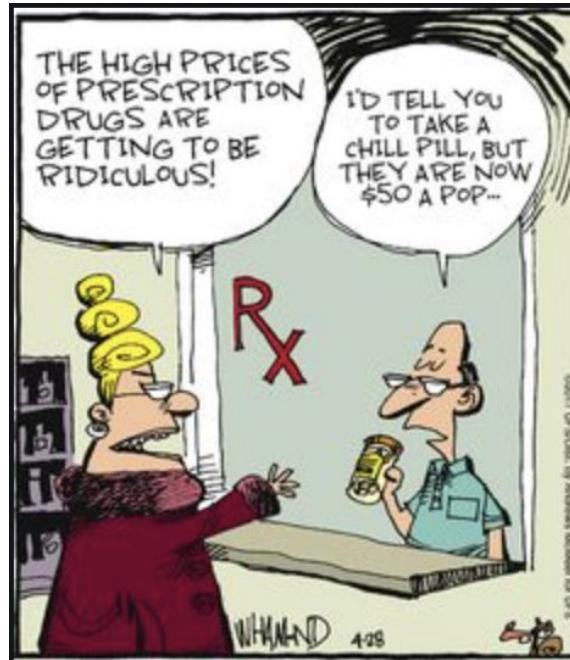


# Pharmaceutical Drug Sales Forecasting



Adam Castillo, Joey Gadbois, Vichetra Lim, Christian Ramirez, Hamzah Sami

# Dataset

The data came from Kaggle and consists of the volume of sales of different classes of pharmaceutical drugs. It spans 6 years (2014-2019). The variables given include:

- **datum:** date
- **M01AB:** anti-inflammatory and antirheumatic products, non-steroids, acetic acid derivatives and related substances
- **M01AE:** anti-inflammatory and antirheumatic products, non-steroids, propionic acid derivatives
- **N02BA:** other analgesics and antipyretics, salicylic acid and derivatives
- **N02BE:** other analgesics and antipyretics, pyrazolones and anilides
- **N05B:** psycholeptic drugs, anxiolytic drugs
- **N05C:** psycholeptic drugs, hypnotics and sedative drugs
- **R03:** drugs for obstructive airway diseases
- **R06:** antihistamines for systemic use

# Dataset

## Transformed Data

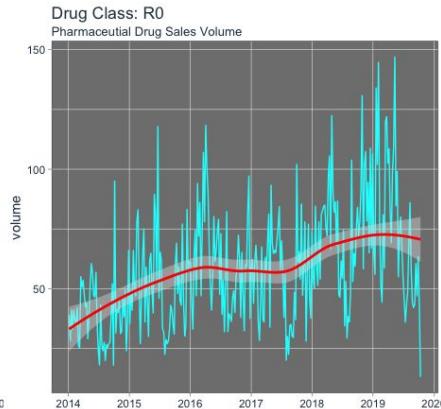
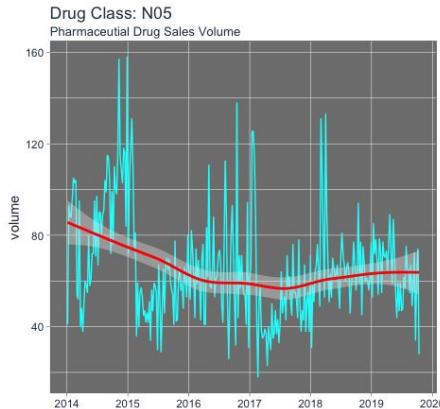
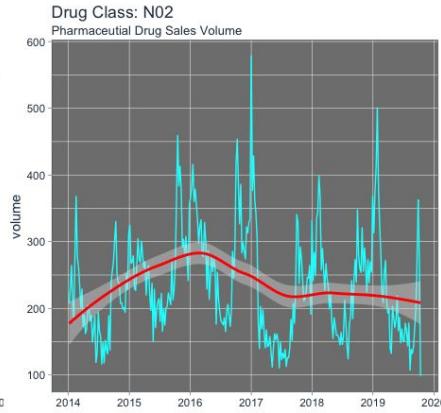
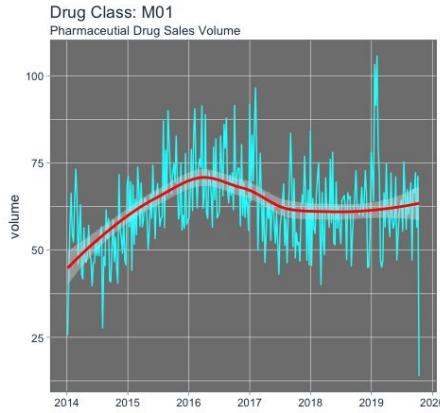
The data has been transformed into their broader classes of drugs, given by:

- **total:** sales volume for all classes of drugs
- **M01:** anti-inflammatory and antirheumatic products
- **N02:** other analgesics and antipyretics
- **N05:** psycholeptic drugs
- **R0:** antihistamines and drugs for obstructive airway diseases

# Research Questions

- 1) How can we deal with outliers in a time series?
- 2) For each class of drugs, the last 11 weeks of 2019 have been forecasted. Are the forecasts reasonable?
- 3) For each class of drugs, are there any life events that could affect how the volume of pharmaceutical drug sales?

# Time Series Plots



- Time series plots for all four general drug classes
- Trend estimation with Loess regression
- **M01:** upward trend from 2014 to 2016, flattens out around mid-2017
- **N02:** very inconsistent regarding a trend
- **N05:** initial downward trend from 2014 to 2016, flattens out after decline
- **R0:** fairly consistent upward trend except for 2016 and the start of 2017

# Data Preparation and Outlier Removal

## Train and Test Data

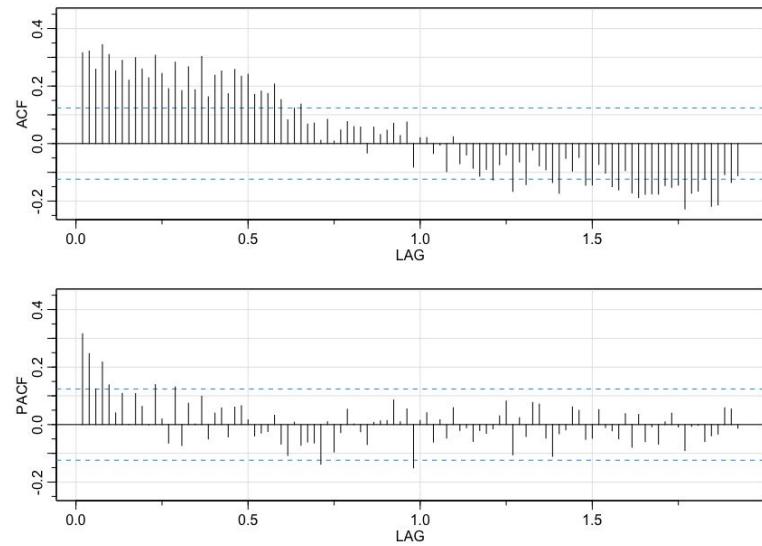
- **Train data:** observations starting on January 1st, 2014 through December 30th, 2018
- **Test data:** observations starting on January 6th, 2019 through October 13th 2019
- **Forecasting horizon:** forecast the weekly drug sales volume for the last 11 weeks of 2019

