sized wild cat with a short tail that is mostly found in the snowy mountainous regions of Alaska and Canada. During the late 19th century and early 20th century, these animals were highly hunted in Canada for their thick fur which was used in the manufacturing of different items.

The current study briefly discusses the possibilities of fitting a seasonal ARIMA time series model to the lynx trappings data that was collected by (Brockwell and Davis 1991). The data comprises of the total number of lynxes that were trapped annually in the Mackenzie River district of Northwest Canada over a period of 114 years from 1821 - 1934.

## 2. Data Exploration

Table 1 shows the summary of the lynx trappings data that was used for the analysis. On first glance, a huge difference was observed between the mean and median values. Hence, a histogram with a normal curve was plotted as shown in Figure 1.

| Minimum | 1st Quartile | Median | Mean | 3rd Quartile | Maximum |
|---|---|---|---|---|---|
| 39 | 348.2 | 771 | 1538 | 2566.8 | 6991 |

**Table 1:** Summary of Lynx trappings between 1821 – 1934

It is clear from Figure 1 that the data is highly skewed to the right (positively skewed) which is the reason behind an extremely high difference between the mean and median values. As the mean is highly influenced by the extreme values in the data, median explains the data better in this case and hence, is a more important measure for this data (Citron 2015).

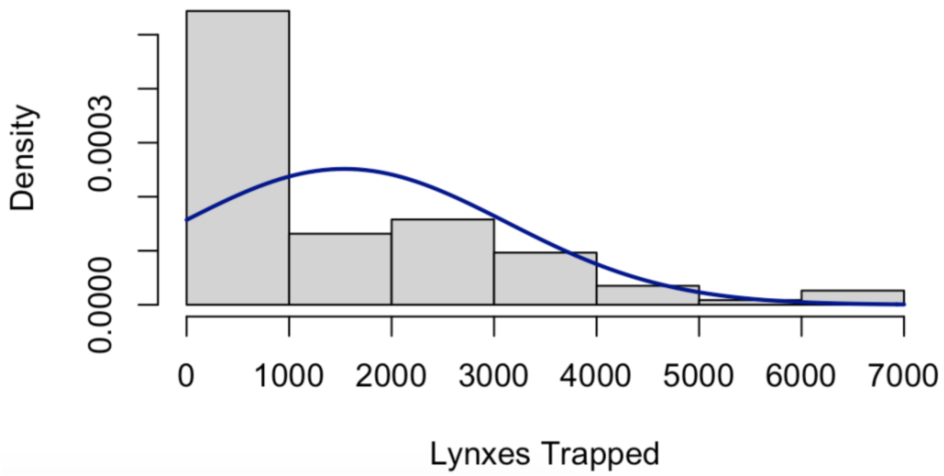



**Figure 1:** Histogram with a fitted normal curve of the lynx trappings data

Let us now visually explore the data and find out why the mean is highly affected by these extreme values. Figure 2 shows a time series plot of the data, and a lot of crests and troughs can be observed from the visual which are clearly impacting the mean. A seasonality trend is also found to be apparent from the plot. However, it is hard to tell the exact years where the local peaks were observed from the base plot.

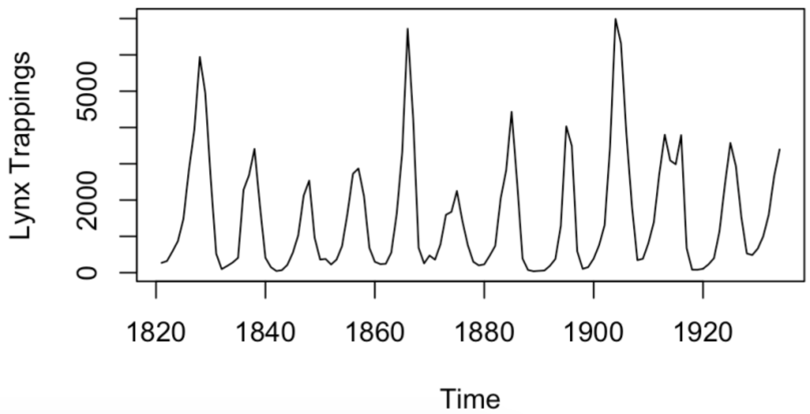



**Figure 2:** Time series plot of lynx trappings

Hence a more detailed graph visualizing the local peaks was plotted as shown in Figure 3 and most of the peaks were spread between 9 – 10 years apart hinting that the period for the fitted time series model might be any of these two values.

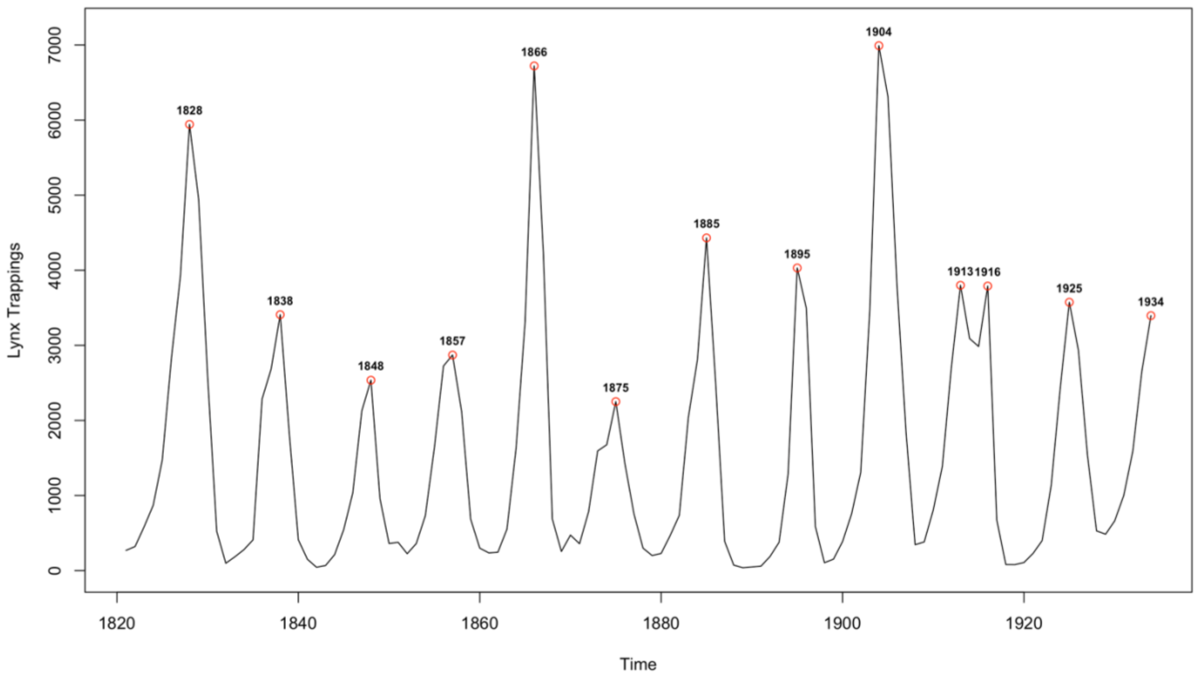



**Figure 3:** Observation of the years of local peaks in the data

## 3. Components of a Time Series Analysis

**What are seasonal effects, how do we identify them and why do they have to be adjusted?**

A seasonal effect is a timely and calendar-related effect. The dramatic rise in the sales of some of the products in supermarkets that occur around December in response to the Christmas holiday, or a surge in water usage in summer due to warmer climate, are two of the instances. According to (Time Series Analysis: The Basics, N.A.), once this trend is observed in the data, it must be carefully approximated and removed from a time series as it was found to not only mask the underlying movements in the data but also hides a few of the non-seasonal traits that might be helpful for fitting an appropriate model.

