IE 360

STATISTICAL FORECASTING AND TIME SERIES

TERM PROJECT

Instructor: Mustafa Gökçe Baydoğan

06.07.2020

GROUP 23:

RAMAZAN ERASLAN (2014402210)

MİRAÇ GÖKTUĞ ALTUNTAŞ (2016402198)

EYÜP TAŞCI (2016402321)

Spring 2020

Table of Content

Table of Content

1. Introduction

Nowadays, as it is knowns, many people shop at online retail sites. Because of this extreme demand, they try to meet the excess demands of retail sites, but this is not easy at all. One of the big obstacles to this is to estimate the sales amount of any product. Forecasting is largely based on historical data and modeling skills. Trendyol is one of these online retail sites. What is required of us in this project is to estimate the next day's sales amounts of Trendyol products shown in **Table 1**. The data covers daily sales of eight products of Trendyol and includes the product, category level, brand and site-level details.

In this project, our aim is to develop a forecasting product. The products developed can be used in various business areas, such as creating appropriate inventory or service levels. Accurate forecasts help businesses reduce waste and understand uncertainty and risk consequences.

Product Id	Top Hier.	Bottom Hier.	Brand
85004	Kozmetik	Yüz Temizleyici	La Roche Posay
4066298	Süpermarket	Islak Mendil & Havlu	Sleepy
6676673	Elektronik	Telefon Bluetooth Kulaklık	Xiaomi
7061886	Elektronik	Süpürge	Fakir
31515569	Giyim	Tayt	TRENDYOLMİLLA
32939029	Süpermarket	Şarj Edebilir Diş Fırçası	Oral-B
5926527	Giyim	Bikini Üstü	TRENDYOLMİLLA
3904356	Giyim	Mont	Koton

Table 1: Product characteristics

2. Approach

In this project, 1-year daily sales data of 8 different product are given. Also, in the data there are price of the product, category of the product, visit count of the category and the visit count of the Trendyol information are given. Even though this information is given, only the sales count and price information for all products are used in the analysis. Information about 8 products are stored 8 different sub data. After that, these subdatas are filtered from outliers. Linear regression, auto regression and dynamic regression models are constructed and evaluated for each product. Models that give minimum MAPE (mean absolute percentage error) value are selected to forecast the sales quantity in the future.

3. Time Series Models

a) Bikini

Daily sales plot and the sales quantity vs price scatter plot for product "5926527" can be seen in figure 1 and 2. Before starting the forecasting, days with negative prices and the extreme daily sale quantities which were decided to higher than 20 were omitted from the data. Afterwards, auto regression with sales quantity and dynamic regression with price models were constructed. Autoregression and dynamic regression model give almost identical AIC and MAPE values for this product. Dynamic regression model was used for forecasting.
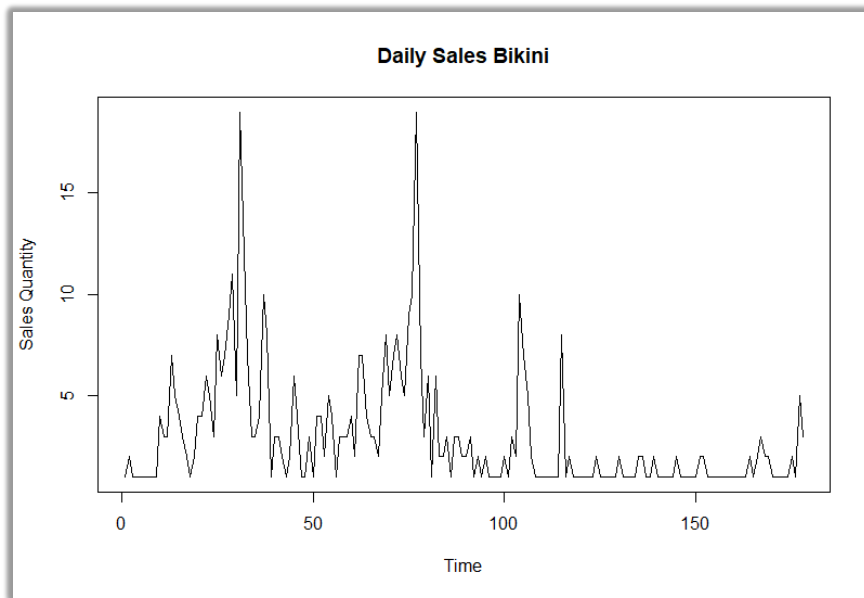

**Figure 1**. Daily Sales Bikini
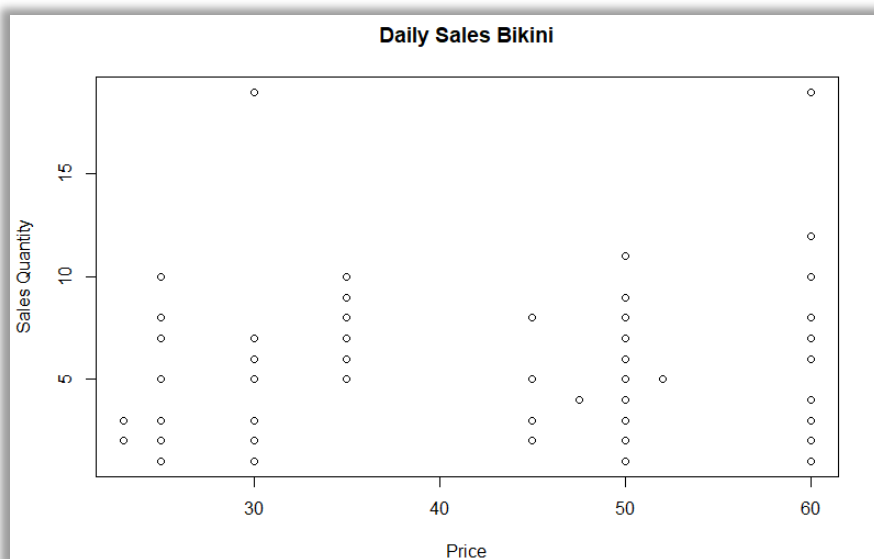

**Figure 2**. Daily Sales Bikini
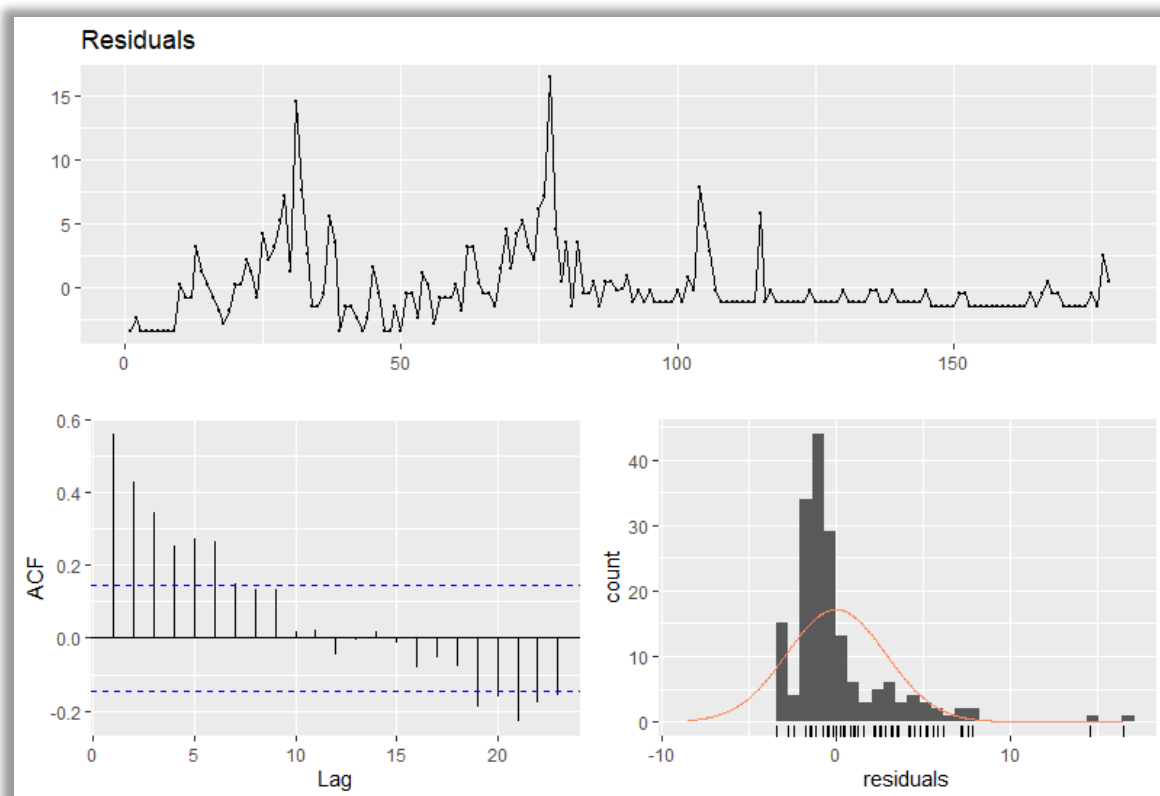
Figure 3. Residuals
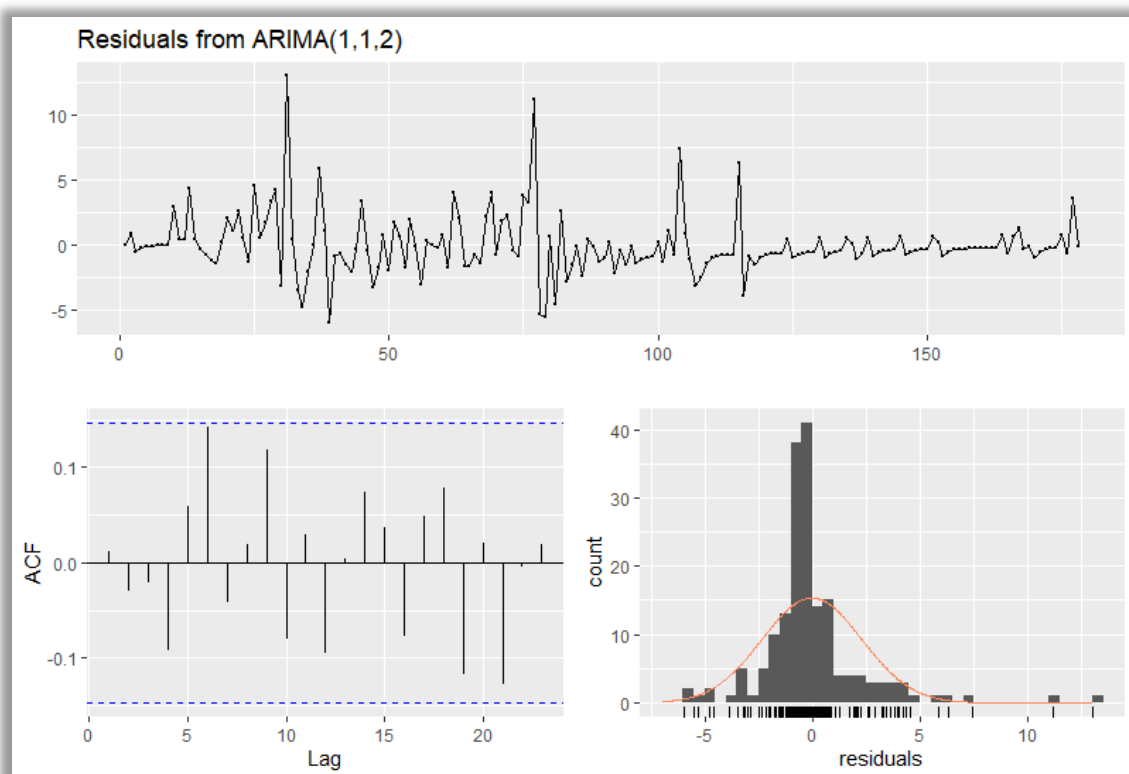
Figure 4. Residuals from ARIMA(1,1,2)
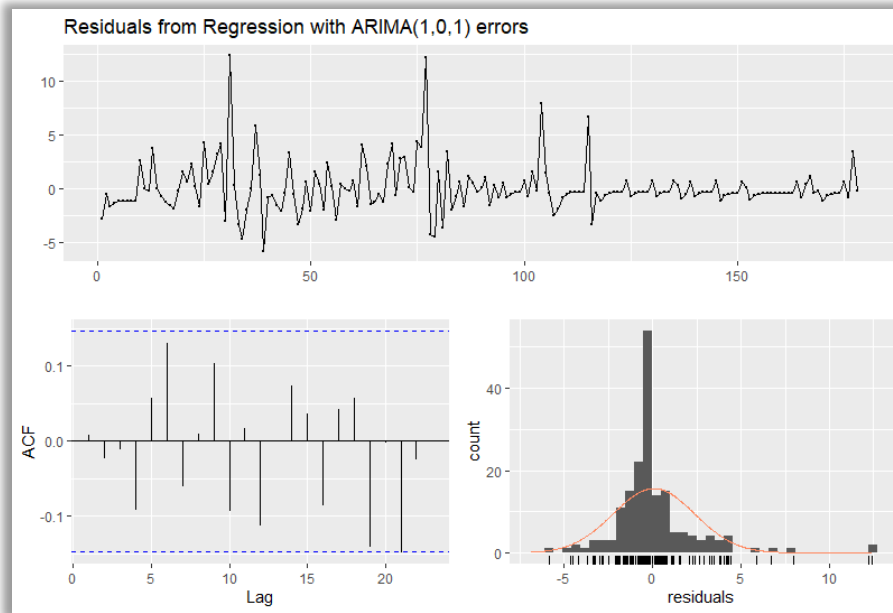
Figure 5. Residuals from Regression with ARIMA(1,0,1) errors