## Outlier Removal

- All time series had outliers removed via robust STL decomposition
- **STL:** Seasonal and Trend Decomposition using Loess
- Can handle any periodicity including a seasonal component that changes over time
- Robust to outliers due to local regression, outliers won't effect trend-cycle or seasonal component estimates

# M01 Exploratory Data Analysis

Sample ACF and PACF

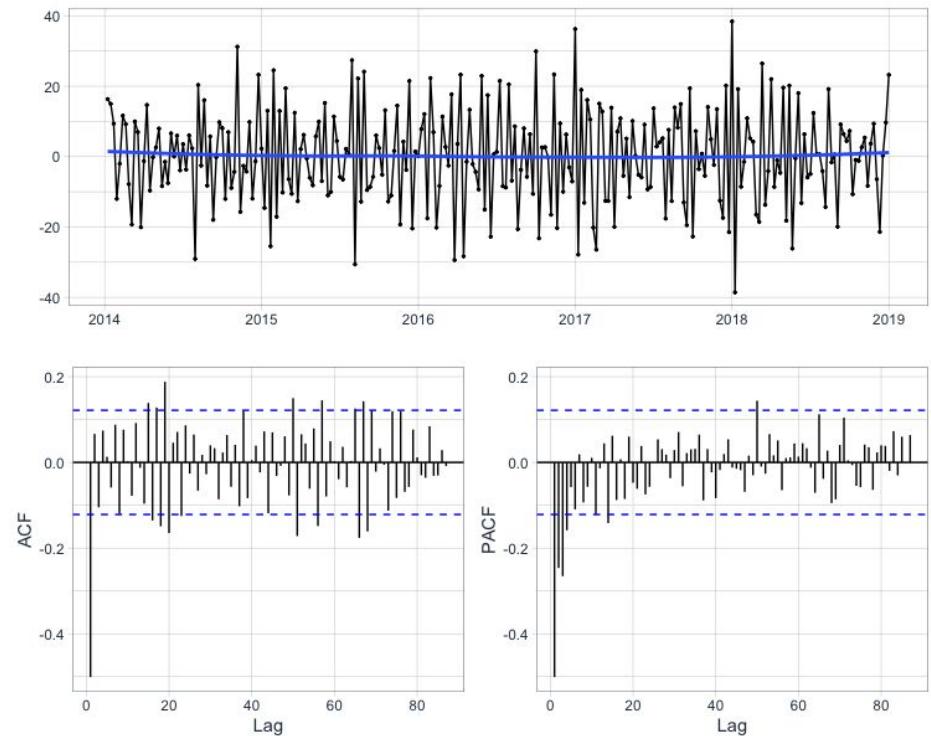


## ADF Test:

The augmented Dickey-Fuller test allows us to conclude that the M01 series is not stationary because:

P-value = 0.1724 > 0.05

First-Order Difference: Plot, ACF, and PACF

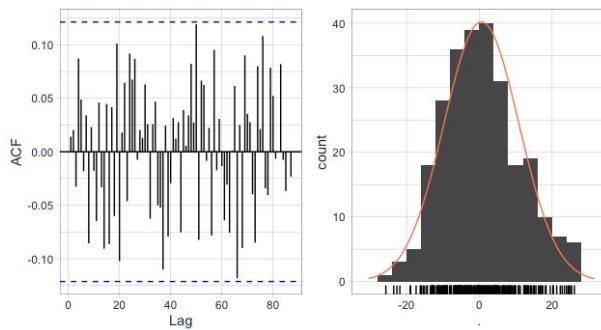
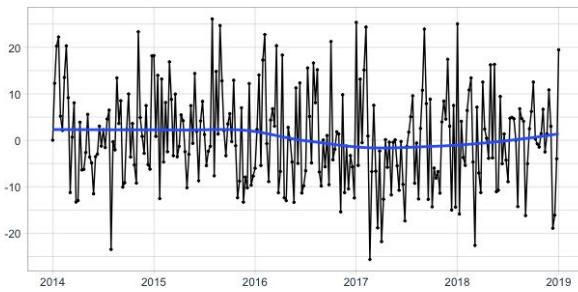


# M01 Model Selection

$Y \sim ARIMA(2, 1, 1)$

AIC: 1962 | AICc: 1962 | BIC: 1977

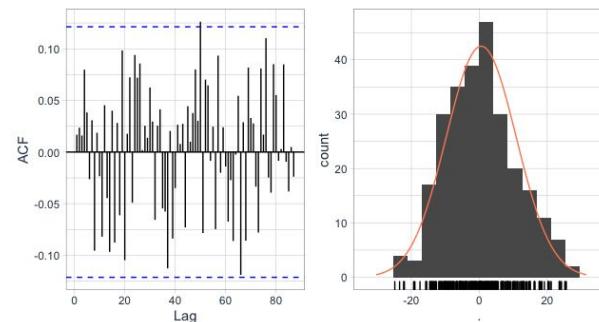
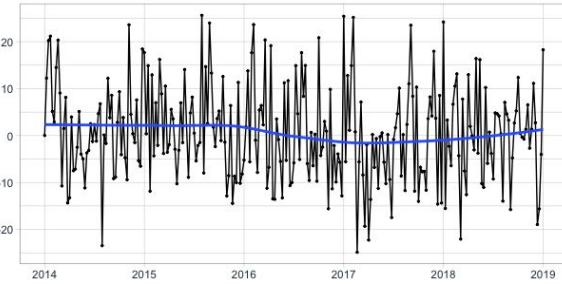
Residuals: stationary by ADF test



$Y \sim ARIMA(3, 1, 1)$

AIC: 1964 | AICc: 1964 | BIC: 1981

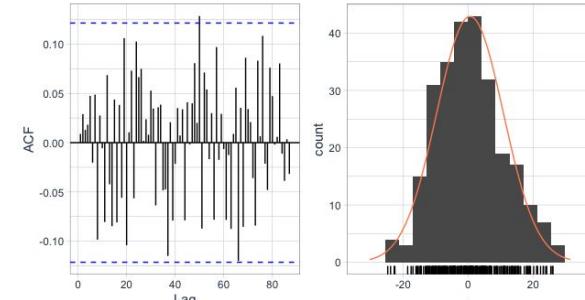
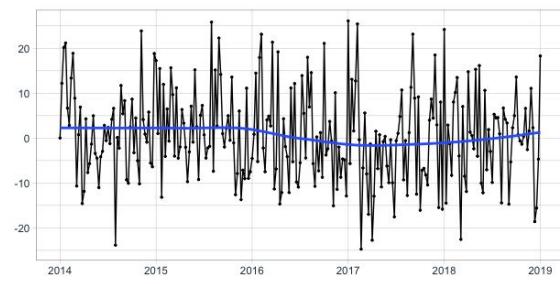
Residuals: stationary by ADF test