These effects can be identified quite easily as the peaks and lows are systematically spread between one another with the magnitude being the same relative to the trend.

It is also worth noting that the year-to-year comparisons is inappropriate in a time series analysis because the trading days for both the years might be completely different and most of the public holidays such as Easter and Chinese New Year also fall on different days which might lead to misinterpretations.

**What is an irregular component?**

The irregular component or residual is the leftover after approximating and eliminating the seasonal and trend components (Time Series Analysis: The Basics, N.A.). The short-term fluctuations such as snowfall during summer, low road traffic due to COVID etc., that are experienced only once in a few years are the examples of this component. It is important to note that these fluctuations are temporary and are not periodic. However, these fluctuations are found to influence the movements which in-turn hides the seasonality and trend.

**What is a trend in a time series?**

The long-term movement of a time series can be explained as a trend which can either be positive or negative. For example, let us suppose we assume the annual population of a country to have exploded over time. When this data is plotted using a line graph, a linearly upward movement can be observed which shows that there is a positive trend in population growth with time.

Two types of trends namely deterministic and stochastic trends exist in a real-world scenario where the former is uniform with consistent increases and decreases while the fluctuations in the stochastic trend are not uniform (Time Series Analysis: The Basics, N.A., Kenton 2021).

Detrending is the process of eliminating the effects of a trend which helps in identifying the cyclic and other patterns in a dataset. However, before applying the detrending techniques, the type of trend involved must be identified.

**What is a correlogram?**

According to (Glen 2016), a correlogram or an autocorrelation function (ACF) plot is a visual representation of serial correlation in the time series data that changes over time. In other words, an autocorrelation function calculates the correlation of a time series observations with the values of the same series but at previous times and a visual examination of this task gives a correlogram (Lee 2021). This trend also carries forward for wrongly predicted values. For example, let us suppose we overestimated the stock prices for the upcoming month based on previous year's data, then these overestimations are also carried forward for the next period which can be observed using a correlogram plot. However, it is important to note that a correlogram cannot show the magnitude of the autocorrelation.

**What is a partial correlogram?**

A partial autocorrelation function performs almost like an autocorrelation function. However, the associations between interfering observations are removed in this case. In other words, all the indirect associations are eliminated (Lee 2021). A clear understanding of both the functions can be obtained by viewing the appendix section of the report.

## 4. Time Series Analysis

**Approach 1:**

In the first approach, the data was considered as it is, and the period was chosen to be 9. **The steps involved in obtaining the period and the mathematics involved in fitting the initial model is briefly explained in the (approach 1) section of the appendix of the report**. The final fitted ARIMA model was found to be (2, 0, 0) $(1, 1, 0)_9$ which produced perfect-white noise residuals as observed from Figure 4 whose AIC score was found to be 1757.93. The model also does not seem to have any convergence problems in the calculations as observed from the coefficients of Table 2.

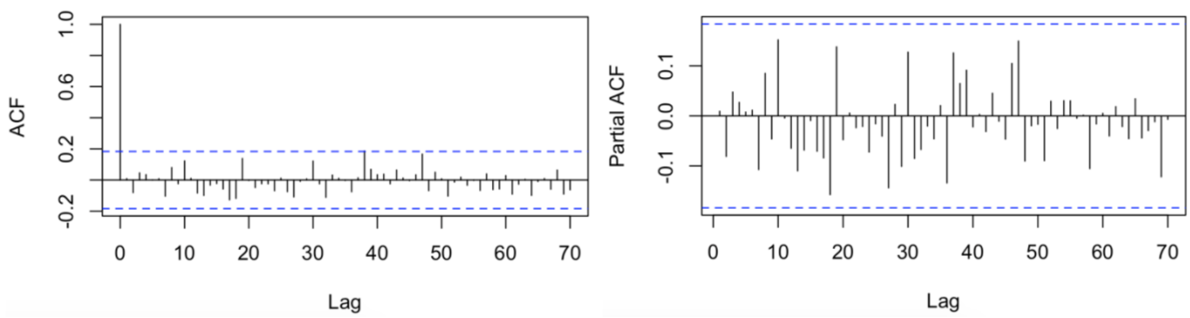



**Figure 4:** ACF and PACF plots of the ARIMA model **(2, 0, 0) $(1, 1, 0)_9$**

| | ar1 | ar2 | sar1 |
|---|---|---|---|
| | 0.9513 | -0.3986 | -0.3086 |
| s.e. | 0.0924 | 0.0923 | 0.0996 |

**Table 2:** Coefficients of the fitted ARIMA model **(2, 0, 0) $(1, 1, 0)_9$**

Based on the coefficients, the equation of the fitted model is as follows:

$$(1 + 0.3086B^9)\ (1 - 0.9513B + 0.3986B^2)\ (1 - B^9)\ X_t = W_t\ ,$$

Where $w_t$ is the white noise term with a variance of 994982. On further simplification, the final equation of the model is as follows:

$$X_t = 0.9513\ X_{t-1} - 0.3986\ X_{t-2} + 0.6914\ X_{t-9} - 0.6577\ X_{t-10} + 0.2756\ X_{t-11} + 0.3086\ X_{t-18} - 0.2936\ X_{t-19} + 0.1230\ X_{t-20} + W_t.$$

The lynx trappings for the next two seasonal cycles (18 years) were forecasted using the fitted ARIMA model as seen from Table 3. A more meaningful representation of Table 3 can be observed from Figure 5 where the forecasts were also found to follow a seasonal trend.

| Year | Cycle 1 Trappings | Year | Cycle 2 Trappings |
|---|---|---|---|
| 1935 | 1983 | 1944 | 2273 |
| 1936 | 943 | 1945 | 1124 |
| 1937 | 354 | 1946 | 408 |
| 1938 | 392 | 1947 | 421 |
| 1939 | 565 | 1948 | 596 |
| 1940 | 840 | 1949 | 890 |
| 1941 | 1458 | 1950 | 1499 |
| 1942 | 2587 | 1951 | 2608 |
| 1943 | 3446 | 1952 | 3431 |

**Table 3:** Forecasts for the next 2 seasonal cycles (1935 – 1952)

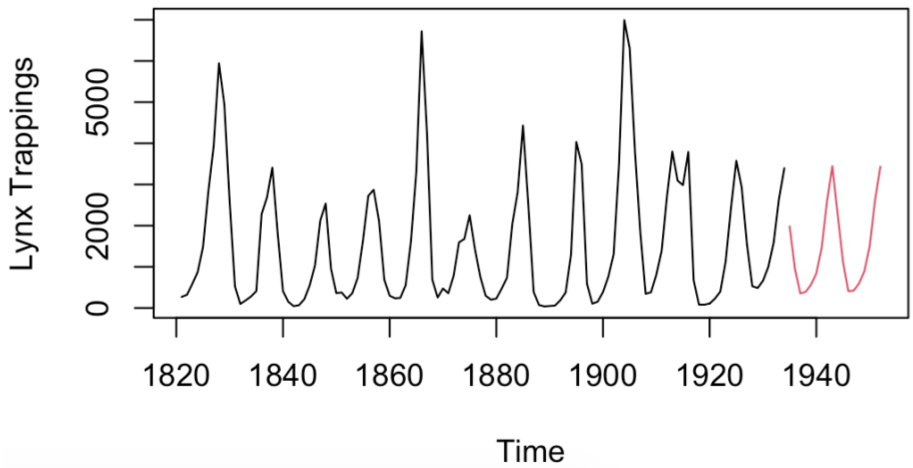



**Figure 5:** A plot of the forecasts for the next 2 seasonal cycles (1935 – 1952)

**Approach 2:**

From Figure 3, it can be noticed that the cycle is extremely asymmetric with pointed and elongated peaks and a comparatively smooth trough. In other words, lynx trappings had increased significantly for the 4$^{th}$ straight seasonal cycle after every 3 cycles. So, the numbers have been drastically increasing and decreasing for every 4 seasonal cycles which might possibly relate to the gestation period of the animal, the number of local populations, and the laws surrounding the protection of these animals. On log transforming the data, the series was found to vary uniformly around the mean as observed from Figure 6. Hence, it might be a good idea to proceed with the analysis and check whether the results are matching or even better compared to those of the previous approach. Articles by (Moran 1953, Tong 1990, Lai 1995) also supports the claim to apply the log transformation to the data.