b) Fakir

Daily sales plot and the sales quantity vs price scatter plot for product "7061886" can be seen in figure 6 and 7. Days with negative price values and days with extreme sales quantities which were decided higher than 200 were omitted from the data. Observations with prices higher than 250 are too few so these days were also omitted from the data. Afterwards, auto regression with sales quantity and dynamic regression with price models were constructed. Autoregression and dynamic regression model give almost identical AIC and MAPE values for this product. Dynamic regression model was used for forecasting.
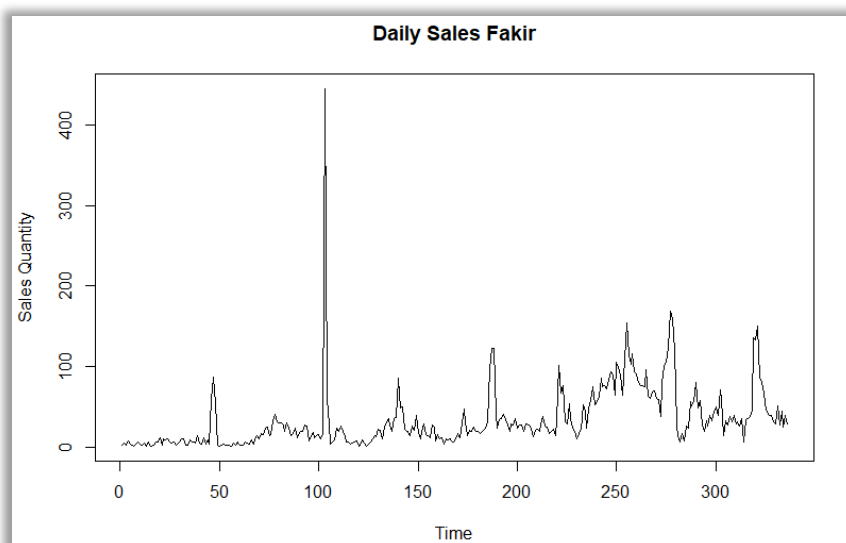
Figure 6. Daily Sales Fakir
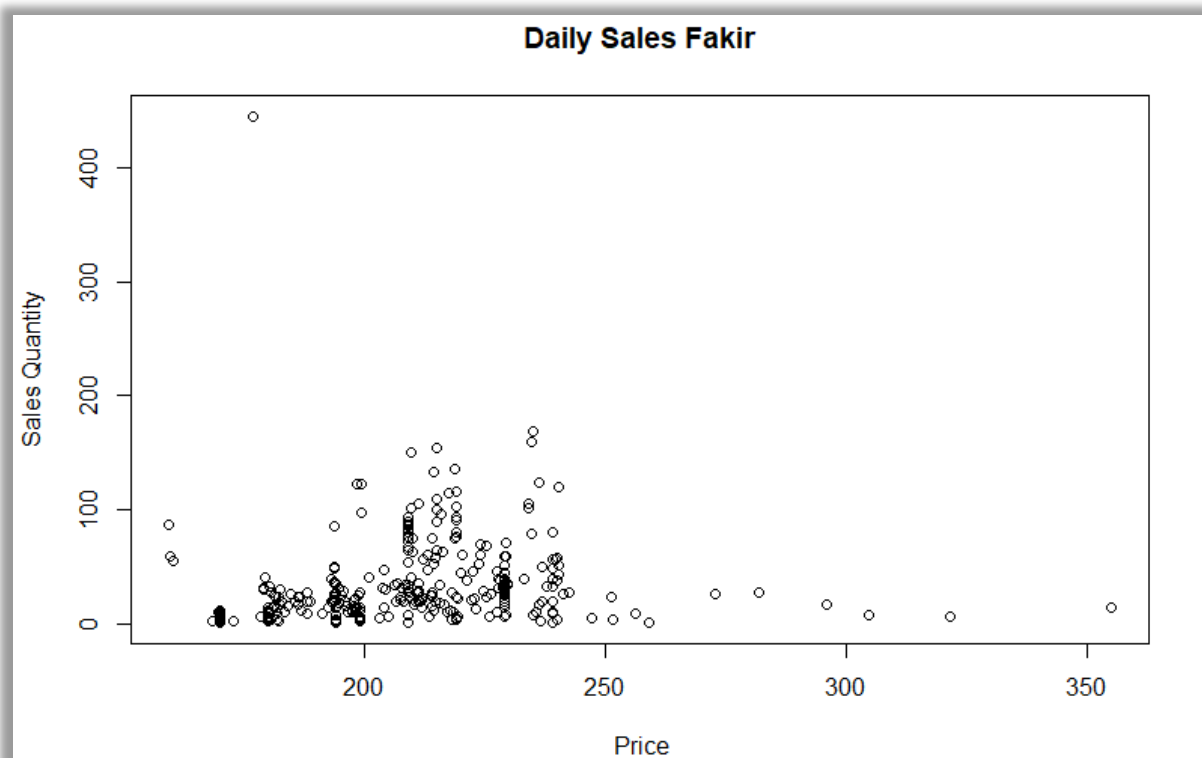
Figure 7. Daily Sales Fakir
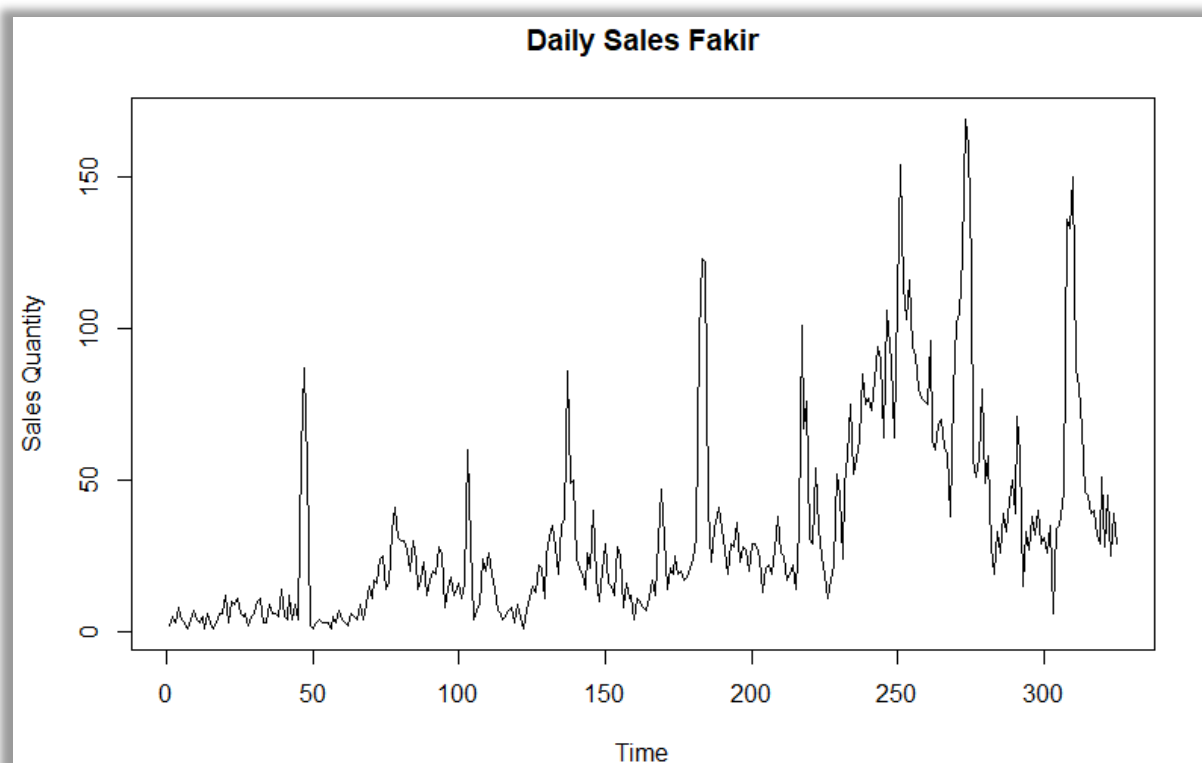
Figure 8. Daily Sales Fakir
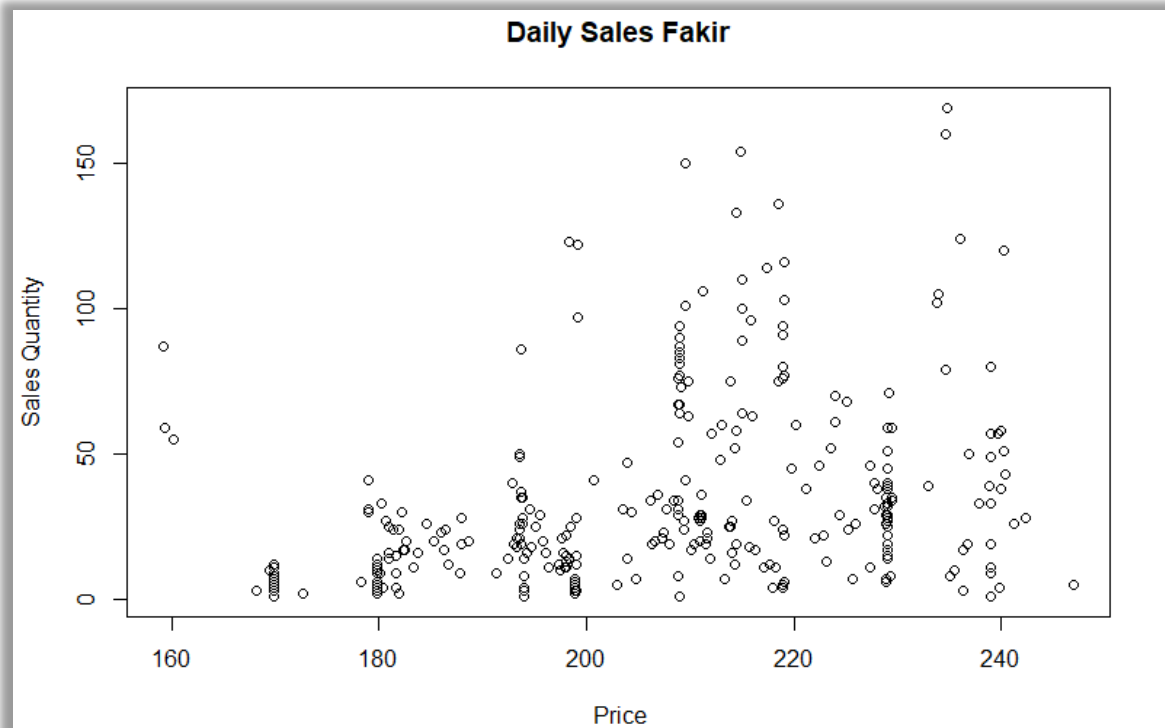
Figure 9. Daily Sales Fakir
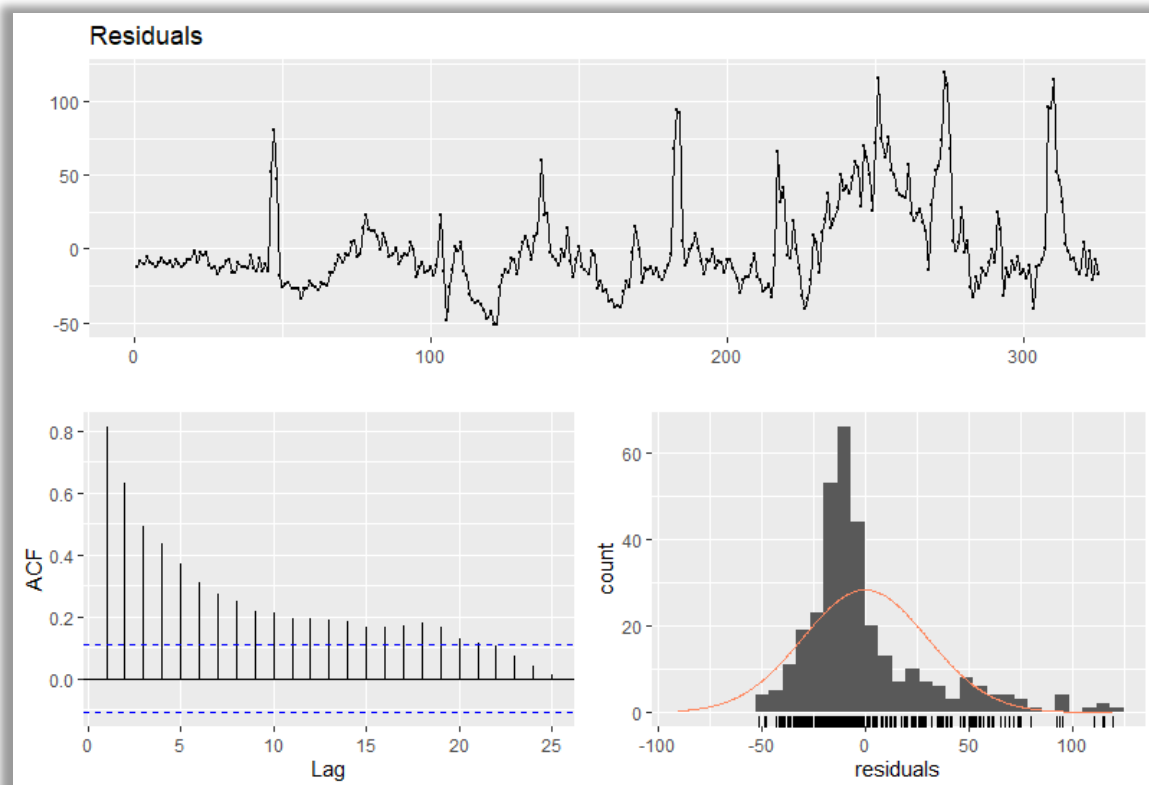
Figure 10. Residuals
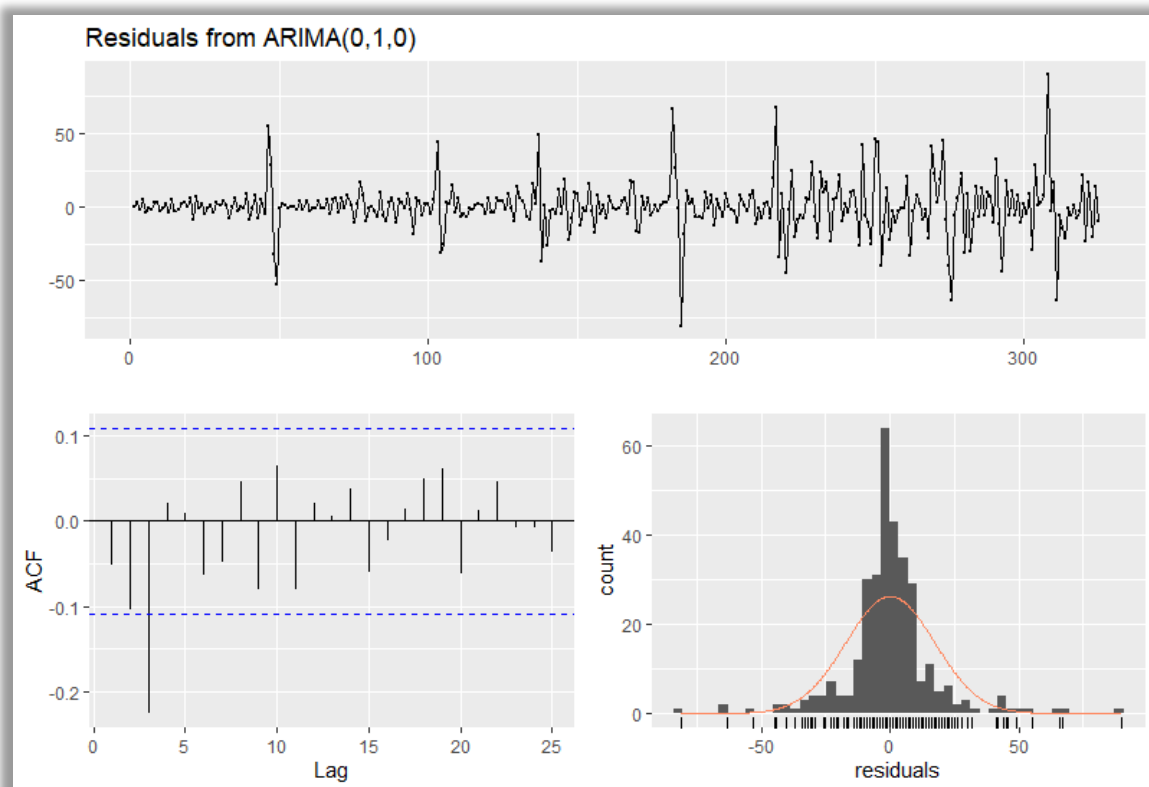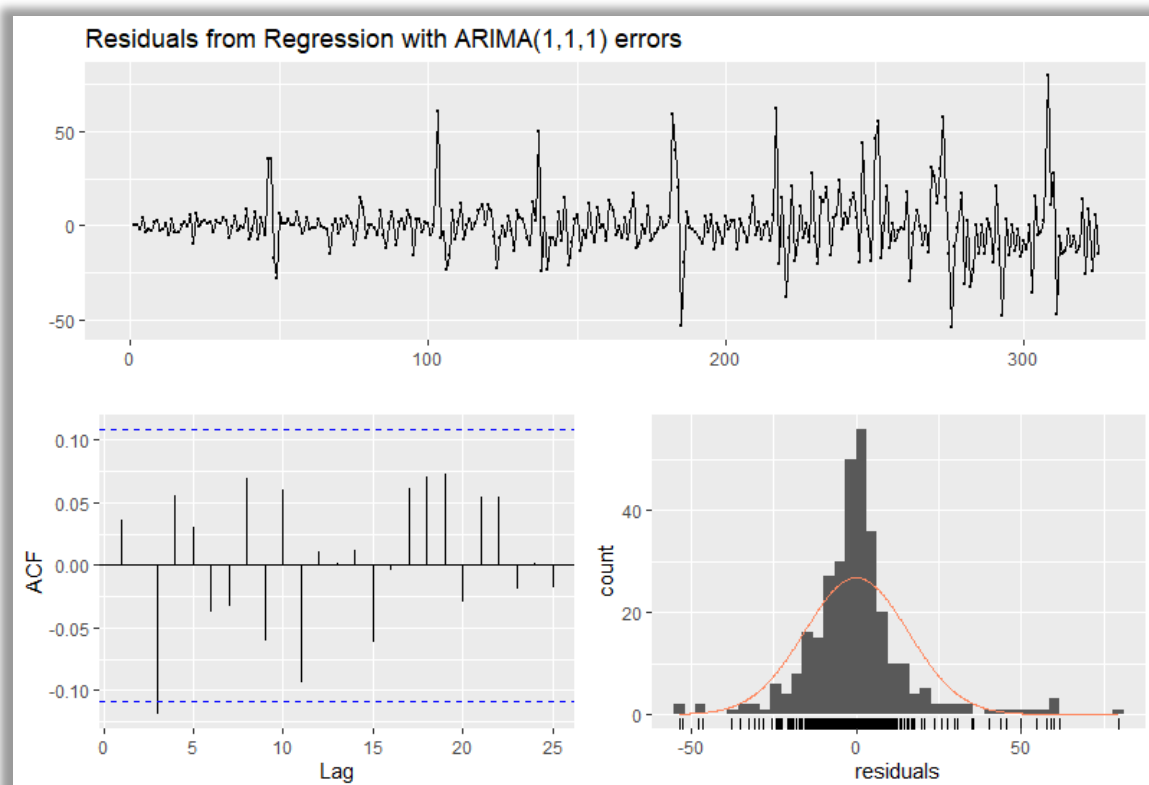
Figure 11. Residuals from ARIMA(0,1,0)