$Y \sim ARIMA(4, 1, 1)$

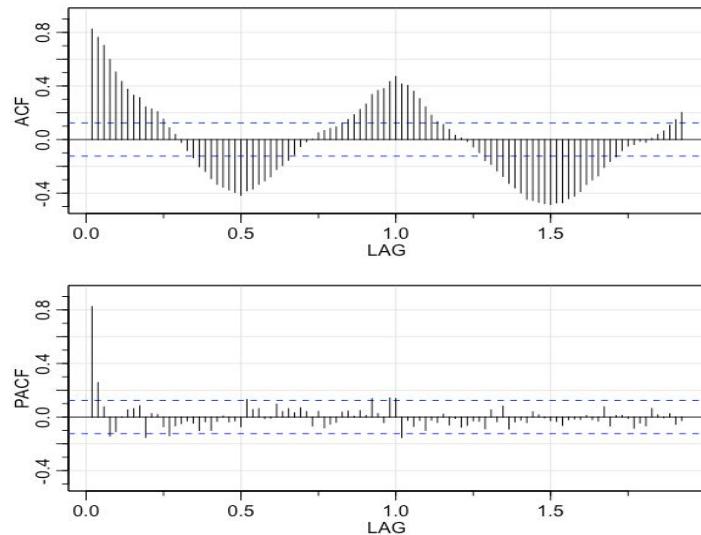
AIC: 1964 | AICc: 1964 | BIC: 1985

Residuals: stationary by ADF test



# N02 Exploratory Data Analysis

Sample ACF and PACF

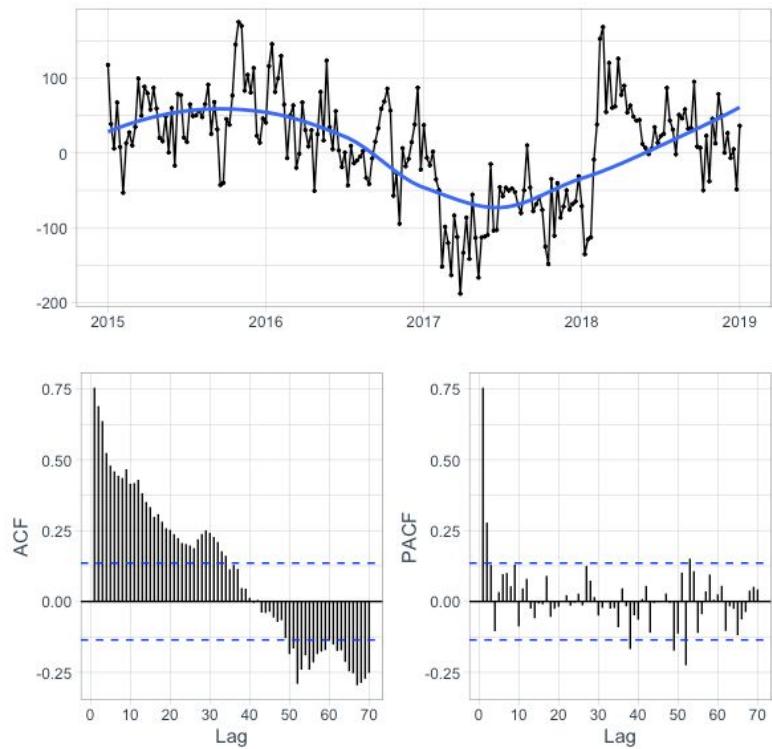


ADF Test:

The augmented Dickey-Fuller test allows us to conclude that the N02 series is not stationary because:

$$\text{P-value} = 0.05403 > 0.05$$

First-Order Difference: Plot, ACF, and PACF

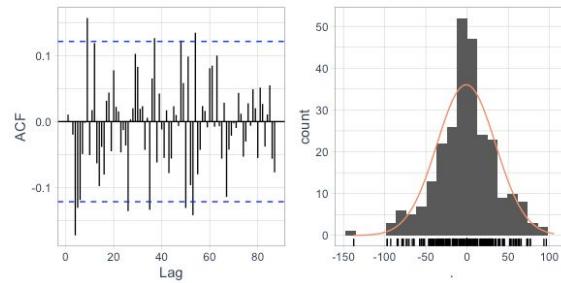
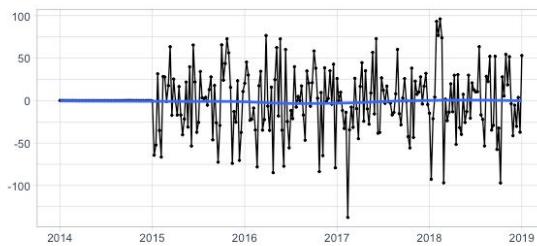


# N02 Model Selection

$Y \sim ARIMA(2, 1, 0)(1, 1, 0)[52]$

AIC: 2144 | AICc: 2144 | BIC: 2157.31

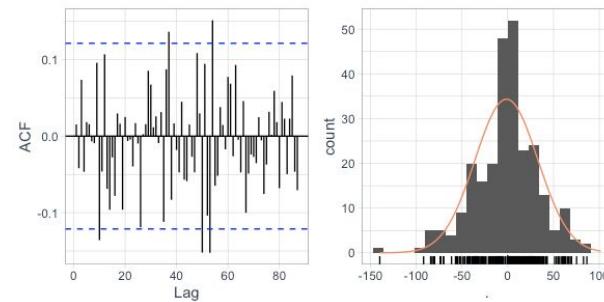
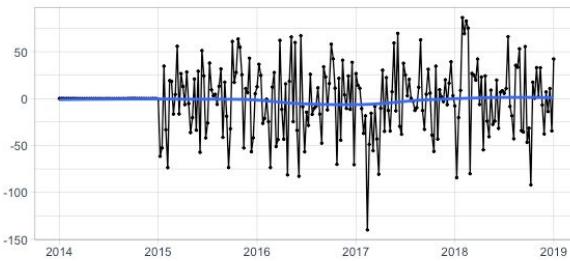
Residuals: stationary by ADF test



$Y \sim ARIMA(2, 1, 3)(1, 1, 0)[52]$

AIC: 2136 | AICc: 2137 | BIC: 2160

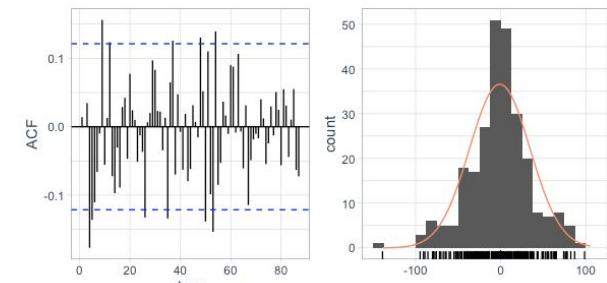
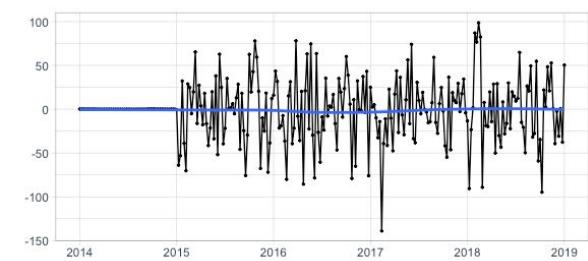
Residuals: stationary by ADF test