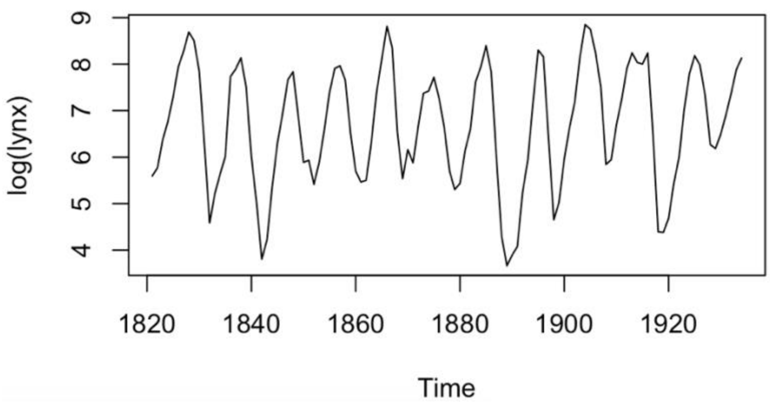



**Figure 6:** Symmetrical distribution of the data around mean after log transformation

The period was chosen to be 10. **The steps involved in obtaining the period and the mathematics involved in fitting the initial model is briefly explained in the (approach 2) section of the appendix of the report**. After twisting and tweaking the model a little, the final chosen ARIMA model that best fits the data with perfect white noise residuals and an AIC score of 170.78 was found to be (2, 0, 3) $(1, 1, 0)_{10}$. The model also does not seem to have any convergence problems in the calculations as observed from the coefficients of Table 4.

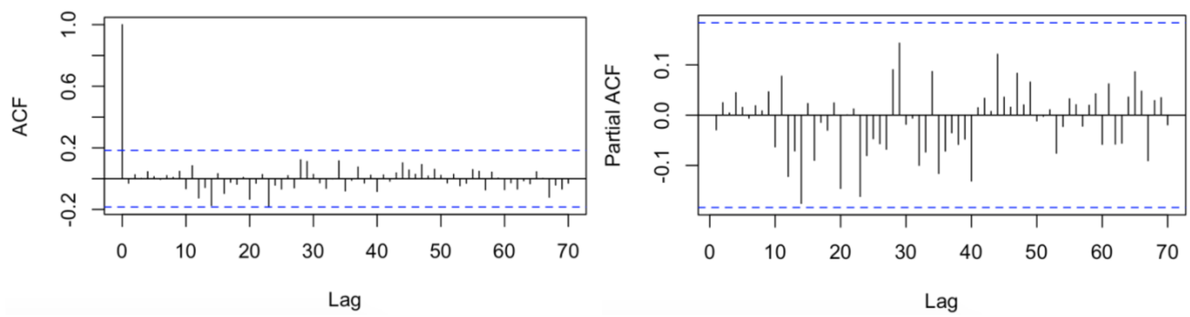



**Figure 7:** ACF and PACF plots of the ARIMA model **(2, 0, 3) $(1, 1, 0)_{10}$**

| | **ar1** | **ar2** | **ma1** | **ma2** | **ma3** | **sar1** |
|---|---|---|---|---|---|---|
| | 1.1794 | - 0.5207 | -0.0179 | 0.0569 | 0.3468 | -0.4704 |
| **s.e.** | 0.2516 | 0.2141 | 0.2516 | 0.1401 | 0.1289 | 0.0938 |

**Table 4:** Coefficients of the fitted ARIMA model **(2, 0, 3) $(1, 1, 0)_{10}$**

Based on the coefficients, the equation of the fitted model on $\log(x_t)$ is as follows:

$$(1 + 0.4704B^{10})\,(1 - 1.1794B + 0.5207B^2)\,(1 - B^{10}) \log X_t = (1 - 0.0179B + 0.0569B^2 + 0.3468B^3)\, W_t\,,$$

Where $w_t$ is the white noise term with a variance of 0.2537. Further simplification of the equation of the model yields

$$\log X_t = 1.1794 \log X_{t-1} - 0.5207 \log X_{t-2} + 0.5296 \log X_{t-10} - 0.6246 \log X_{t-11} + 0.2758 \log X_{t-12} + 0.4704 \log X_{t-20} - 0.5548 \log X_{t-21} + 0.2449 \log X_{t-22} + W_t - 0.0179W_{t-1} + 0.0569W_{t-2} + 0.3468W_{t-3}.$$

After removing logs by taking exponential, we obtain the final equation as follows:

$$X_t = ((1.1794\, X_{t-1} * 0.5296\, X_{t-10} * 0.2758\, X_{t-12} * 0.4704\, X_{t-20} * 0.2449\, X_{t-22}) / (0.5207\, X_{t-2} * 0.6246\, X_{t-11} * 0.5548\, X_{t-21}))\ \exp\,(W_t - 0.0179W_{t-1} + 0.0569W_{t-2} + 0.3468W_{t-3})$$

$$X_t = \frac{1.1794\, X_{t-1} * 0.5296\, X_{t-10} * 0.2758\, X_{t-12} * 0.4704\, X_{t-20} * 0.2449\, X_{t-22}}{0.5207\, X_{t-2} * 0.6246\, X_{t-11} * 0.5548\, X_{t-21}}\ \exp\left(W_t - 0.0179 W_{t-1} + 0.0569 W_{t-2} + 0.3468 W_{t-3}\right)$$

The forecasts for the next 2 seasonal cycles (20 years) were first made for log $x_t$ and then converted to normal values by using the exponential function which can be found from Table 5 and its visual is available from Figure 8.

| Year | Cycle 1 Trappings | Year | Cycle 2 Trappings |
|---|---|---|---|
| 1935 | 3950 | 1945 | 3777 |
| 1936 | 3765 | 1946 | 3364 |
| 1937 | 1161 | 1947 | 1330 |
| 1938 | 232 | 1948 | 343 |
| 1939 | 211 | 1949 | 312 |
| 1940 | 278 | 1950 | 418 |
| 1941 | 488 | 1951 | 683 |
| 1942 | 813 | 1952 | 1113 |
| 1943 | 1757 | 1953 | 2133 |
| 1944 | 2892 | 1954 | 3119 |

**Table 5:** Forecasts for the next 2 seasonal cycles (1935 – 1954)

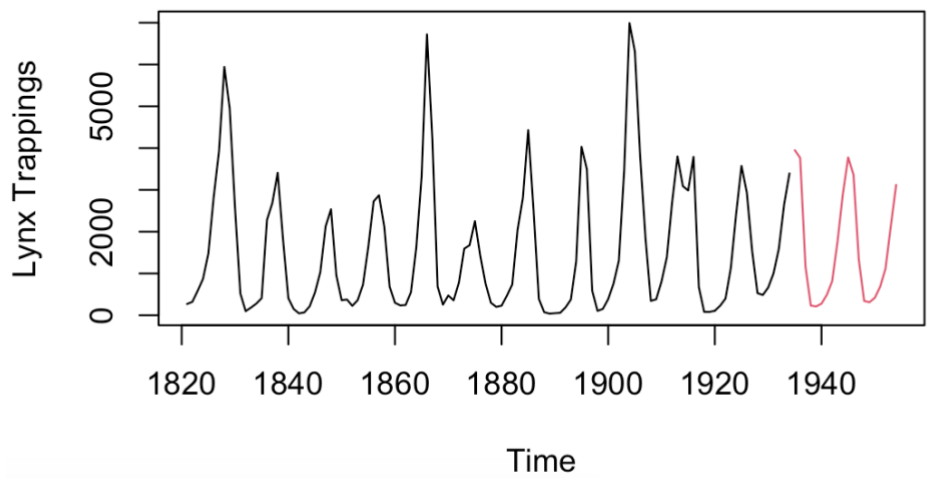



**Figure 8:** A plot of the forecasts for the next 2 seasonal cycles (1935 – 1954)

## 5. Conclusion

To sum up, prior investigation of the collected data using line plots showed that the lynx trappings in Canada followed a seasonal trend with each season ranging between 9 – 10 years.

