Time Series Analysis for Monthly Federal Reserve Board Production Index data from 1948 to 1978

Aliza Aziz Lakho

April 2022

1 Abstract

This report aims to conduct a time series analysis of the data of the Monthly Federal Reserve Board Production Index data from 1948 to 1978. This analysis can help us understand the underlying pattern in the data and aid in predicting future production indexes. The data used is taken from the available package "astsa" in R. For this analysis, SARIMA models were used to predict and perform spectral analysis. As a result, it was found that the production index would continue to increase in the upcoming ten months. The useful keywords for this report are Differencing of Data, ACF and PACF of data, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, SARIMA models, ARIMA models, Standardized Residuals, ACF of residuals, Normal QQ plot of standardized residuals, and p-values for Ljung-Box statistic, Spectral Analysis, Dominant Frequencies, Spectrum, and Confidence Intervals. Towards the end of the report, you will find a conclusion, some limitations, and directions for further research.

2 Introduction

We will be looking at Monthly Federal Reserve Board Production Index data from 1948 to 1978. The Federal Reserve Board of the United States publishes their Production Index monthly. It measures the gross production return on industries that affect a country's economy. These industries are manufacturing, construction, electric and gas utilities, and mining. The primary purpose of the Federal Reserve (The Fed) is to act as the central banking system of the United States. Therefore, The Fed compiles the gross production index monthly to highlight any short-term production changes. The production index reflects the growth or decline of the industry. Consequently, it depends on the increases or decreases of the index as compared to the last month. More can be found about this at https://www.federalreserve.gov/releases/g17/IpNotes.htm.

Throughout this report, we want to be able to find a model that best interprets the variation of the production index depending on the past trends. The goal is to use statistical methods to understand if the factor that contributes to the economy has trends or patterns and how closely we can predict them. For our purposes, the prediction would be based on the assumption that the future production index depicts similar trends as the past production index does. In conclusion, we want to look at how good the past trends in the production index are in predicting the future production index.

3 Statistical Methods

3.1 Exploring the data

After loading the data, it is plotted against time to observe the variation of the production index with time. We can see this plot in Figure 1. We can essentially see an upward trend with a few non-significant drops. One of the most significant drops can be seen around 1975 which can be because of the end of the Vietnamese War. After this dip, we can see the production index rise again.

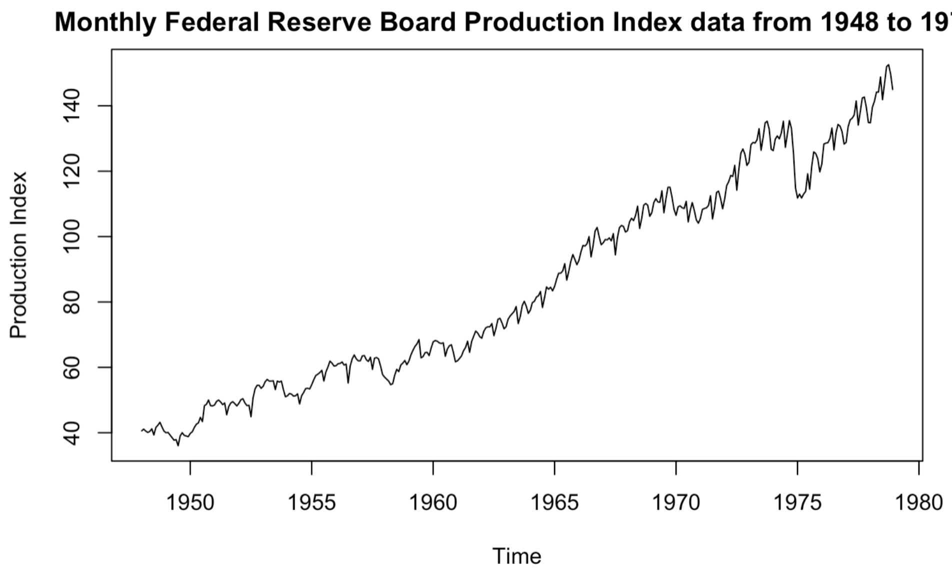

Figure 1: Plot of the data "prodn" from the astsa package

3.2 Differencing the data

This plot does not indicate a constant mean or variance. Further, observing the gradual decay in the data's ACF plot in the Figure 2 as the lag h increases proves that the process is currently not stationary and some sort of transformation or differencing is necessary. Figure 2 shows the plot for both ACF and PACF.

