2024/2025
MSCI570: Forecasting and Predictive Analytics
Group Coursework

Group A
36686277, 36668150, 36894640, 36689206

Table of contents

Executive Summary 3
Introduction 4
1. Exploratory Data Analysis of the Time Series 5
  A. Cleaning the Data (Missing Values) 5
  B. Outliers and General Descriptive Statistics: 5
    B.1. Outliers 5
    B.2. General Descriptive Statistics 6
  C. Exploring the Time Series 7
    C.1. Time Series Plots 7
    C.2. Decomposition of Time Series 9
2. Normality 10
3. Irregular Components 11
  A. Visual Observations 11
  B. Statistical Tests 12
4. Trend 14
  A. Visual Observations 14
  B. Stationary Tests 15
  C. Trend Tests 16
  D. Exploration of Detrended Series 17
5. Seasonality 19
  A. Visual Observations 19
  B. Seasplot and Other Tests 21
  C. Removing the Seasonality 23
6. ACF and PACF Analysis of Time Series 24
    A. Analysis of Raw Time Series 24
    B. Analysis of Time Series after First Order Differencing 26
    C. Analysis of Time Series after Seasonal Differencing 28
Conclusion 30
List of Graphs (in main body of text) 32
List of Tables (in main body of text) 33
Appendix 34
    A.1 - Time Series Plots at Different Intervals 34
    A.2 - BIC and RSS Plots from Bai-Perron Tests 35
    A.3 - R Code for Analysis and Plots 35

Executive Summary

Our analysis identified distinct patterns across all four time series. Each time series followed an additive model with yearly, quarterly, and weekly seasonality. NN5-024 and NN5-025 display a slight upward trend, whereas NN5-036 shows a slight downward trend. NN5-042 exhibits no apparent trend. NN5-024 and NN5-025 each have a structural break, with the latter containing outliers which could require further attention.

Having applied seasonal differencing to make the time series stationary, we feel an ARIMA model would be best going forward. However, a trend-seasonal exponential smoothing model could also be considered due to the additive seasonality of all four time series. Below is a summary of our recommended models for each time series:

Table 0 - Summary Table of Proposed Model Recommendations

Time Series	ARIMA By Hand	Auto ARIMA Function	ETS
NN5-024	(0,1,2)	(1,1,2)	ETS(A,A,A)
NN5-025	(1,1,0)	(1,1,0)	ETS(A,A,A)
NN5-036	(1,1,0)	(0,0,3)	ETS(A,A,A)
NN5-042	(0,0,1)	(0,0,1)	ETS(A,N,A)

Introduction

This report analyses the daily withdrawal transactions from cash machines across the northwest region, represented by four distinct time series: NN5-024, NN5-025, NN5-036, and NN5-042. Each time series reflects transaction data for a specific machine, with unique patterns, trends and seasonal fluctuations that require careful analysis and modeling.

The primary objective of this report is to conduct a comprehensive time series analysis for these four series and recommend suitable forecasting models. Forecasting transaction patterns is crucial for optimising the management of cash levels in ATMs, ensuring sufficient funds are available without overstocking. By accurately predicting transaction volumes, banks can improve operational efficiency, reduce costs, and enhance customer satisfaction.

We begin with exploratory data analysis (EDA) to clean the data, handle missing values, and assess key statistical properties such as skewness and outliers. Additive decomposition is then applied to separate each time series into its trend, seasonal, and irregular components, providing insights into the underlying structure.

Next, we apply statistical tests including the Shapiro-Wilk test, Kolmogorov-Smirnov test, and Cox-Stuart test to verify the assumptions related to normality, structural breaks, and stationarity. The KPSS and ADF tests are conducted to assess whether the series are stationary or non-stationary. For non-normal series that exhibit trends, we conduct Cox-Stuart and Mann-Kendall tests and detrend the data to isolate randomness and seasonal patterns.

Finally, we transform the series into stationary datasets by removing trends and seasonality through first-order and seasonal differencing. Autocorrelations are then analysed using ACF and PACF plots to ensure the series are appropriately modeled before making forecasting recommendations about which models and techniques to use depending on the time series specifities.

1. Exploratory Data Analysis of the Time Series

As part of this report we will be analysing multiple series of cash money withdrawals at cash-machines with the following IDs: NN5-024, NN5-025, NN5-036, NN5-042.

A. Cleaning the Data (Missing Values)

We begin our analysis by first making sure that the data is cleaned so that it does not have any missing values. Following states the number of missing values present in each of the datasets:
- Number of Missing Values in NN5-024: **13**
- Number of Missing Values in NN5-025: **13**
- Number of Missing Values in NN5-036: **17**
- Number of Missing Values in NN5-042: **12**

We will not be directly removing the rows with the missing values in order to maintain the seasonality aspect of the datasets. Instead, we will be using linear interpolation to estimate the missing values.

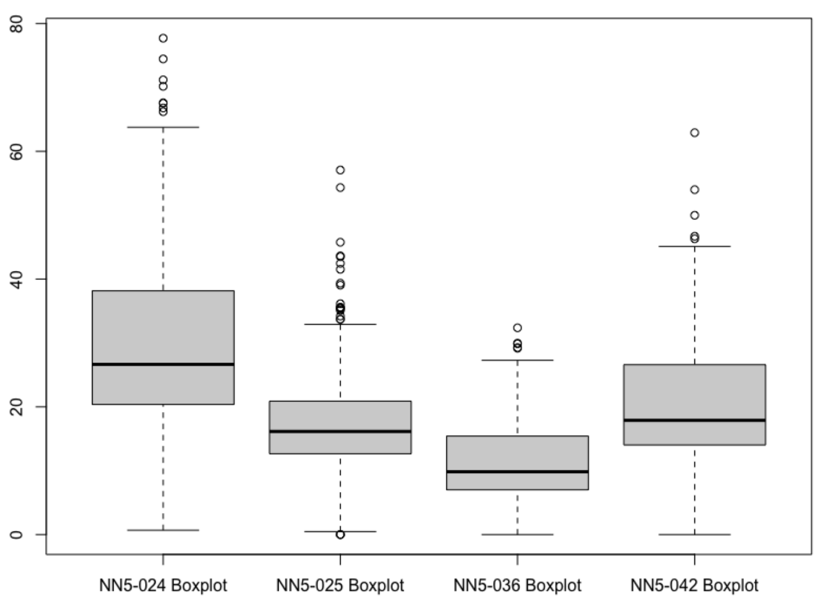
Now that we have cleaned our data, we can continue with the analysis.

B. Outliers and General Descriptive Statistics:

B.1. Outliers

The other irregular components present in the datasets are outliers. In order to understand and check for outliers, we plot boxplots.

Figure 1 - Box Plot of Time Series



From Figure 1, we can see that each of the datasets have outliers. These outliers might indicate peaks of values in the dataset.

- Number of Outliers in NN5-024: **8**
- Number of Outliers in NN5-025: **21**
- Number of Outliers in NN5-036: **5**
- Number of Outliers in NN5-042: **5**

Clearly, dataset NN5-025 has the highest number of outliers. This may be due to multiple seasonal peaks. However, for the other datasets, since their number of outliers is lesser than NN5-025, it may indicate that their seasonal peaks may not be that extreme.

B.2. General Descriptive Statistics

We shall now look at some basic descriptive statistics. For the purposes of this analysis we will be looking at mean, median, standard deviation and skewness. Following table captures the mentioned descriptive statistics for each of the datasets.

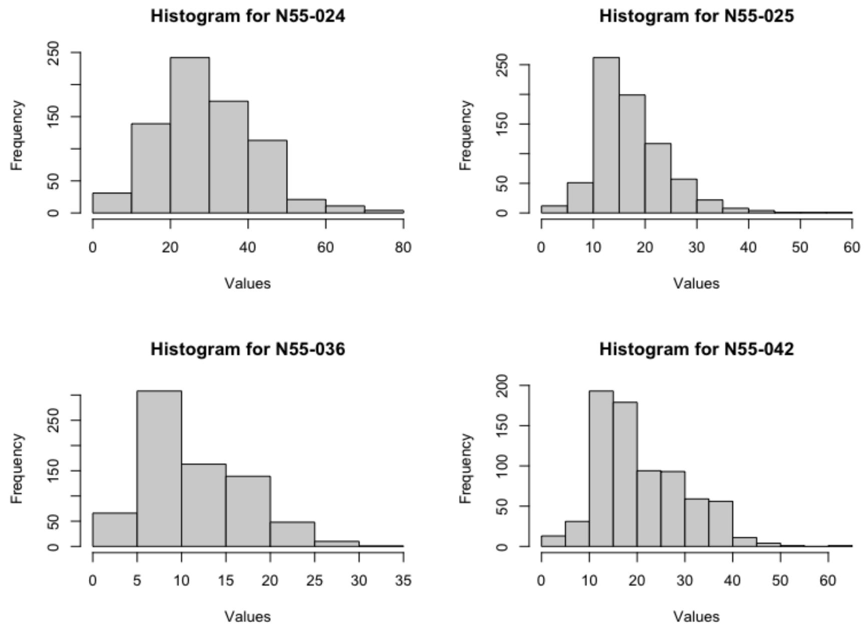
Table 1 - Summary Table for Descriptive Statistics for all Time Series

Dataset/Criteria	**NN5 - 024**	**NN5 - 025**	**NN5 - 036**	**NN5 - 042**
Minimum	0.00	0.00	0.00	0.00
Mean	29.21	17.32	11.31	20.75
Median	26.64	16.14	9.84	17.89
Standard Deviation	12.55	7.01	5.71	9.29
Maximum	77.71	57.07	32.37	62.91
Skewness	0.571	1.128	0.608	0.675

The standard deviation calculated here shows that the NN5 - 024 dataset has the highest variability most likely from the outliers present. The NN5 - 036 dataset, on the other hand, has the lowest variability indicating that the values are closer to the mean.

From the calculated statistics we can see that the mean is higher than the one in all of the datasets which indicates that there is some skewness in the datasets. The skewness values for all datasets also suggests that they are skewed and since the values are all positive, we can conclude that the skewness is positive. The values for NN5-024,NN5-036 and NN5-042 also indicate that they are not as heavily skewed as NN5-025, which has a skewness value of almost double as compared to the other datasets. To understand the skewness better and to look at the distributions of the datasets, we can also plot histograms.

Figure 2 - Histogram for all Datasets



From the graph we can clearly see that all the plots have a long right tail which further confirms the initial analysis that the datasets have positive skewness. Further, from the histogram we can see the bell-curve shape but due to the skewness we cannot conclude that the datasets follow a normal distribution.

C. Exploring the Time Series

C.1. Time Series Plots

We have now converted our given datasets into time series and plotted the graphs for them. Plots for the quarterly and weekly data can be found in Appendix A.1.

Figure 3.1 - Time Series Plot for NN5 - 024

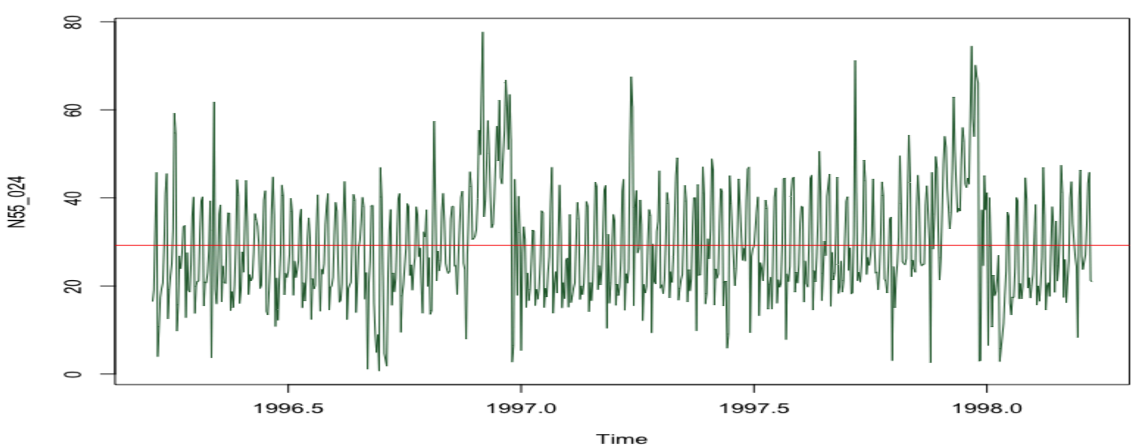


Figure 3.2 - Time Series Plot for NN5 - 025

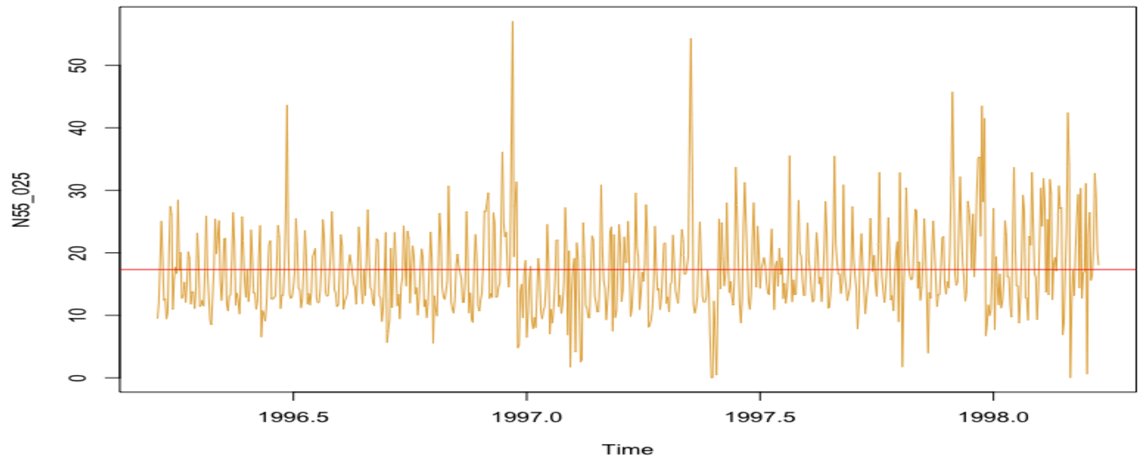


Figure 3.3 - Time Series Plot for NN5 - 036

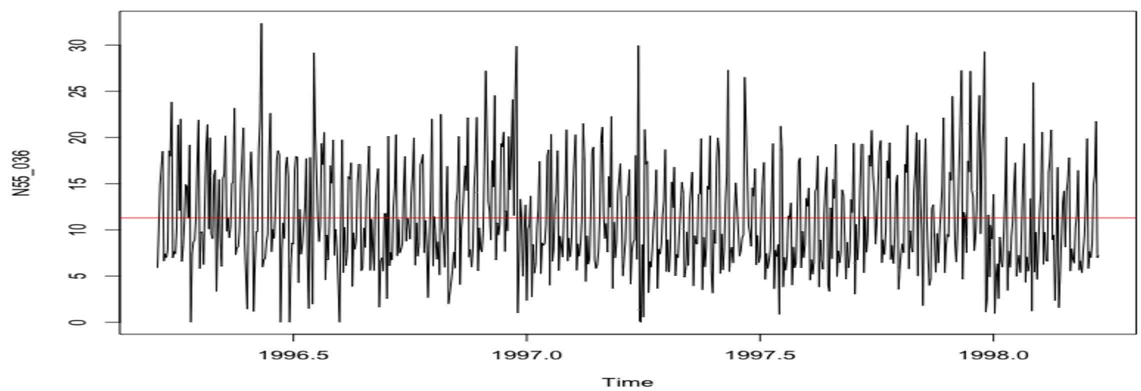
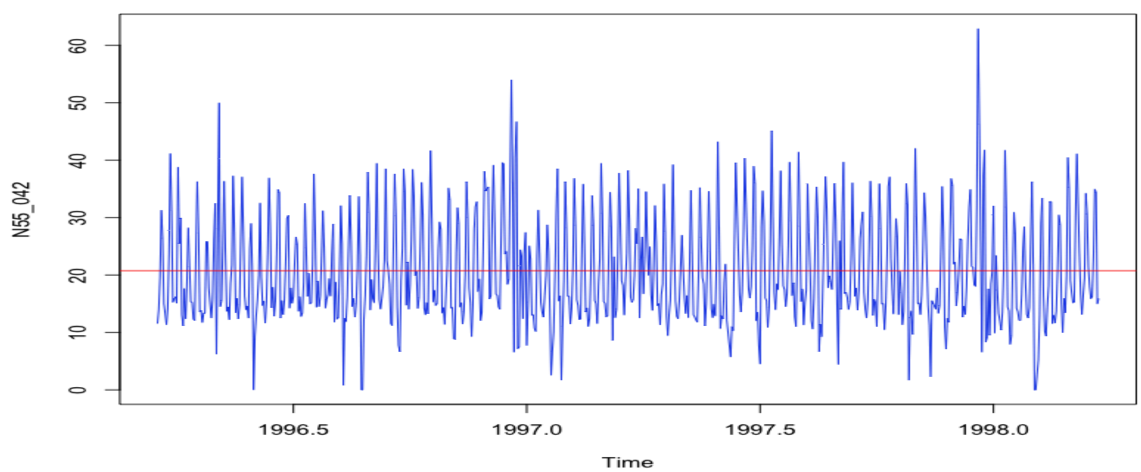


Figure 3.4 - Time Series Plot for NN5 - 042



Initial inspection of all the time series plots indicate that we do not seem to have very obvious trend lines. While visually, we can conclude that there is no trend line, we will be conducting further analysis in the later part of the report using statistical tests to better understand the presence/absence of the trend line.