An increase in the animal trappings on the 4th consecutive cycle after every 3 seasonal cycles was also observed.

Research by (Stenseth et.al. 1997) states that the population of lynx is greatly influenced by the hare populations in the area and that the seasonal cycles in the population of hare are responsible for causing an increase in the lynx populations thereby allowing the trapping restrictions to ease after every 3 cycles. However, the current research is only limited to the lynx trappings alone and the local hare populations are not taken into consideration.

The predictions as observed from both the Tables 4 and 5 seems to be valid and acceptable. However, the results after applying the log transformation might be more appropriate because the peak as observed from Figure 8 was a little higher compared to that of Figure 5 and follows the general trend of the data where after every 3 seasonal cycles, the number of trappings for the 4th consecutive cycle was on a rise which steeply declined again the following year. It is also important to note that the AIC scores between both the models are non-comparable as one the datasets consist of log transformed data while the other does not.

## 6. References


Brockwell, PJ & Davis, RA (1991), 'Time Series and Forecasting Methods, Second edition', *Springer*, *Series G*, pp. 557.

Citron, E 2015, 'Lynx Trappings: An analysis of the lynx dataset in R', *RStudio*, 13 April, viewed 18 November 2021, <https://rpubs.com/rez/72572>.

Glen, S 2016, 'Correlogram / Auto Correlation Function ACF Plot: Definition', *Statistics How to*, 23 August, viewed 9 November 2021, <https://www.statisticshowto.com/correlogram/>.

Kenton, W 2021, 'Detrend', *Investopedia*, 17 November, viewed 21 November 2021, <https://www.investopedia.com/terms/d/detrend.asp>.

Lai, D 1995, 'Comparison study of AR models of the Canadian lynx data: A close look at BDS statistic', *Computational Statistics & Data Analytics*, vol. 22, no. 4, pp. 409 – 423.

Lee, M 2021, 'What is the difference between autocorrelation & partial autocorrelation for time series analysis?', *Medium*, 9 March, viewed 19 November 2021, <https://mxplus3.medium.com/interpreting-autocorrelation-partial-autocorrelation-plots-for-time-series-analysis-23f87b102c64>.

Moran, PAP 1953, 'The statistical analysis of the Canadian Lynx cycle', *Australian Journal of Zoology*, vol. 1, no. 3, pp. 291 – 298.

Peixeiro, M 2019, 'The Complete Guide to Time Series Analysis and Forecasting', *Towards Data Science*, 7 August, viewed 22 November 2021, <https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775>.

Ranganathan, K & Dhanan, V 2012, 'Data mining in Canadian lynx time series', *Journal of Reliability and Statistical Studies*, vol. 5, no. 1, pp. 01 – 06.

Stenseth, NC, Falck, W, BjØrnstad, ON, & Krebs, CJ 1997, 'Population regulation in snowshoe hare and Canadian lynx: Asymmetric food web configurations between hare and lynx', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, pp. 5147 – 5152.

"Time Series Analysis: The Basics" N.A., *Australian Bureau of Statistics*, viewed 19 November 2021, <https://www.abs.gov.au/websitedbs/d3310114.nsf/home/time+series+analysis:+the+basics>.

Tong, H 1990, 'Non-linear time series: a dynamical system approach', *Society for Industrial and Applied Mathematics*, vol. 34, no. 1, pp. 149 – 151.

## 7. Appendix

The code related to the entire analysis and a few of the accompanying plots for fitting an initial model and arriving at a final model are briefly explained in this section of the report.

```
options(scipen = 999) # function to stop showing scientific notations

df <- lynx # load the dataset

# Plotting a histogram with a normal curve (Figure 1)
m<-mean(df)
std<-sqrt(var(df))
hist(df, prob = TRUE,
   xlab="Lynxes Trapped",
   main="normal curve over histogram")
curve(dnorm(x, mean=m, sd=std),
    col="darkblue", lwd=2, add=TRUE, yaxt="n")

# basic plot of the lynx trappings with highlighted local maximums (Figure 3)
plot(lynx, ylab='Lynx Trappings', ylim = c(0, 7200))

# Create data vectors
a <- c(1828, 1838, 1848, 1857, 1866, 1875, 1885, 1895, 1904, 1913, 1916, 1925, 1934)
b <- c(window(lynx, 1828, c(1828, 1)), window(lynx, 1838, c(1838, 1)), window(lynx, 1848, c(1848, 1)),
     window(lynx, 1857, c(1857, 1)), window(lynx, 1866, c(1866, 1)), window(lynx, 1875, c(1875, 1)),
     window(lynx, 1885, c(1885, 1)), window(lynx, 1895, c(1895, 1)),window(lynx, 1904, c(1904, 1)),
     window(lynx, 1913, c(1913, 1)), window(lynx, 1916, c(1916, 1)), window(lynx, 1925, c(1925, 1)),
```

```
    window(lynx, 1934, c(1934, 1)))

# Add the local maximum points
points(x = a,
    y = b,
    col = 'red')

# Add id labels
text(x = a,
   y = b,
   labels = a,
   pos = 3,
   cex = 0.7,
   font = 2)

###########################################################################
############################# Approach 1 ################################
###########################################################################
```

The normal correlogram clearly signifies that there is seasonality associated with the data as observed from Figure 9 and as a result, it must be removed before proceeding further and applying an ARIMA model. However, an appropriate period between 9 and 10 must be chosen carefully as observed earlier which is the basis for building the model.

```
acf(lynx, lag.max = 70)
pacf(lynx, lag.max = 70)
```

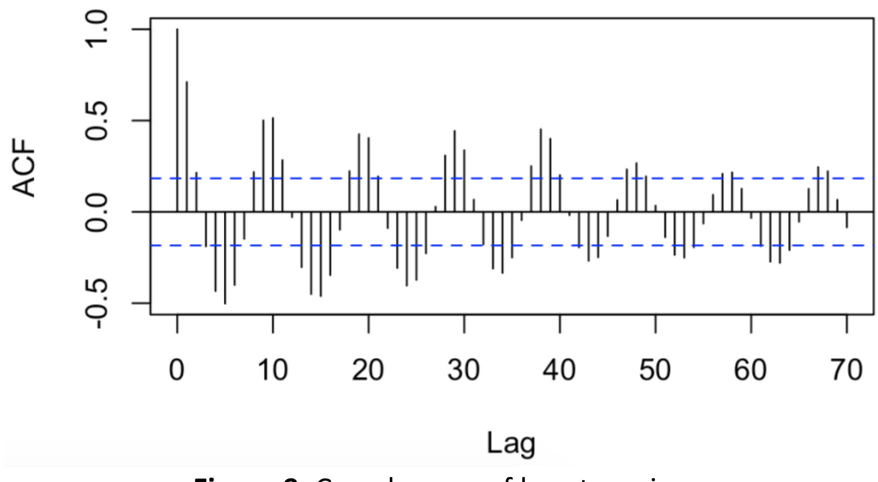



**Figure 9:** Correlogram of lynx trappings

The spectral density estimation function, simply named as spectrum() in r can be used to find the best suitable period for the data. As seen from Figure 10, two peaks can be observed as highlighted by the red circles and their frequency values were observed to be 0.1 and 0.1083 respectively which when divided by 1 gives a period of 10 and 9. However, the autocorrelation and partial-autocorrelation plots using both the values must be tried before fixing to a particular value.

```
# Spectrum function checks the optimum period for the data
spectrum(lynx)
# Add the local maximum points
a <- c(0.1, 0.108333333) # highlighting the two peaks in the plot
b <- c(39718652.61, 37928771.24)
points(x = a,
      y = b,
      col = 'red')

spectrum(lynx)$freq
spectrum(lynx)$spec
```