We proceed by differencing to make our data stationary; the plot below in the Figure 3 shows the differenced data, its ACF and PACF. We can see peaks on lags 1s, 2s, and 3s (we define s=12). We also have a gradual decay in ACF. Hence this indicates the need for seasonal difference.

Below in Figure 4 is the plot of the data that has been differenced with a twelfth-order difference due to the persistence in the months. It goes to show that the trend has been largely removed. The ACF of this differenced data, as seen in Figure 4, decays very quickly to 0. This proves that we will no longer need to transform or difference the data anymore. To confirm if the data is stationary, we perform the KPSS test in R. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test has the null hypothesis that the data is stationary. After performing this test on the differenced data, we get the p-value of 0.1. This result gives us no evidence against the null hypothesis. The results of the KPSS test, along with the month plot of the seasonally differenced data, are seen in Figure 5. The month plot further confirms

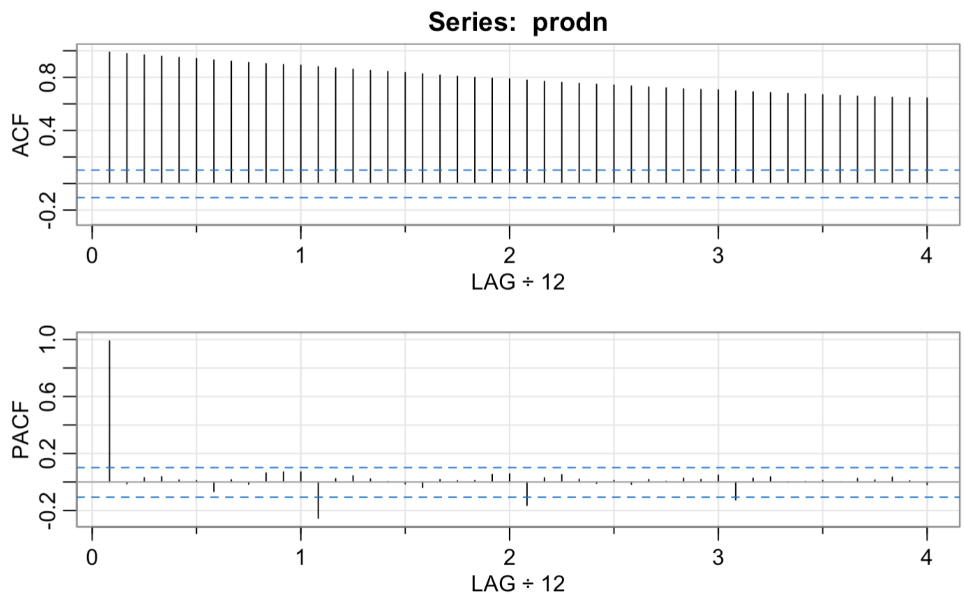

Figure 2: ACF and PACF of the series ”prodn”