For seasonality, however, from Figure 3.1 for NN5 - 024 we do not see any clear evidence of weekly or monthly seasonality but there seem to be spikes around the end of the year which may imply a yearly seasonality - likely due to external factors such as the holiday season. We can conclude that these spikes are likely to be seasonal structural breaks.

Similarly upon inspecting Figure 3.2 for NN5 - 025, Figure 3.3 for NN5 - 036 and Figure 3.4 for NN5 - 042, we are unable to see any distinguishable seasonal patterns. Thus, upon conducting visual analysis, none of the other 3 time series seem to have any seasonality. However, similar to trend, we will be conducting statistical tests and further analysis in the later part of the report to determine if the time series actually follow a seasonality.

C.2. Decomposition of Time Series

To better understand the time series and break it down into its individual components, we use time series decomposition. The seasonal fluctuations around the mean in the time series plots seem to be constant for the most part - until the spikes that appear during the end of the year for time series NN5 - 024, shown in Figure 3.1. As discussed earlier, these spikes are going to be interpreted as seasonal structural breaks and all four time series are going to be assumed to have additive seasonality. The additive decomposition plots for all 4 time series are shown below in Figures 4.1 - 4.4

Figure 4.1 - Decomposition of NN5 - 024

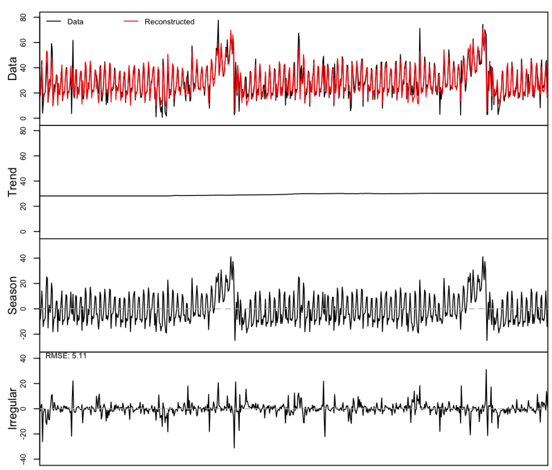


Figure 4.2 - Decomposition of NN5 - 025

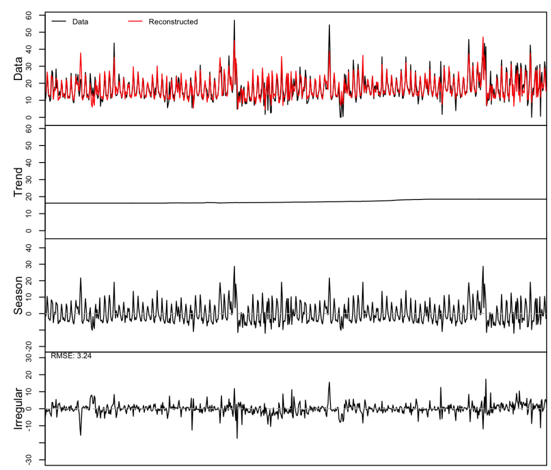


Figure 4.3 - Decomposition of NN5 - 036

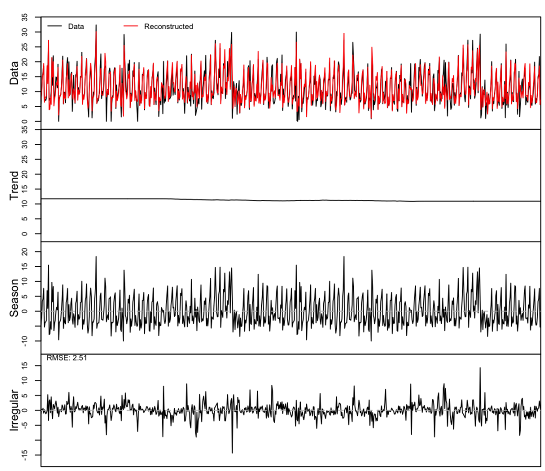
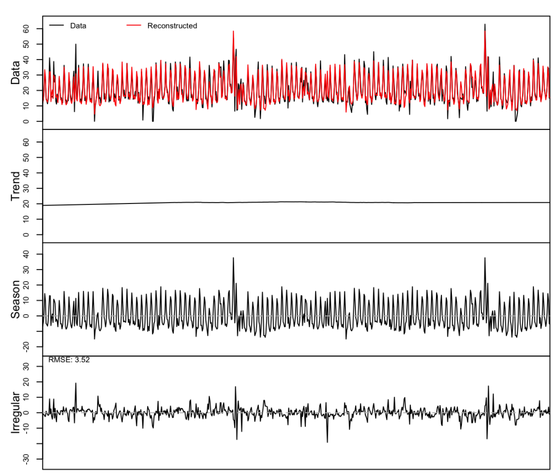


Figure 4.4 - Decomposition of NN5 - 042



The trend component seems to be consistent with our initial observations. From the graphs we can see that the trend line looks relatively stable on average, however, there are slight slopes to be observed.

The NN5 - 036 plot seems to show a slight downward sloping trend.The NN5 - 042 trend component looks very flat and likely has no trend at all. Since the trend lines for NN5 - 024, NN5 - 025 and NN5 - 036, seem to have a small slope, further investigation will be required. Time series NN5 - 024 and NN5 - 025 seem to have minimal upward sloping trend lines, although it's hard to tell for NN5 - 025 and so further investigation is required.

The observed yearly structural break in time series NN5 - 024 is captured in the seasonal component, and in fact NN5 - 025 and NN5 - 036 seem to spike around the end of the year, and all the time series seem to dip around the start of the year. Thus, we can say that all time series seem to have yearly seasonality - although it is too difficult to see any potential weekly/monthly/quarterly seasonality as of now so we will need to look into this further.

2. Normality

In order to see statistical tests we can use later, we decided to test the time series for normality. We used both the Shapiro-Wilk and Kolmogorov-Smirnov tests on the irregular component. The table below shows the p-values produced for all 4 datasets:

Table 2 - Summary Table for the Shapiro-Wilk Test for Normality

Time Series	P-Value	Conclusion
NN5-024	3.358e-06	Reject H_0: **Errors do not follow the normal distribution**
NN5-025	2.714e-10	Reject H_0: **Errors do not follow the normal distribution**
NN5-036	0.001171	Reject H_0: **Errors do not follow the normal distribution**
NN5-042	0.8426	Accept H_0: **Errors follow the normal distribution**

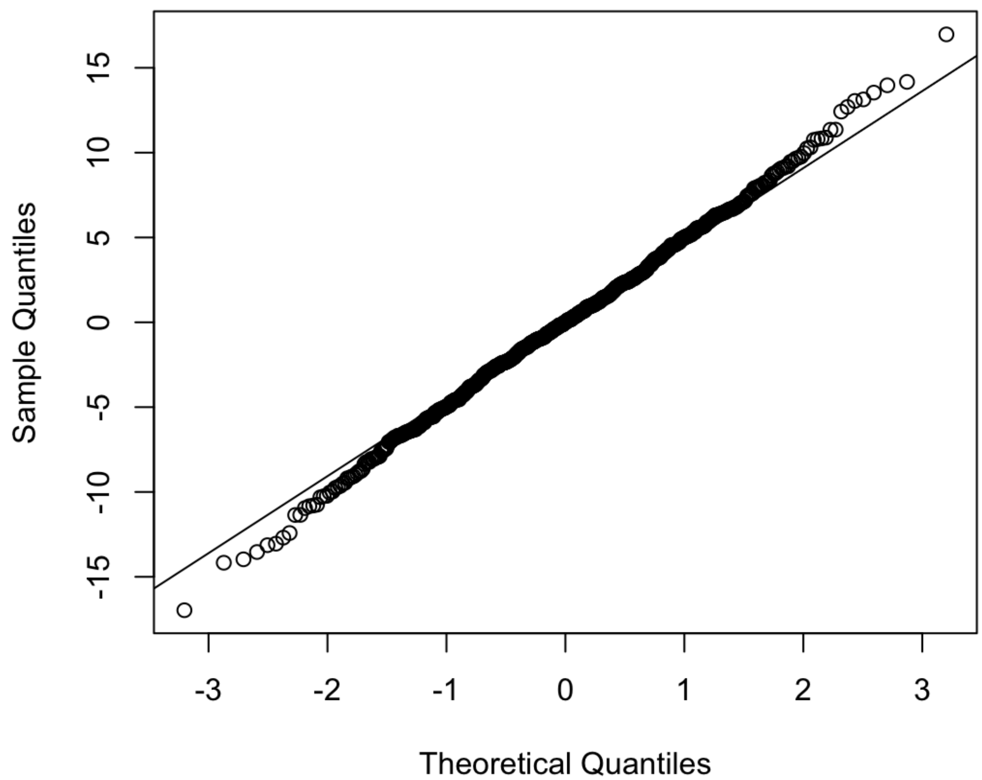
Table 3 - Summary Table for the Kolmogorov-Smirnov Test for Normality

Time Series	P-Value	Conclusion
NN5-024	$< 2.2e-16$	Reject H_0: **Errors do not follow the normal distribution**
NN5-025	$< 2.2e-16$	Reject H_0: **Errors do not follow the normal distribution**
NN5-036	$< 2.2e-16$	Reject H_0: **Errors do not follow the normal distribution**
NN5-042	$< 2.2e-16$	Reject H_0: **Errors do not follow the normal distribution**

There's clear evidence from both tests that the errors don't follow the normal distribution for time series 24, 25, and 36. 42 however, has conflicting results as we can only reject the null in the case of the Kolmogorov-Smirnov test, so further investigation was required. We therefore decided to produce a

QQ-plot of NN5 - 042's errors, shown below in Figure 8. The points mainly follow the straight line and thus our graphical interpretations, backed up by the results of the Shapiro-Wilk test, are the basis for our conclusion that NN5 - 042's errors do follow the normal distribution. So from this point, any statistical test used on the time series 24, 25, and 36 will be non-parametric and not rely on gaussian assumptions, whereas we may use parametric tests on 42.

Figure 5 - QQ Plot of NN5 - 042's Errors

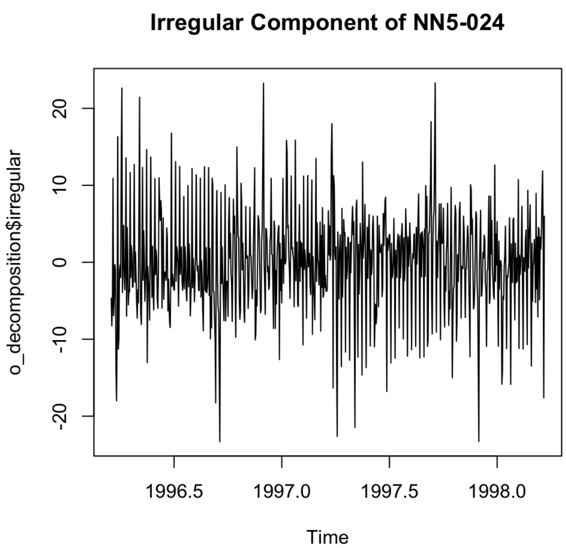


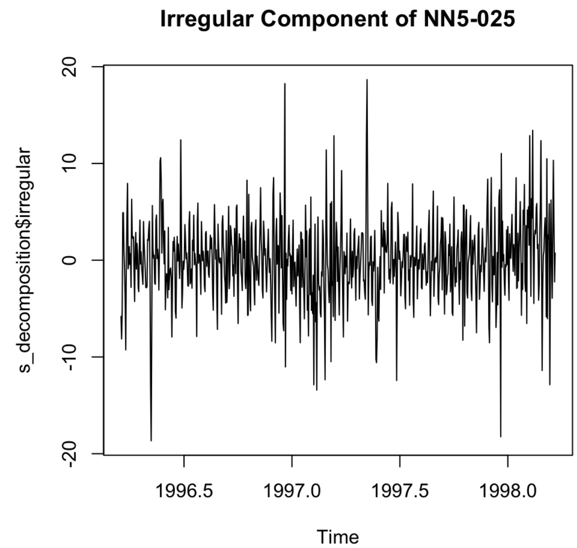
3. Irregular Components

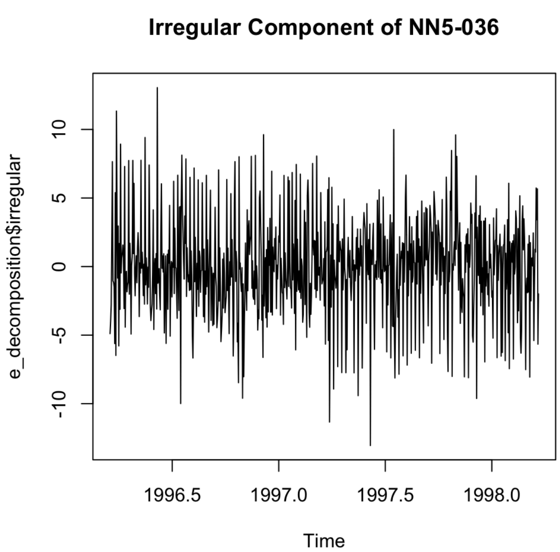
A. Visual Observations

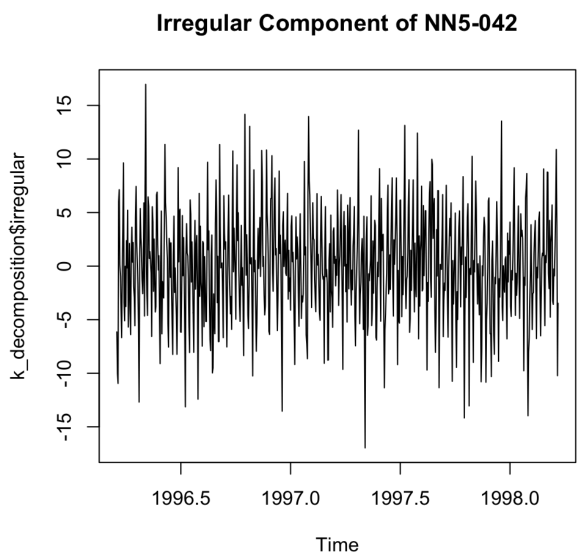
The plots for the irregular components after decomposition are shown below in Figure 6. The outliers are very noticeable, especially in NN5-025. There no longer appears to be a trend or seasonal pattern. All patterns appear to be gone, and the plots now look like white noise, with random fluctuations.