$Y \sim ARIMA(0, 1, 1)(1, 1, 0)[52]$

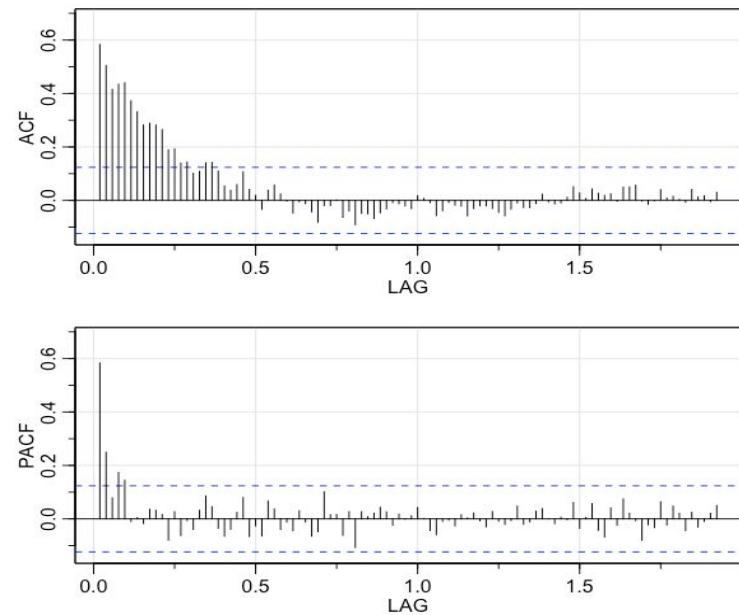
AIC: 2142 | AICc: 2142 | BIC: 2152.16

Residuals: stationary by ADF test



# N05 Exploratory Data Analysis

Sample ACF and PACF

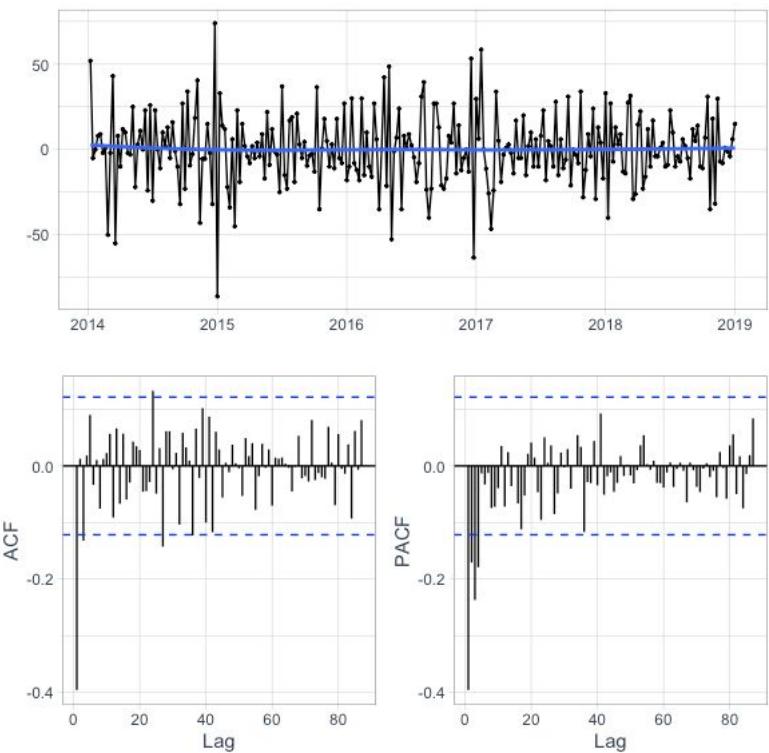


ADF Test:

The augmented Dickey-Fuller test allows us to conclude that the N05 series is not stationary because:

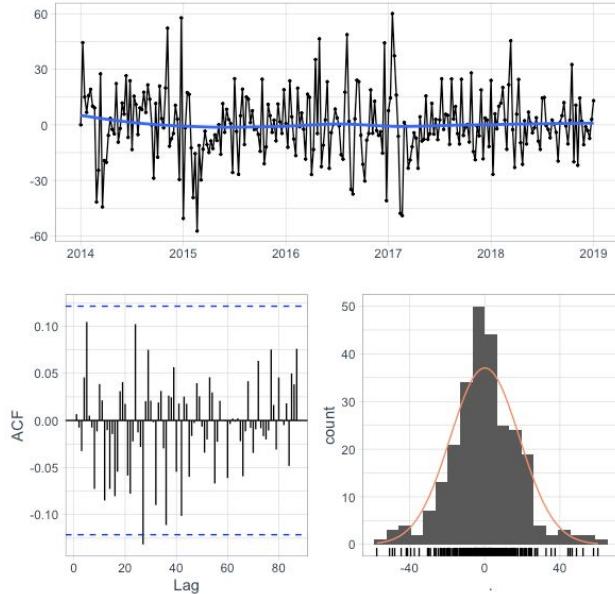
$$\text{P-value} = 0.06014 > 0.05$$

First-Order Difference: Plot, ACF, and PACF

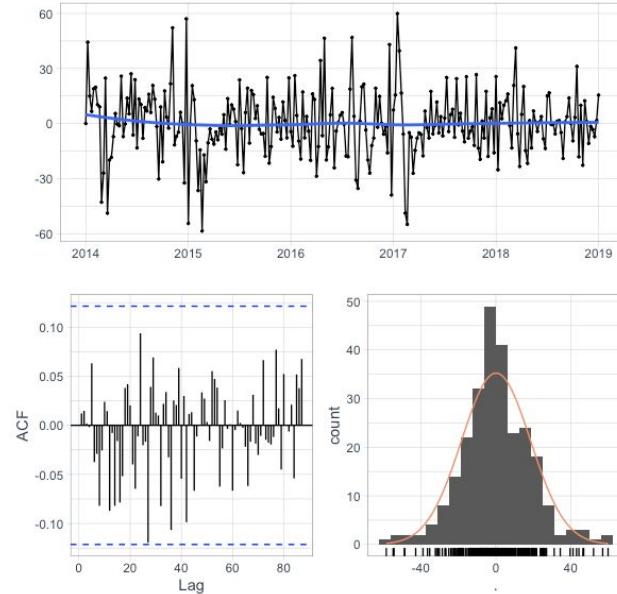


# N05 Model Selection

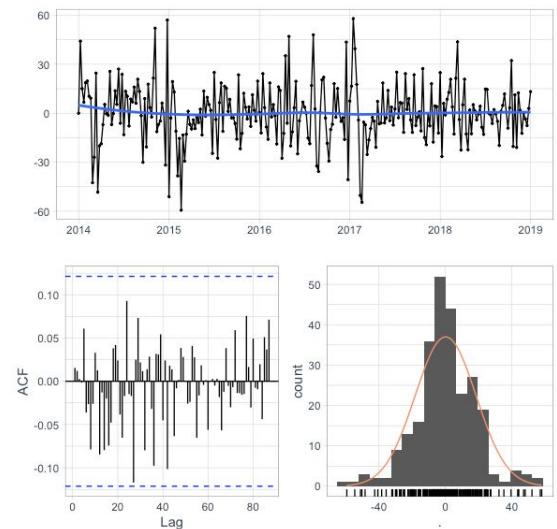
$Y \sim ARIMA(0, 1, 3)(1, 0, 0)[52]$   
AIC: 2263 | AICc: 2263 | BIC: 2280  
Residuals: stationary by ADF test