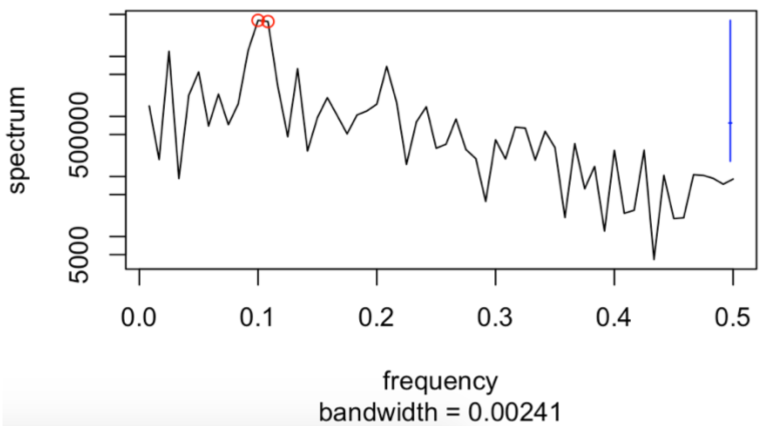


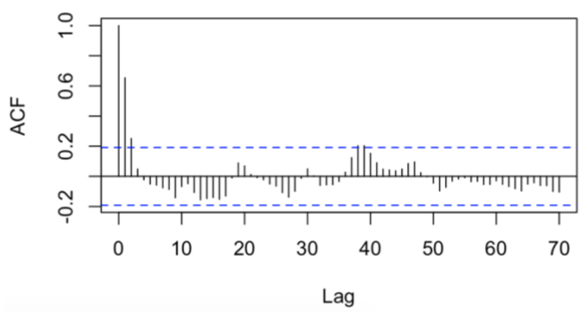

**Figure 10:** Spectral density estimation plot for choosing a period for the data

The correlogram and partial correlogram plots with lags 9 and 10 can be observed from Figures 11 and 12 respectively. The partial correlogram of $(1 - B^9)x_t$ was found to be more stable with a smaller number of significant spikes compared to $(1 - B^{10})x_t$ suggesting that fitting based on period = 9 would result in a less complex model.

```
# Finding the optimum period for the data

acf(diff(lynx, lag = 9), lag.max = 70)
pacf(diff(lynx, lag = 9), lag.max = 70)

acf(diff(lynx, lag = 10), lag.max = 70)
pacf(diff(lynx, lag = 10), lag.max = 70)
```



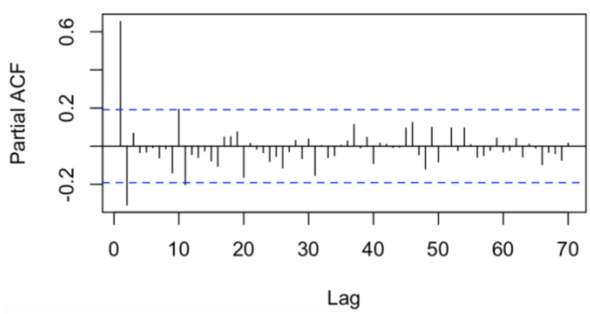



**Figure 11:** ACF and PACF plots with period = 9

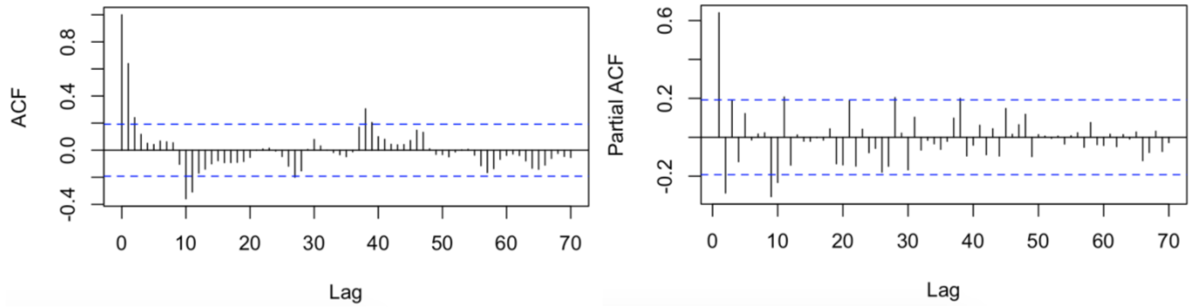



**Figure 12:** ACF and PACF plots with period = 10

Closer inspection of the PACF plot with period = 9 from Figure 13 signifies that a non-seasonal AR(2) model with possibly no components at all for the seasonal AR part would be an appropriate option to start with.

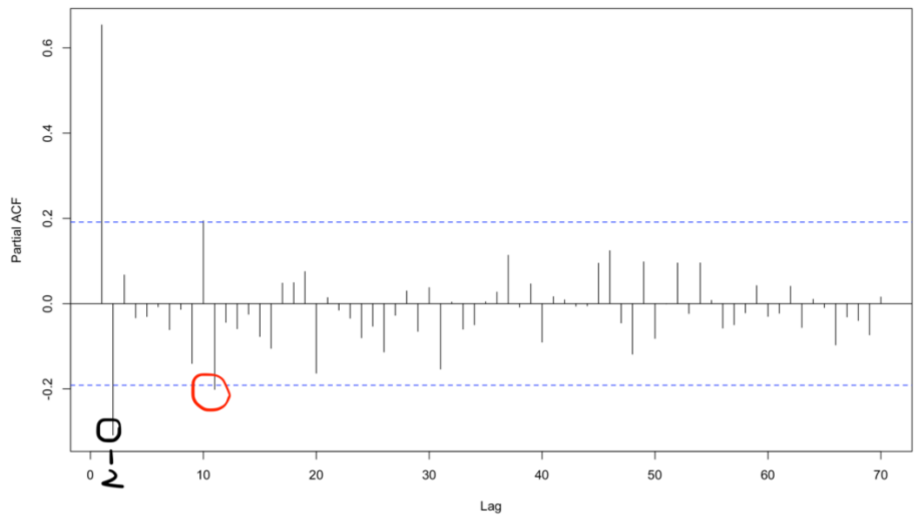



**Figure 13:** Exploring the partial correlogram plot with period = 9

The partial correlogram plot of the residuals of the preliminary fitted model as observed from Figure 14 signifies that including a seasonal AR(1) model would be appropriate.