Figure 6 - Plot of Irregular Component for all Datasets









B. Statistical Tests

We used the non-parametric Cox-Stuart test for dispersion on all four time series. The p-values are summarised in the table below:

Table 4 - Summary Table for the Cox-Stuart Test on Irregular Components

Time Series	**P-Value**	**Conclusion**
NN5-024	0.2806904	Accept H_0: **No significant change in variability**
NN5-025	0.06508884	Accept H_0: **No significant change in variability**
NN5-036	0.06508884	Accept H_0: **No significant change in variability**

NN5-042	0.4537803	Accept H_0: **No significant change in variability**

For all four time series, there's insufficient evidence to reject H_0 and so the tests indicate that there's no significant change in variability in the time series. So, we have evidence to say that there's no structural change/break in the data sets. This coincides well with our previous graphical interpretations of 25, 36, and 42, as they by and large have few outliers relative to the total number of observations and do not seem to have any structural breaks. However, NN5 - 024 shows a clear annual level shift in December which is not being detected as a change in dispersion in the Cox-Stuart results. This is likely because, as we saw earlier, the pattern is being captured as part of the seasonal component in the time series' decomposition rather than a sudden change in the overall spread of the data. However, further investigation would be useful in assessing whether this pattern is a structural break, and potentially find more of them.

Table 5 - Summary Table for the supF Test on Irregular Components

Time Series	**P-Value**	**Conclusion**
NN5-024	0.0001787	Reject H_0: **Evidence of a structural break**
NN5-025	5.319e-07	Reject H_0: **Evidence of a structural break**
NN5-036	0.6445	Accept H_0: **No significant evidence of a structural break**
NN5-042	0.1303	Accept H_0: **No significant evidence of a structural break**

As previously analysed in their decomposition, each time series' seasonal component seem to exhibit some sort of spike and/or dip around the turn of the year. However, we don't know the exact date these occur and if it is seasonal then these structural breaks will be deterministic, hence we use the supF test to test for a single structural break. The p-values are shown in Table 3. Only 24 and 25 have significant evidence of a structural break. To determine the number of breaks, we used the Bai-Perron test. The summary output and plots comparing the BIC and RSS values for different breakpoints are shown in the Appendix. The number and location of breakpoints for each time series is shown in Table 4 below, such that the BIC is minimised. Again, only 24 and 25 have significant evidence of a structural break. The plots comparing the BIC and RSS for different numbers of breakpoints are in Appendix A.2.

Table 6 - Summary Table for the Bai-Perron Statistical Test for Structural Break

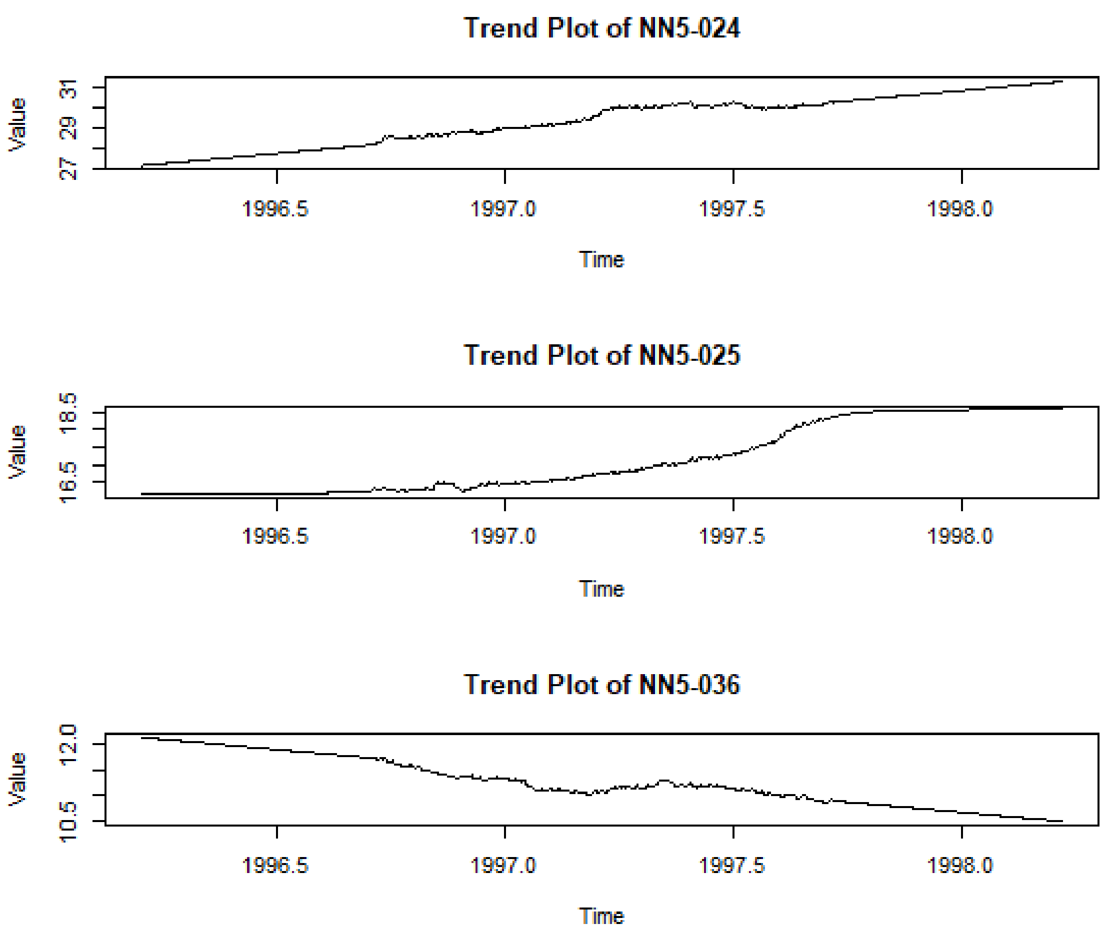
Test	**NN5 - 024**	**NN5 - 025**	**NN5 - 036**	**NN5 - 042**
Number of Breakpoints	1	1	0	0
Observation Number	219	618	N/A	N/A
Date	1996(295)	1997(329)	N/A	N/A
BIC	5810	4942	4659	5375

4. Trend

A. Visual Observations

As mentioned in Section 1.C.2, time series NN5-024, NN5-025, and NN5-036 exhibit slight slopes in their trend components, which are visualised more clearly in the trend plots shown in Figure 7. It is worth noting that the trend plot for time series NN5-042 is omitted, as Figure 4.4 already shows a flat trend with no observable trend component for this time series.

Figure 7 - Trend plot for Time Series



By visual inspection, there is a clear upward trend for time series 24 and 25, and a downward trend for time series 36. These observations align with the earlier decomposition findings.

B. Stationary Tests

To further investigate the trend (and later the seasonality) for each series, stationarity tests are conducted. First, the **Augmented Dickey-Fuller (ADF) test** is performed, and the results are summarised in the following table:

Table 7 - Summary Table of ADF Test

Time Series	P-value	Conclusion
NN5-024	< 0.01	Reject H_0: **Data is stationary**
NN5-025	< 0.01	Reject H_0: **Data is stationary**
NN5-036	< 0.01	Reject H_0: **Data is stationary**
NN5-042	< 0.01	Reject H_0: **Data is stationary**

For time series 42, the conclusion of rejecting the null hypothesis at 5% significance level aligns with the pattern observed in Figure 4.4, confirming that the series is stationary with no noticeable trend. However, for time series 24, 25 and 36, the ADF test results contradict the patterns observed in the trend plots in Figure 7. To gather additional evidence, the **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test** is conducted as a complementary stationarity test, testing the null hypothesis that the series is stationary. A summary is provided in the table below:

Table 8 - Summary Table of KPSS Test More imp

Time Series	P-value	Conclusion
NN5-024	0.037	Reject H_0: Data is likely **non-stationary**
NN5-025	< 0.01	Reject H_0: Data is likely **non-stationary**
NN5-036	> 0.1	Fail to reject H_0: Data is **stationary**
NN5-042	> 0.1	Fail to reject H_0: Data is **stationary**

Both the ADF and KPSS tests yield consistent results for time series 42, confirming that the series is stationary and does not exhibit a trend, aligning with our graphical interpretation earlier. However, although the time series doesn't seem to show a trend, the time series could still be non-stationary if it were to have seasonality and so we'll have to investigate further. On the other hand, while the ADF test suggests that time series 24 and 25 are stationary, the KPSS test indicates non-stationarity at 5% significance level. Given the reliability of the KPSS test in detecting level stationarity, as well as our earlier graphical observations, we chose to trust its conclusion. The ADF test might be less accurate in this context as it may overlook seasonality or structural changes detected in the time series, leading to a false positive for stationarity. For time series 36, the conflicting results from both tests do not align with the trend patterns observed in the graphical analysis earlier. This inconsistency calls into question the reliability of the tests in this case, prompting us to conduct further analysis using non-parametric tests.

C. Trend Tests

To assess the trend in time series 24, 25 and 36 more rigorously, we conducted the **Cox-Stuart test**. A summary of the results for each time series is provided in the following table:

Table 9 - Summary Table of Cox-Stuart Test

Time Series	P-value	Conclusion
NN5-024	0.358	Fail to reject H_0: Data **does not exhibit a trend**
NN5-025	0.007	Reject H_0: Data **exhibits a trend**
NN5-036	0.016	Reject H_0: Data **exhibits a trend**

The results indicate that for time series 25 and 36, the null hypothesis (H_0) of no trend is rejected at the 5% significance level, confirming the presence of a trend in both series. This aligns with the graphical interpretation that both series display a trend. However, for time series 24, the test suggests that there is no trend, which contradicts the earlier graphical analysis. This discrepancy could stem from the test's relatively weak power of capturing weak trends. Given the conflicting results, further investigation using the **Mann-Kendall test** is conducted as it does not require that the data be normally distributed, and also helps determine the direction of the trend for the time series. The summary of the findings is presented in the table below:

Table 10 - Summary Table of Mann-Kendall Test

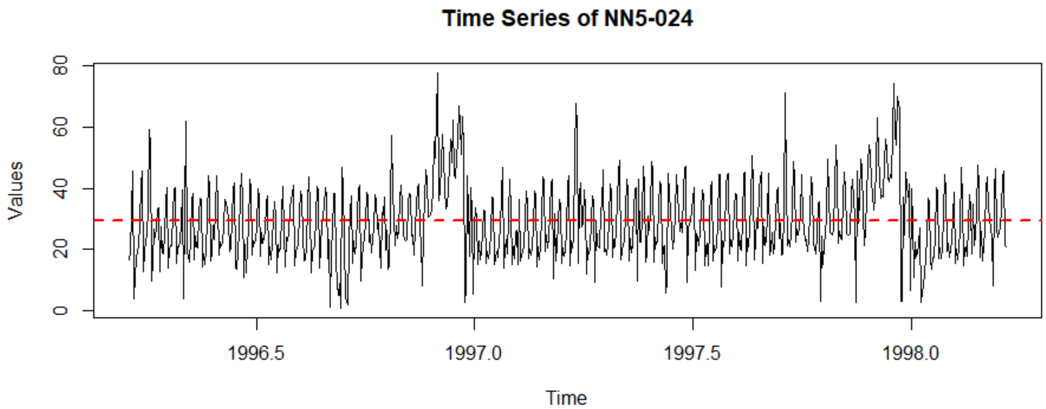
Time Series	P-value	Tau value	Conclusion
NN5-024	0.002	0.076	- Reject H_0: **Data exhibits a trend** - Positive tau: **Weak upward trend** in the data
NN5-025	5.484e-06	0.112	- Reject H_0: **Data exhibits a trend** - Positive tau: **Weak upward trend** in the data
NN5-036	0.038	-0.051	- Reject H_0: **Data exhibits a trend** - Negative tau: **Weak downward trend** in the data

The results of the Mann-Kendall test indicate the presence of weak trends in all three time series, as the null hypothesis (H_0) of no trend is rejected at the 5% significance level for each series. On top of that, the direction of these trends aligns with the graphical interpretations derived from the trend plots presented in Section 3A. Specifically, time series 24 and 25 exhibit weak upward trends, while time series 36 shows a weak downward trend.

D. Exploration of Detrended Series

For time series 42, detrending is deemed unnecessary as the series is confirmed to be stationary with no visible trend. In contrast, for time series 24, 25 and 36, the detrending process removed the upward/ downward trends, as shown in the following figures:

Figure 8 - Time Series of NN5-024 Before and After Detrending



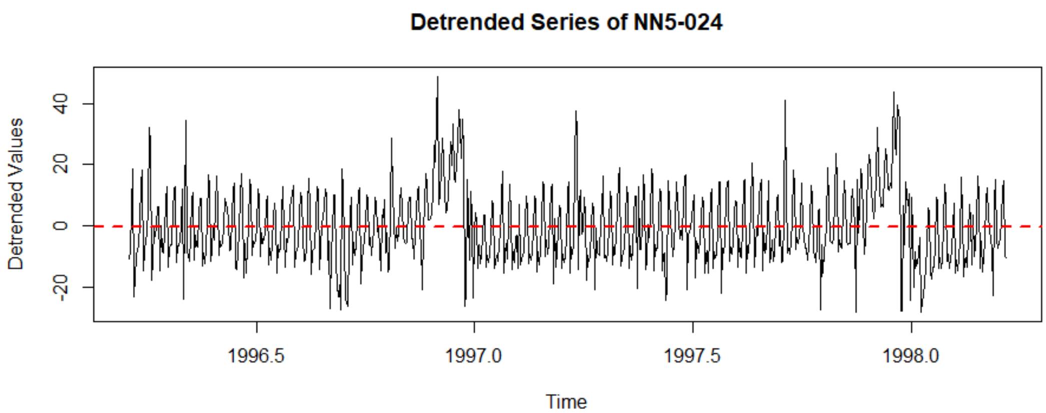
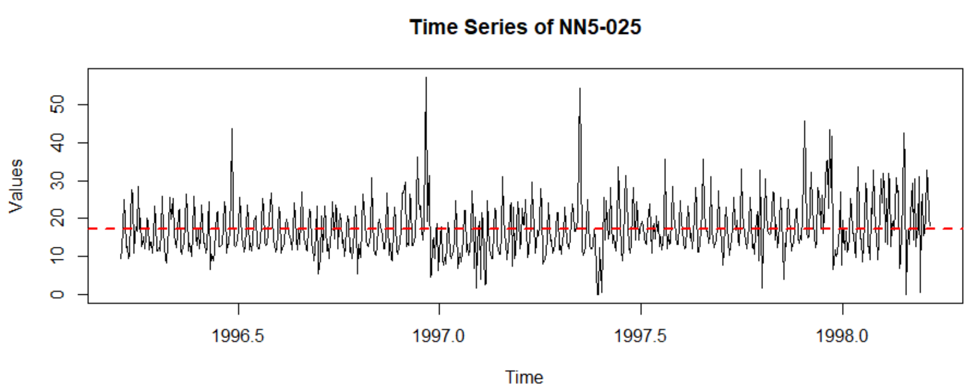