$Y \sim ARIMA(3, 1, 2)$   
AIC: 2265 | AICc: 2266 | BIC: 2287  
Residuals: stationary by ADF test

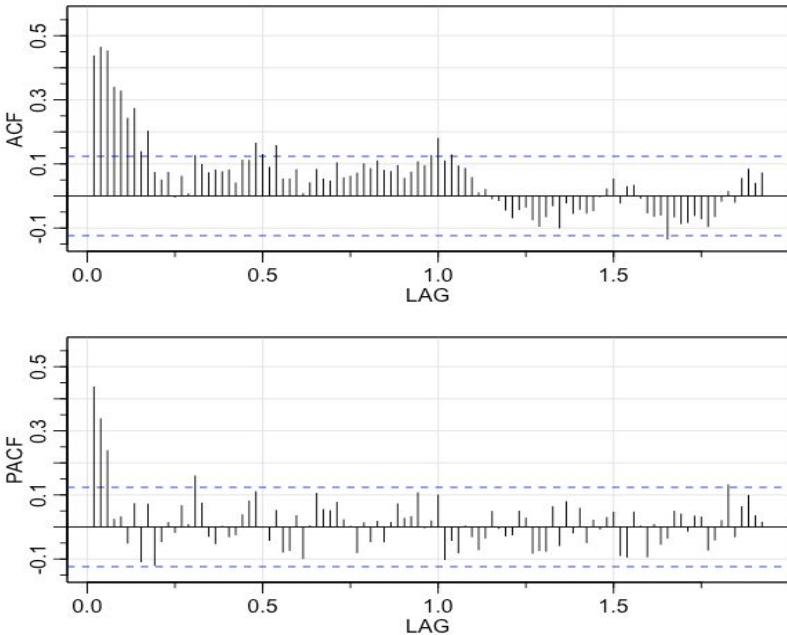


$Y \sim ARIMA(3, 1, 2)(1, 0, 1)[52]$   
AIC: 2268 | AICc: 2268 | BIC: 2296  
Residuals: stationary by ADF test



# R0 Exploratory Data Analysis

Sample ACF and PACF

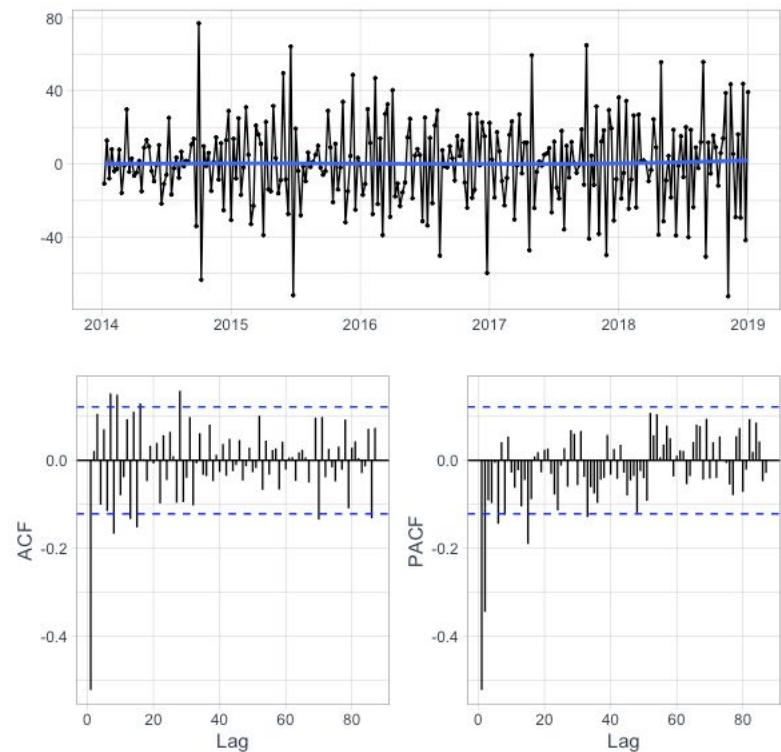


ADF Test:

The augmented Dickey-Fuller test allows us to conclude that the R0 series is stationary because:

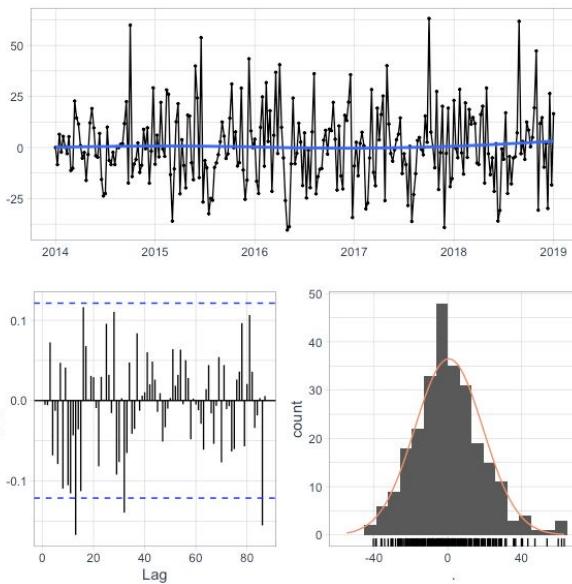
P-value = 0.01 < 0.05

First-Order Difference: Plot, ACF, and PACF

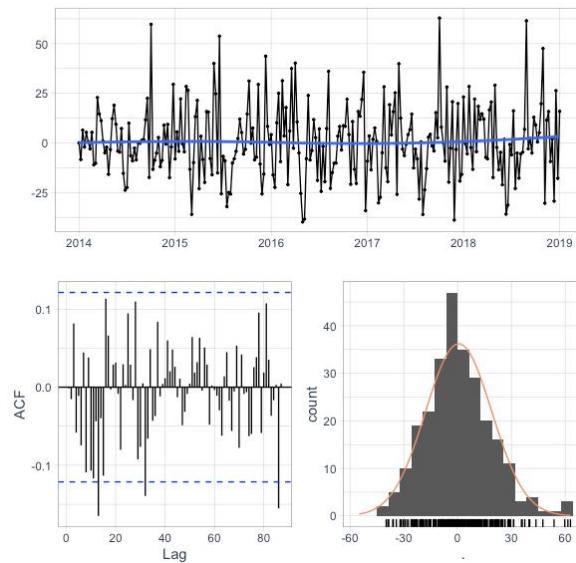


# R0 Model Selection

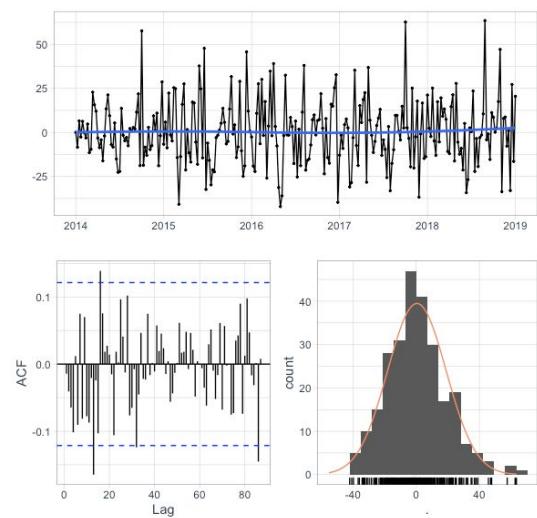
$Y \sim ARIMA(0, 1, 2)(0, 0, 1)[52]$   
AIC: 2265 | AICc: 2265 | BIC: 2279  
Residuals: stationary by ADF test