**# Preliminary ARIMA model according to the PACF plot with period = 9**

```
acf(arima(lynx,order=c(2,0,0),seasonal=list(order=c(0,1,0),period=9))$resid, lag.max=70)
pacf(arima(lynx,order=c(2,0,0),seasonal=list(order=c(0,1,0),period=9))$resid, lag.max=70)
```

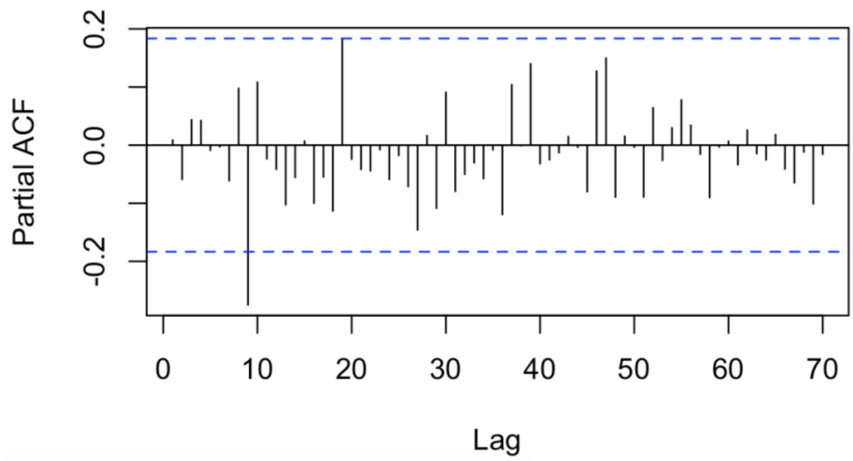



**Figure 14:** Inspecting the PACF plot of the preliminary model

**# Final ARIMA model with white-noise residuals**
arima(lynx,order=c(2,0,0),seasonal=list(order=c(1,1,0),period=9))

Call:
arima(x = lynx, order = c(2, 0, 0), seasonal = list(order = c(1, 1, 0), period = 9))

Coefficients:

```
         ar1      ar2     sar1
      0.9513  -0.3986  -0.3086
s.e.  0.0924   0.0923   0.0996
```

sigma^2 estimated as 994982:  log likelihood = -874.96,  aic = **1757.93**

acf(arima(lynx,order=c(2,0,0),seasonal=list(order=c(1,1,0),period=9))$resid, lag.max=70)
pacf(arima(lynx,order=c(2,0,0),seasonal=list(order=c(1,1,0),period=9))$resid, lag.max=70)

Although the model is simple and produced white noise residuals, we will try to find some other white noise models and check their AIC scores. The aim here is to minimize the AIC score as much as possible while keeping the model simple. So, let us begin by increasing the seasonal AR part, adding non-seasonal MA part, and reducing the non-seasonal AR part.

**# Scores of other tested ARIMA white noise models**

**# Test Model 1**
> arima(lynx,order=c(1,0,1),seasonal=list(order=c(1,1,0),period=9))

Call:
arima(x = lynx, order = c(1, 0, 1), seasonal = list(order = c(1, 1, 0), period = 9))

Coefficients:

```
         ar1     ma1     sar1
      0.4685  0.4926  -0.2943
s.e.  0.1046  0.0945   0.1011
```

sigma^2 estimated as 1004090:  log likelihood = -875.39,  aic = **1758.79**

**# Test Model 2**

```
> arima(lynx,order=c(2,0,2),seasonal=list(order=c(1,1,1),period=9))

Call:
arima(x = lynx, order = c(2, 0, 2), seasonal = list(order = c(1, 1, 1), period = 9))

Coefficients:
         ar1      ar2      ma1      ma2    sar1     sma1
      1.2863  -0.5818  -0.2294  -0.2323  0.3499  -0.9999
s.e.  0.1561   0.1103   0.1784   0.1440  0.1214   0.1478
```

sigma^2 estimated as 744331:  log likelihood = -867.94,  aic = **1749.88**

Even though, a few other models were tried whose AIC scores were below that of the final model, they were quite complex and the difference in the AIC scores was also not found to be too significant. Hence, **(2, 0, 0) (1, 1, 0)$_9$** will be the final model for this dataset.

**# Making predictions for the next 2 seasonal cycles**

```
predictions <-
round(predict(arima(lynx,order=c(2,0,0),seasonal=list(order=c(1,1,0),period=9)),
n.ahead=18)$pred, 0)
ts.plot(df, predictions, lty = c(1,1), col=c(1,2), ylab="Lynx Trappings")
```

**# Comparing the results generated by those of the predict function (turns out to be the same)**

```
library(forecast)
plot(forecast(arima(lynx,order=c(2,0,0),seasonal=list(order=c(1,1,0),period=9)), h=18))
```

**##########################################################################**
**############################ Approach 2 ################################**
**##########################################################################**

**# Log transforming and plotting the lynx trappings data**

```
log_lynx <- log(lynx)
plot(log_lynx, ylab = "log(lynx)")

acf(log_lynx, lag.max = 70)
pacf(log_lynx, lag.max = 70)
```

As seen from Figure 15, the frequency values 0.1 and 0.1083 remain as the peak points which when divided by 1 gives a period of 10 and 9. However, as with the previous case the correlograms and partial-correlograms of both the cases must be checked.

```
spectrum(log_lynx)
```

**# Add the local maximum points**

```
a <- c(0.1, 0.108333333)
b <- c(30.123923112, 31.709863295)
points(x = a,
       y = b,
       col = 'red')

spectrum(log_lynx)$freq
spectrum(log_lynx)$spec
```

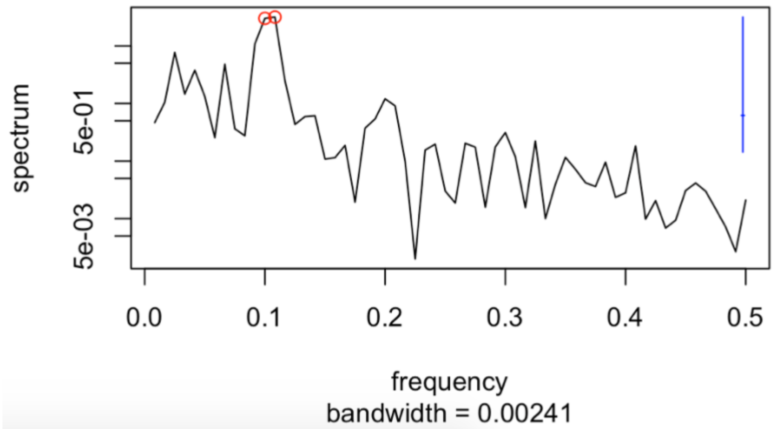



**Figure 15:** Spectral density estimation plot for choosing a period for the log data

From Figures 16 and 17, the partial correlogram of $(1 - B^{10})x_t$ was found to be more stable this time with a smaller number of significant spikes compared to $(1 - B^{9})x_t$ suggesting that fitting based on period = 10 would result in a less complex model.

```
acf(diff(log_lynx, lag = 9), lag.max = 70)
pacf(diff(log_lynx, lag = 9), lag.max = 70)

acf(diff(log_lynx, lag = 10), lag.max = 70)
pacf(diff(log_lynx, lag = 10), lag.max = 70)
```