Figure 12. Residuals from Regression with ARIMA(1,1,1) errors

c) La Roche

Daily sales plot and the sales quantity vs price scatter plot for product "85004" can be seen in figure 13 and 14. Before starting forecasting, days with negative prices were omitted from the data. Then the extreme sales were also omitted from the data. Sales of the product "85004" has two different patterns. In the first 300 observation, sales are relatively low and almost stable. It was decided to divide data into two group which are the first 300 days and the last 126 days. Analysis were made according to the two sub datasets. Dynamic regression model gives the best result in terms of AIC and MAPE fort his product.
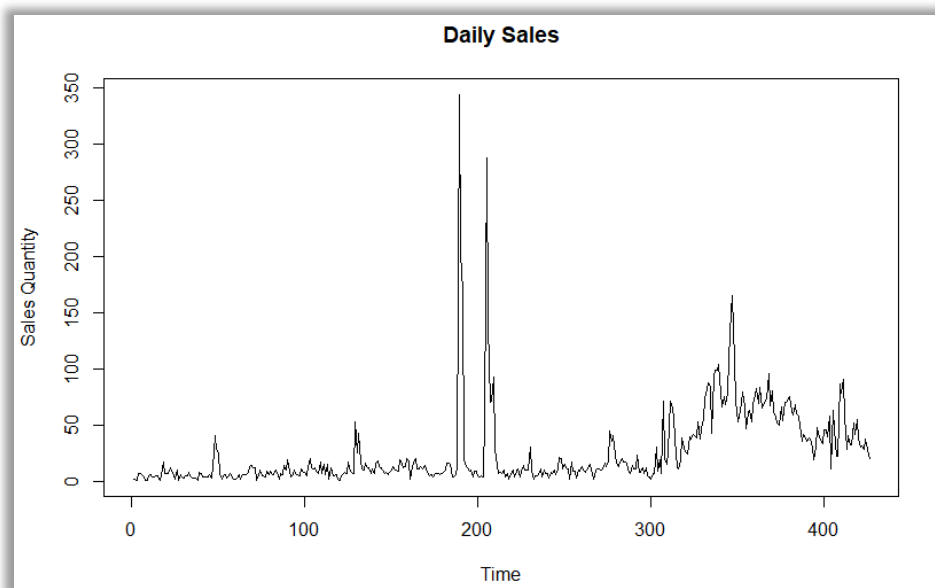
Figure 13. Daily Sales of La Roche
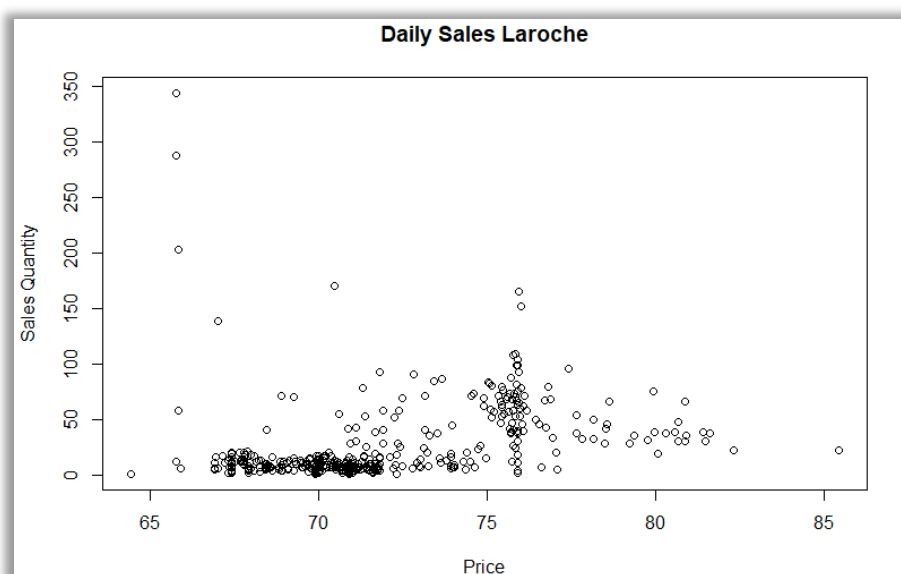
Figure 14. Daily Sales of La Roche
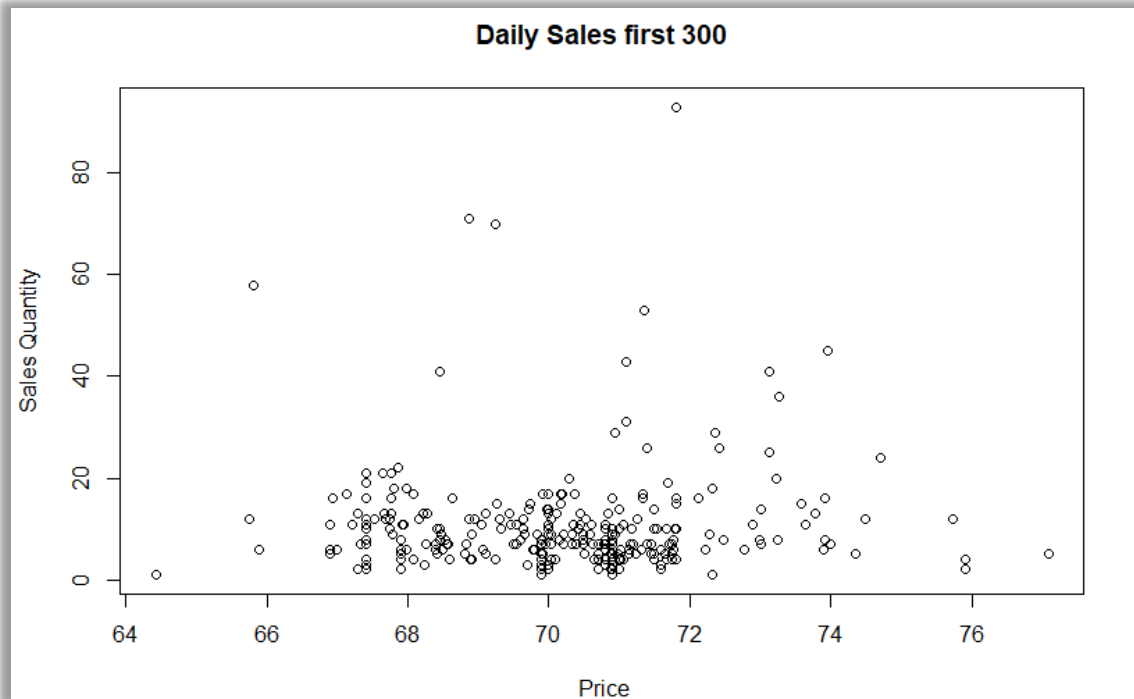
Figure 15. Daily Sales First 300
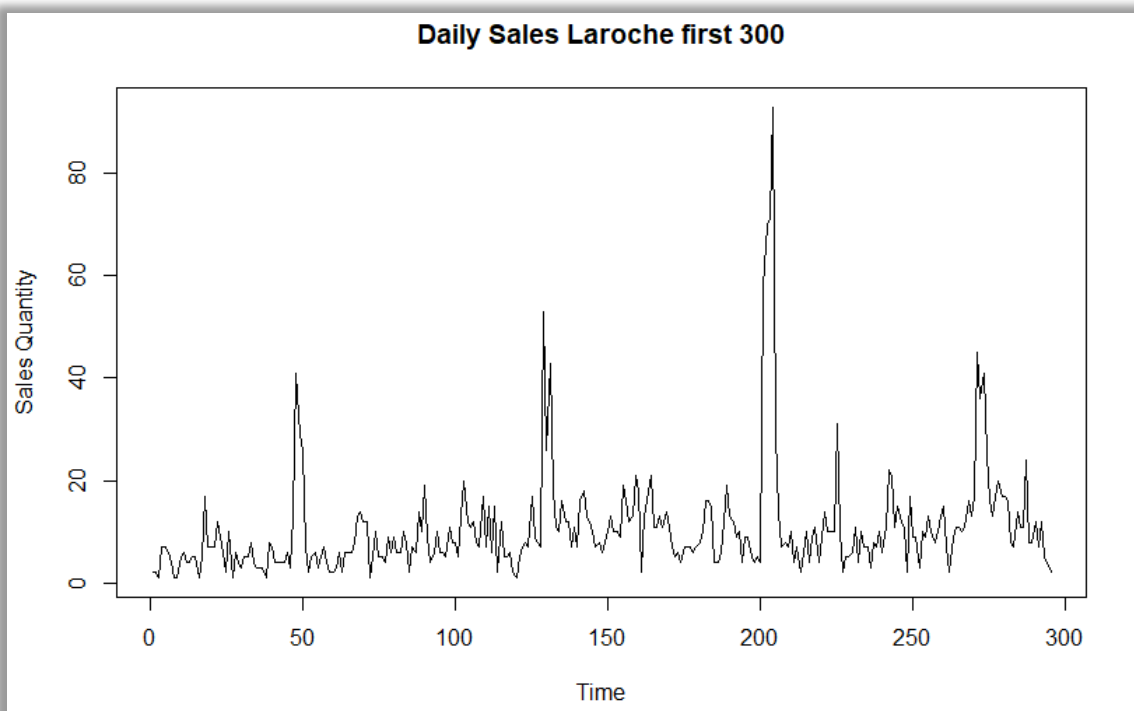
Figure 16. Daily Sales Laroche first 300
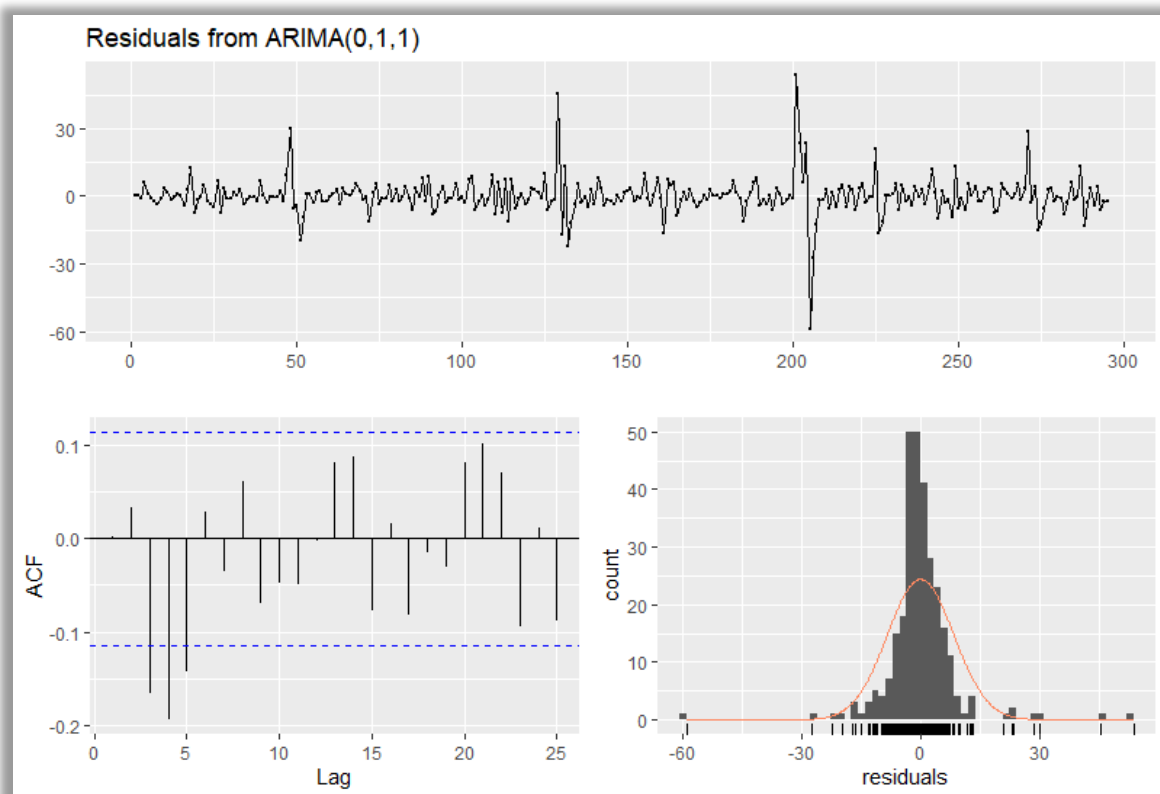
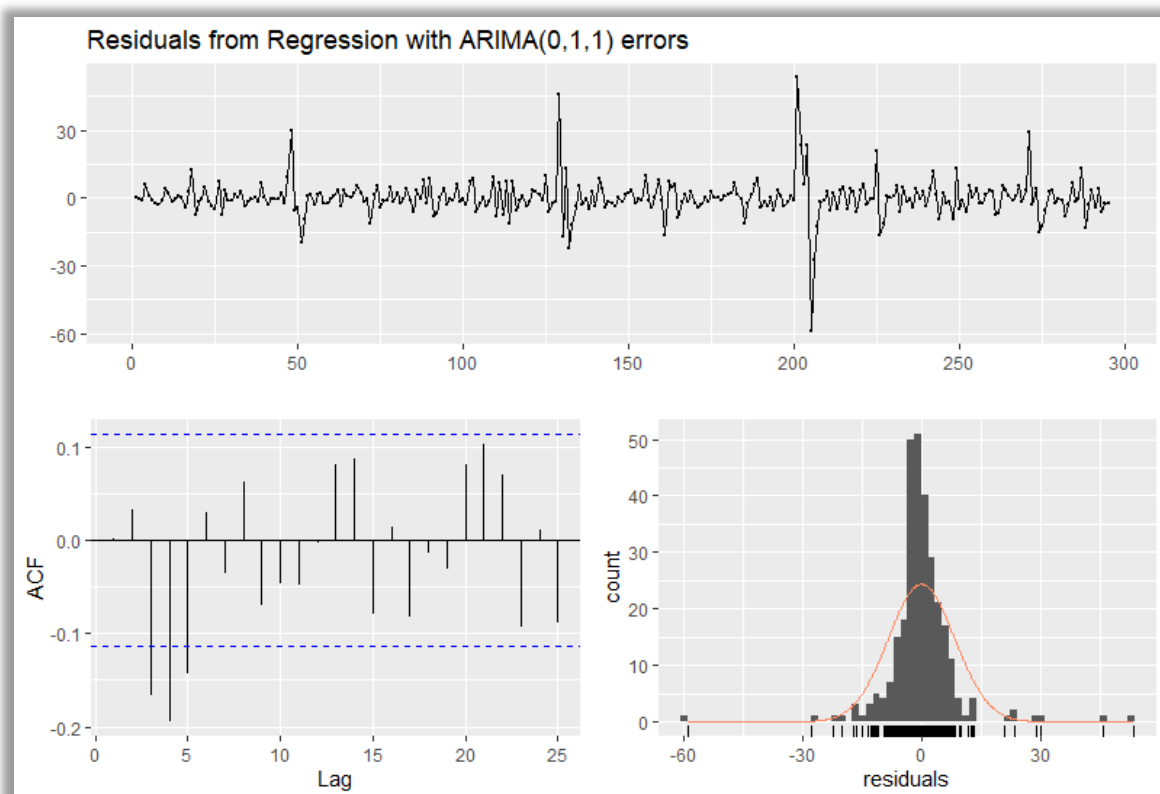**Figure 17**. Residuals from ARIMA (0,1,1)

Figure 18. Residuals from Regression with ARIMA(0,1,1) errors
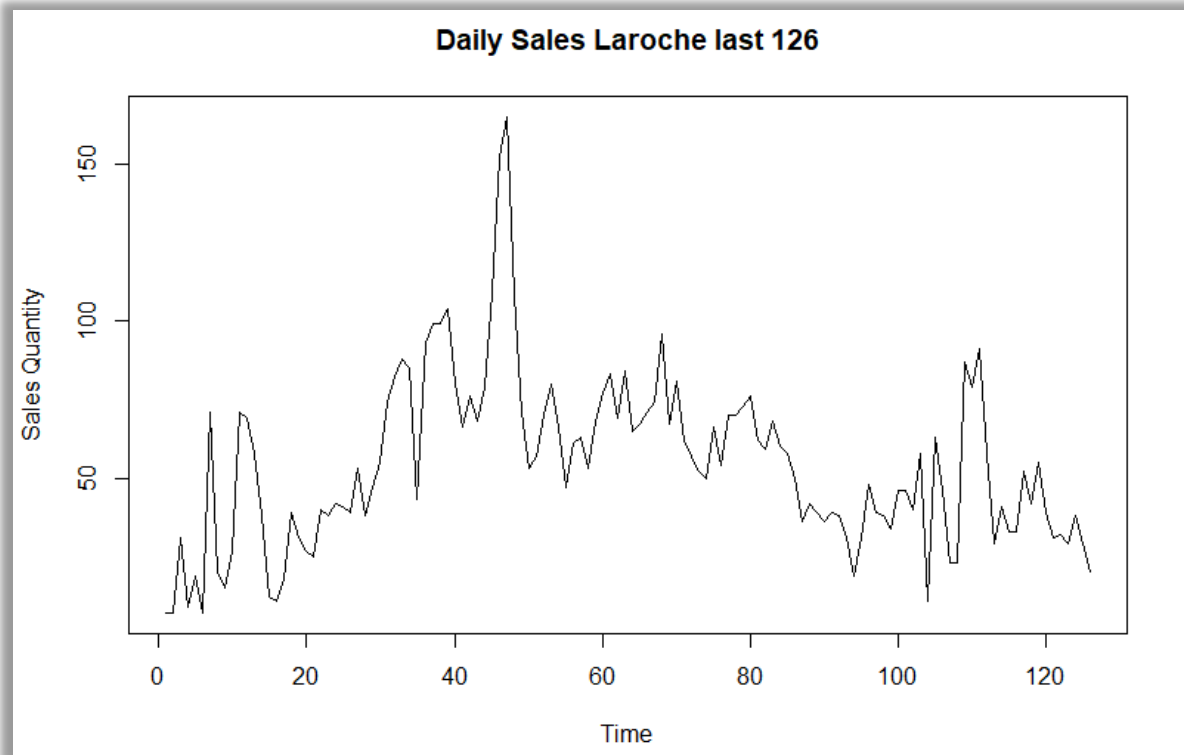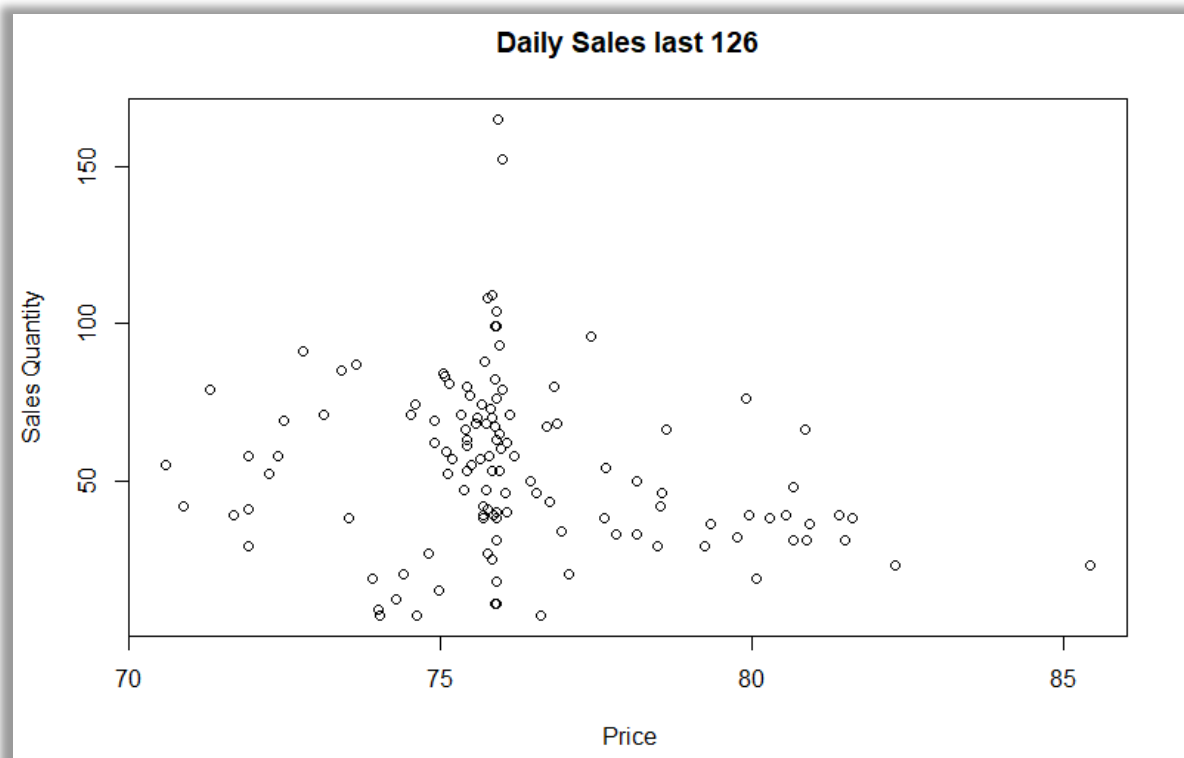
Figure 19. Daily Sales Laroche last 126