Figure 9 - Time Series of NN5-025 Before and After Detrending



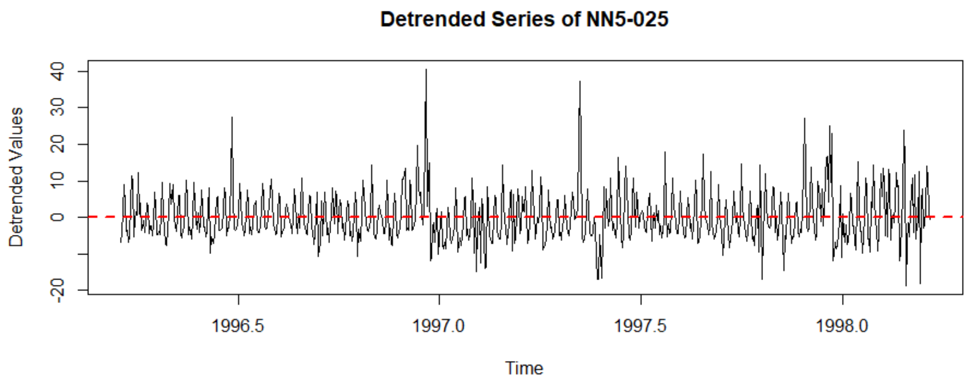
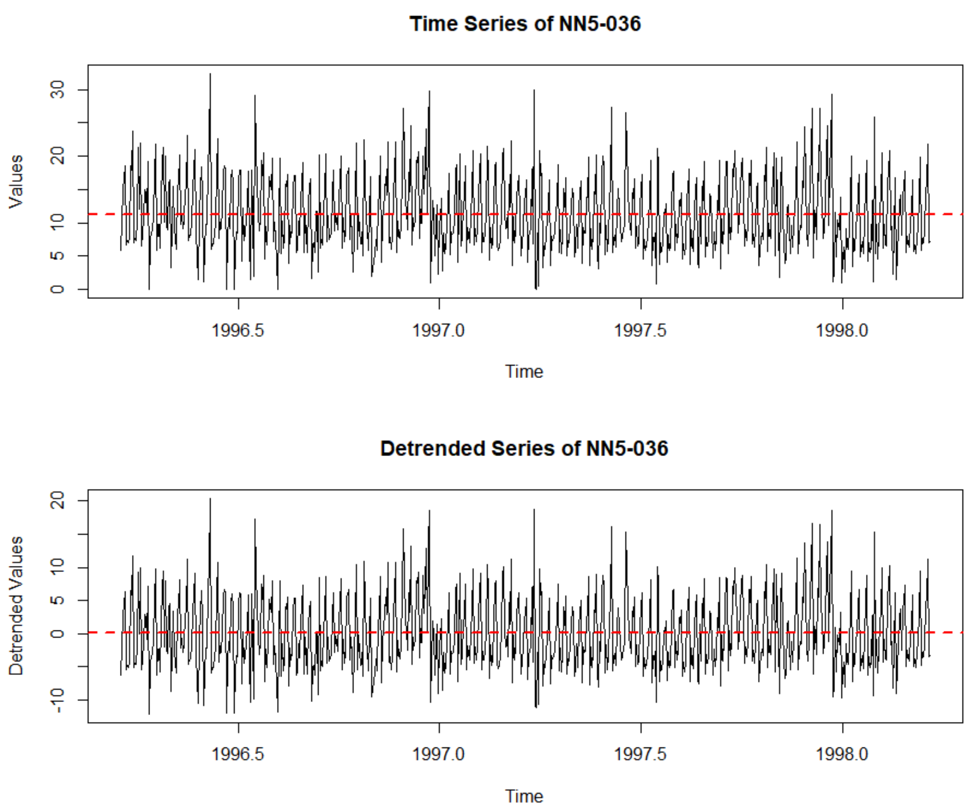


Figure 10 - Time Series of NN5-036 Before and After Detrending



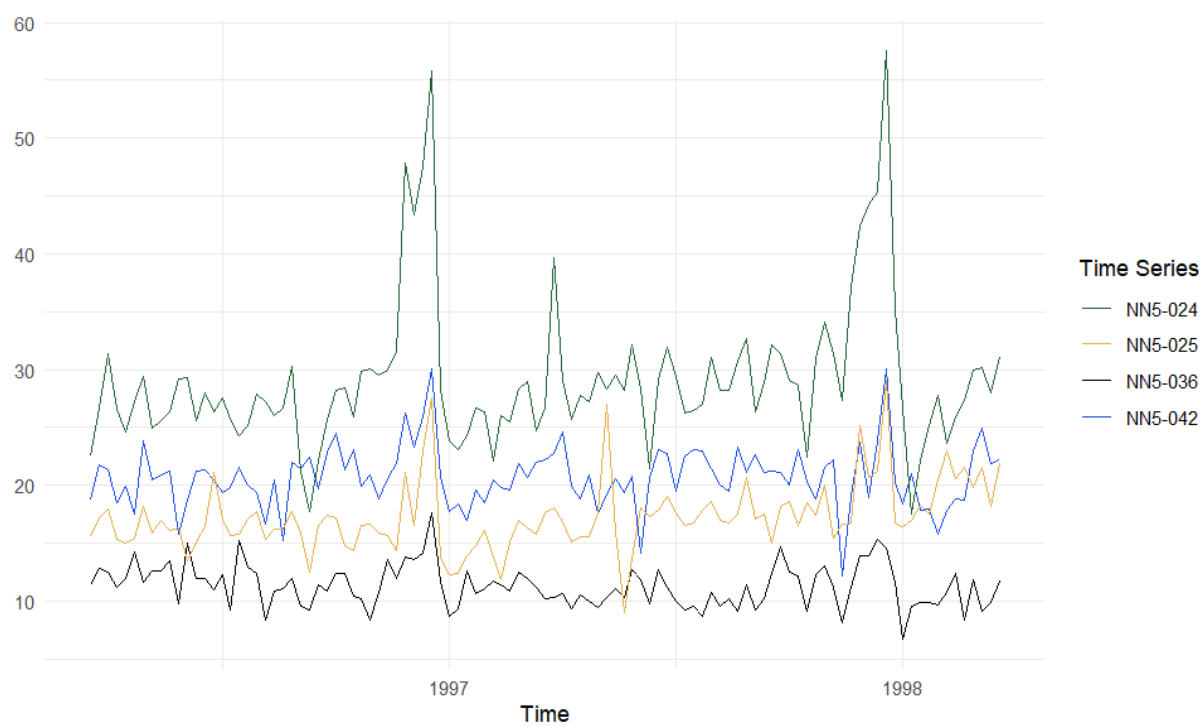
After detrending the time series, all three series fluctuate around a constant mean, making the seasonal patterns more prominent and easier to analyse. This removal of the trend allows subsequent analyses to focus on cyclical and random fluctuations, which will be explored in detail in the next section.

5. Seasonality

A. Visual Observations

Before conducting any statistical tests, we first try to identify seasonal patterns by looking at the graphs. As we stated before, it is hard to see anything on the daily plots because of the high number of points. Thus, we decide to take a look at seasonality in the weekly plots below. [Fig 11]

Figure 11 - Weekly Time series



Here we can clearly observe the effect of the winter holidays with great spikes at the end of both years (1997 and 1998) for the 4 data sets. The cash-machine associated with the time series NN5-042 seems particularly affected. Each of the 4 data sets seem to be exhibiting yearly seasonality.

As previously mentioned, the data sets NN5-025, NN5-024 and NN5-036 possess a trend that could diminish our ability to track seasonal patterns. Thus, we plotted the season part of the decomp() function which automatically isolates the desired component. [Fig 12.1-12.4]
By examining them, we can recognise the yearly seasonality mentioned earlier as well as the surge in demand during the Christmas period.

Figure 12.1 - Seasonal component for NN5-024

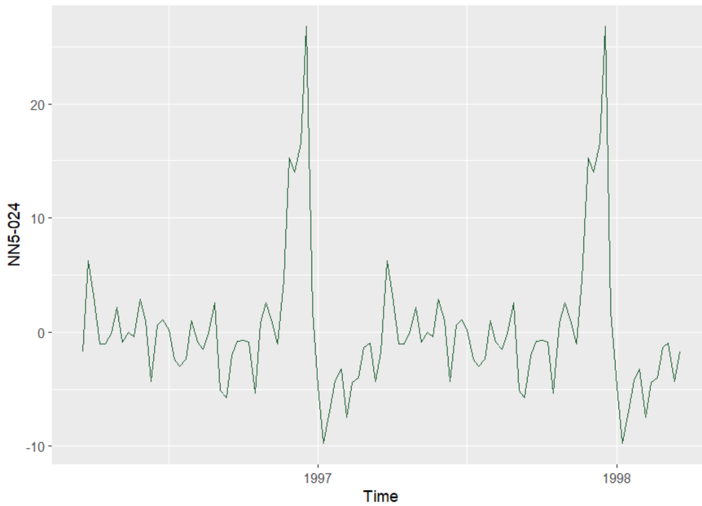


Figure 12.2 - Seasonal component for NN5-025

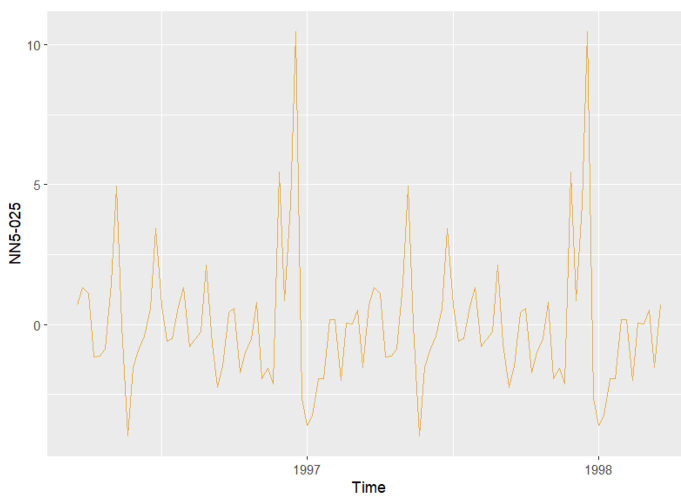


Figure 12.3 - Seasonal component for NN5-036

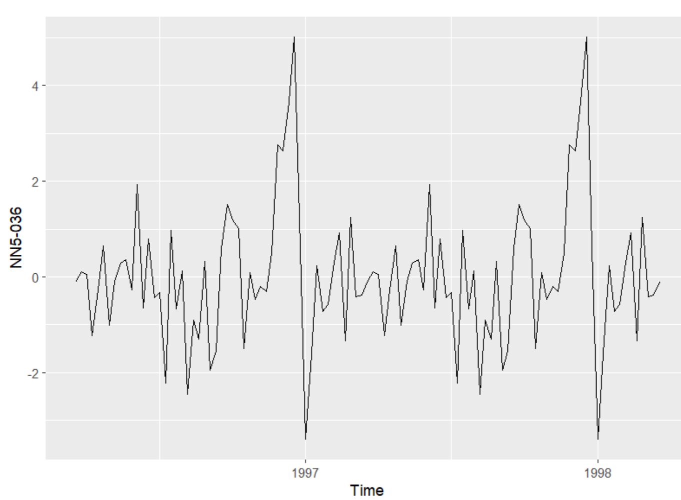
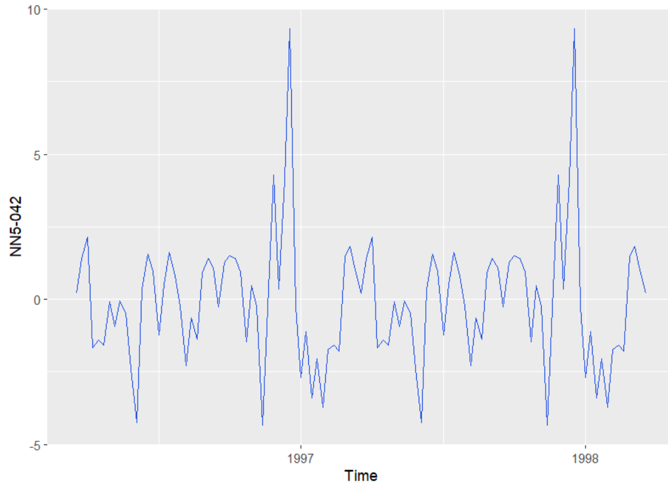


Figure 12.4 - Seasonal component for NN5-042



On NN5-042 and NN5-024 we can clearly observe a repetition every quarter of year. Similarly, a closer look at NN5-025 and NN5-036 time series brings us to the same conclusion. The 4 time series seem to exhibit quarterly seasonality.

B. Seasplot and Other Tests

Before using seasplot, we need to define which frequency is going to be tested as every year is not composed of exactly 365 days. However, overall the number of days is mostly 365 so we will define the yearly frequency to be 365. As the number of years where the number of weeks is 53 instead of 52 is less than a fifth of the years, we will set the weekly frequency at 52. We apply the same reasoning to the quarters hence quarterly frequency is 91 days per quarter.

Because each month is different and does not have the same number of days, we found it not relevant to evaluate it, thus we will not be focusing on monthly seasonality. The graph can however still be found in Appendix Figure A.1.

Using the function seasplot, we obtain a p-value of 0 on each data set for yearly, quarterly, and weekly seasonality. Therefore, we can reject the H_0 hypothesis (which is having no seasonality) and conclude that there are seasonal patterns in all of the previously cited time ranges. The test aligns with our previous assumptions on yearly and quarterly seasonality and also raises a new question on potential weekly seasonality.

To further reinforce this theory, we decided to test for seasonality using the nonparametric Kruskal-Wallis test. We are forced to use non-parametric tests as NN5-024, NN5-025 and NN5-024 do not have a normality of the residuals, as previously demonstrated. NN-042 has residuals following a normal distribution. Therefore, if the following tests are not conclusive, we could use parametric tests to conduct further experiments (e.g the F-test) on NN5-042 only.
For Kruskal-Wallis, the hypothesis H_0 is that the median values of the datasets are the same across all seasons (e.g no seasonal effect). On the table underneath [Table 11.1], we can see the p-value extracted from those tests.

Table 11.1 - Kruskal-Wallis Test on Yearly Seasonality

Yearly Time Series	P-value	Conclusion
NN5-024	5.664e-07	Reject H_0: **Data is seasonal**
NN5-025	0.0001886	Reject H_0: **Data is seasonal**
NN5-036	0.04009	Reject H_0: **Data is seasonal**
NN5-042	1.1e-07	Reject H_0: **Data is seasonal**

This test reinforces our previous observations and other testings that the four datasets exhibit yearly seasonality.

Table 11.2 - Kruskal-Wallis Test on Quarterly Seasonality

Quarterly Time Series	P-value	Conclusion
NN5-024	0.7108	Fail to reject H_0: **Data is probably not seasonal**
NN5-025	0.1341	Fail to reject H_0: **Data is probably not seasonal**
NN5-036	0.2865	Fail to reject H_0: **Data is probably not seasonal**
NN5-042	0.2474	Fail to reject H_0: **Data is probably not seasonal**

Table 11.3 - Kruskal-Wallis Test on Weekly Seasonality

Weekly Time Series	P-value	Conclusion
NN5-024	0.01124	Reject H_0: **Data is weekly seasonal**
NN5-025	0.3732	Fail to reject H_0: **Data is probably not weekly seasonal**
NN5-036	0.2111	Fail to reject H_0: **Data is probably not weekly seasonal**
NN5-042	0.06933	Fail to reject H_0: **Data is probably not weekly seasonal**

For all time series, the test does not have enough evidence to prove the presence of the quarterly and weekly seasonality previously visually observed (except for the weekly seasonality in NN5-024). Thus we decide to run further tests with the non-parametric Friedman test.

Table 12.1 - Friedman Test on Quarterly Seasonality

Quarterly Time Series	P-value	Conclusion
NN5-024	0.334	Fail to reject H_0: **Data is probably not quarterly seasonal**
NN5-025	0.8964	Fail to reject H_0: **Data is probably not quarterly seasonal**
NN5-036	0.5319	Fail to reject H_0: **Data is probably not quarterly seasonal**
NN5-042	0.9776	Fail to reject H_0: **Data is probably not quarterly seasonal**

The Friedman test does not have enough evidence to reject the null hypothesis and therefore conclude on the possibility of a quarterly seasonality across the four datasets. This could be due to the low number of points as we can only have 9 quarters in the time span of 1.5 years. In addition, by looking at the raw data, we can see that the changes are rather small. Thus, as this test is non-parametric, it could potentially lack the sensitivity to accurately pick them up.
Given our visual observations and seasplot, we choose to conclude that those 4 time series all possess quarterly seasonality.

Table 12.2 - Friedman Test on Weekly Seasonality

Weekly Time Series	P-value	Conclusion
NN5-024	0.0001936	Reject H_0: **Data is weekly seasonal**
NN5-025	0.02919	Reject H_0: **Data is weekly seasonal**
NN5-036	0.05163	Fail to reject H_0: **Data is probably not weekly seasonal**
NN5-042	0.001842	Reject H_0: **Data is weekly seasonal**

There is enough evidence to prove that the data is weekly seasonal for NN5-024, NN5-025 and NN5-042 as stated by both seasplot and Friedman test as well as our visual diagnostic.
Accordingly, we can say that the previous contradictory values obtained with the Kruskal-Wallis test for weekly seasonality were probably due to the test being too low in sensitivity to detect a mean variation. Moreover, the fact that it is unable to take into account the variance and skewness could be why it only detected the weekly seasonality in NN5-024 as it is the least skewed time series of the four (around 0.57).