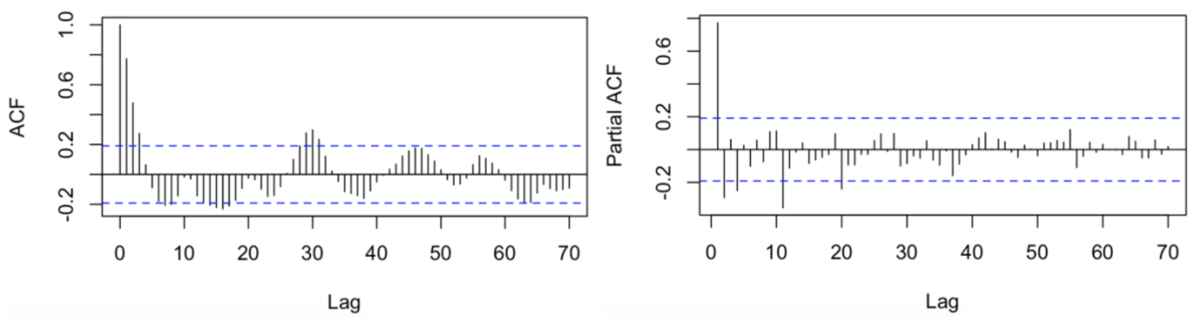



**Figure 16:** ACF and PACF plots of log(lynx) with period = 9

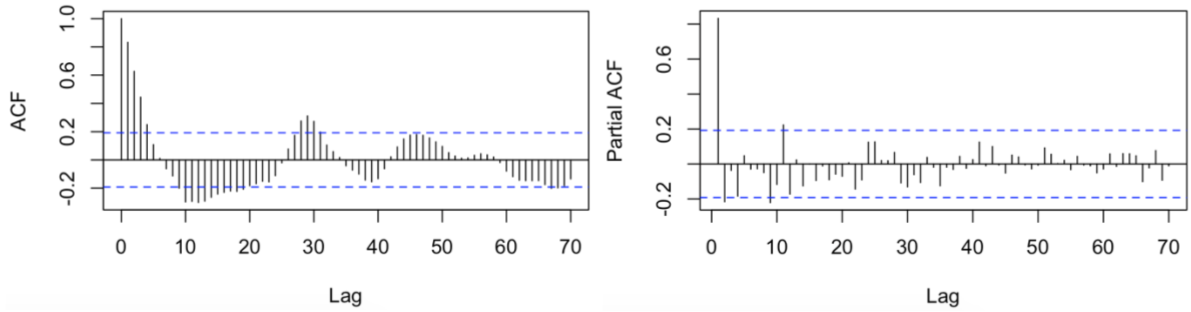



**Figure 17:** ACF and PACF plots of log(lynx) with period = 10

Closer inspection of the PACF plot with period = 10 from Figure 18 signifies that a non-seasonal AR(9) model with no seasonal AR part would be an appropriate model to begin with.

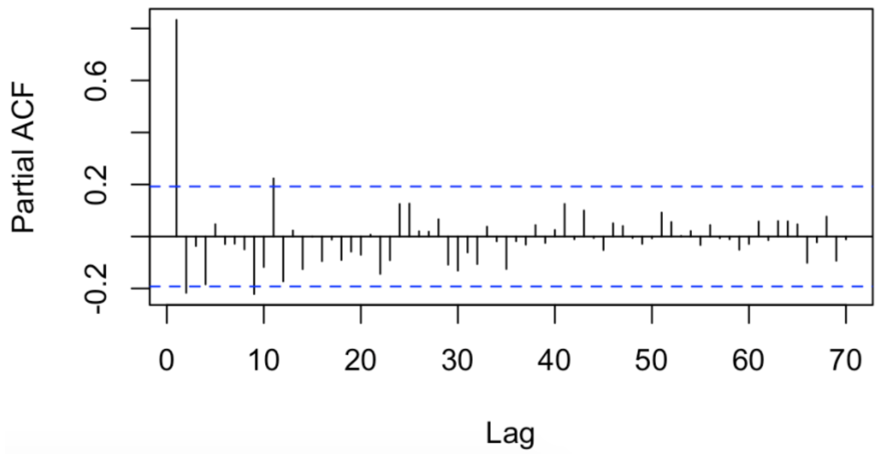



**Figure 18:** Exploring the partial correlogram plot of log(lynx) with period = 10

The partial correlogram of the residuals of the preliminary fitted model as observed from Figure 19 signifies that including a seasonal AR(1) model would be appropriate.

```
# Preliminary ARIMA model according to the PACF plot with period = 10
acf(arima(log_lynx,order=c(9,0,0),seasonal=list(order=c(0,1,0),period=10))$resid,
lag.max=70)
pacf(arima(log_lynx,order=c(9,0,0),seasonal=list(order=c(0,1,0),period=10))$resid,
lag.max=70)
```

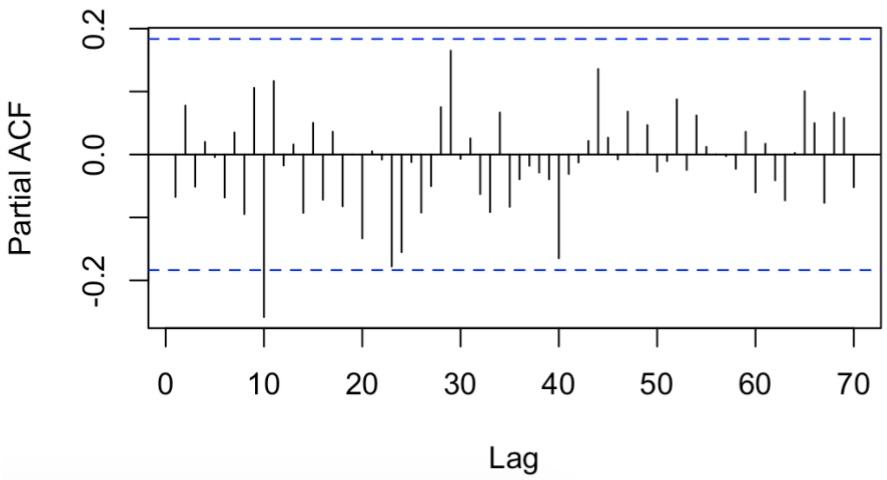



**Figure 19:** Inspecting the PACF plot of the preliminary model fitted to log(lynx)

The above fitted (9, 0, 0) (1, 1, 0)$_{10}$ model produced perfect white-noise residuals with an AIC score of 176.3 as observed from the partial-correlogram from Figure 20. However, the model looks too complex, and it would be a good idea to try adding non-seasonal MA and reduce non-seasonal AR terms.

```
# 2nd fitted model from the results of the PACF plot above (turns out to be white noise)
# However, model too complex
acf(arima(log_lynx,order=c(9,0,0),seasonal=list(order=c(1,1,0),period=10))$resid,
lag.max=70)
pacf(arima(log_lynx,order=c(9,0,0),seasonal=list(order=c(1,1,0),period=10))$resid,
lag.max=70)

> arima(log_lynx,order=c(9,0,0),seasonal=list(order=c(1,1,0),period=10))

Call:
arima(x = log_lynx, order = c(9, 0, 0), seasonal = list(order = c(1, 1, 0),
   period = 10))

Coefficients:
       ar1     ar2    ar3     ar4    ar5    ar6     ar7    ar8
     1.1680  -0.4494  0.2694  -0.3203  0.0230  0.0730  -0.0033  0.1455
s.e.  0.1023   0.1582  0.1596   0.1601  0.1641  0.1627   0.1613  0.1550
        ar9    sar1
     -0.1617  -0.4432
s.e.   0.1046   0.1022

sigma^2 estimated as 0.247:  log likelihood = -77.15,  aic = 176.3
```

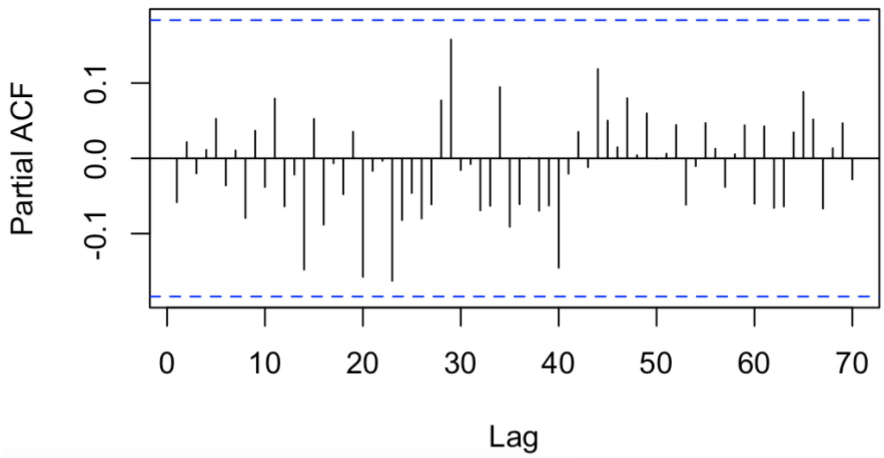


**Figure 20:** Partial-correlogram of the ARIMA (9, 0, 0) (1, 1, 0)$_{10}$ model

let us now begin fitting other models and test their AIC scores by increasing the seasonal AR part, adding non-seasonal MA part, and reducing the non-seasonal AR part.