Figure 20. Daily Sales last 126

d) Mont

Daily sales plot and the sales quantity vs price scatter plot for product 3904356 can be seen in figure 21 and 22. Before starting the forecasting, days with negative prices and the extreme daily sale quantities which were decided to higher than 20 were omitted from the data. Afterwards, linear regression with sales quantity and price, autoregression with sales quantity and dynamic regression with price models were constructed. Autoregression and dynamic regression model give almost identical AIC and MAPE values for this product. Dynamic regression model was used for forecasting.
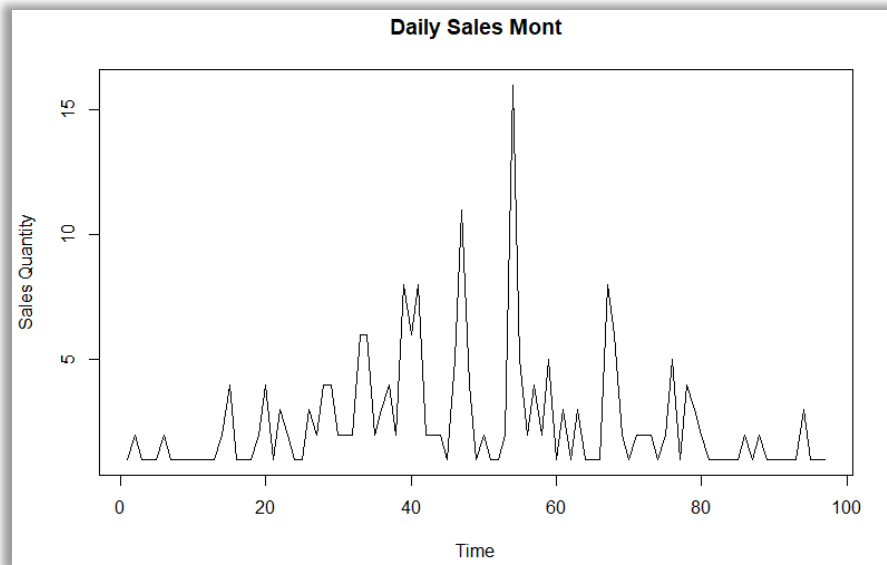
Figure 21. Daily Sales Mont
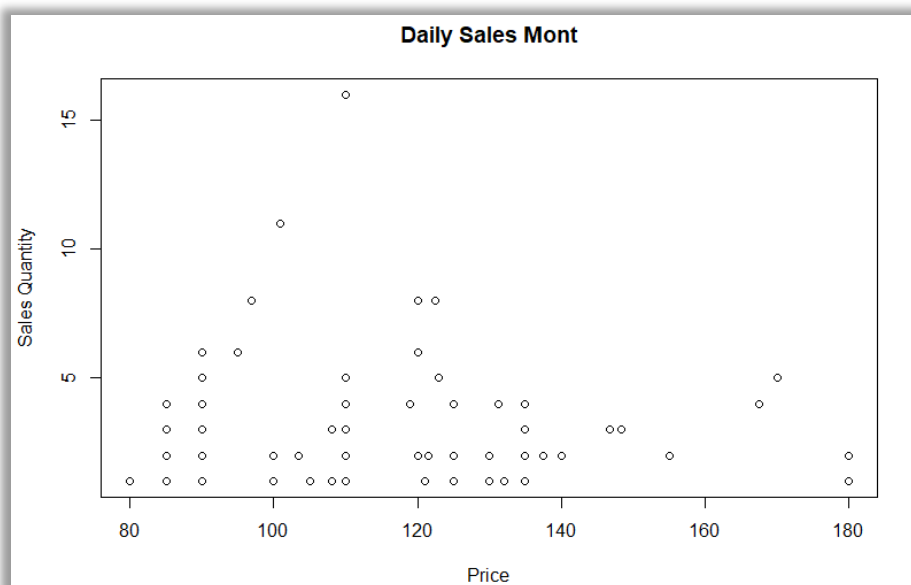
Figure 22. Daily Sales Mont
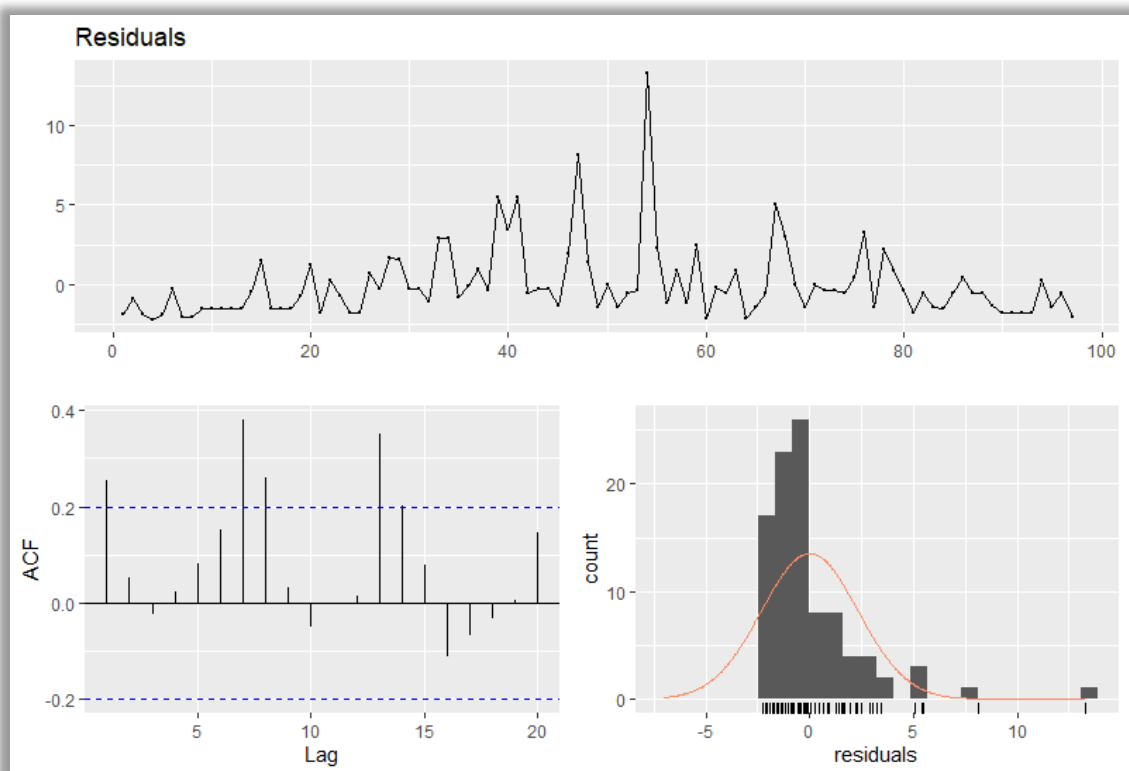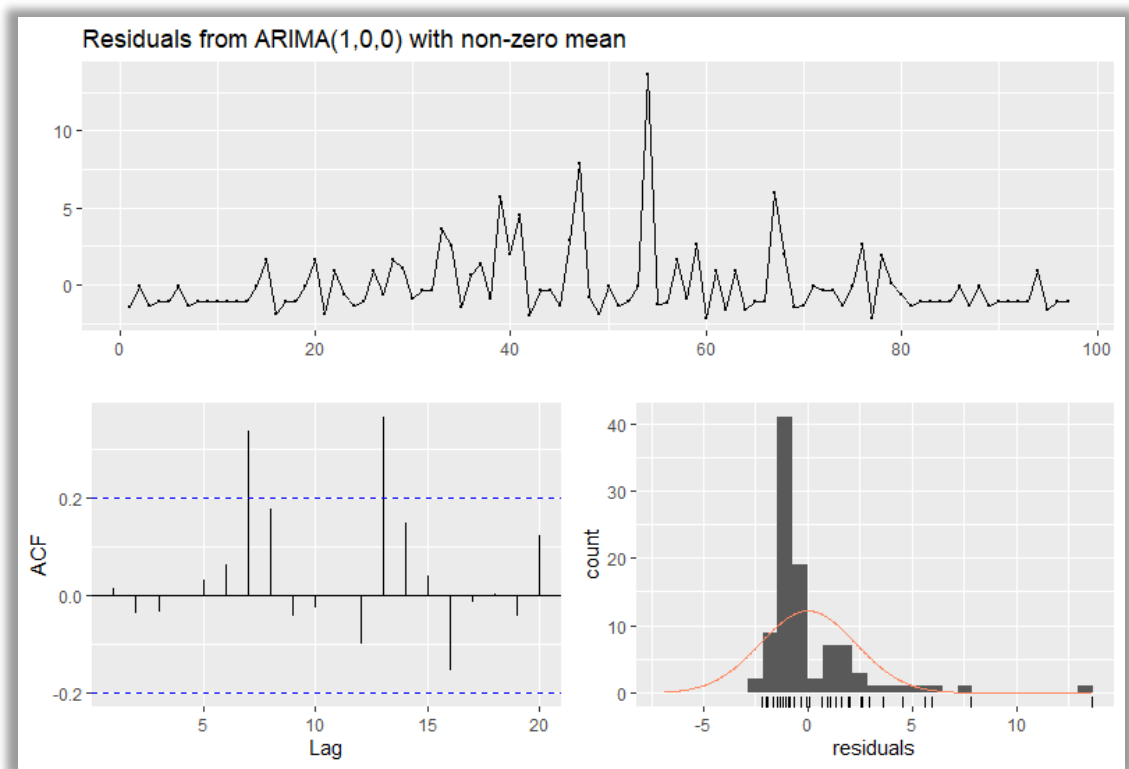
Figure 23. Residuals

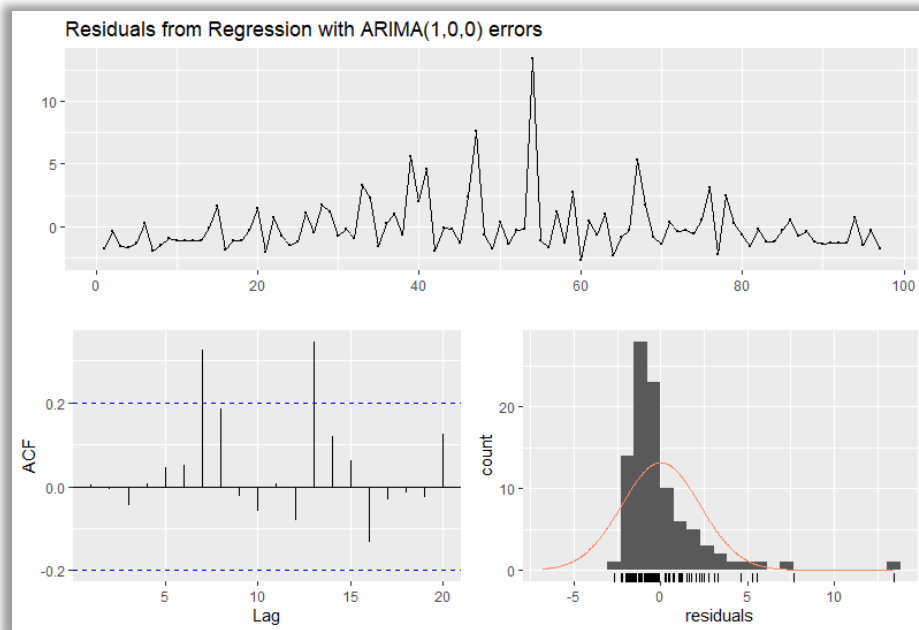**F**igure 24. Residuals from ARIMA(1,0,0) with non-zero mean