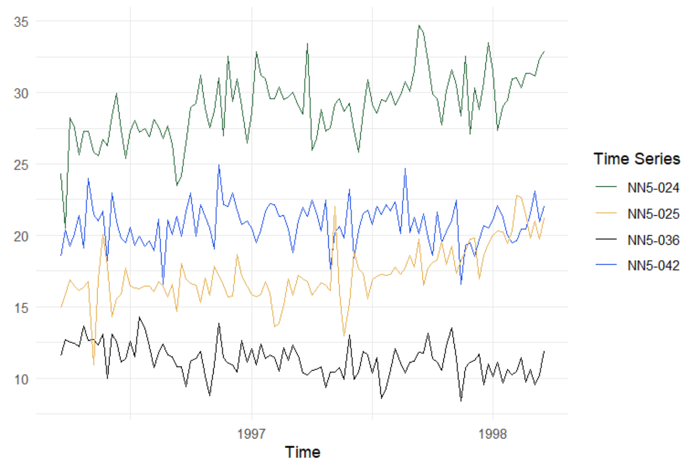
Upon calculating the weekly standard deviation for all time series, we can observe that NN5-036 has the smallest with 1.8 compared to 2.7, 6.6 and 3.1 for NN5-042, NN5-024 and NN5-025 respectively. This could be why the weekly seasonality in NN5-036 was not detected by the Friedman test.

Overall, we have determined that all time series possess a yearly, quarterly and weekly seasonality.

C. Removing the Seasonality

Now that we have analysed the seasonality, we are going to remove it to look at the remaining graphs.

Figure 13 - Time series without the Seasonal Component



As expected, we can notice the trend more clearly without the seasonal pattern. The trends have the same tendencies as described above. Other than that, none of the time series exhibit obvious patterns apart from those already mentioned. However, there is a possibility that some cycles cannot be seen as the data span is too short.

6. ACF and PACF Analysis of Time Series

A. Analysis of Raw Time Series

We begin the ACF and PACF analysis of the time series by first plotting ACF diagrams and PACF diagrams for each of the raw time series. Following are the ACF and PACF graphs of all four raw time series. We show the PACF graph only for the first 91 days so that it is easier to see the patterns.

Figure 14.1 - ACF and PACF Plots for raw NN5 - 024

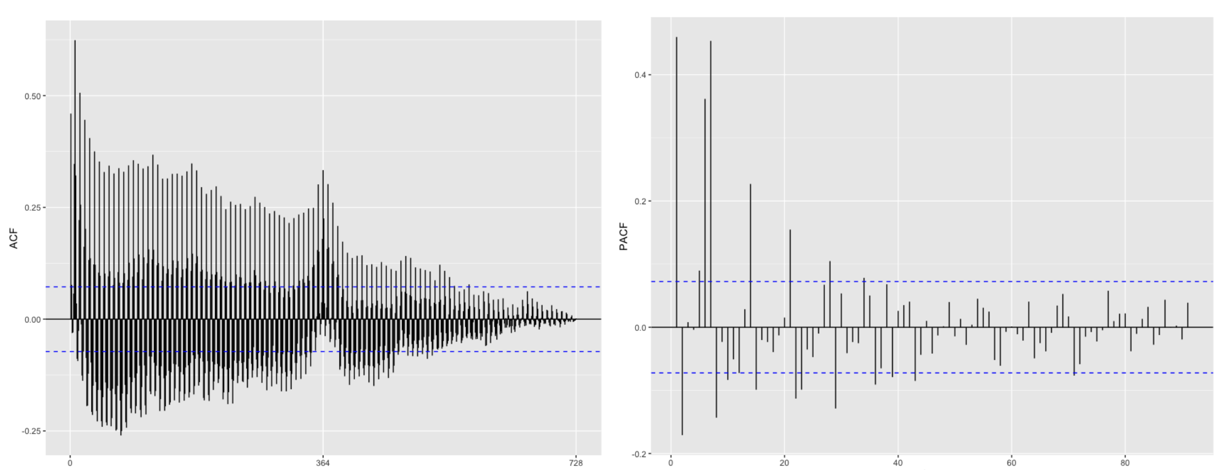


Figure 14.2 - ACF and PACF Plots for raw NN5 - 025

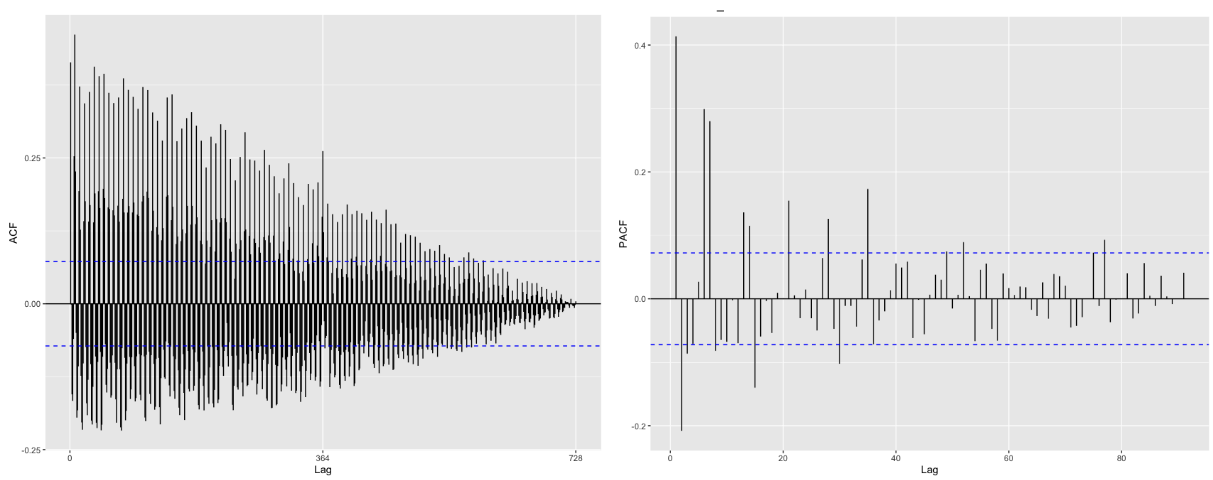


Figure 14.3 - ACF and PACF Plots for raw NN5 - 036

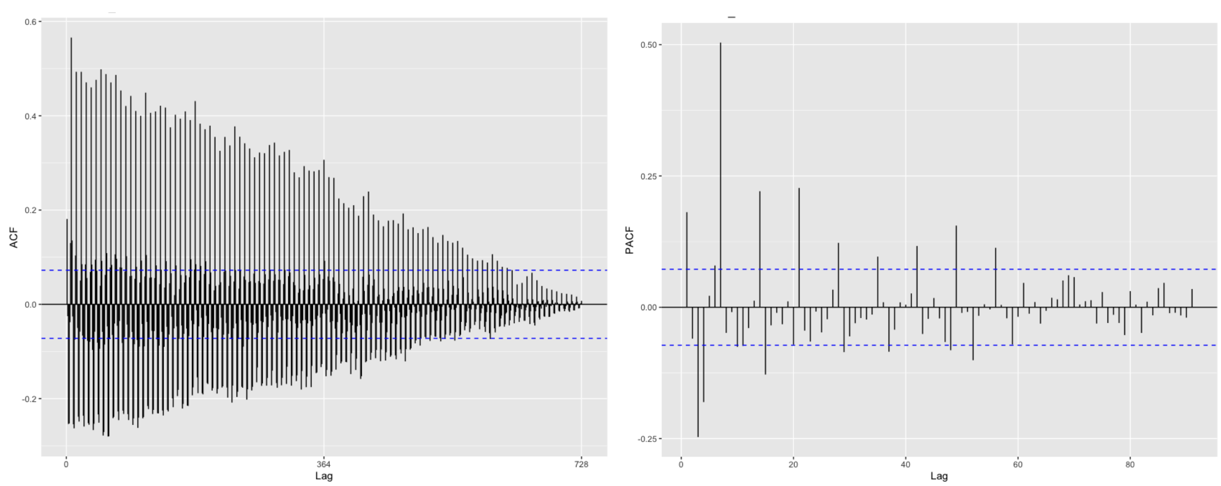
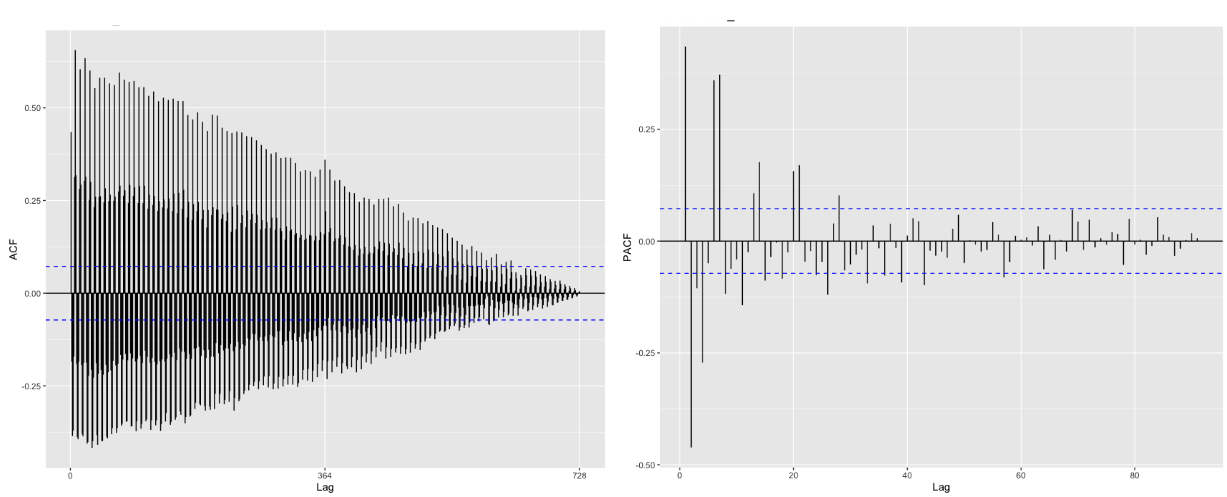


Figure 14.4 - ACF and PACF Plots for raw NN5 - 042



From these ACF graphs we can see that Figure 14.1, Figure 14.2 and Figure 14.3 have a gradual decay which indicates that the series is non-stationary and a possible trend line exists. However, for Figure 14.4, we see that the values largely fit within the confidence interval and rarely go beyond it, hence indicating that the time series does not follow a trend. This visual analysis of ACF graphs for raw time series data is consistent with the visual analysis on decomposition of time series conducted at the beginning of the report. In addition to visual analysis, these results are also consistent with the statistical tests conducted earlier on in the report which clearly indicated that NN5 - 024, NN5 - 025 and NN5 - 036 do follow a slight trend whereas NN5 - 042 does not.

The conclusions from the PACF graphs for NN5 - 024 and NN5 - 025 are also consistent with the conclusions from the ACF graphs. However, for NN5 - 036 we do not see a significant first lag in the PACF which indicates no trend line. This is contradicting our initial visual analysis, statistical analysis and ACF analysis but since ACF is a more direct indicator of trend we follow our initial conclusion of non-stationarity for NN5 - 036 time series.

Similar to the NN5 - 036 time series, the NN5 - 042 time series also shows inconsistencies between PACF analysis and initial visual analysis, statistical tests and ACF analysis. Using the same logic of ACF being a more direct indicator of trend, we follow our conclusion of stationarity in terms of trend for NN5 - 042 time series.

Since we can conclude that there is non-stationarity due to a trend line in NN5 - 024, NN5 - 025 and NN5 - 036 time series, we perform **First Order Differencing** on these time series to get rid of the trend.

B. Analysis of Time Series after First Order Differencing

We conduct first order differencing on the time series NN5 - 024, NN5 - 025 and NN5 - 036 and following as the ACF and PACF plots after differencing. Similar to the previous section, we will be seeing an interval of 91 days on our PACF graphs for clarity purposes.

Figure 15.1 ACF and PACF Plots after First Order Differencing for NN5 - 024 Series

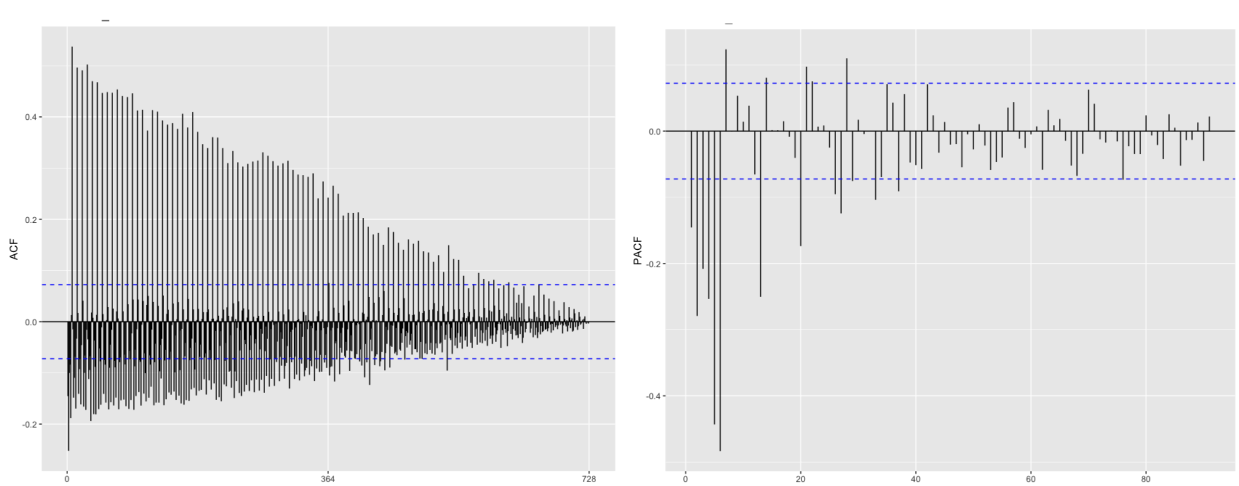


Figure 15.2 ACF and PACF Plots after First Order Differencing for NN5 - 025 Series

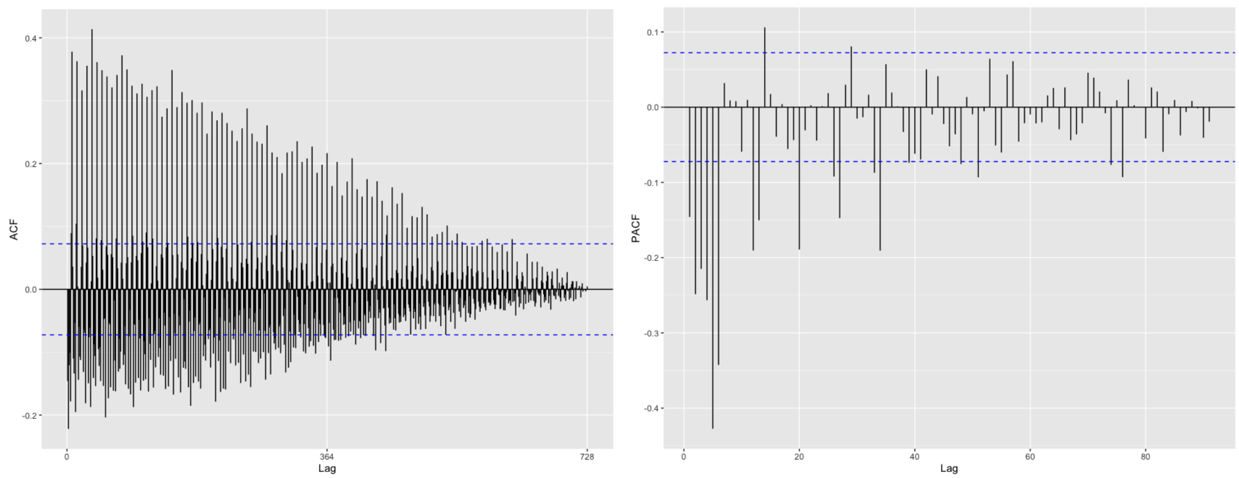
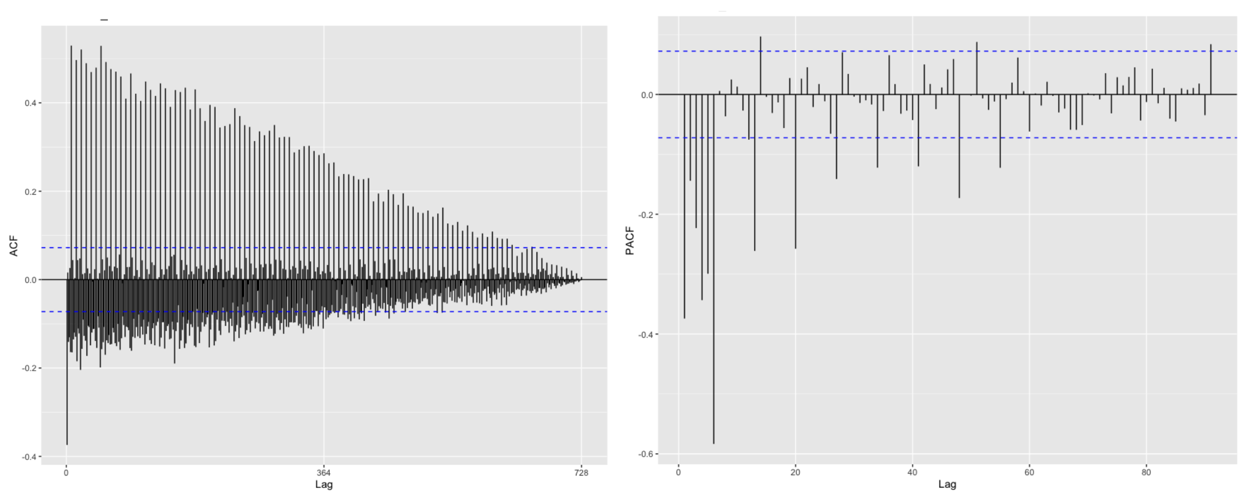