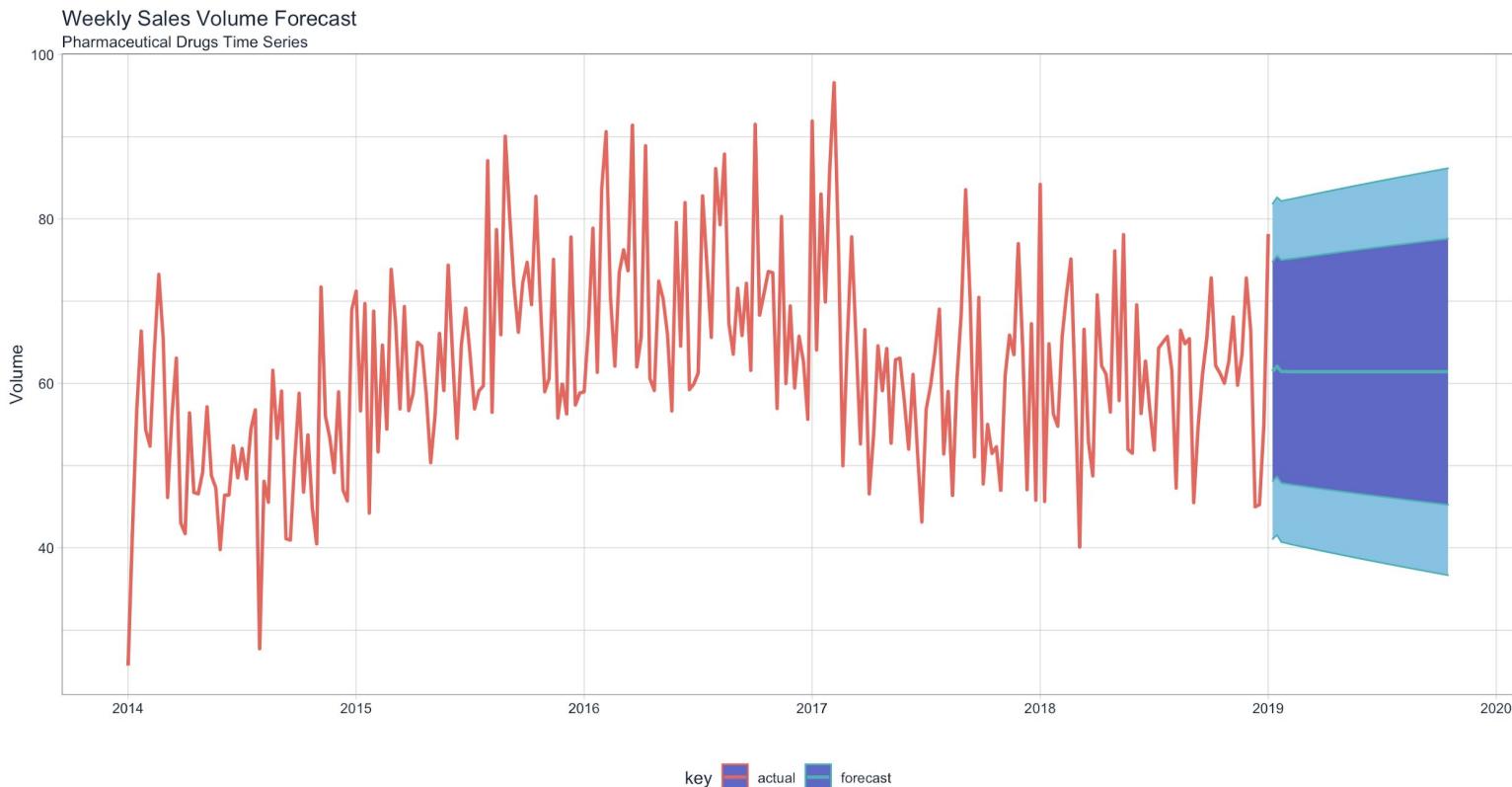
$Y \sim ARIMA(0, 1, 3)(0, 0, 1)[52]$   
AIC: 2267 | AICc: 2267 | BIC: 2285  
Residuals: stationary by ADF test



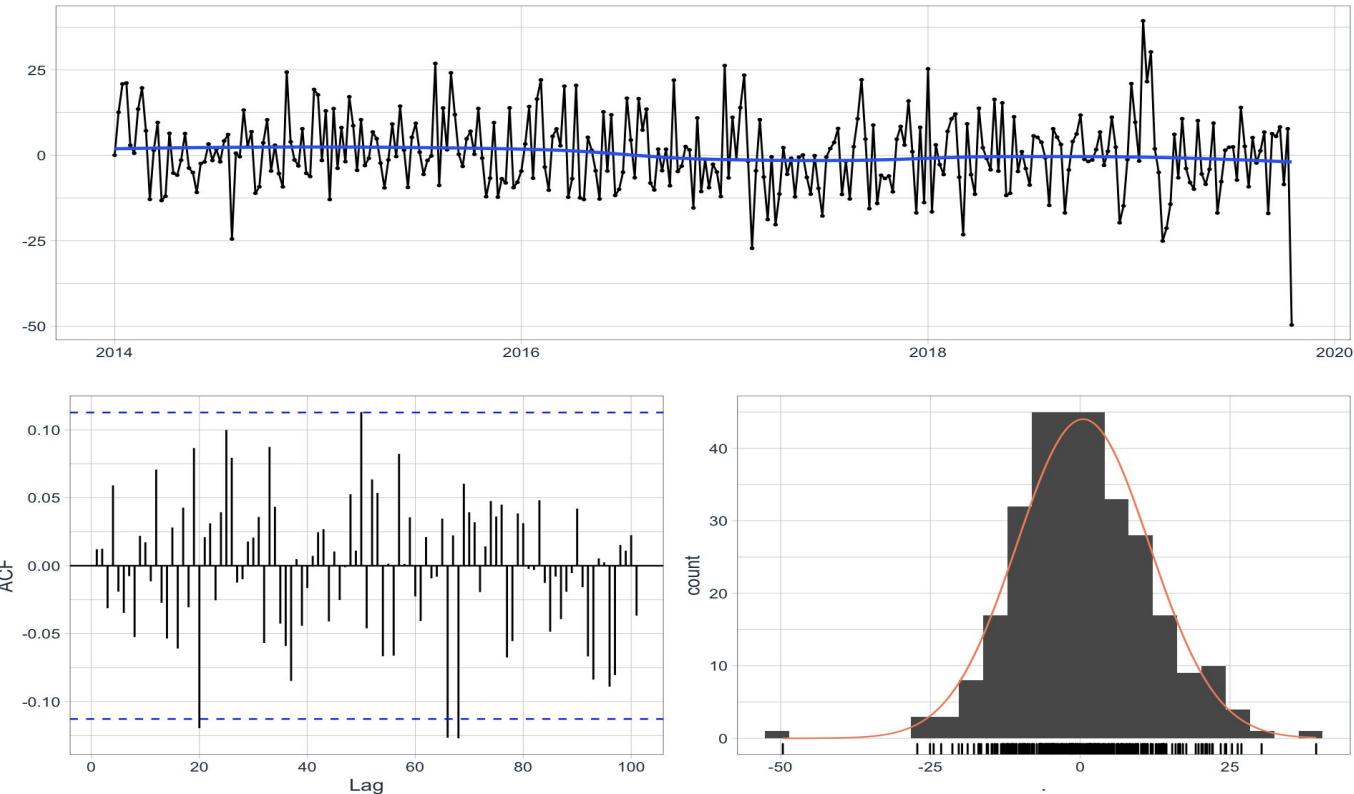
$Y \sim ARIMA(3, 1, 0)(1, 0, 0)[52]$   
AIC: 2269 | AICc: 2269 | BIC: 2287  
Residuals: stationary by ADF test