Figure 25. Residuals from Regression with ARIMA(1,0,0) errors

e) OralB

Daily sales plot and the sales quantity vs price scatter plot for product 32939029 can be seen in figure 26 and 27. Before starting the forecasting, days with negative prices and the extreme daily sale quantities which were decided to higher than 600 were omitted from the data. Afterwards, linear regression with sales quantity and price, auto regression with sales quantity and dynamic regression with price models were constructed. Autoregression model gives the best MAPE value but dynamic regression model gives better AIC value for this product. ARIMA model was used for forecasting.
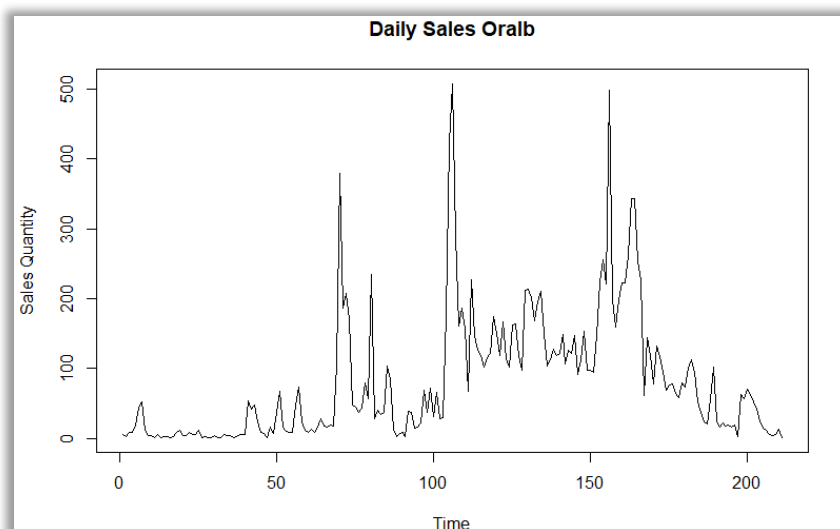
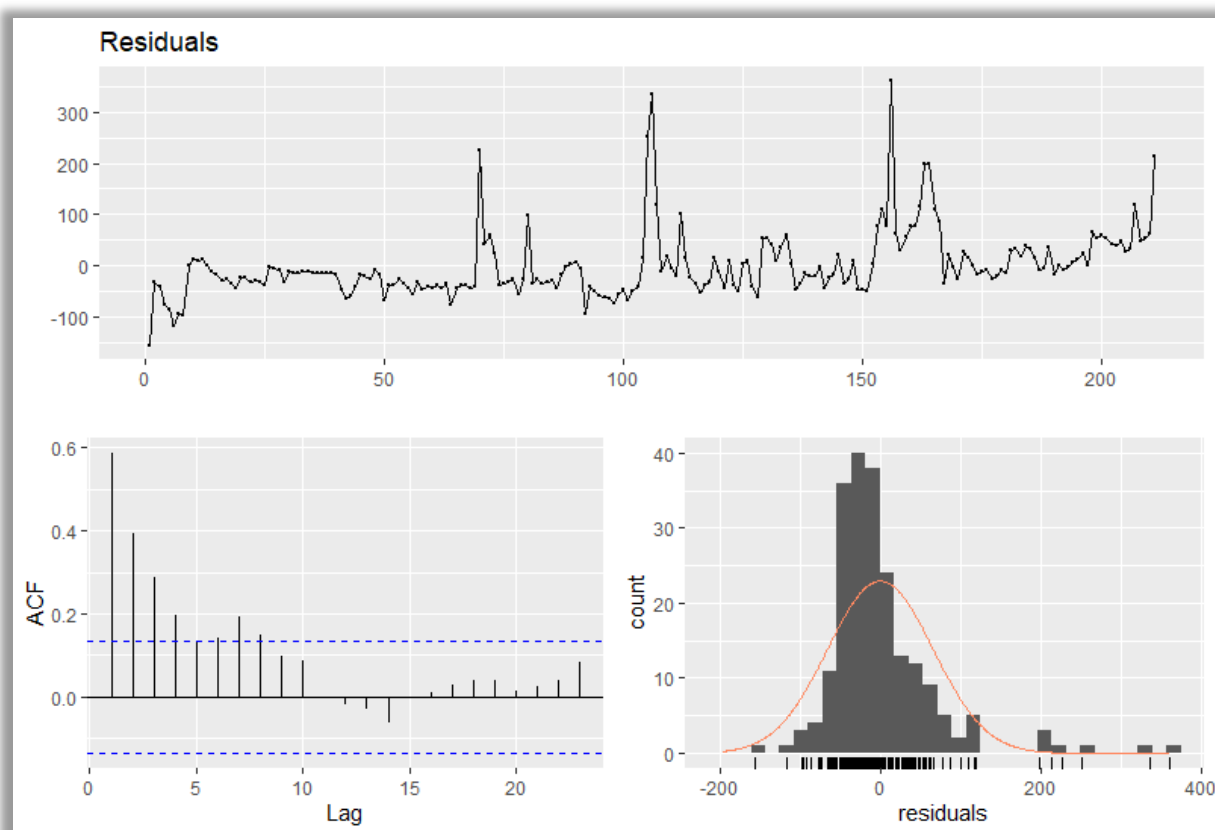**Figure 26**. Daily Sales OralB
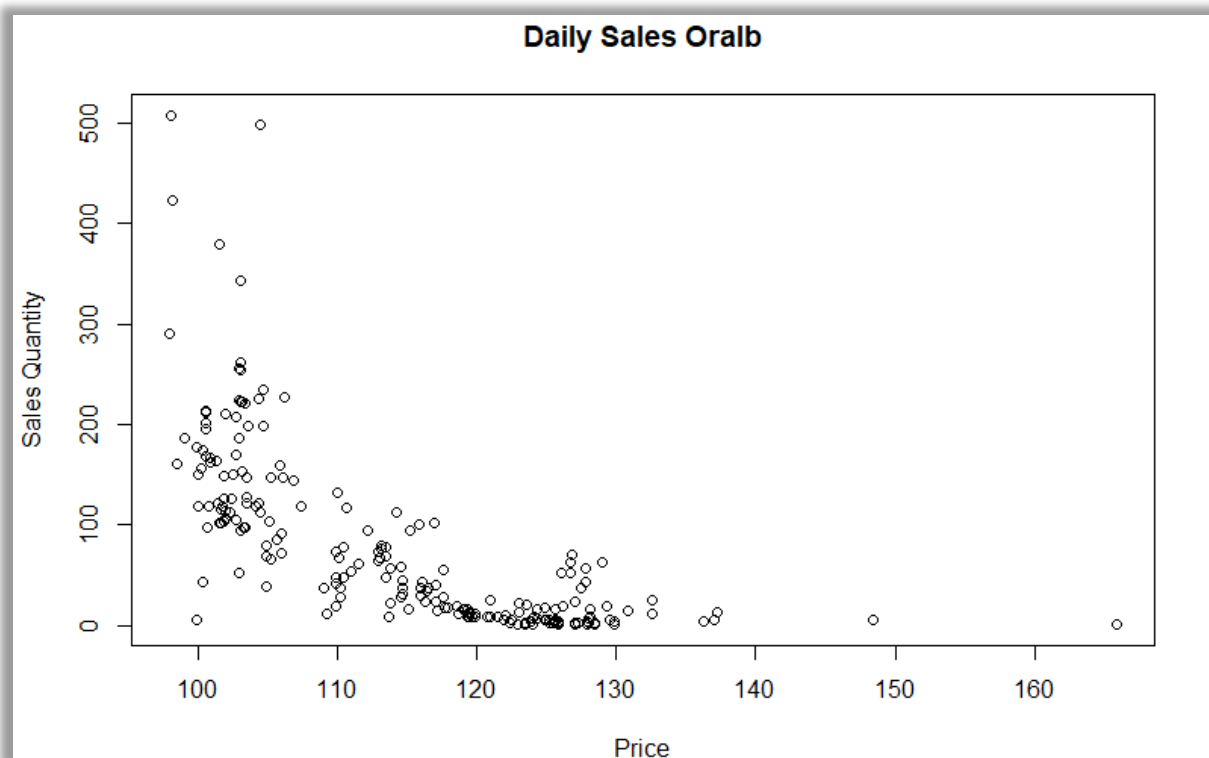
Figure 27. Daily Sales OralB
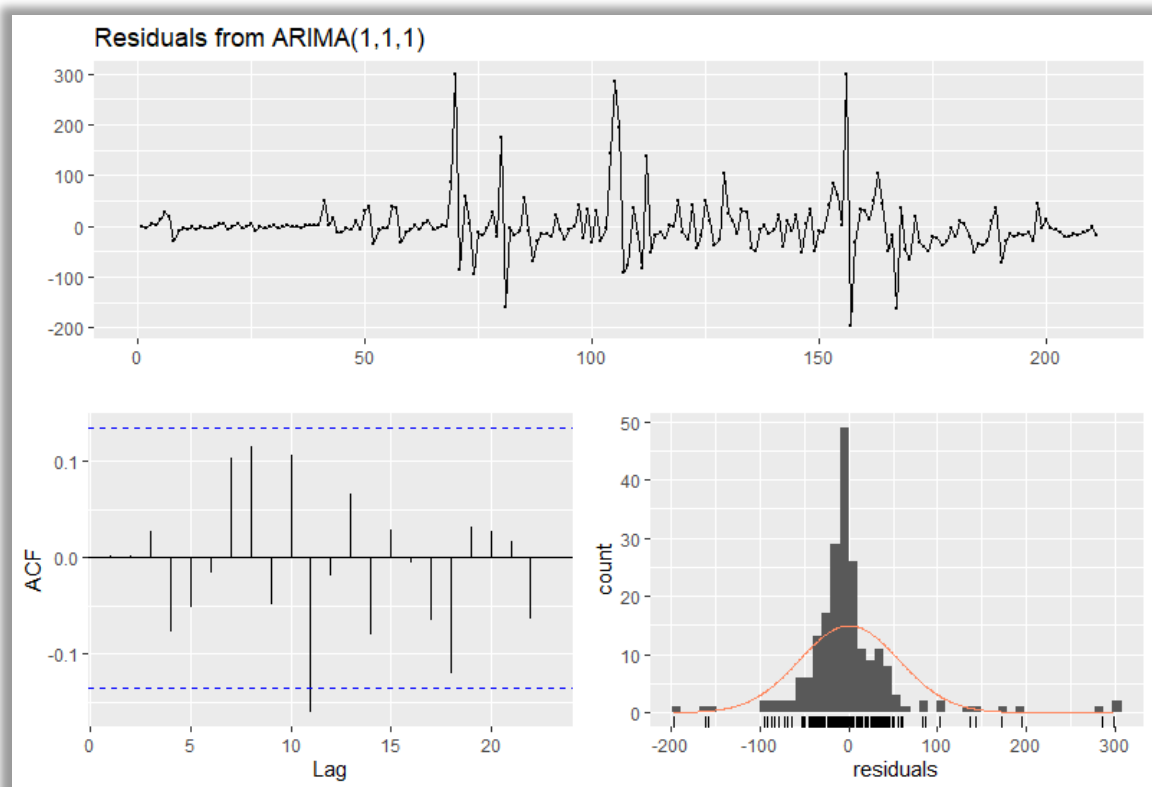

Figure 28. Residuals

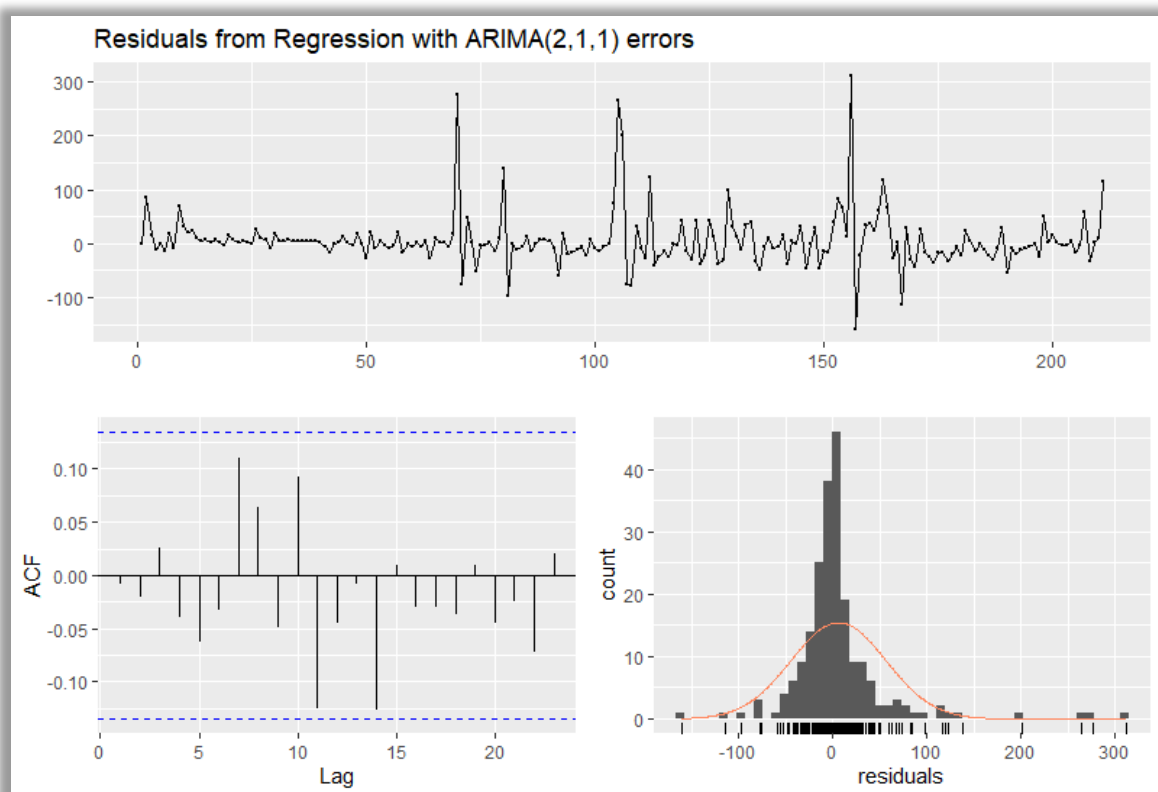**F**igure 29. Residuals from ARIMA(1,1,1)

Figure 30. Residuals from Regression with ARIMA(2,1,1) errors

f) Sleepy

Daily sales plot and the sales quantity vs price scatter plot for product "4066298" can be seen in figure 31 and 32. Before starting the forecasting, days with prices higher than 60 and lower than 60 were omitted from the data. Extreme daily sale quantities which were decided to higher than 20 were also omitted from the data. Afterwards, linear regression with sales quantity and price, auto regression with sales quantity and dynamic regression with price models were constructed. Dynamic regression model gives the best result in terms of AIC and MAPE fort his product.
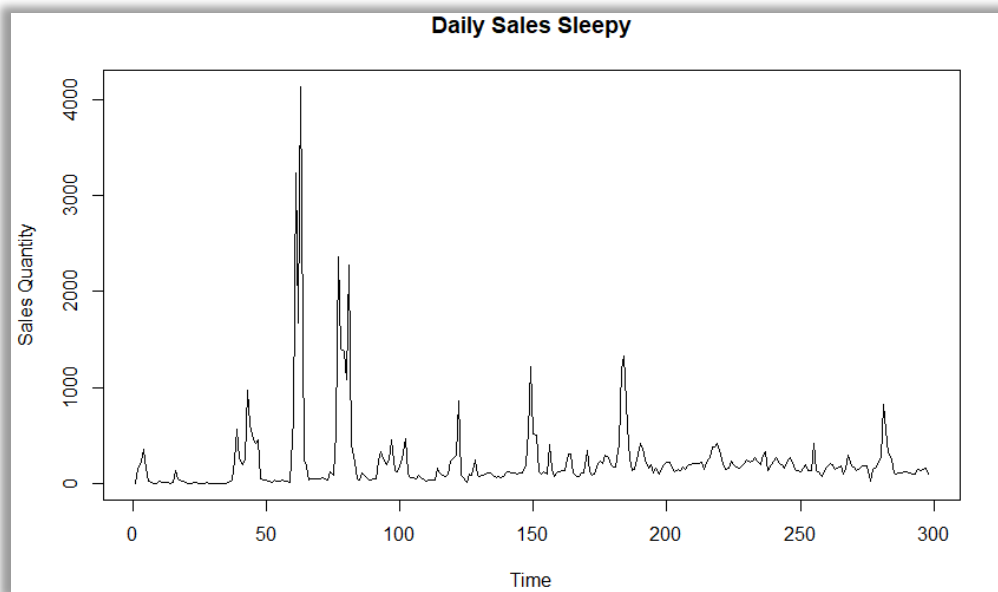
Figure 31. Daily Sales Sleepy
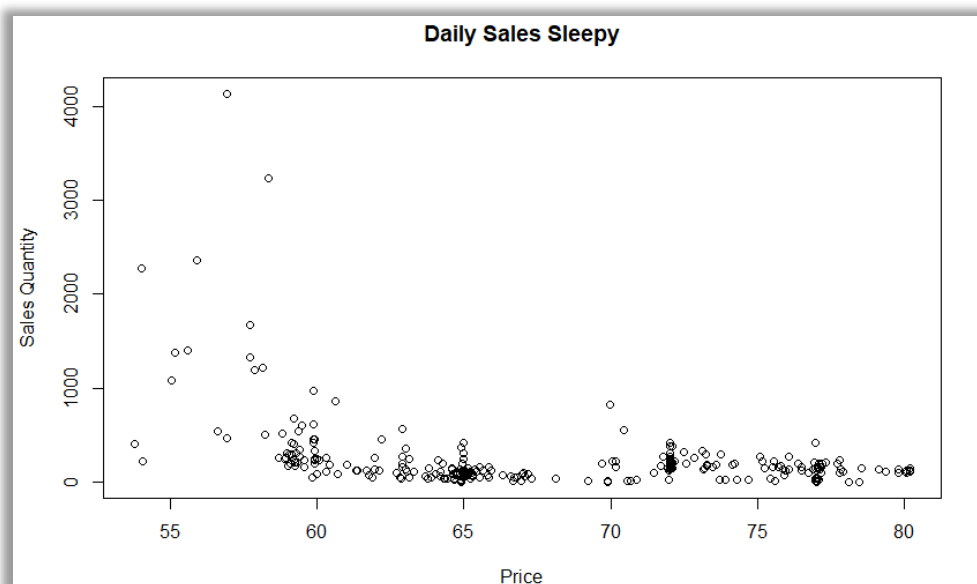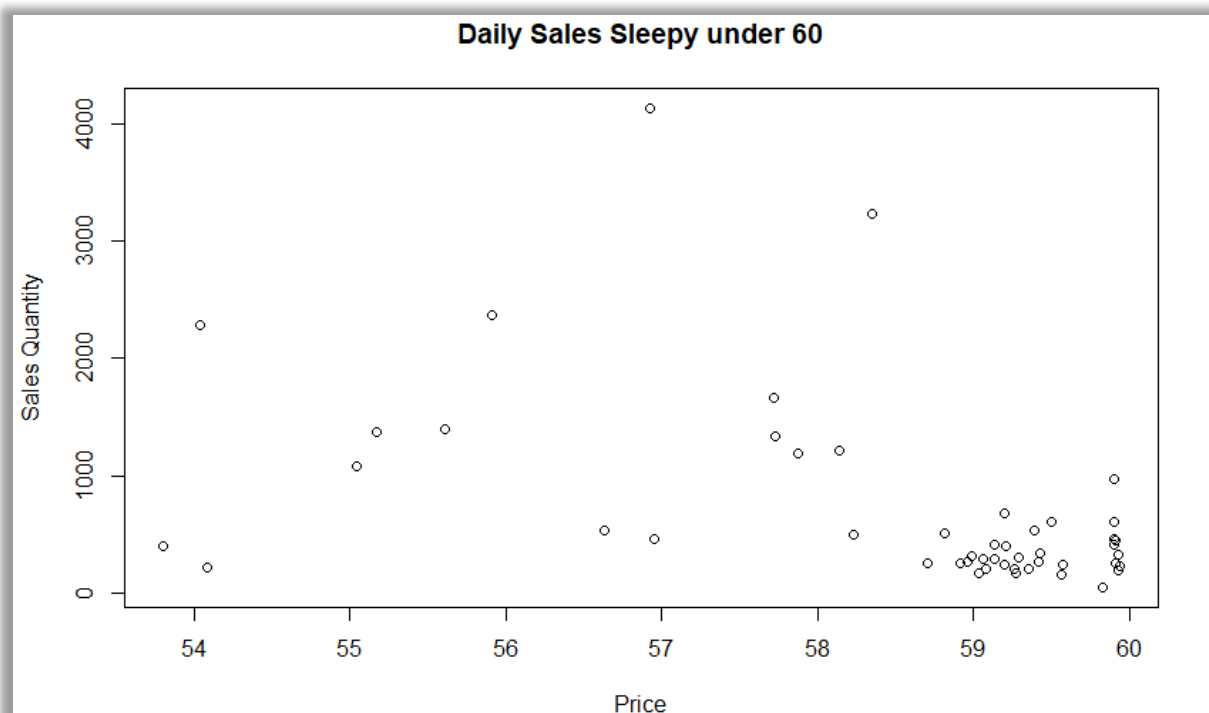

Figure 32. Daily Sales Sleepy

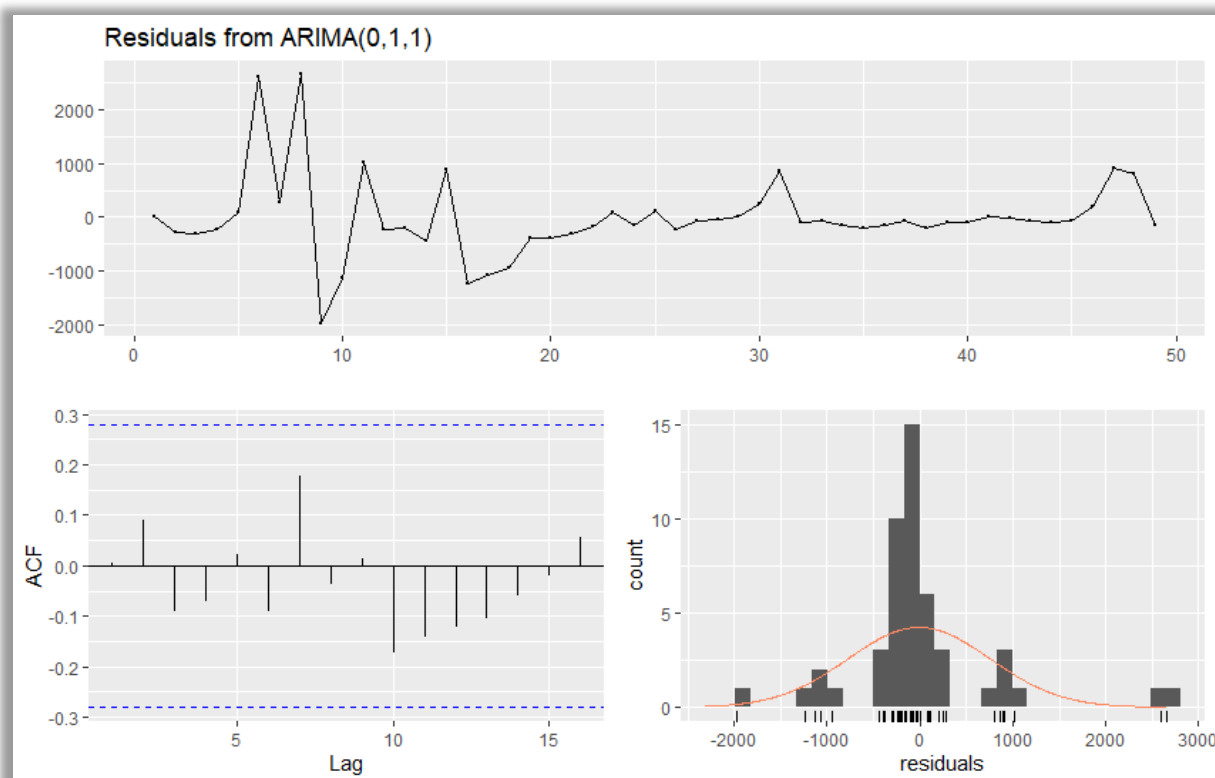**Figure 33**. Daily Sales Sleepy Under 60