Figure 15.3 ACF and PACF Plots after First Order Differencing for NN5 - 036 Series



After conducting first order differencing, we can see from the ACF graphs that the 3 time series are much more concentrated within the confidence interval which indicates that there is now stationarity from trend in these time series. To further support our claim, we run the ADF and KPSS test of stationarity. The following table shows the conclusions of the statistical tests:

Table 13 - Summary Table for Stationarity Tests

Time Series	**ADF Test (P-Value)**	**KPSS Test (P-Value)**	**Conclusion**
NN5-024	<0.01	>0.1	ADF Test - Reject H_0: **Data is Stationary** KPSS Test - Fail to reject H_0: **Data is Stationary**
NN5-025	<0.01	>0.1	ADF Test - Reject H_0: **Data is Stationary** KPSS Test - Fail to reject H_0: **Data is Stationary**
NN5-036	<0.01	>0.1	ADF Test - Reject H_0: **Data is Stationary** KPSS Test - Fail to reject H_0: **Data is Stationary**

Thereby, based on analysis of the ACF graph and ADF and KPSS statistical tests after first order differencing, we conclude that all the 3 time series are now stationary for trend as well

On the other hand, when we have a look at the PACF graphs for NN5 - 024, NN5 - 025 and NN5 - 036 after first order differencing and PACF graph for NN5 - 042 raw time series, we can see that every 7th peak is significant. This indicates that there is weekly seasonality in the time series and thus we need to do seasonal differentiation to get rid of it and ensure complete stationarity from trend and seasonality.

C. Analysis of Time Series after Seasonal Differencing

We conduct seasonal differencing on the time series first order differentiated time series of NN5 - 024, NN5 - 025 and NN5 - 036 and raw time series of NN5 - 042 to remove the weekly seasonality in the time series. Similar to the previous section, we will be seeing an interval of 91 days on our PACF graphs for clarity purposes.

Figure 16.1 ACF and PACF Plots after Seasonal Differencing for NN5 - 024 Series

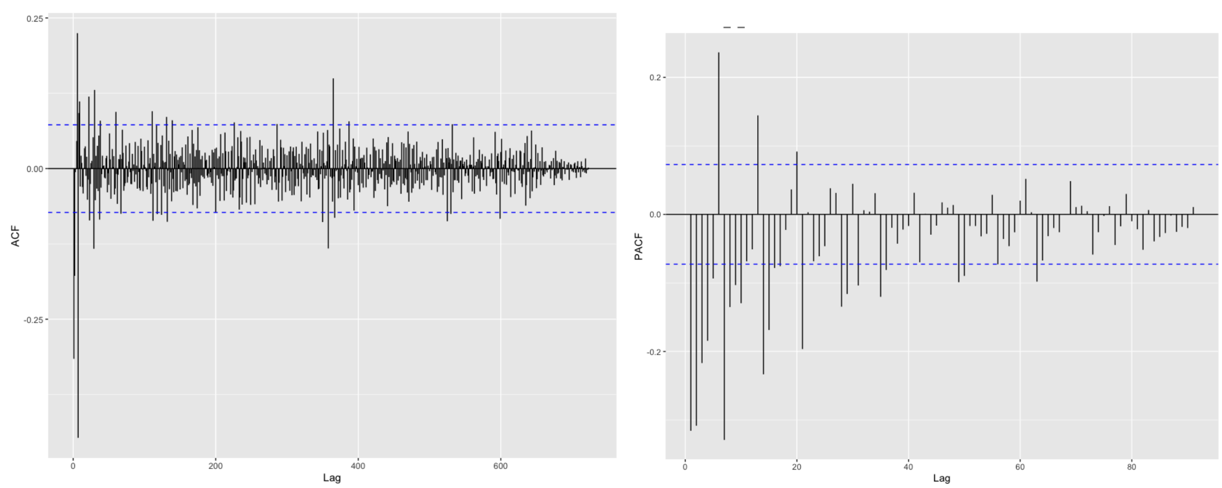


Figure 16.2 ACF and PACF Plots after Seasonal Differencing for NN5 - 025 Series

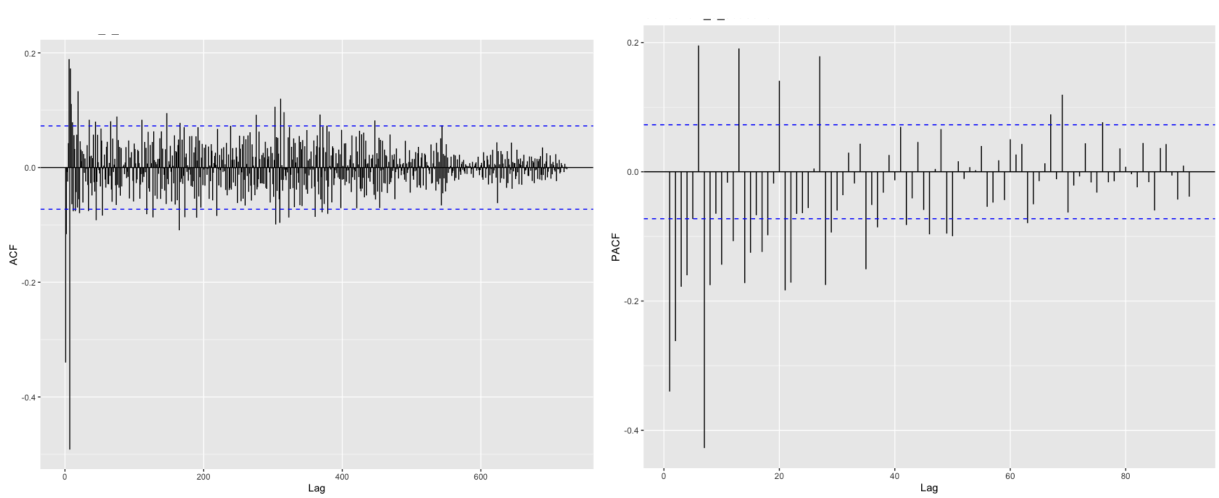


Figure 16.3 ACF and PACF Plots after Seasonal Differencing for NN5 - 036 Series

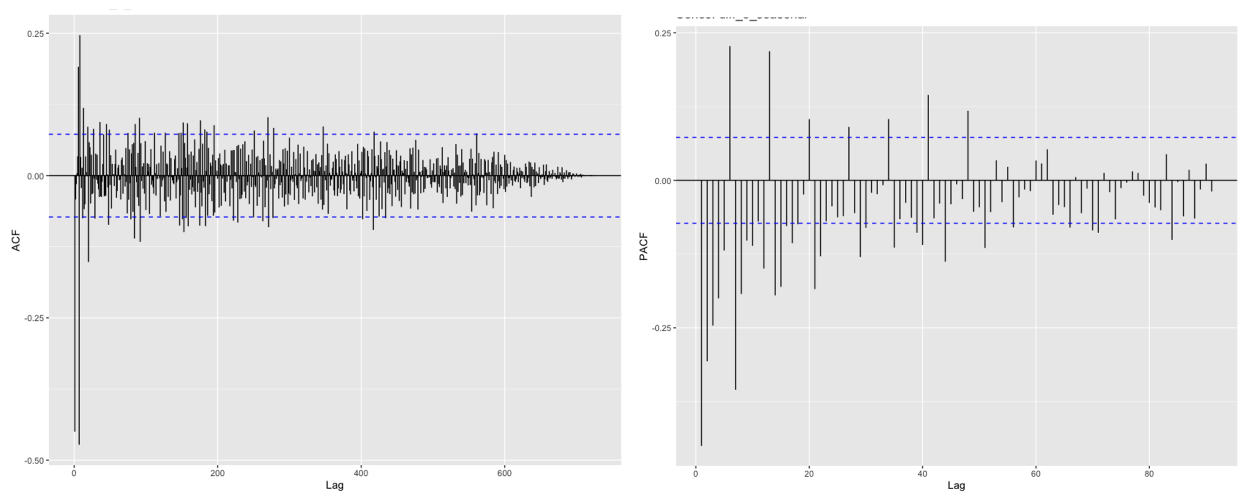
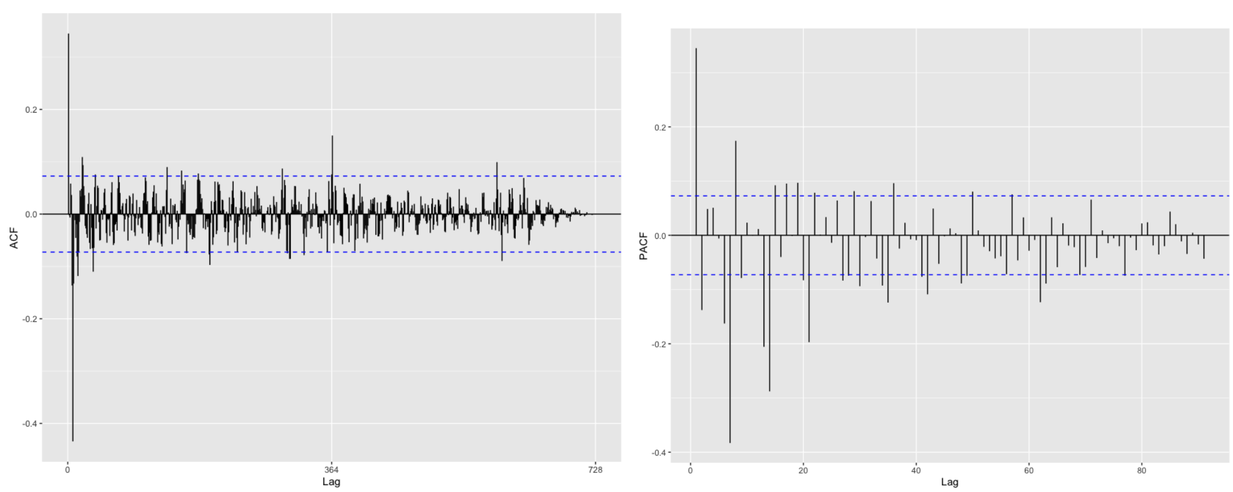


Figure 16.4 ACF and PACF Plots after Seasonal Differencing for NN5 - 042 Series



After conducting seasonal differencing, we can see from the ACF graphs that the lags are now much more concentrated within the confidence intervals which indicates that there is stationarity from trend in these time series. To further support our claim, we run the ADF and KPSS test of stationarity. The following table shows the conclusions of the statistical tests:

Table 14 - Summary Table for Stationarity Tests

Time Series	**ADF Test (P-Value)**	**KPSS Test (P-Value)**	**Conclusion**
NN5-024	<0.01	>0.1	ADF Test - Reject H_0: **Data is Stationary** KPSS Test - Fail to reject H_0: **Data is Stationary**
NN5-025	<0.01	>0.1	ADF Test - Reject H_0: **Data is Stationary** KPSS Test - Fail to reject H_0: **Data is Stationary**
NN5-036	<0.01	>0.1	ADF Test - Reject H_0: **Data is Stationary** KPSS Test - Fail to reject H_0: **Data is Stationary**

When we now analyse our PACF graphs as well, we can see that the lags do not follow any seasonality since there are no significant lags at certain intervals.

Thus, based on the ACF plots, PACF plots and statistical tests for seasonality, we can conclude that we have successfully managed to make all the 4 time series stationary for trend and seasonality. These time series can now be used for time series analysis using ARIMA.

Conclusion

To conclude, we can see that each time series is different, reflecting how cash machines can behave differently depending on where they are.

NN5-024 follows an additive model, it has a weak upward trend as well as yearly, quarterly and weekly seasonality. One structural break can also be found in that time series.

NN5-025 also follows the same additive model as well as the same trend and seasonal patterns as NN5-024. It has a structural break but also a significant amount of outliers that could require additional manipulation before forecasting.

NN5-036 has the same model and seasonal patterns as the two time series before but its weak trend is a downward trend. This time series does not exhibit any structural breaks.

NN5-042 is again similar in seasonal patterns to the other time series. However, this additive model has residuals that follow a normal distribution and no apparent trend nor structural breaks.

As time series patterns are not always clear, several models would seem suitable for forecasting them. As we already made the data series stationary by seasonally differencing, ARIMA would be the first to be suitable. However because of the differences in the time series behaviors, they would not have the same models. After conducting the previous PACF and ACF analysis, here is a summary of our findings.

Table 15 - Summary of Recommended ARIMA Models

ARIMA model	By hand	Auto Arima function
NN5-024	(0,1,2)	(1,1,2)
NN5-025	(1,1,0)	(1,1,0)
NN5-036	(1,1,0)	(0,0,3)
NN5-042	(0,0,1)	(0,0,1)

Given that we made the data stationary, we could also use time series regression models that would give us explanations on the factors influencing the data. However, those may be more complex to build and subject to our judgments. Finally, we could also think of using trend-seasonal exponential smoothing (ETS(A,A,A)) for NN5-024, NN5-025 and NN5-036 as they exhibit an additive trend and strong additive seasonal effects. For NN5-042 it would be best to use ETS(A,N,A) as this time series does not have a trend.

To conclude, even though those time series have some similarities, each of them have particular characteristics that make it difficult to forecast them together automatically. Moreover, the patterns are not always clear to see, forcing us to test multiple models in order to find the best one (e.g the simplest and best fitted).