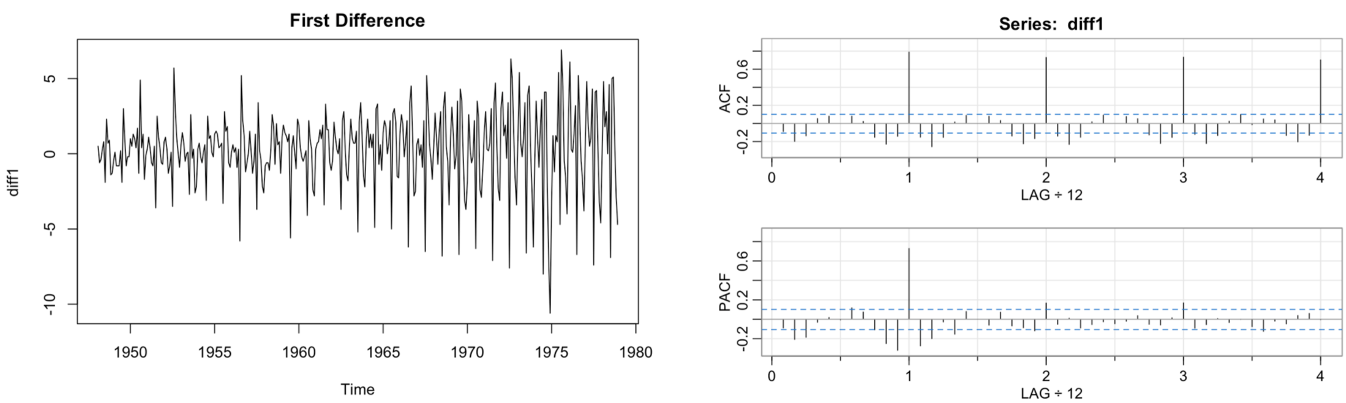

Figure 3: Plot, ACF and PACF of the First Difference

the differenced data is stationary. Therefore, our differenced data is stationary, and we can move to the next step of finding models.

3.3 Proposing SARIMA models

We choose to use a SARIMA model since our data has a seasonal aspect to it. Since we differenced our data only once, we propose D=d=2 in our SARIMA model. To find the other dependence orders, we want to look at the Seasonal and Non-Seasonal Components.

Seasonal Component

1. Both ACF and PACF seem to be tailing off with two spikes seen in PACF. This suggests SARMA with P=2, Q=1.
2. The ACF primarily cuts off after lag 1s, but the PACF tails off. This suggests SMA with P=0, Q=1.

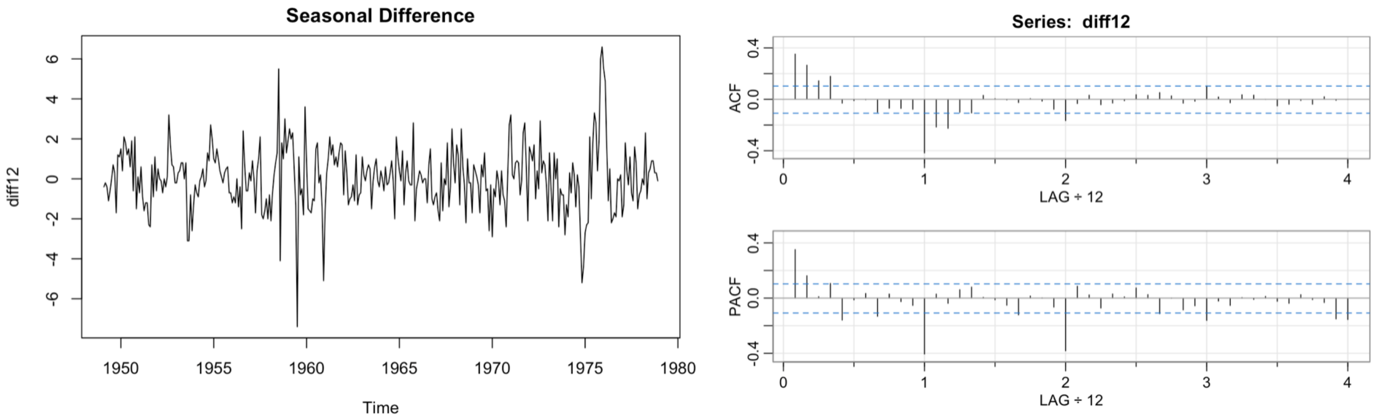

Figure 4: Plot, ACF and PACF of the seasonally differneced data

```
        KPSS Test for Level Stationarity

data:  diff12
KPSS Level = 0.018477, Truncation lag parameter = 5, p-value = 0.1
```