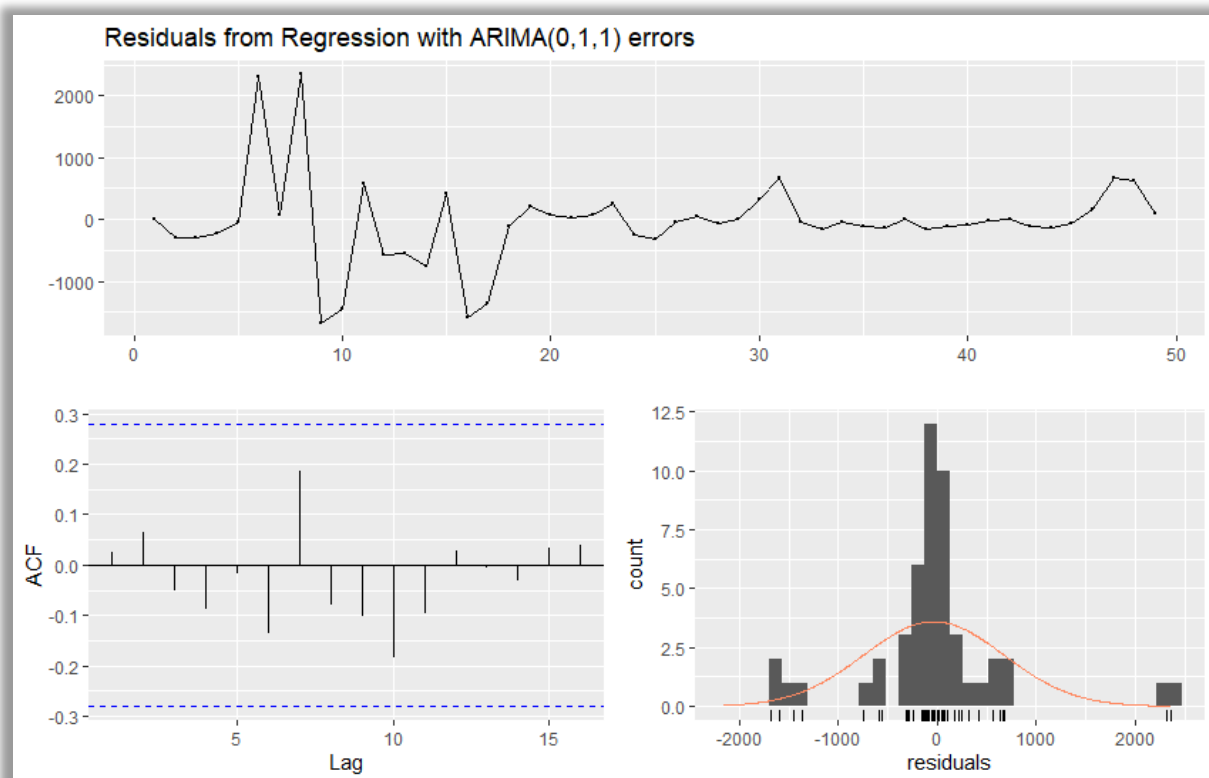**Figure 34**. Residuals from ARIMA(0,1,1)

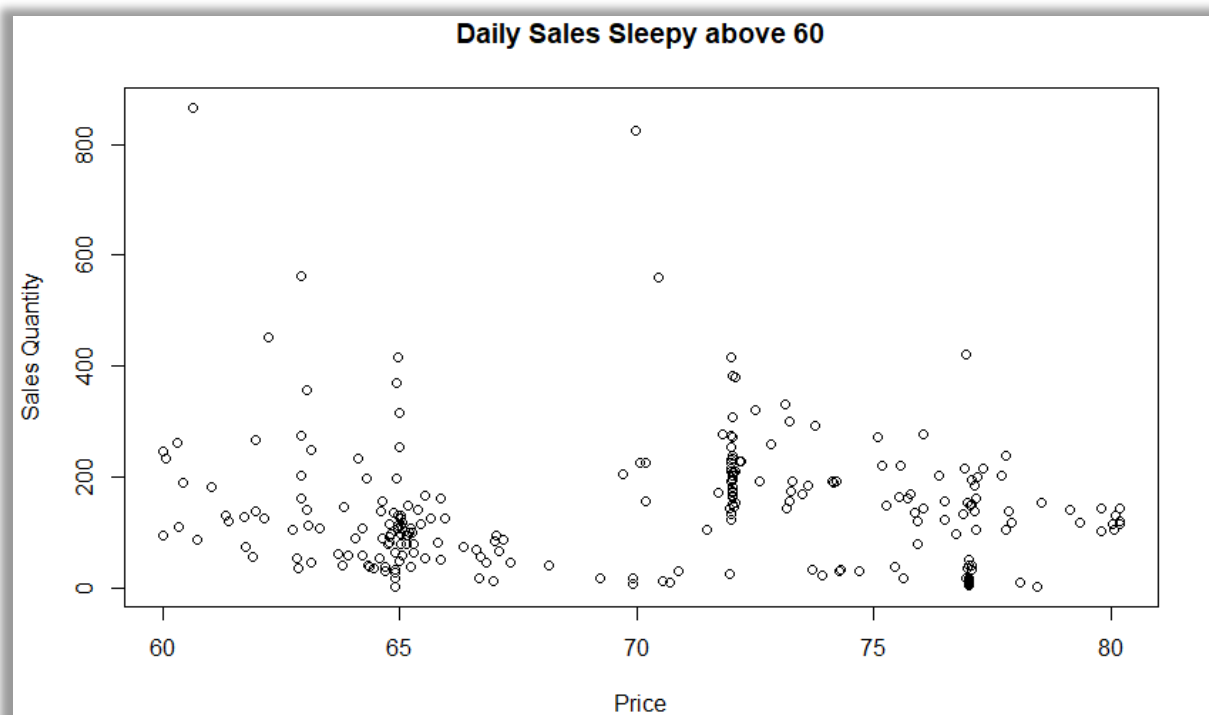**Figure 35**. Residuals from Regression with ARIMA(0,1,1) errors
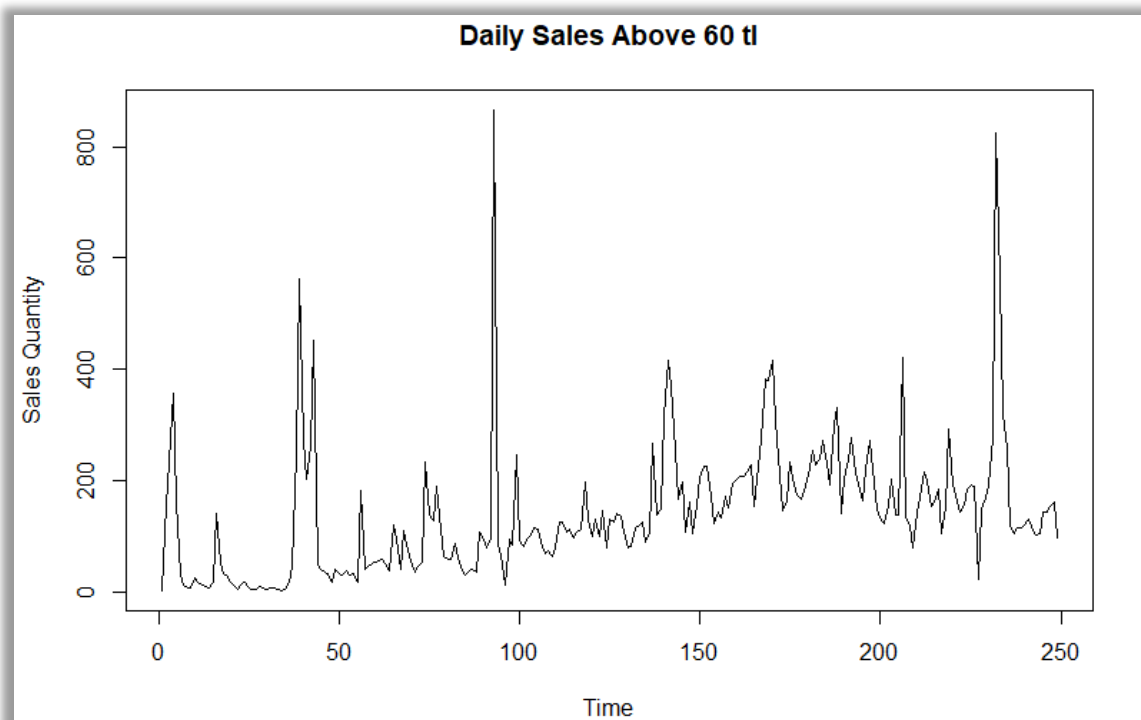

Figure 36. Daily Sales Sleepy Above 60
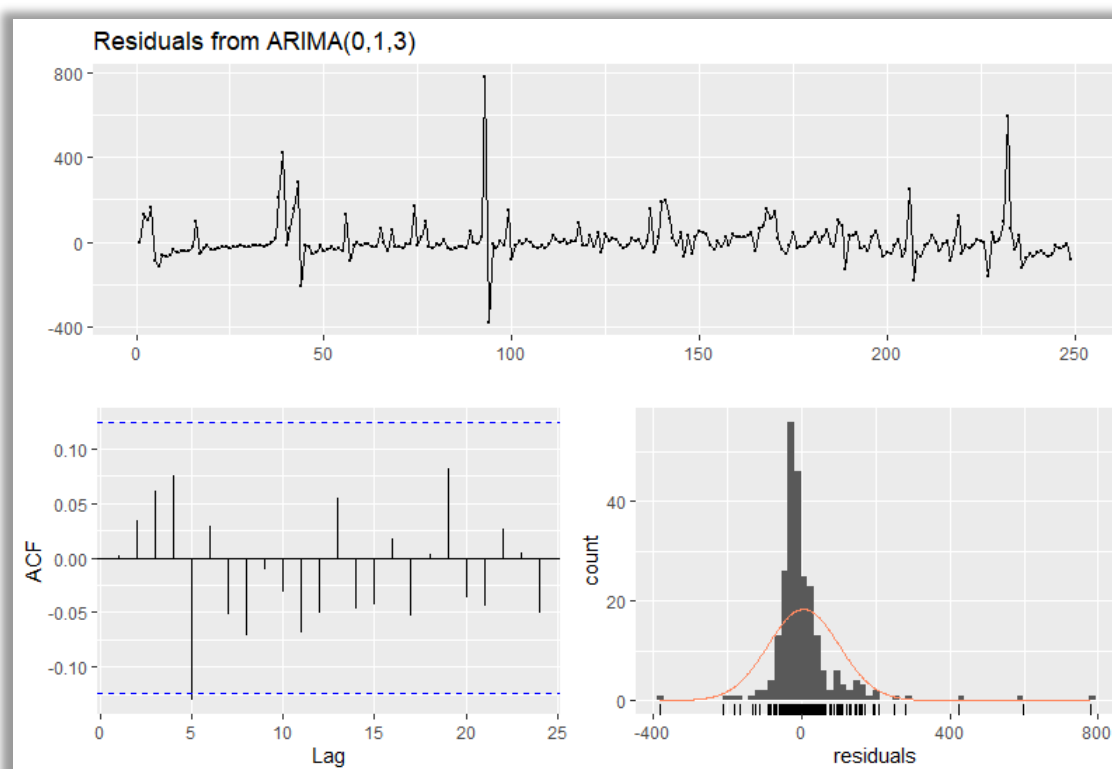
Figure 37. Daily Sales Sleepy Above 60 tl

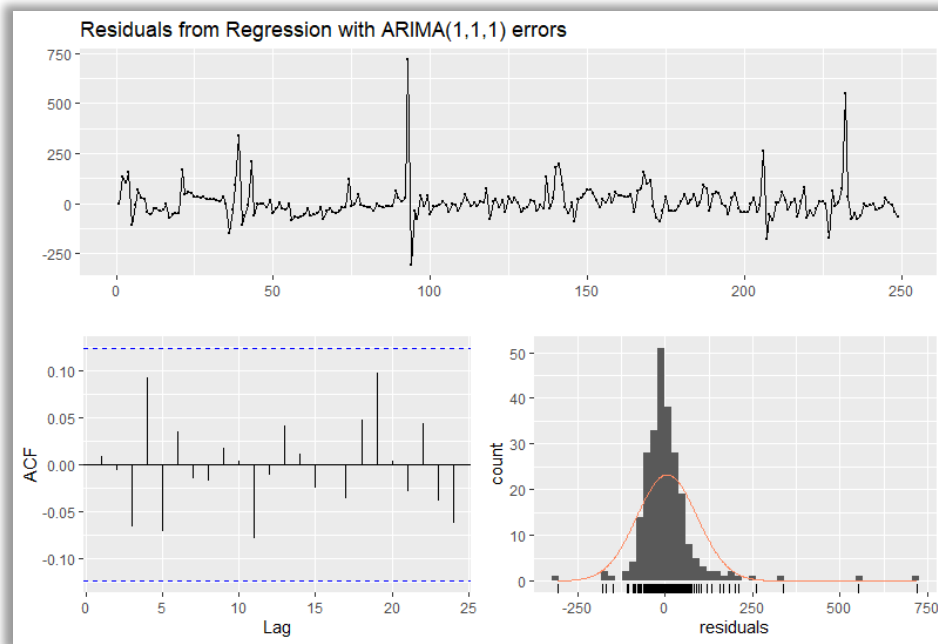
Figure 38. Residuals from ARIMA(0,1,3)

Figure **39**. Residuals from Regression with ARIMA(1,1,1) errors

g) Tayt

Daily sales plot and the sales quantity vs price scatter plot for product "31515569" can be seen in figure 40 and 41. Before starting the forecasting, days with negative prices and the extreme daily sale quantities which were decided to lower than 20 were omitted from the data. Afterwards, linear regression with sales quantity and price, auto regression with sales quantity and dynamic regression with price models were constructed. Autoregression and dynamic regression model give almost identical AIC and MAPE values for this product. Dynamic regression model was used for forecasting.
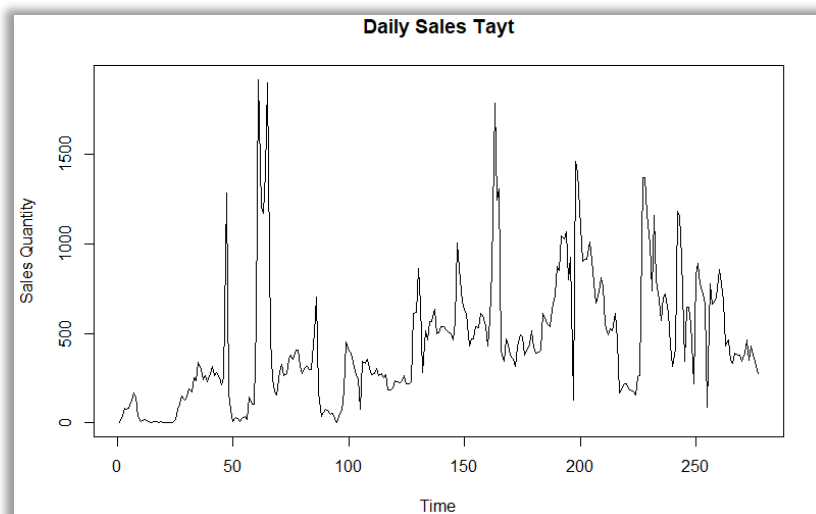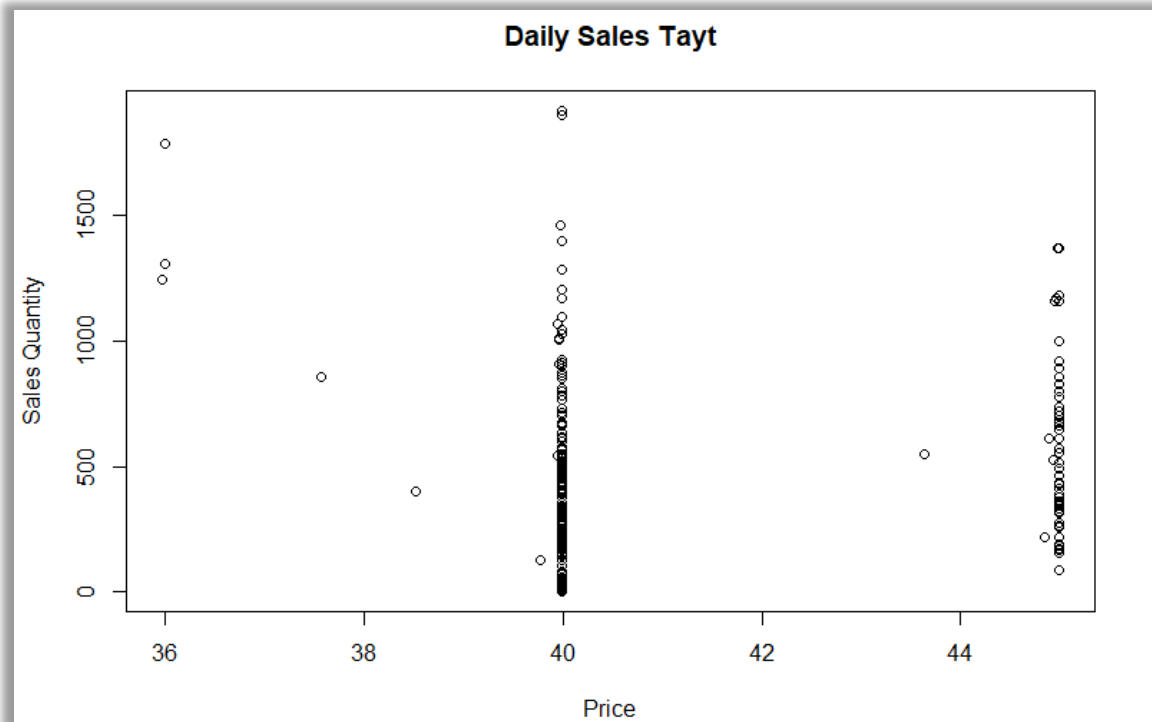