List of Graphs (in main body of text)

Figure 1 - Boxplot Comparison of Datasets
Figure 2 - Histogram for all Datasets
Figure 3 - Time Series Plots for all Datasets
    Figure 3.1 - Time Series Plot for NN5 - 024
    Figure 3.2 - Time Series Plot for NN5 - 025
    Figure 3.3 - Time Series Plot for NN5 - 036
    Figure 3.4 - Time Series Plot for NN5 - 042
Figure 4 - Decomposition Plots of all Datasets
    Figure 4.1 - Additive Decomposition of NN5 - 024
    Figure 4.2 - Additive Decomposition of NN5 - 025
    Figure 4.3 - Additive Decomposition of NN5 - 036
    Figure 4.4 - Additive Decomposition of NN5 - 042
Figure 5 - QQ Plot of NN5 - 036's Errors
Figure 6 - Plot of Irregular Component for all Datasets
Figure 7 - Trend plot for Time Series
Figure 8 - Time Series of NN5-024 Before and After Detrending
Figure 9 - Time Series of NN5-025 Before and After Detrending
Figure 10 - Time Series of NN5-036 Before and After Detrending
Figure 11 - Weekly Time series
Figure 12 - Seasonal Components for all Time Series
    Figure 12.1 - Seasonal component for NN5-024
    Figure 12.2 - Seasonal component for NN5-025
    Figure 12.3 - Seasonal component for NN5-036
    Figure 12.4 - Seasonal component for NN5-042
Figure 13 - Time series without the Seasonal Component
Figure 14 - ACF and PACF Plots for all raw Time Series
    Figure 14.1 - ACF and PACF Plots for Raw NN5 - 024
    Figure 14.2 - ACF and PACF Plots for Raw NN5 - 025
    Figure 14.3 - ACF and PACF Plots for Raw NN5 - 036
    Figure 14.4 - ACF and PACF Plots for Raw NN5 - 042
Figure 15 - ACF and PACF Plots after First Order Differencing for all Time Series
    Figure 15.1 - ACF and PACF Plots after First Order Differencing for Raw NN5 - 024
    Figure 15.2 - ACF and PACF Plots after First Order Differencing for Raw NN5 - 025
    Figure 15.3 - ACF and PACF Plots after First Order Differencing for Raw NN5 - 036
Figure 16 - ACF and PACF Plots after First Order Differencing for all Time Series
    Figure 16.1 - ACF and PACF Plots after Seasonal Differencing for Raw NN5 - 024
    Figure 16.2 - ACF and PACF Plots after Seasonal Differencing for Raw NN5 - 025
    Figure 16.3 - ACF and PACF Plots after Seasonal Differencing for Raw NN5 - 036
    Figure 16.4 - ACF and PACF Plots after Seasonal Differencing for Raw NN5 - 042

List of Tables (in main body of text)

Table 0 - Summary Table of Proposed Model Recommendations
Table 1 - Summary Table for Descriptive Statistics for all Time Series
Table 2 - Summary Table for the Shapiro-Wilk Test for Normality
Table 3 - Summary Table for the Kolmogorov-Smirnov Test for Normality
Table 4 - Summary Table for the Cox-Stuart Test on Irregular Components
Table 5 - Summary Table for the supF Test on Irregular Components
Table 6 - Summary Table for the Bai-Perron Statistical Test for Structural Breaks
Table 7 - Summary Table of ADF Test
Table 8 - Summary Table of KPSS Test
Table 9 - Summary Table of Cox-Stuart Test
Table 10 - Summary Table of Mann-Kendall Test
Table 11 - Kruskal-Wallis Test Results on Seasonality
  Table 11.1 - Kruskal-Wallis Test on Yearly Seasonality
  Table 11.2 - Kruskal-Wallis Test on Quarterly Seasonality
  Table 11.3 - Kruskal-Wallis Test on Weekly Seasonality
Table 12 - Friedman Test Results on Seasonality
  Table 12.1 - Friedman Test on Quarterly Seasonality
  Table 12.2 - Friedman Test on Weekly Seasonality
Table 13 - Summary Table for Stationarity Tests after First Order Differencing
Table 14 - Summary Table for Stationarity Tests after Seasonal Differencing
Table 15 - Summary of Recommended ARIMA Models

Appendix

A.1 - Time Series Plots at Different Intervals

Fig.A. Monthly Time series

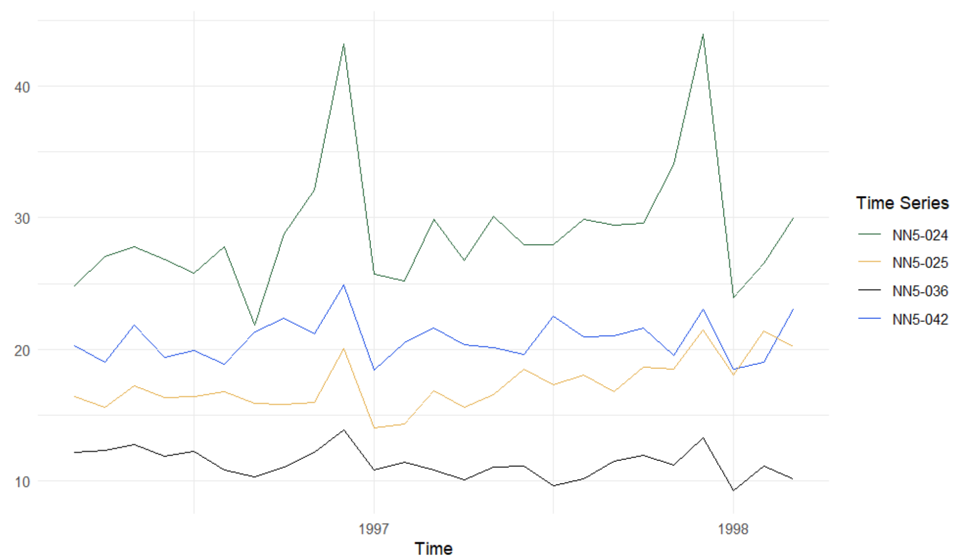
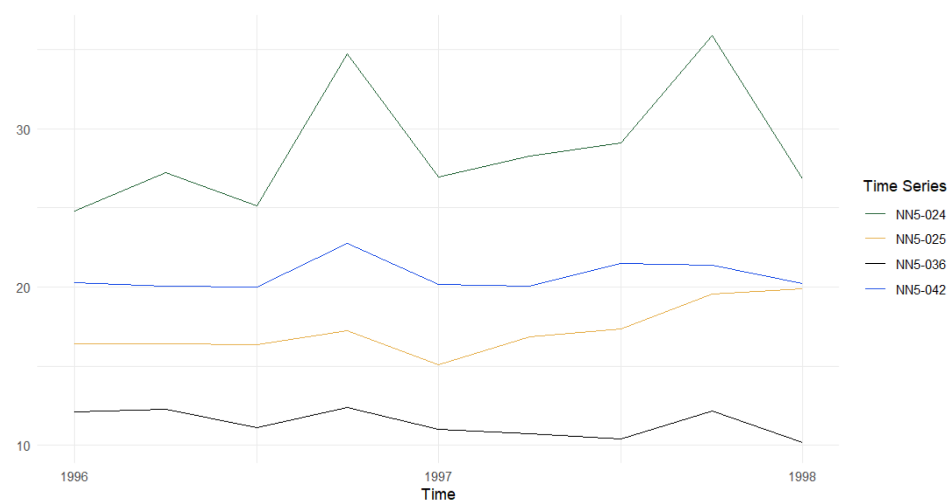


Fig.B. Quarterly Time series



A.2 - BIC and RSS Plots from Bai-Perron Tests

Fig.C. NN5 - 024

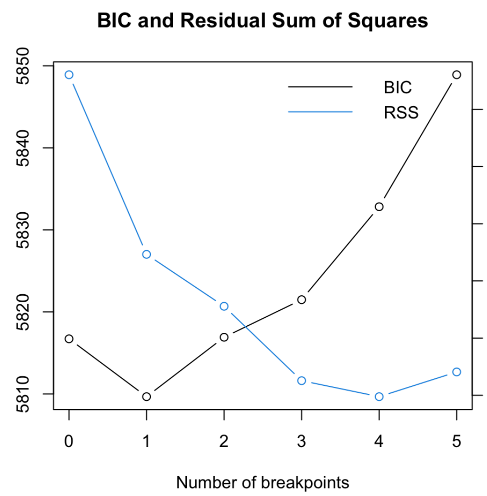


Fig.D. NN5 - 025

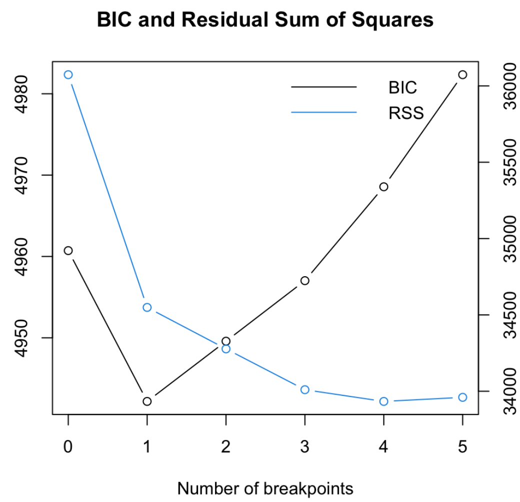


Fig.E. NN5 - 036

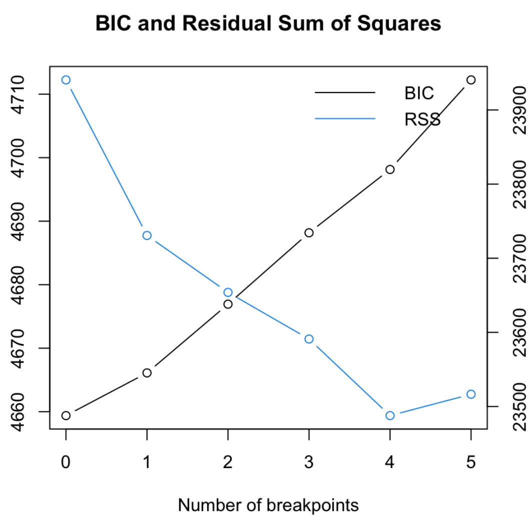
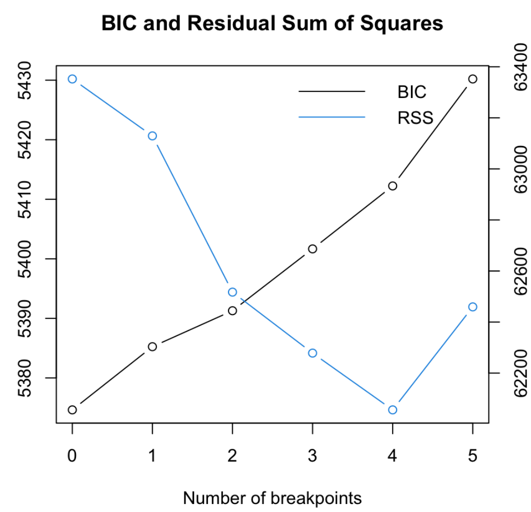


Fig.F. NN5 - 042



A.3 - R Code for Analysis and Plots

Section 1:

```
library(forecast)
library(readxl)
library(ggplot2)
library(Kendall)
library(zoo)
library(tsutils)
```

```
library(tseries)
library(TTR)
library(e1071)
library(strucchange)

#Importing the datasets into R - daily
#------------------------------------------------------------

data_daily=read_xlsx('DataCW.xlsx',sheet = "Year", col_names = TRUE)

#Counting the number of missing values per dataset
#------------------------------------------------------------
sum(is.na(data_daily$O))
sum(is.na(data_daily$S))
sum(is.na(data_daily$E))
sum(is.na(data_daily$K))
data_o_daily<-na.approx(data_daily$O)
data_s_daily<-na.approx(data_daily$S)
data_e_daily<-na.approx(data_daily$E)
data_k_daily<-na.approx(data_daily$K)

#Importing and approximating the data weekly
#------------------------------------------------------------

data_weekly <-read_excel("DataCW.xlsx", sheet = "Week",col_names =
TRUE)
data_o_weekly <-na.approx(data_weekly$O)
data_s_weekly <-na.approx(data_weekly$S)
data_e_weekly <-na.approx(data_weekly$E)
data_k_weekly <-na.approx(data_weekly$K)


#Importing and approximating the data monthly
#------------------------------------------------------------

data_monthly <-read_excel("DataCW.xlsx", sheet = "Month",col_names =
TRUE)
data_o_monthly <-na.approx(data_monthly$O)
data_s_monthly <-na.approx(data_monthly$S)
data_e_monthly <-na.approx(data_monthly$E)
data_k_monthly <-na.approx(data_monthly$K)


#Importing and approximating the data quarterly
#------------------------------------------------------------

data_quarterly <-read_excel("DataCW.xlsx", sheet = "Quarterly",
col_names = TRUE)
```

```
data_o_quarterly <-na.approx(data_quarterly$O)
data_s_quarterly <-na.approx(data_quarterly$S)
data_e_quarterly <-na.approx(data_quarterly$E)
data_k_quarterly <-na.approx(data_quarterly$K)


#Identifying Outliers using boxplots
#-------------------------------------------------------------

par(mfrow = c(1, 1))
boxplot(data_o_daily,data_s_daily, data_e_daily, data_k_daily, names
= c("NN5-024 Boxplot", "NN5-025 Boxplot", "NN5-036 Boxplot", "NN5-042
Boxplot"), main = "Fig 1. Boxplot Comparison of Datasets")
length(boxplot.stats(data_daily$O)$out)
length(boxplot.stats(data_daily$S)$out)
length(boxplot.stats(data_daily$E)$out)
length(boxplot.stats(data_daily$K)$out)

#Generating Descriptive Statistics
#-------------------------------------------------------------

summary(data_o_daily)
summary(data_s_daily)
summary(data_e_daily)
summary(data_k_daily)
sd(data_o_daily)
sd(data_s_daily)
sd(data_e_daily)
sd(data_k_daily)
skewness(data_o_daily)
skewness(data_s_daily)
skewness(data_e_daily)
skewness(data_k_daily)

#Histograms
#-------------------------------------------------------------

par(mfrow = c(2, 2), oma = c(2, 2, 4, 2))
hist(data_o_daily, main = "Histogram for N55-024", xlab = "Values")
hist(data_s_daily, main = "Histogram for N55-025", xlab = "Values")
hist(data_e_daily, main = "Histogram for N55-036", xlab = "Values")
hist(data_k_daily, main = "Histogram for N55-042", xlab = "Values")
par(mfrow = c(1, 1))


#Convert to Time Series - daily
#-------------------------------------------------------------
```

```
data_o=ts(data_o_daily, frequency =365, start = c(1996,77))
data_s=ts(data_s_daily, frequency =365, start = c(1996,77))
data_e=ts(data_e_daily, frequency =365, start = c(1996,77))
data_k=ts(data_k_daily, frequency =364, start = c(1996,77))

#Convert to Time Series - weekly
#-------------------------------------------------------------

data_e_w <-ts(data_e_weekly,frequency=52,start=c(1996,12))
data_k_w <-ts(data_k_weekly,frequency=52,start=c(1996,12))
data_o_w <-ts(data_o_weekly,frequency=52,start=c(1996,12))
data_s_w <-ts(data_s_weekly,frequency=52,start=c(1996,12))

#Convert to Time Series - monthly
#-------------------------------------------------------------

data_e_m <-ts(data_e_monthly,frequency=12,start=c(1996,3))
data_k_m <-ts(data_k_monthly,frequency=12,start=c(1996,3))
data_o_m <-ts(data_o_monthly,frequency=12,start=c(1996,3))
data_s_m <-ts(data_s_monthly,frequency=12,start=c(1996,3))
#Convert to Time Series - quarterly
#-------------------------------------------------------------

data_e_q <-ts(data_e_quarterly,frequency=4,start=c(1996,1))
data_k_q <-ts(data_k_quarterly,frequency=4,start=c(1996,1))
data_o_q <-ts(data_o_quarterly,frequency=4,start=c(1996,1))
data_s_q <-ts(data_s_quarterly,frequency=4,start=c(1996,1))


#Plot Time Series Plots - daily
#-------------------------------------------------------------

plot(data_o, ylab="N55_024", col="#246739")
abline(h=mean(data_o),col="red")
plot(data_s, ylab="N55_025", col="#e7b24d")
abline(h=mean(data_s),col="red")
plot(data_e, ylab="N55_036", col="#000000")
abline(h=mean(data_e),col="red")
plot(data_k, ylab="N55_042", col="#2152e9")
abline(h=mean(data_k),col="red")

#Plot additive decomposition plots - daily
#-------------------------------------------------------------

o_decomposition<-decomp(data_o,decomposition="additive",outplot=TRUE)
s_decomposition<-decomp(data_s,decomposition="additive",outplot=TRUE)
e_decomposition<-decomp(data_e,decomposition="additive",outplot=TRUE)
k_decomposition<-decomp(data_k,decomposition="additive",outplot=TRUE)
```

```
#Plot additive decomposition plots - weekly
#------------------------------------------------------------

o_w_decomposition<-decomp(data_o_w,decomposition="additive",outplot=T
RUE)
s_w_decomposition<-decomp(data_s_w,decomposition="additive",outplot=T
RUE)
e_w_decomposition<-decomp(data_e_w,decomposition="additive",outplot=T
RUE)
k_w_decomposition<-decomp(data_k_w,decomposition="additive",outplot=T
RUE)
```