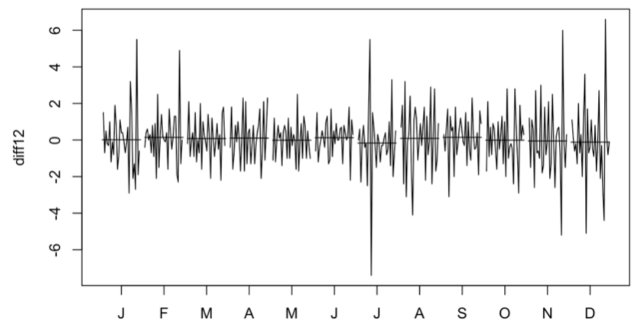

Figure 5: Results of the KPSS test and the monthplot for the seasonally differneced data

3. Similarly, ACF primarily cuts off after lag 3s, but the PACF tails off. This suggests SMA with P=0, Q=3.

Non-Seasonal Component: Here we observe how the ACF and PACF behave within seasonal lags

1. The PACF seems to be cutting off at lag 2. This suggests p=2 and q=0.
2. Both ACF and PACF seem to be tailing off. This suggests p=1 and q=1.

Therefore, our proposed models become:

1. $ARIMA(2,1,1) \times (0,1,3)_{12}$
2. $ARIMA(2,1,0) \times (0,1,1)_{12}$
3. $ARIMA(2,1,0) \times (0,1,3)_{12}$
4. $ARIMA(2,1,1) \times (2,1,1)_{12}$
5. $ARIMA(2,1,0) \times (2,1,1)_{12}$

4 Results

4.1 Model Diagnostics

The models' parameter estimates are given below in Table 1.Looking at the first model $ARIMA(2,1,1) \times (0,1,3)_{12}$, we can see that the parameters are not significant. Hence we can try to drop one parameter. So, $ARIMA(2,1,0) \times (0,1,3)_{12}$ is proposed. This model is significant so that we can continue with this model. Next, if we look at $ARIMA(2,1,0) \times (0,1,1)_{12}$, we see that the model parameters are again insignificant; hence we drop it. Finally, we get to observe the two last model proposed which are $ARIMA(2,1,1) \times (2,1,1)_{12}$ and $ARIMA(2,1,0) \times (2,1,1)_{12}$. The parameters of these models are significant, so we can continue to consider these models. Now we move toward model diagnostics. We have three finalised models $ARIMA(2,1,0) \times (0,1,3)_{12}$, $ARIMA(2,1,1) \times (2,1,1)_{12}$ and $ARIMA(2,1,0) \times (2,1,1)_{12}$.

Model	Parameter Estimates
$ARIMA(2,1,1) \times (0,1,3)_{12}$	ar1: -0.2818, ar2: 0.3039, ma1:0.5976
$ARIMA(2,1,0) \times (0,1,1)_{12}$	ar1: 0.2970, ar2: 0.1001
$ARIMA(2,1,0) \times (0,1,3)_{12}$	ar1: 0.3038, ar2: 0.1077
$ARIMA(2,1,1) \times (2,1,1)_{12}$	ar1: -0.3005, ar2: 0.3058, ma1: 0.6126
$ARIMA(2,1,0) \times (2,1,1)_{12}$	ar1: 0.2992, ar2: 0.1086

Table 1: Proposed models and their parameter estimates

Looking at $ARIMA(2,1,0) \times (0,1,3)_{12}$, the standardized residuals, ACF of residuals, Normal QQ plot of standardized residuals, and p-values for Ljung-Box statistic are shown in the Figure 6. If we look at the standardized residuals, we do not see any trend in particular. So we can say that it seems like white noise, which works in our favour. Next, looking at the ACF of residuals, we do not see any particular significance either. As for the QQ plot, the residual points seem to follow the normal distribution since they align very well with the straight line. Finally, as for the p-values for the Ljung-Box statistic, some of the values are insignificant, but since most of them are significant, it is acceptable to have this model.