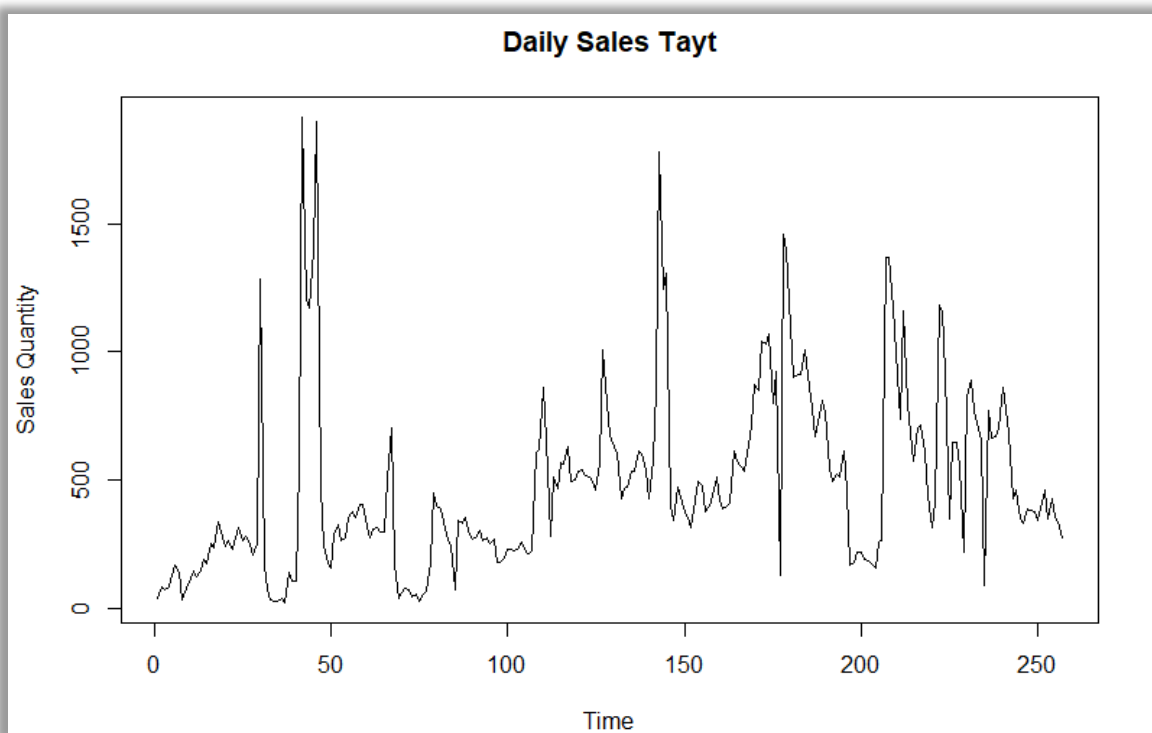**F**igure **40**. Daily Sales Tayt

Figure 41. Daily Sales Tayt

Figure 42. Daily Sales Tayt
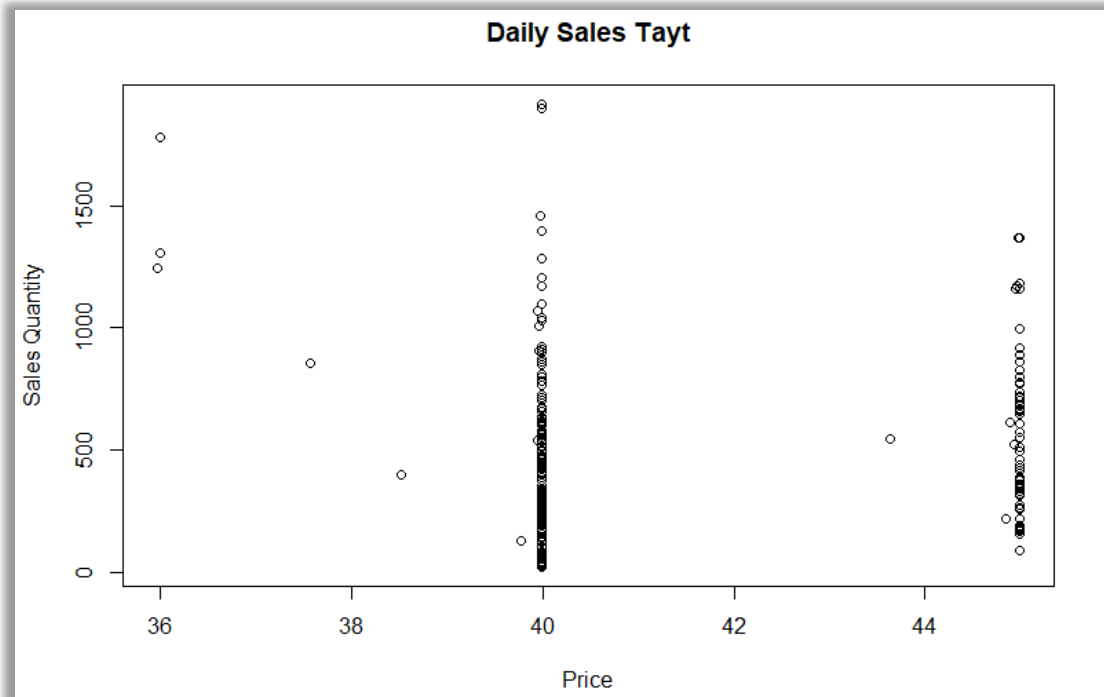
Figure 43. Daily Sales Tayt

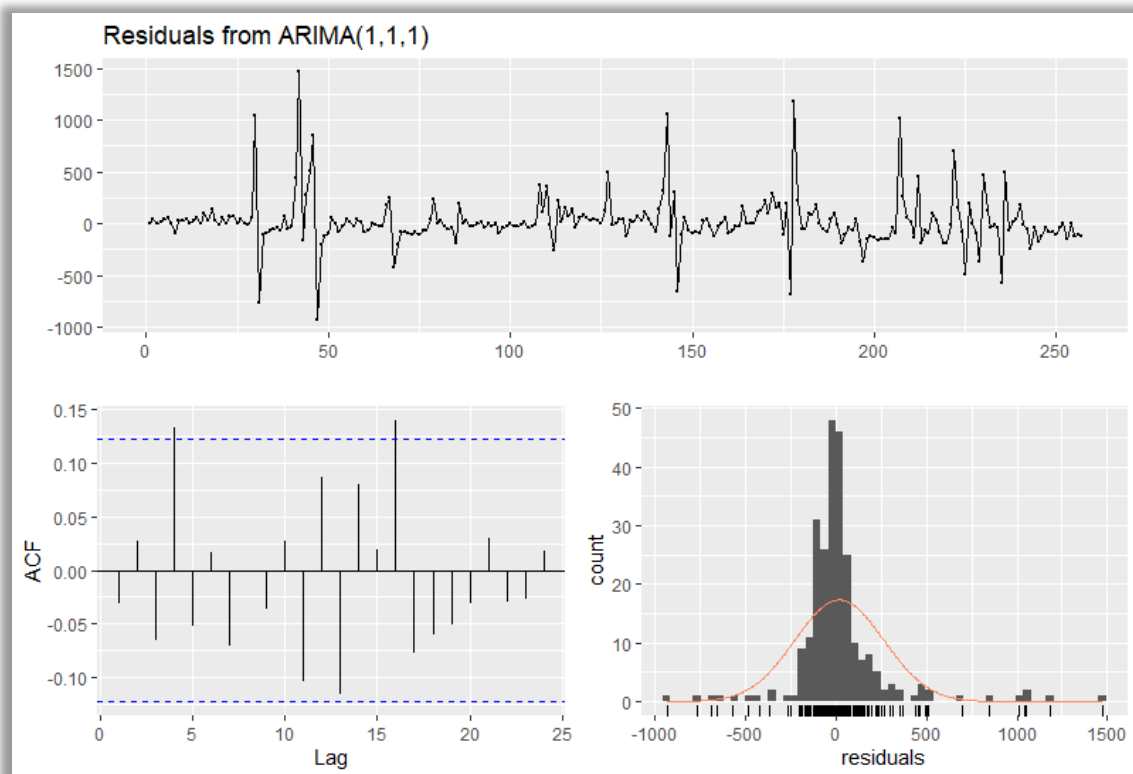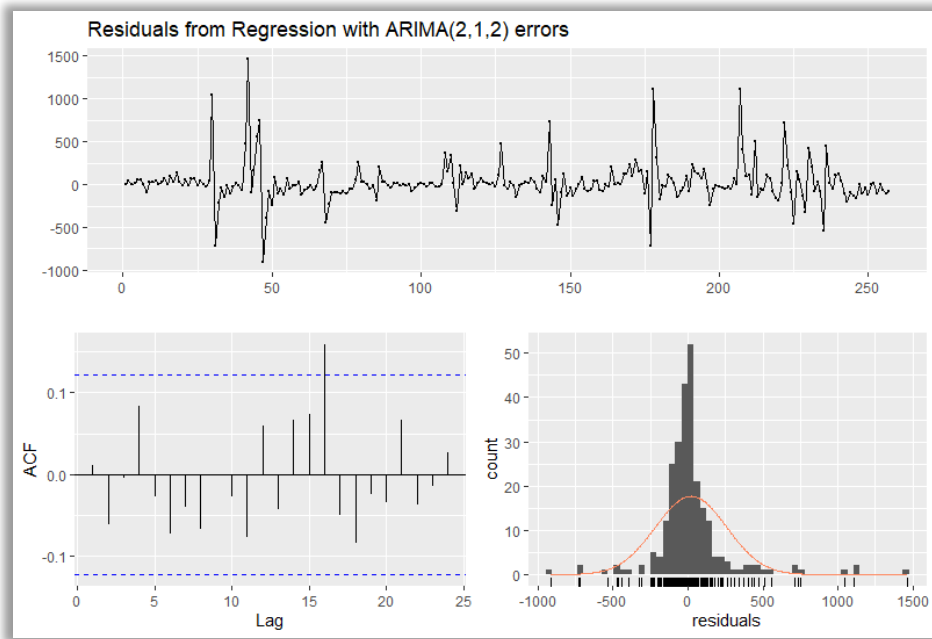**Figure 44**. Residuals from ARIMA (1,1,1)

Figure 45. Residuals from ARIMA (1,1,1)

h) Xiaomi

Daily sales plot and the sales quantity vs price scatter plot for product "6676673" can be seen in figure 46 and 47. Before starting the forecasting, days with negative prices and the extreme daily sale quantities which were decided to higher than 2000 were omitted from the data. Afterwards, linear regression with sales quantity and price, auto regression with sales quantity and dynamic regression with price models were constructed. ARIMA model gives the best result in terms of AIC and MAPE fort his product.
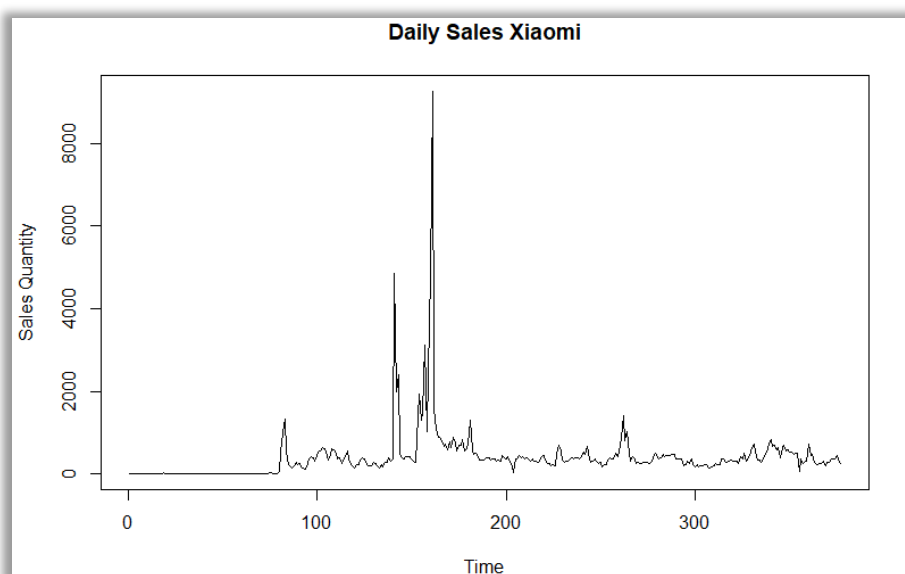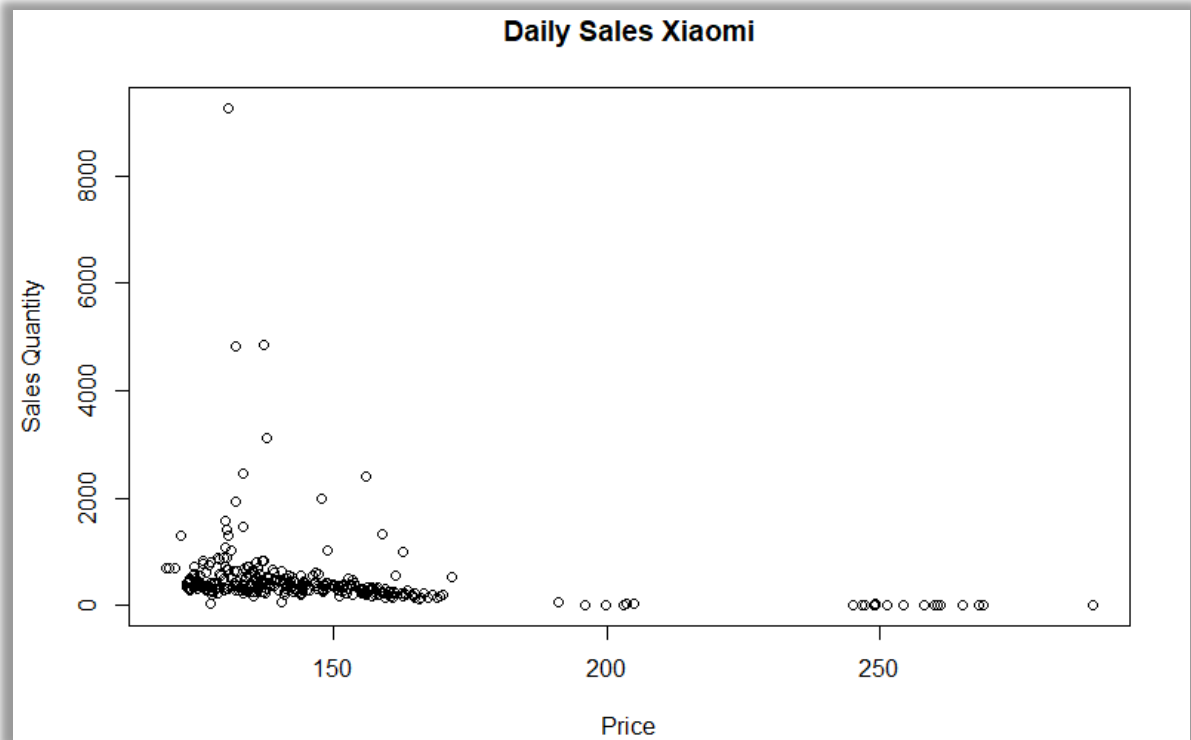