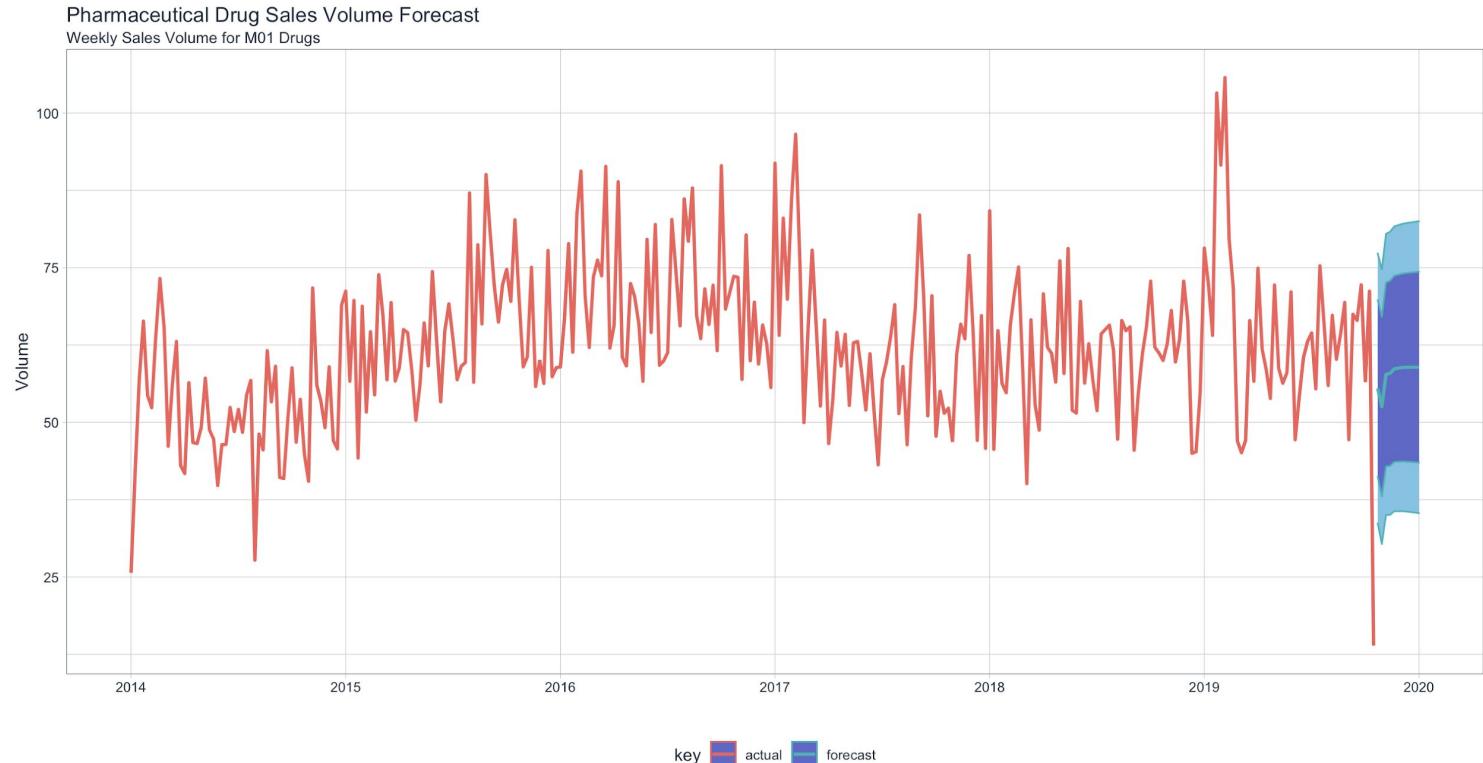
# Analysis of M01



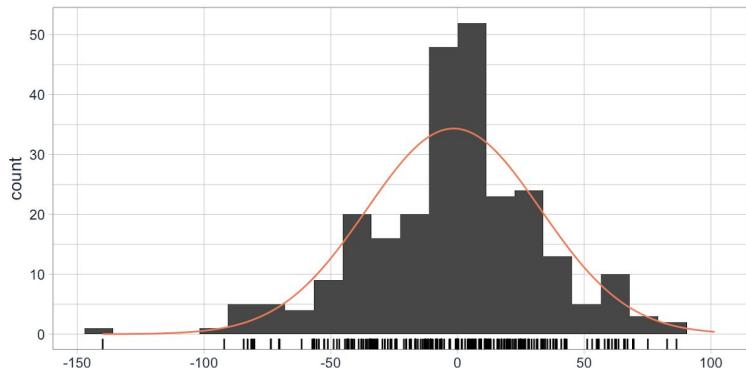
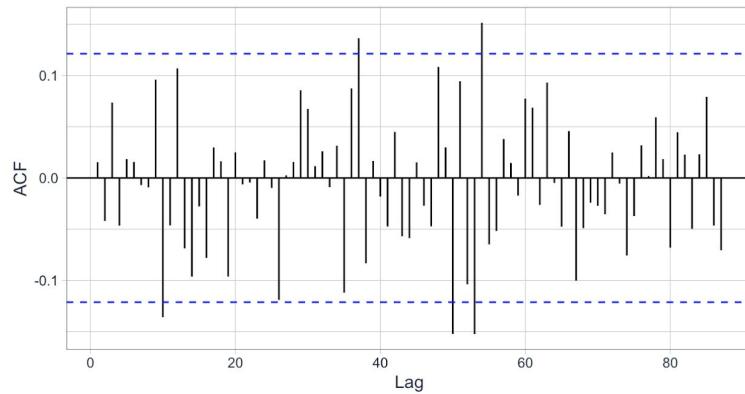
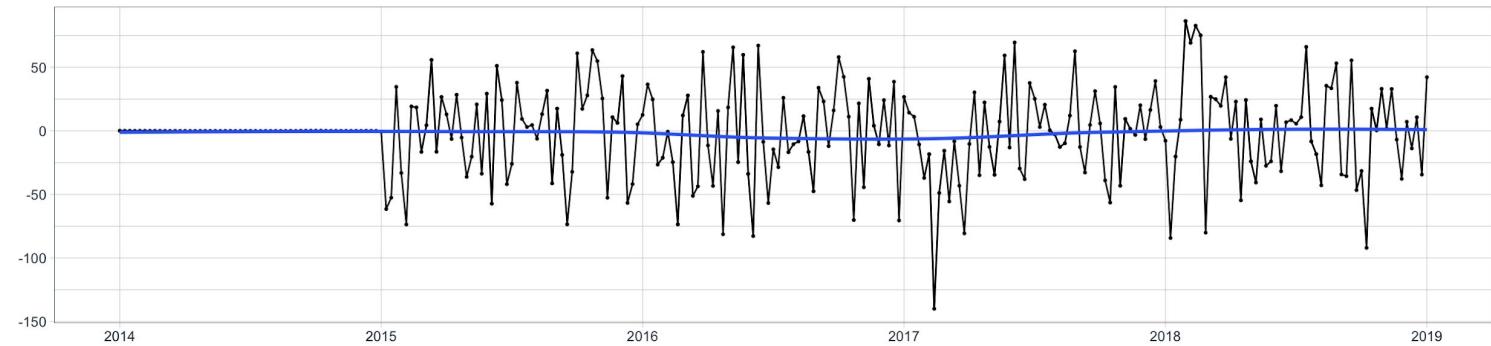
# Analysis of M01