Next, looking at $ARIMA(2,1,1) \times (2,1,1)_{12}$, the standardized residuals, ACF of residuals, Normal QQ plot of standardized residuals, and p-values for Ljung-Box statistic are shown in the Figure 7. Looking at the standardized residuals, we do not see any trend in particular. Therefore, it looks like white noise. Next, looking at the ACF of residuals, we do not see any particular significance either. As for the QQ plot, the residual points seem to follow the normal distribution as they align very well with the straight line. As for the p-values for the Ljung-Box statistic, some of the values are insignificant, but some are significant. Hence we would prefer the previous model to this one. The model $ARIMA(2,1,0) \times (2,1,1)_{12}$ has the same results as they can be seen in the Figure 8.

Next, we will check the AIC and BIC values to make a more informed choice. The proposed models' AIC and BIC values are posted in Table 2. After looking at this table, we have concluded that the best model is $ARIMA(2,1,0) \times (0,1,3)_{12}$ since its AIC and BIC are the least as compared to all other models.

4.2 Prediction

So, the final model is $ARIMA(2,1,0) \times (0,1,3)_{12}$. Next, we predict the next ten months of "prodn" data using the abovementioned model. The plot below in Figure 9 shows the prediction for the next

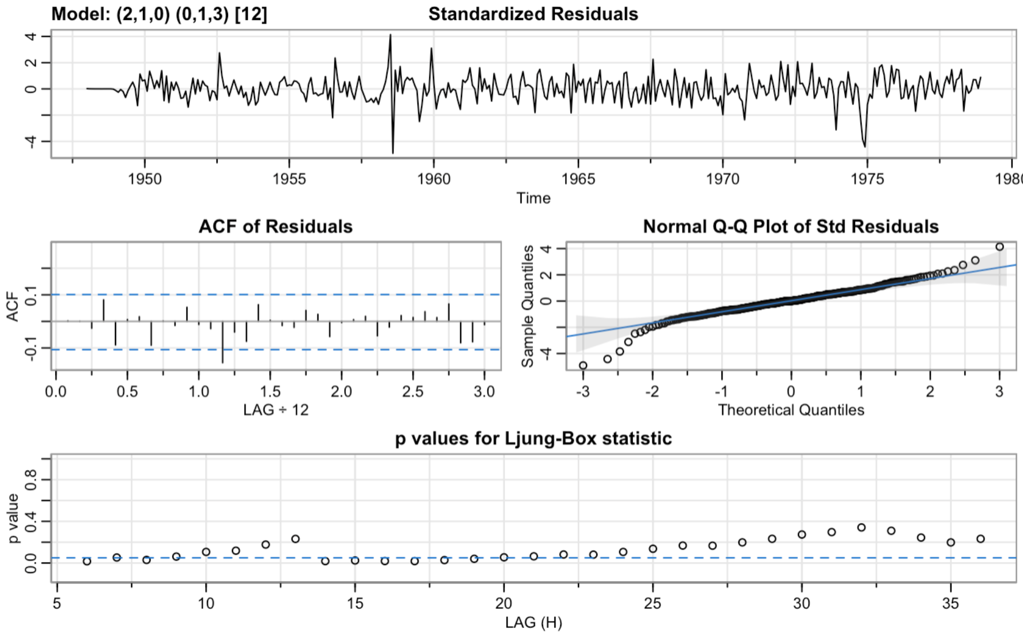

Figure 6: $ARIMA(2,1,0) \times (0,1,3)_{12}$ model diagnostics

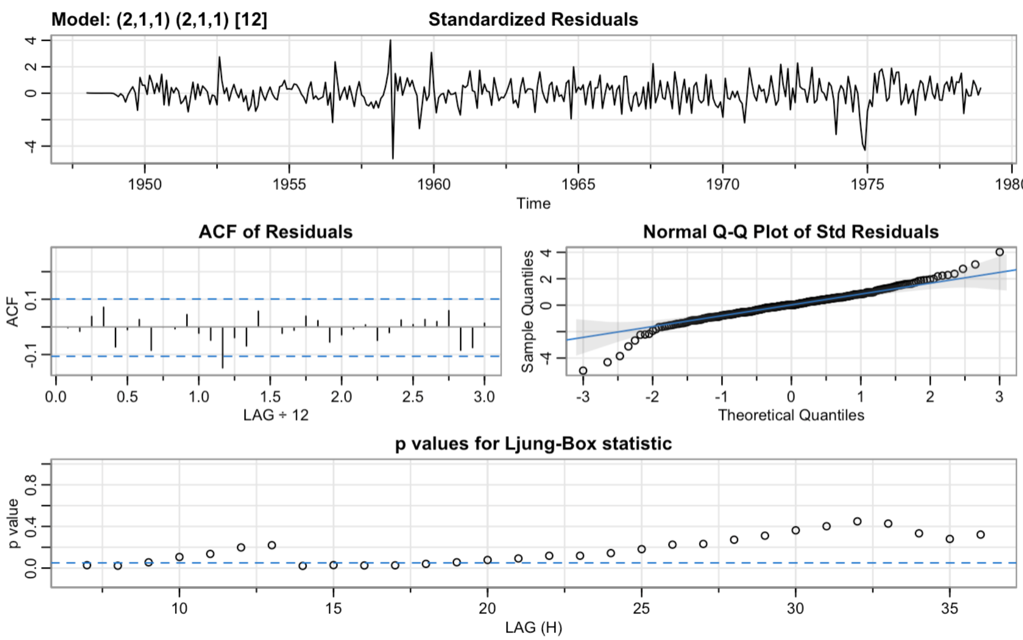

Figure 7: $ARIMA(2,1,1) \times (2,1,1)_{12}$ model diagnostics

Model	AIC	BIC
$ARIMA(2,1,1) \times (0,1,3)_{12}$	3.176308	3.252027
$ARIMA(2,1,0) \times (0,1,1)_{12}$	3.237700	3.280969
$ARIMA(2,1,0) \times (0,1,3)_{12}$	3.175390	3.240292
$ARIMA(2,1,1) \times (2,1,1)_{12}$	3.198316	3.274035
$ARIMA(2,1,0) \times (2,1,1)_{12}$	3.197700	3.262602

Table 2: Proposed models and their AICs and BICs

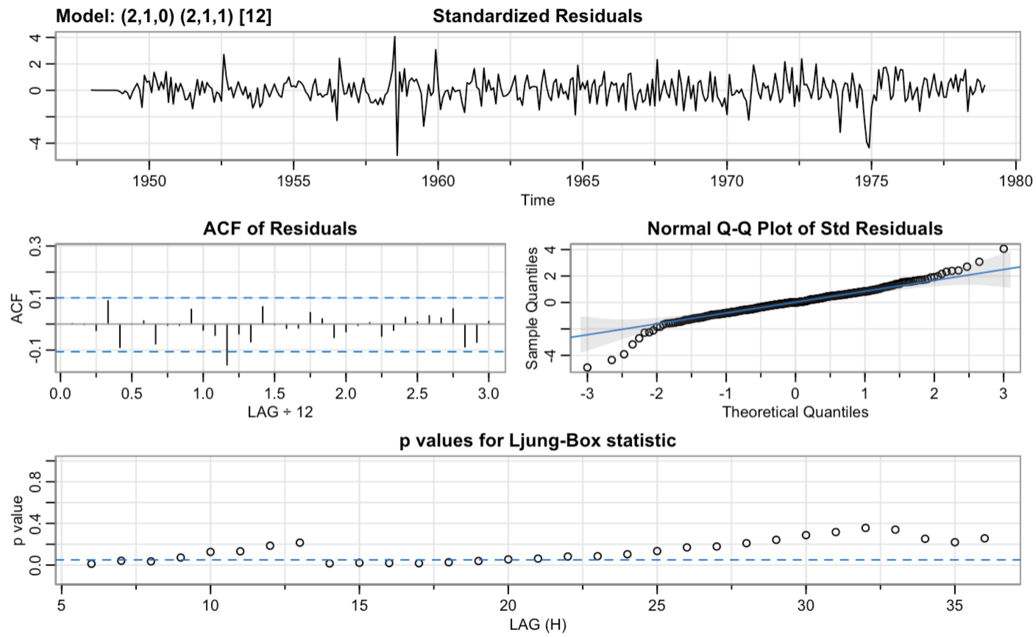

Figure 8: $ARIMA(2,1,0) \times (2,1,1)_{12}$ model diagnostics

ten months. As we can see from the plot shows an increasing trend with only one month that sees a decline. The predicted values and their 95% intervals are shown below in Table 3.