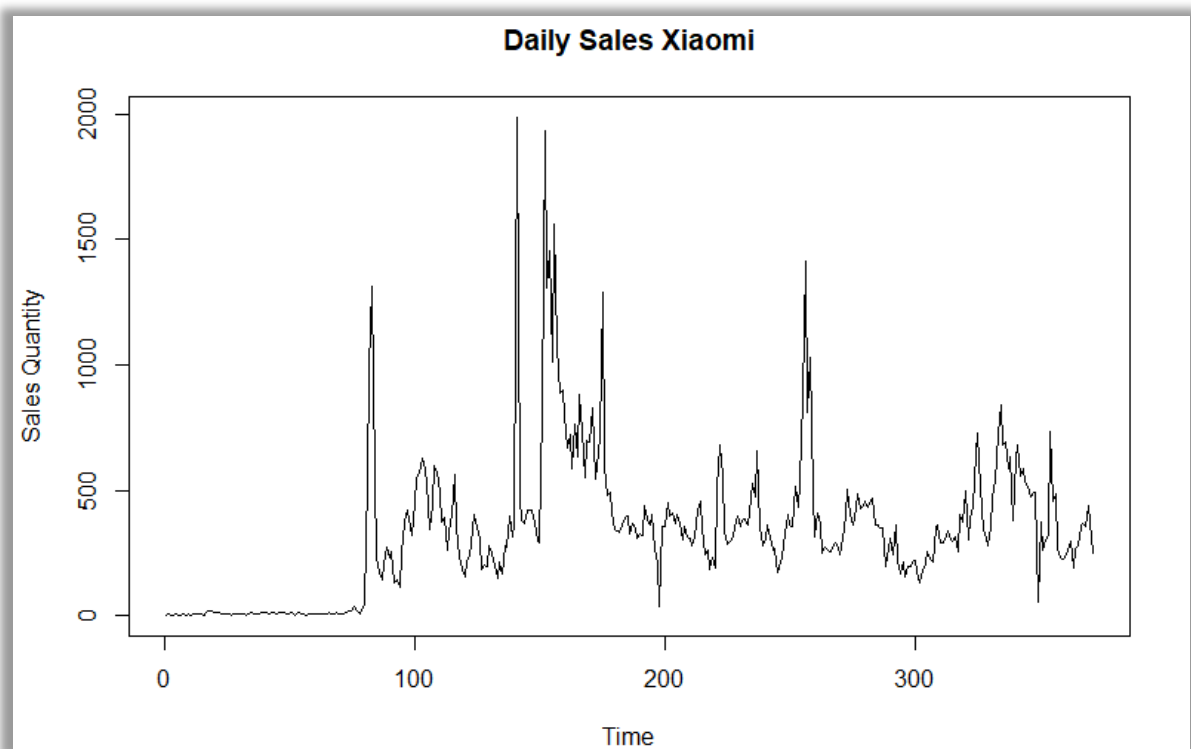**Figure 46**. Daily Sales Xiaomi

Figure 47. Daily Sales Xiaomi

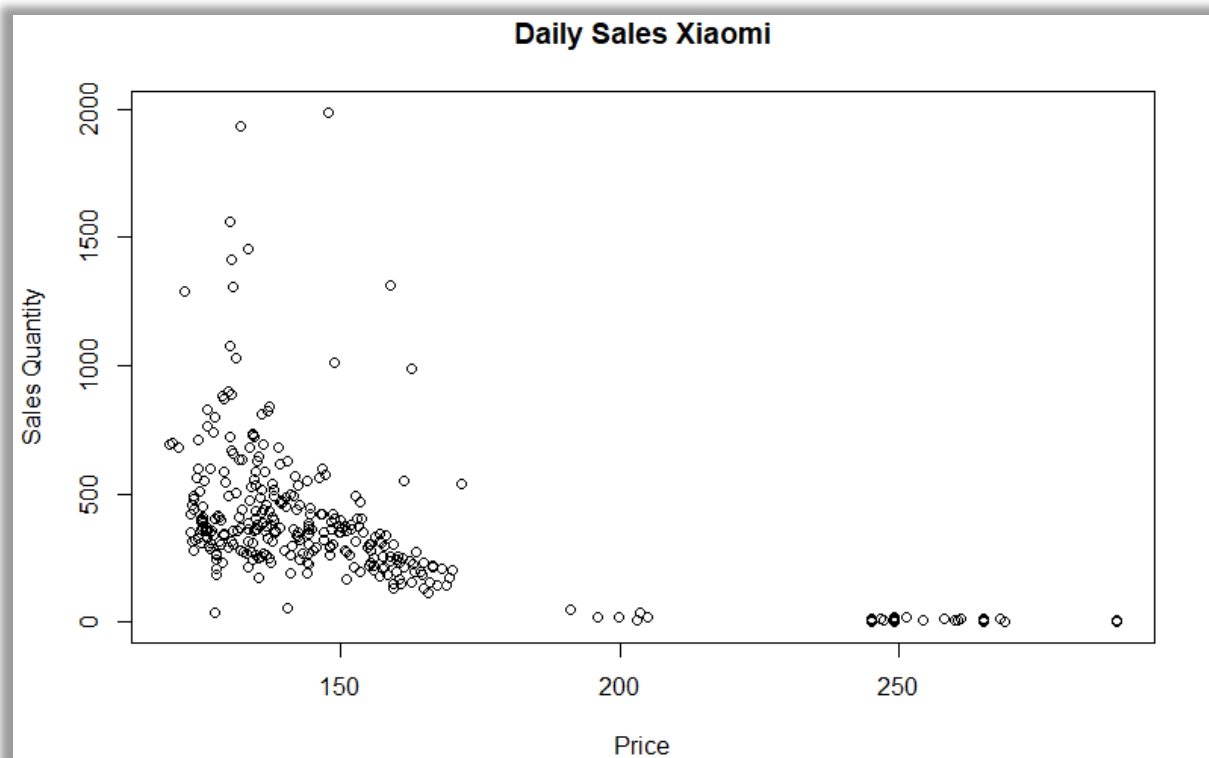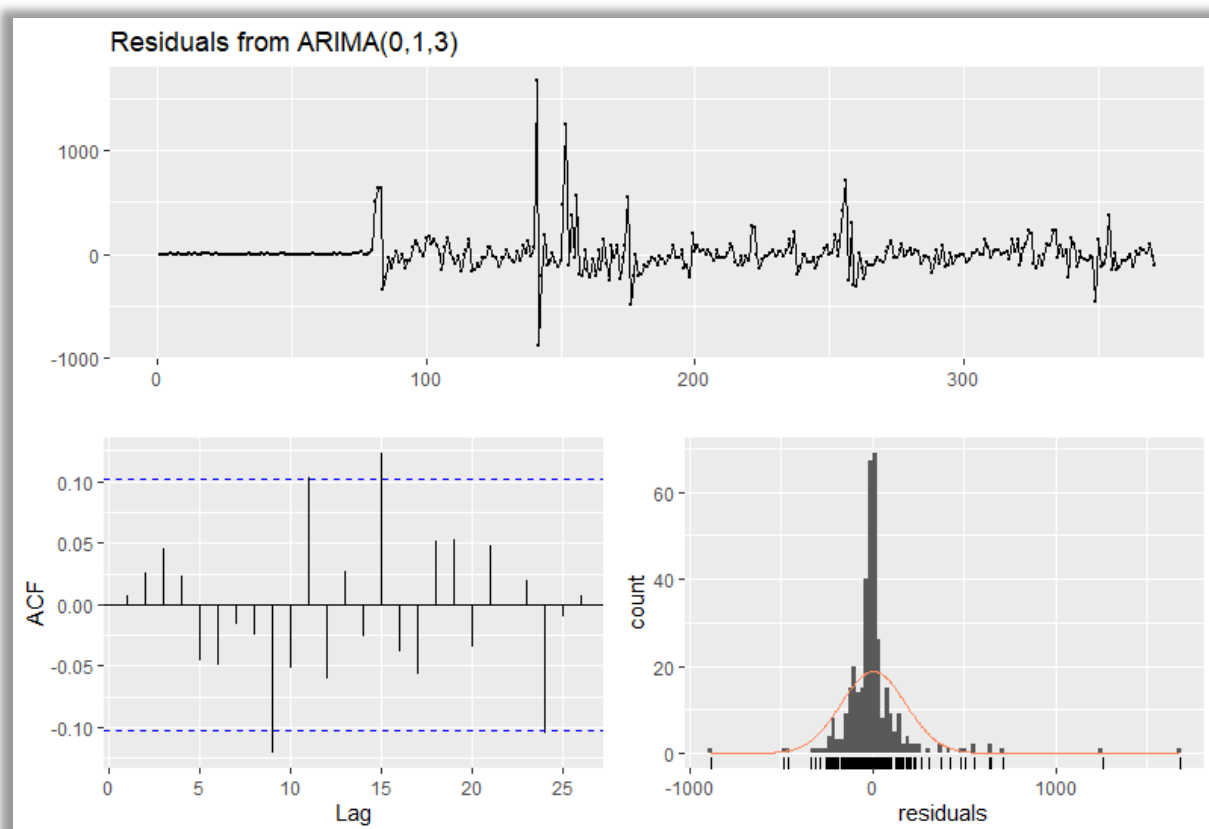**Figure 48**. Daily Sales Xiaomi

Figure 49. Daily Sales Xiaomi

Figure 50. Residuals from ARIMA(0,1,3)
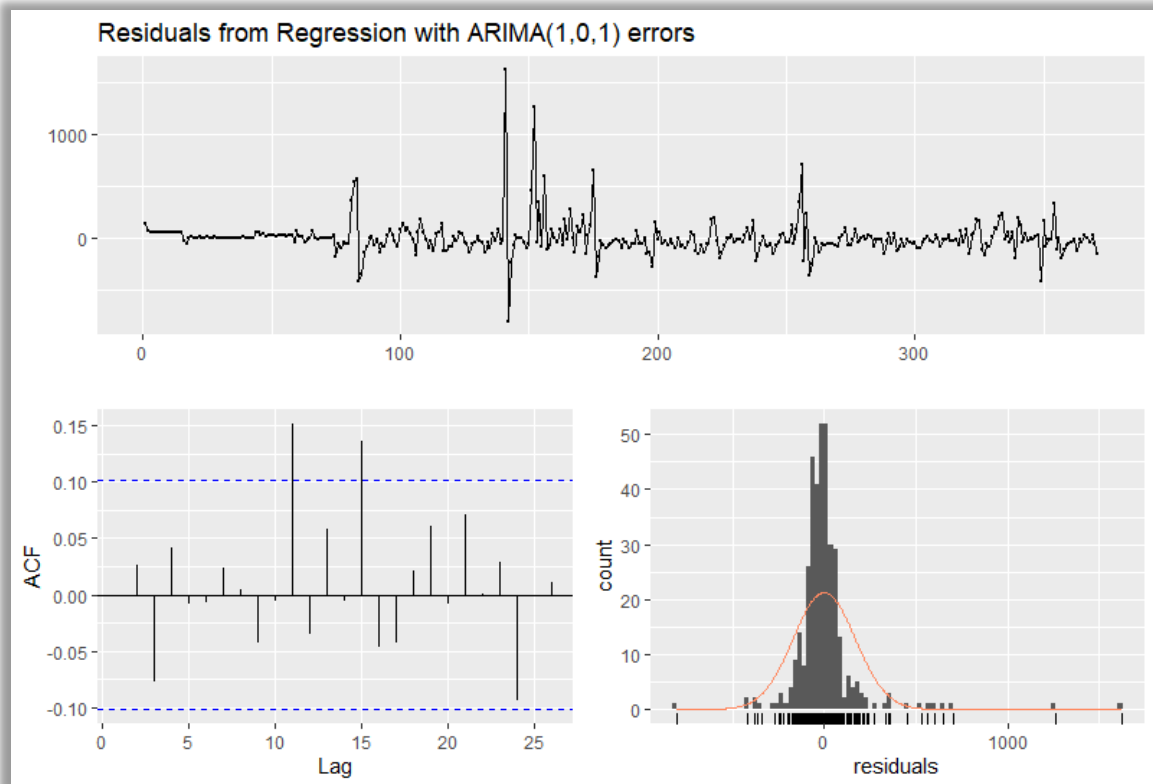
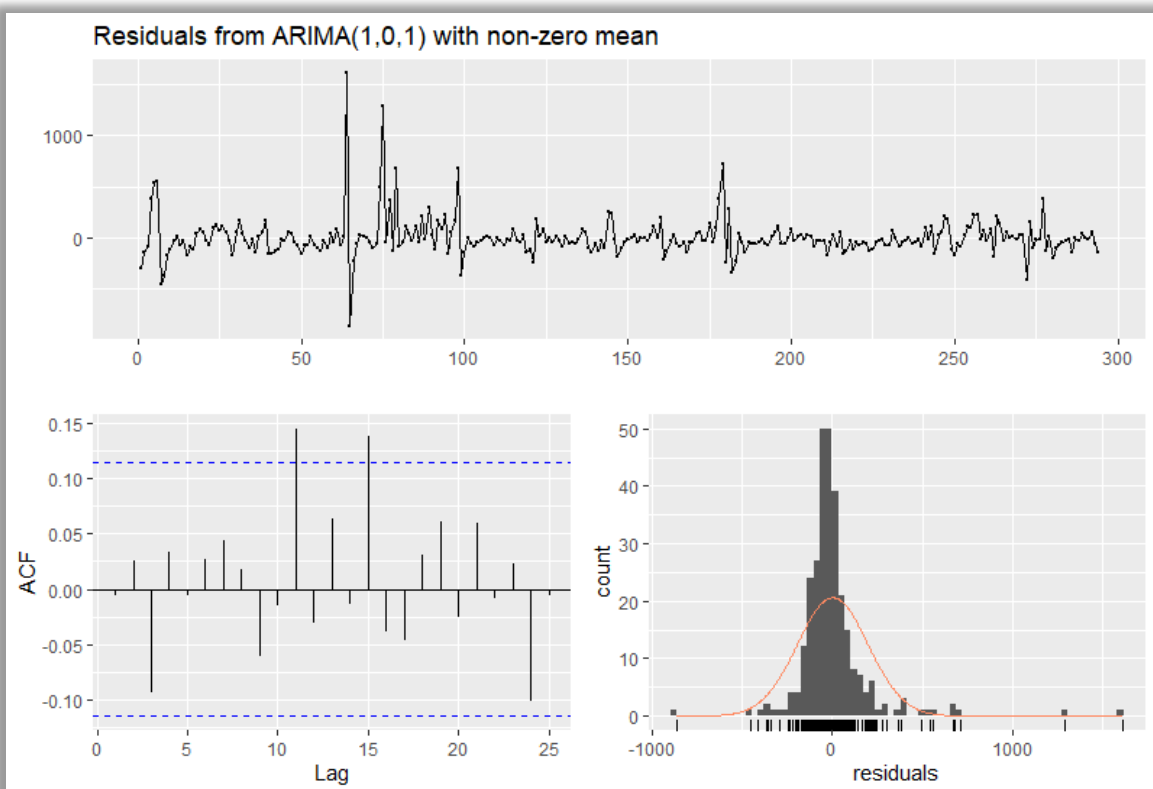**F**igure **51**. Residuals from Regression with ARIMA(1,0,1) errors

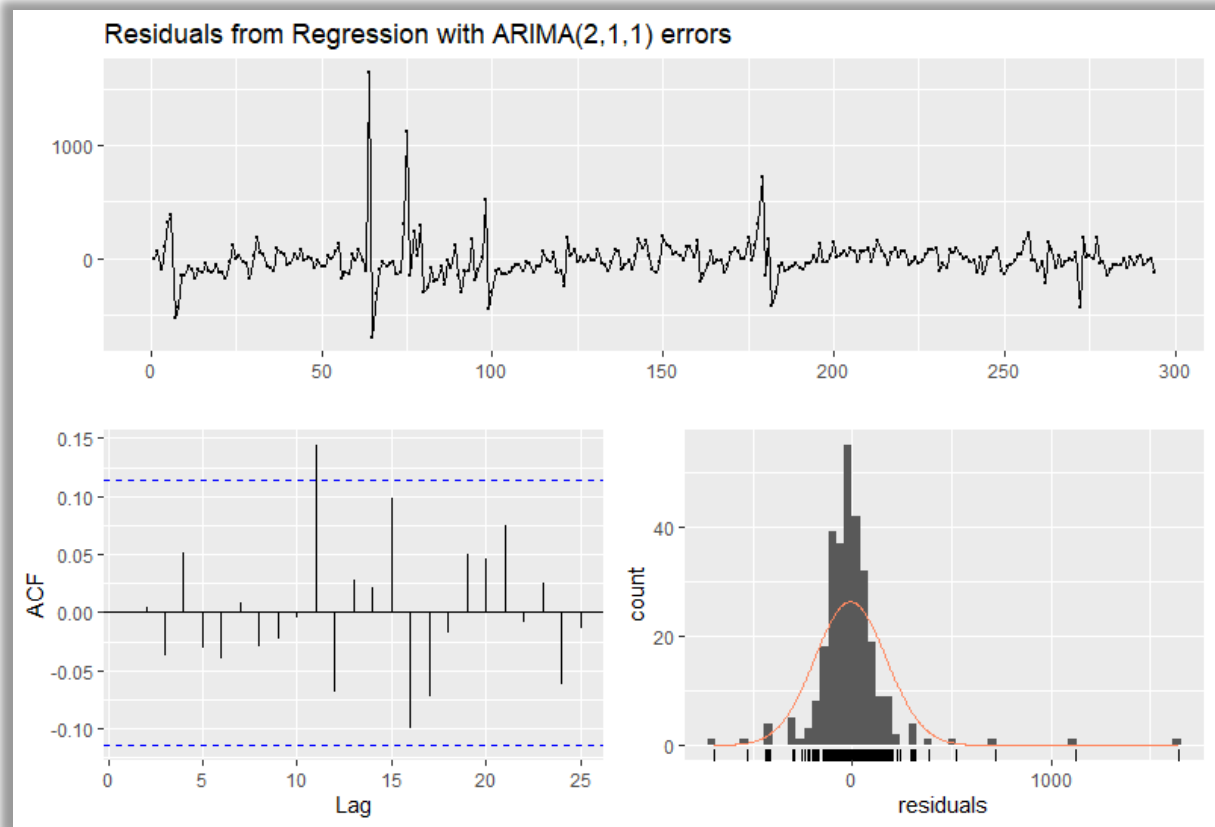**F**igure **52**. Residuals ARIMA(1,0,1) with non-zero mean

Figure 53. Residuals from Regression with ARIMA(2,1,1) errors

4. Conclusion

In this project, sales forecasts for 8 different product type were made. Alternative time series models used for finding the closest value to the real sales quantity. For the products "Bikini", "Fakir" "LaRoche", "Mont", "Tayt" and "Sleepy" dynamic regression model was used. For the "OralB" and "Xiaomi" products ARIMA model was used. GitHub link of R code of the project was given in the Appendix.

5. Appendix

https://github.com/BU-IE-360/spring20-MiracGoktugAltuntas.git