Section 2:

```
#The Shapiro-Wilk Test for Normality
#------------------------------------------------------------

shapiro.test(e_decomposition$irregular)
#p-value = 0.001171 so reject null - don't follow the normal
shapiro.test(k_decomposition$irregular)
#p-value = 0.8426 so not reject - follows the normal
shapiro.test(o_decomposition$irregular)
#p-value = 3.358e-06 so reject null - don't follow the normal
shapiro.test(s_decomposition$irregular)
#p-value = 2.714e-10 so reject null - don't follow the normal

#The Kolmogorov-Smirnov Test for Normality
#------------------------------------------------------------

ks.test(e_decomposition$irregular, y = "rnorm")
#p-value < 2.2e-16 so reject H0, errors don't follow a normal
distribution
ks.test(k_decomposition$irregular, y = "rnorm")
#p-value < 2.2e-16 so reject H0, errors don't follow a normal
distribution
ks.test(o_decomposition$irregular, y = "rnorm")
#p-value < 2.2e-16 so reject H0, errors don't follow a normal
distribution
ks.test(s_decomposition$irregular, y = "rnorm")
#p-value < 2.2e-16 so reject H0, errors don't follow a normal
distribution

#QQ Plot
#------------------------------------------------------------

qqnorm(k_decomposition$irregular, main = "QQ Plot of NN5-042's
Errors")
```

```
qqline(k_decomposition$irregular)
```

Section 3:

```
#Plot of Irregular Components
#------------------------------------------------------------

plot(e_decomposition$irregular, main = "Irregular Component of
NN5-036")
plot(k_decomposition$irregular, main = "Irregular Component of
NN5-042")
plot(o_decomposition$irregular, main = "Irregular Component of
NN5-024")
plot(s_decomposition$irregular, main = "Irregular Component of
NN5-025")

#Cox Stuart Test for Dispersion
#------------------------------------------------------------

coxstuart(data_e,"dispersion")
#0.06508884
coxstuart(data_k,"dispersion")
#0.4537803
coxstuart(data_o,"dispersion")
#0.06508884
coxstuart(data_s,"dispersion")
#0.2806904

#supF Test and Plot of F-Statistics
#------------------------------------------------------------

fs_e <- Fstats(data_e ~ 1)
plot(fs_e)
break_test_e <- sctest(fs_e)
break_test_e

fs_k <- Fstats(data_k ~ 1)
plot(fs_k)
break_test_k <- sctest(fs_k)
break_test_k

fs_o <- Fstats(data_o ~ 1)
plot(fs_o)
break_test_o <- sctest(fs_o)
break_test_o

fs_s <- Fstats(data_s ~ 1)
plot(fs_s)
```

```
break_test_s <- sctest(fs_s)
break_test_s

#Bai-Perron Test
#------------------------------------------------------------

breakpoints_model_e <- breakpoints(data_e ~ 1)
summary(breakpoints_model_e)
plot(breakpoints_model_e)

breakpoints_model_k <- breakpoints(data_k ~ 1)
summary(breakpoints_model_k)
plot(breakpoints_model_k)

breakpoints_model_o <- breakpoints(data_o ~ 1)
summary(breakpoints_model_o)
plot(breakpoints_model_o)

breakpoints_model_s <- breakpoints(data_s ~ 1)
summary(breakpoints_model_s)
plot(breakpoints_model_s)
```

Section 4:

```
# Plot trend plots of time series 24, 25, 36
#------------------------------------------------------------

plot(o_decomposition$trend, main = "Trend Plot of NN5-024", xlab =
"Time", ylab = "Value")
plot(s_decomposition$trend, main = "Trend Plot of NN5-025", xlab =
"Time", ylab = "Value")
plot(e_decomposition$trend, main = "Trend Plot of NN5-036", xlab =
"Time", ylab = "Value")



# Carry out stationarity tests (KPSS & ADF)
#------------------------------------------------------------

# Running KPSS test (H0: data is stationary)
kpss.test(data_e) # do not reject H0: data is stationary
kpss.test(data_k) # do not reject H0: data is stationary
kpss.test(data_o) # reject H0: data is not stationary
kpss.test(data_s) # reject H0: data is not stationary


# Running ADF test (H0: data is not stationary)
adf.test(data_e) # reject H0: data is stationary
```

```
adf.test(data_k) # reject H0: data is stationary
adf.test(data_o) # reject H0: data is stationary
adf.test(data_s) # reject H0: data is stationary



# Carry out trend tests (Cox-Stuart & Mann-Kendall)
#-------------------------------------------------------------

# Running cox-Stuart test (H0: no trend)
coxstuart(data_e, "trend")
coxstuart(data_o, "trend")
coxstuart(data_s, "trend")

# Running Mann-Kendall test (H0: no trend in data)
MannKendall(data_e)
MannKendall(data_o)
MannKendall(data_s)



# Plot time series before and after detrending
#-------------------------------------------------------------

# Removing trend component for time series 24, 25, and 36
detrended_data_o = data_o - o_decomposition$trend
detrended_data_s = data_s - s_decomposition$trend
detrended_data_e = data_e - e_decomposition$trend

# Set plotting layout: 2 rows and 1 column
par(mfrow = c(2, 1))


# Time series 24 before detrending
plot(data_o, main = "Time Series of NN5-024", ylab = "Values", xlab =
"Time")
mean_data_o = mean(data_o)
abline(h = mean_data_o, col = "red", lwd = 2, lty = 2)

# Time series 24 after detrending
plot(detrended_data_o, main = "Detrended Series of NN5-024", ylab =
"Detrended Values", xlab = "Time")
mean_detrended_o = mean(detrended_data_o, na.rm = TRUE)
abline(h = mean_detrended_o, col = "red", lwd = 2, lty = 2)


# Time series 25 before detrending
```

```
plot(data_s,  main = "Time Series of NN5-025", ylab = "Values", xlab
= "Time")
mean_data_s = mean(data_s)
abline(h = mean_data_s, col = "red", lwd = 2, lty = 2)

# Time series 25 after detrending
plot(detrended_data_s, main = "Detrended Series of NN5-025", ylab =
"Detrended Values", xlab = "Time")
mean_detrended_s = mean(detrended_data_s, na.rm = TRUE)
abline(h = mean_detrended_s, col = "red", lwd = 2, lty = 2)


# Time series 36 before detrending
plot(data_e,  main = "Time Series of NN5-036", ylab = "Values", xlab
= "Time")
mean_data_e = mean(data_e)
abline(h = mean_data_e, col = "red", lwd = 2, lty = 2)

# Time series 36 after detrending
plot(detrended_data_e, main = "Detrended Series of NN5-036", ylab =
"Detrended Values", xlab = "Time")
mean_detrended_e = mean(detrended_data_e, na.rm = TRUE)
abline(h = mean_detrended_e, col = "red", lwd = 2, lty = 2)
```

Section 5:

```
 #Plot Time Series Plots - weekly
#------------------------------------------------------------

autoplot(data_e_w, ylab="", series = "NN5-036") +
  autolayer(data_k_w, series = "NN5-042") +
  autolayer(data_o_w, series = "NN5-024") +
  autolayer(data_s_w, series = "NN5-025") +
  scale_color_manual(
     name = "Time Series",  # Title for the legend
     values = c("NN5-036" = "black", "NN5-042" = "#2152e9",
"NN5-024" = "#246739", "NN5-025" = "#e7b24d")
  ) +
  theme_minimal()

#Plot Time Series Plots - monthly
#------------------------------------------------------------

autoplot(data_e_m, ylab="", series = "NN5-036") +
  autolayer(data_k_m, series = "NN5-042") +
  autolayer(data_o_m, series = "NN5-024") +
  autolayer(data_s_m, series = "NN5-025") +
  scale_color_manual(
```

```
      name = "Time Series",  # Title for the legend
      values = c("NN5-036" = "black", "NN5-042" = "#2152e9",
"NN5-024" = "#246739", "NN5-025" = "#e7b24d")
  ) +
  theme_minimal()

 #Plot Time Series Plots - quarterly
#------------------------------------------------------------

autoplot(data_e_q, ylab="", series = "NN5-036") +
  autolayer(data_k_q, series = "NN5-042") +
  autolayer(data_o_q, series = "NN5-024") +
  autolayer(data_s_q, series = "NN5-025") +
  scale_color_manual(
      name = "Time Series",  # Title for the legend
      values = c("NN5-036" = "black", "NN5-042" = "#2152e9",
"NN5-024" = "#246739", "NN5-025" = "#e7b24d")
  ) +
  theme_minimal()

 #Plot Seasonal component - weekly
#------------------------------------------------------------
autoplot(e_w_decomposition$season, ylab="", series = "NN5-036") +
  autolayer(k_w_decomposition$season, series = "NN5-042") +
  autolayer(o_w_decomposition$season, series = "NN5-024") +
  autolayer(s_w_decomposition$season, series = "NN5-025") +
  scale_color_manual(
      name = "Time Series",  # Title for the legend
      values = c("NN5-036" = "black", "NN5-042" = "#2152e9",
"NN5-024" = "#246739", "NN5-025" = "#e7b24d")
  ) +
  theme_minimal()

 #Plot Seasonal component - NN5-024
autoplot(o_w_decomposition$season$season, xlab =
"Time",ylab="NN5-024")

 #Plot Seasonal component - NN5-025
autoplot(s_w_decomposition$season$season, xlab =
"Time",ylab="NN5-025")

 #Plot Seasonal component - NN5-036
autoplot(e_w_decomposition$season$season, xlab =
"Time",ylab="NN5-036")

 #Plot Seasonal component - NN5-042
```

```
autoplot(k_w_decomposition$season$season, xlab =
"Time",ylab="NN5-042")

#Carry out seasplot test
#------------------------------------------------------------
#Weekly
seasplot(data_o,m=7)
seasplot(data_s,m=7)
seasplot(data_e,m=7)
seasplot(data_k,m=7)

#Monthly
seasplot(data_o,m=28)
seasplot(data_s,m=28)
seasplot(data_e,m=28)
seasplot(data_k,m=28)

#Quarterly
seasplot(data_o,m=91)
seasplot(data_s,m=91)
seasplot(data_e,m=91)
seasplot(data_k,m=91)

# Carry out seasonality tests (Kruskal-Wallis and Friedman)
#------------------------------------------------------------

#Kruskal-Wallis
#Daily

data_o_s<-cycle(data_o)
kruskal.test(data_o~factor(data_o_s))
data_s_s<-cycle(data_s)
kruskal.test(data_s~factor(data_s_s))
data_e_s<-cycle(data_e)
kruskal.test(data_e~factor(data_e_s))
data_k_s<-cycle(data_k)
kruskal.test(data_k~factor(data_k_s))

#Weekly
data_o_w_s<-cycle(data_o_w)
kruskal.test(data_o_w~factor(data_o_w_s))
data_s_w_s<-cycle(data_s_w)
kruskal.test(data_s_w~factor(data_s_w_s))
data_e_w_s<-cycle(data_e_w)
kruskal.test(data_e_w~factor(data_e_w_s))
data_k_w_s<-cycle(data_k_w)
kruskal.test(data_k_w~factor(data_k_w_s))
```

```
#Quarterly
data_e_q_s<-cycle(data_e_q)
kruskal.test(data_e_q~factor(data_e_q_s))
data_k_q_s<-cycle(data_k_q)
kruskal.test(data_k_q~factor(data_k_q_s))
data_o_q_s<-cycle(data_o_q)
kruskal.test(data_o_q~factor(data_o_q_s))
data_s_q_s<-cycle(data_s_q)
kruskal.test(data_s_q~factor(data_s_q_s))

#Friedman test
#Weekly
week_matrix_o <- matrix(data_o_w, ncol = 52, byrow = TRUE)
friedman.test(week_matrix_o)
week_matrix_s <- matrix(data_s_w, ncol = 52, byrow = TRUE)
friedman.test(week_matrix_s)
week_matrix_e <- matrix(data_e_w, ncol = 52, byrow = TRUE)
friedman.test(week_matrix_e)
week_matrix_k <- matrix(data_k_w, ncol = 52, byrow = TRUE)
friedman.test(week_matrix_k)

#Monthly
month_matrix_o <- matrix(data_o_m, ncol = 12, byrow = TRUE)
friedman.test(month_matrix_o)
month_matrix_s <- matrix(data_s_m, ncol = 12, byrow = TRUE)
friedman.test(month_matrix_s)
month_matrix_e <- matrix(data_e_m, ncol = 12, byrow = TRUE)
friedman.test(month_matrix_e)
month_matrix_k <- matrix(data_k_m, ncol = 12, byrow = TRUE)
friedman.test(month_matrix_k)

#Quarterly
quarter_matrix_o <- matrix(data_o_q, ncol = 4, byrow = TRUE)
friedman.test(quarter_matrix_o)
quarter_matrix_s <- matrix(data_s_q, ncol = 4, byrow = TRUE)
friedman.test(quarter_matrix_s)
quarter_matrix_e <- matrix(data_e_q, ncol = 4, byrow = TRUE)
friedman.test(quarter_matrix_e)
quarter_matrix_k <- matrix(data_k_q, ncol = 4, byrow = TRUE)
friedman.test(quarter_matrix_k)
```

Section 6:

```
#Plotting ACF for daily time series
#-------------------------------------------------------------

ggAcf(data_o)
ggAcf(data_s)
ggAcf(data_e)
```

```
ggAcf(data_k)

#Plotting PACF for daily time series
#-------------------------------------------------------------

ggPacf(data_o, lag=91)
ggPacf(data_s, lag=91)
ggPacf(data_e, lag=91)
ggPacf(data_k, lag=91)

#First Order Differencing
#-------------------------------------------------------------

diff_o=diff(data_o, lag=1)
diff_s=diff(data_s, lag=1)
diff_e=diff(data_e, lag=1)

#ACF Plots after First Order Differencing
#-------------------------------------------------------------

ggAcf(diff_o)
ggAcf(diff_s)
ggAcf(diff_e)

#PACF Plots after First Order Differencing
#-------------------------------------------------------------

ggPacf(diff_o, lag=91)
ggPacf(diff_s, lag=91)
ggPacf(diff_e, lag=91)

#Stationary Test after First Order Differencing ADF Test (H0:non
stationary, H1:stationary)
#-------------------------------------------------------------

adf.test(diff_o)
adf.test(diff_s)
adf.test(diff_e)

#Stationary Test after First Order Differencing KPSS Test
(H0:stationary, H1:non stationary)
#-------------------------------------------------------------

kpss.test(diff_o)
kpss.test(diff_s)
kpss.test(diff_e)

#Seasonal Differencing
```

```
#-------------------------------------------------------------

diff_o_seasonal=diff(diff_o,lag=7)
diff_s_seasonal=diff(diff_s,lag=7)
diff_e_seasonal=diff(diff_e,lag=7)
diff_k_seasonal=diff(data_k,lag=7)

#ACF Plots after Seasonal Differencing
#-------------------------------------------------------------

ggAcf(diff_o_seasonal)
ggAcf(diff_s_seasonal)
ggAcf(diff_e_seasonal)
ggAcf(diff_k_seasonal)

#PACF Plots after Seasonal Differencing
#-------------------------------------------------------------

ggPacf(diff_o_seasonal, lag=91)
ggPacf(diff_s_seasonal, lag=91)
ggPacf(diff_e_seasonal, lag=91)
ggPacf(diff_k_seasonal, lag=91)

#Stationary Test after Seasonal Differencing ADF Test (H0:non
stationary, H1:stationary)
#-------------------------------------------------------------

adf.test(diff_o_seasonal)
adf.test(diff_s_seasonal)
adf.test(diff_e_seasonal)
adf.test(diff_k_seasonal)

#Stationary Test after Seasonal Differencing KPSS Test
(H0:stationary, H1:non stationary)
#-------------------------------------------------------------

kpss.test(diff_o_seasonal)
kpss.test(diff_s_seasonal)
kpss.test(diff_e_seasonal)
kpss.test(diff_k_seasonal)
```