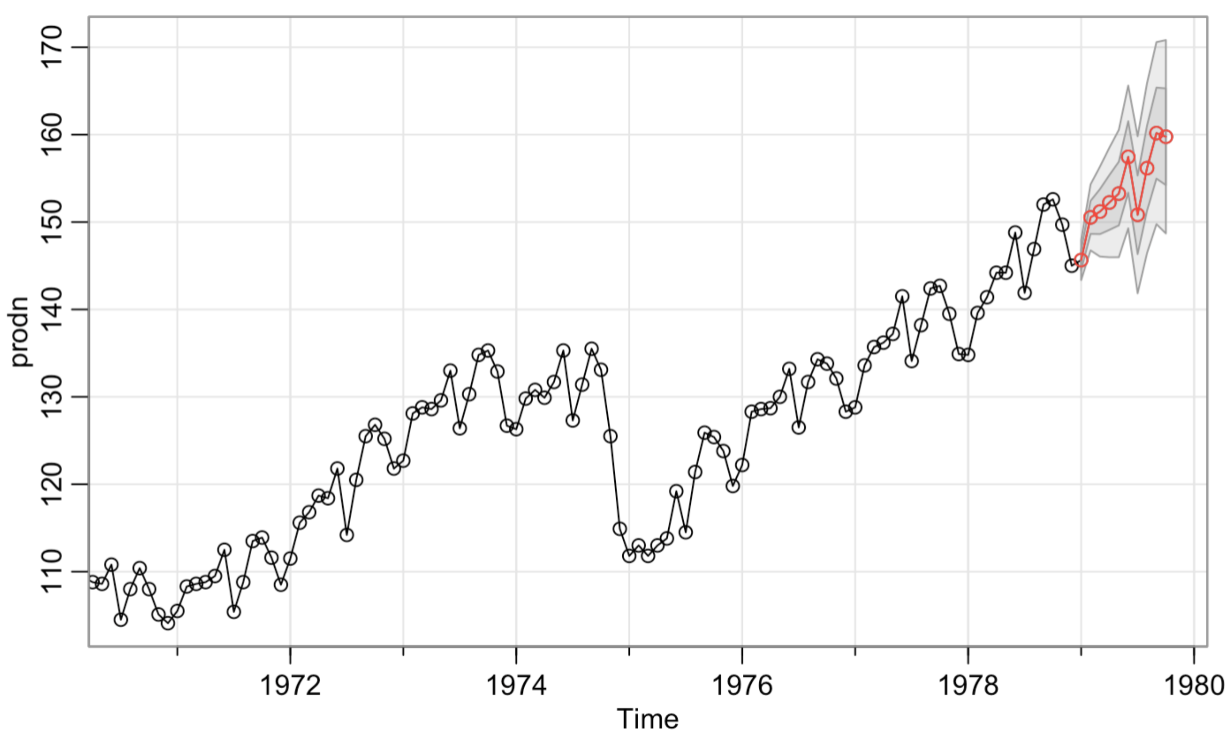

Figure 9: $ARIMA(2,1,0) \times (0,1,3)_{12}$ model forecast

Prediction	Lower Bound	Upper Bound
145.6479	143.4077	147.8881
150.5339	146.8317	154.2362
151.1982	146.1520	156.2444
152.2393	146.0934	158.3852
153.2584	146.1194	160.3973
157.4510	149.4461	165.4559
150.8310	142.0282	159.6338
156.1692	146.6406	165.6977
160.1752	149.9665	170.3838
159.7604	148.9173	170.6036

Table 3: Predicted values and their 95% intervals

4.3 Spectral analysis

After performing the spectral analysis, we identified the first three predominant periods with their 90% confidence intervals for the specified periods, shown in the table below in Table 4. By looking at this table, we can say that we cannot establish the significance of the first, second, and third peaks since the periodogram ordinate of 240.9410 lies in the confidence intervals of the second and third peaks. The periodogram ordinate of 33.49 lies in the confidence interval of the first and third peaks. Lastly, the periodogram ordinate of 23.69 lies in the confidence interval of the first and second peaks. The plot for the periodogram is shown in Figure 10.