**# Scores of other tested ARIMA white noise models**

**# Test Model 1**

```
> arima(log_lynx,order=c(8,0,1),seasonal=list(order=c(2,1,0),period=10))

Call:
arima(x = log_lynx, order = c(8, 0, 1), seasonal = list(order = c(2, 1, 0),
    period = 10))

Coefficients:
        ar1      ar2     ar3      ar4     ar5      ar6     ar7
      2.0968  -1.5808  0.7406  -0.6473  0.5187  -0.2477  0.2685
s.e.  0.1098   0.2457  0.2854   0.2827  0.2840   0.2917  0.2451
         ar8      ma1     sar1     sar2
      -0.1878  -0.9720  -0.6428  -0.2743
s.e.   0.1044   0.0666   0.1155   0.1165

sigma^2 estimated as 0.2174:  log likelihood = -72.1,  aic = 168.2
```

**# Test Model 2**

```
> arima(log_lynx,order=c(7,0,1),seasonal=list(order=c(1,1,1),period=10))

Call:
arima(x = log_lynx, order = c(7, 0, 1), seasonal = list(order = c(1, 1, 1),
    period = 10))

Coefficients:
        ar1      ar2     ar3      ar4     ar5      ar6     ar7
      1.2172  -0.5351  0.2447  -0.3534  0.1629  -0.0882  0.1387
s.e.  0.5230   0.6471  0.3064   0.1854  0.2356   0.1748  0.1136
```

```
       ma1     sar1     sma1
     -0.0088  -0.1329  -0.5691
s.e.   0.5202   0.1755   0.1682

sigma^2 estimated as 0.2321:  log likelihood = -74.82,  aic = 171.65
```

**# Test Model 3**

```
> arima(log_lynx,order=c(6,0,1),seasonal=list(order=c(2,1,1),period=10))

Call:
arima(x = log_lynx, order = c(6, 0, 1), seasonal = list(order = c(2, 1, 1),
    period = 10))

Coefficients:
        ar1      ar2     ar3      ar4     ar5     ar6      ma1
     1.5718  -0.9486  0.4049  -0.4032  0.2025  0.0326  -0.3771
s.e.  0.3274   0.4267  0.2666   0.2177  0.2215  0.1210   0.3132
        sar1     sar2     sma1
     -0.3910  -0.1631  -0.2710
s.e.   0.5588   0.3280   0.5846

sigma^2 estimated as 0.2355:  log likelihood = -75.29,  aic = 172.58
```

**# Test Model 4**

```
> arima(log_lynx,order=c(5,0,1),seasonal=list(order=c(2,1,2),period=10))

Call:
arima(x = log_lynx, order = c(5, 0, 1), seasonal = list(order = c(2, 1, 2),
    period = 10))

Coefficients:
        ar1      ar2     ar3      ar4     ar5      ma1     sar1
     1.6370  -1.0224  0.4335  -0.4361  0.2614  -0.4450  -0.8631
s.e.  0.2524   0.3380  0.2369   0.1865  0.0987   0.2429   0.4492
        sar2    sma1     sma2
     -0.2224  0.2038  -0.2775
s.e.   0.1962  0.4572   0.3303

sigma^2 estimated as 0.234:  log likelihood = -75.1,  aic = 172.21
```

After a few trails and errors, the final model which is simple and had the lowest possible AIC score was **(2, 0, 3) (1, 1, 0)**$_{\mathbf{10}}$.

**# Final ARIMA model with white-noise residuals**

```
arima(log_lynx,order=c(2,0,3),seasonal=list(order=c(1,1,0),period=10))

Call:
```

```
arima(x = log_lynx, order = c(2, 0, 3), seasonal = list(order = c(1, 1, 0),
    period = 10))

Coefficients:
         ar1      ar2      ma1     ma2     ma3     sar1
      1.1794  -0.5207  -0.0179  0.0569  0.3468  -0.4704
s.e.  0.2516   0.2141   0.2516  0.1401  0.1289   0.0938

sigma^2 estimated as 0.2537:  log likelihood = -78.39,  aic = 170.78
```

```
acf(arima(log_lynx,order=c(2,0,3),seasonal=list(order=c(1,1,0),period=10))$resid,
lag.max=70)
pacf(arima(log_lynx,order=c(2,0,3),seasonal=list(order=c(1,1,0),period=10))$resid,
lag.max=70)
```

**# Making predictions for the next 2 seasonal cycles (20 years ahead)**

```
predictions <-
round(exp(predict(arima(log_lynx,order=c(2,0,3),seasonal=list(order=c(1,1,0),period=10)),
n.ahead=20)$pred), 0)
ts.plot(df, predictions, lty = c(1,1), col=c(1,2), ylab="Lynx Trappings")
```

Simplification of the model equation for Approach 1 (2, 0, 0) $(1, 1, 0)_9$ is as follows :

$$(1+0.3086B^9-0.9513B+0.3986B^2-0.2936B^{10}+0.1230B^{11})(1-B^9)X_t=W_t$$

$$\begin{aligned}(1-B^9+0.3086B^9-0.3086B^{18}-0.9513B+0.9513B^{10}+0.3986B^2-0.3986B^{11}\\ -0.2936B^{10}+0.2936B^{19}+0.1230B^{11}-0.1230B^{20})X_t=W_t\end{aligned}$$

$$\begin{aligned}(1-0.9513B+0.3986B^2-0.6914B^9+0.6577B^{10}-0.2756B^{11}-0.3086B^{18}\\ +0.2936B^{19}-0.1230B^{20})X_t=W_t\end{aligned}$$

$$\begin{aligned}X_t=0.9513X_{t-1}-0.3986X_{t-2}+0.6914X_{t-9}-0.6577X_{t-10}+0.2756X_{t-11}\\ +0.3086X_{t-18}-0.2936X_{t-19}+0.1230X_{t-20}+W_t\end{aligned}$$

Simplification of the model equation for Approach 2 (2, 0, 3) $(1, 1, 0)_{10}$ is as follows:

$$\text{LHS} = \left(1 - 1.1794B + 0.5207B^2 + 0.4704B^{10} - 0.5548B^{11} + 0.2449B^{12}\right)\left(1 - B^{10}\right)\log x_t$$

$$\Rightarrow \left(1 - 1.1794B + 0.5207B^2 + 0.4704B^{10} - 0.5548B^{11} + 0.2449B^{12} - B^{10}\right.$$

$$\left. + 1.1794B^{11} - 0.5207B^{12} - 0.4704B^{20} + 0.5548B^{21} - 0.2449B^{22}\right)\log x_t$$

$$\Rightarrow \left(1 - 1.1794B + 0.5207B^2 - 0.5296B^{10} + 0.6246B^{11} - 0.2758B^{12}\right.$$

$$\left. - 0.4704B^{20} + 0.5548B^{21} - 0.2449B^{22}\right)\log x_t$$

$$\log x_t = 1.1794\log x_{t-1} - 0.5207\log x_{t-2} + 0.5296\log x_{t-10} - 0.6246\log x_{t-11}$$

$$+ 0.2758\log x_{t-12} + 0.4704\log x_{t-20} - 0.5548\log x_{t-21} + 0.2449\log x_{t-22}$$

$$+ w_t - 0.0179w_{t-1} + 0.0569w_{t-2} + 0.3468w_{t-3}$$