Dominant Frequency	Spectrum	Lower Bound	Upper Bound
0.032	240.9410	80.42810	4697.3197
0.064	107.6371	35.93010	2098.4634
0.224	45.3174	15.12732	883.4956

Table 4: First three Dominant Frequencies with their Spectrums and 90% Confidence Intervals

5 Discussion

The data provided for Monthly Federal Reserve Board Production Index data from 1948 to 1978 was not stationary. Stationary data is required; hence, the trend had to be removed from the given data to make it stationary. By doing so, we ended up with three models: $ARIMA(2,1,0) \times (0,1,3)_{12}$, $ARIMA(2,1,1) \times (2,1,1)_{12}$ and $ARIMA(2,1,0) \times (2,1,1)_{12}$. The model $ARIMA(2,1,0) \times (0,1,3)_{12}$ described our data the best, and it was used to predict the production index for the next ten months. Finally, spectral analysis was performed to find the most dominant frequencies, spectrums, and 95% confidence intervals. This concludes that the production index would see an upward trend, decline briefly in July of 1979 and then grow again until October 1979.

One of the limitations experienced in this report was the lack of fit by the use of the model $ARIMA(2,1,0) \times (0,1,3)_{12}$. Even though this model was chosen after model diagnostics, AIC and BIC comparisons, and testing the significance of parameters, the QQ plot showed a lack of fit. The data points towards the ends deviated a lot from the straight line. The other limitation experienced was that the significance of the first, second, and third peaks was not established. Hence, we could not get useful information from their spectrums.

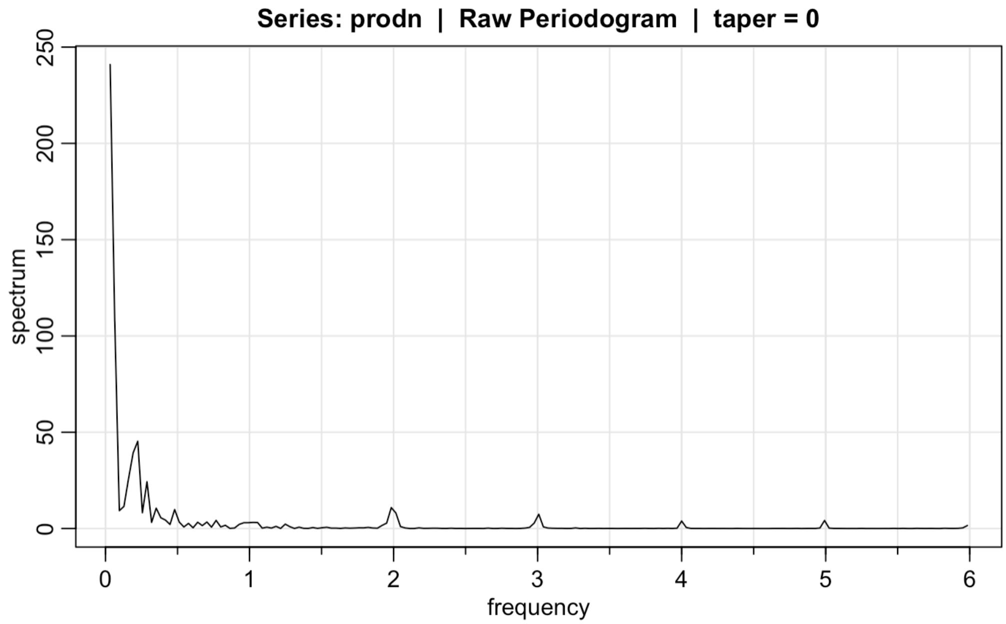

Figure 10: Spectral Analysis

For future research purposes, a good way would be to consider transformations (such as log transformations) instead of just differencing.