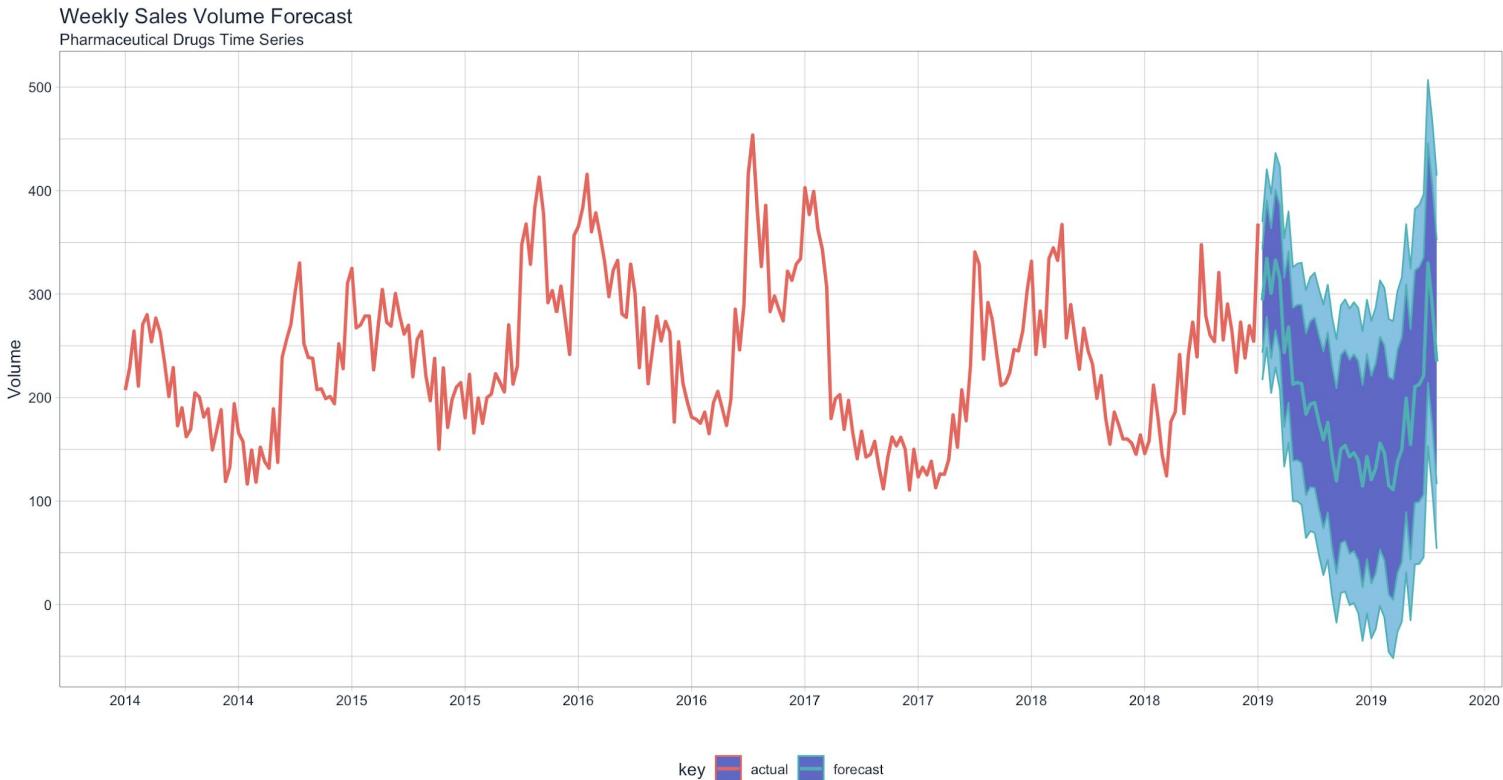
# Analysis of M01



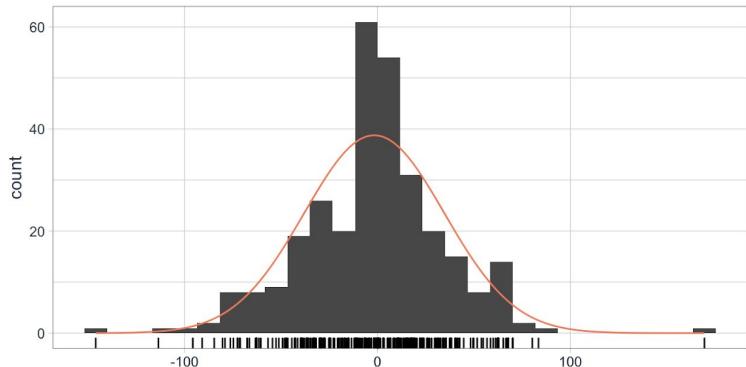
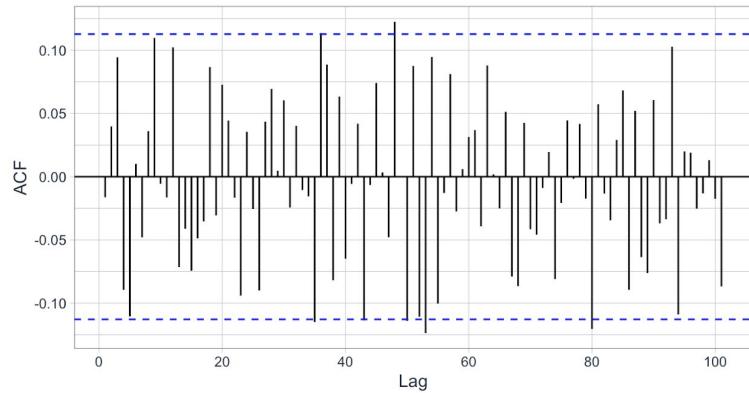
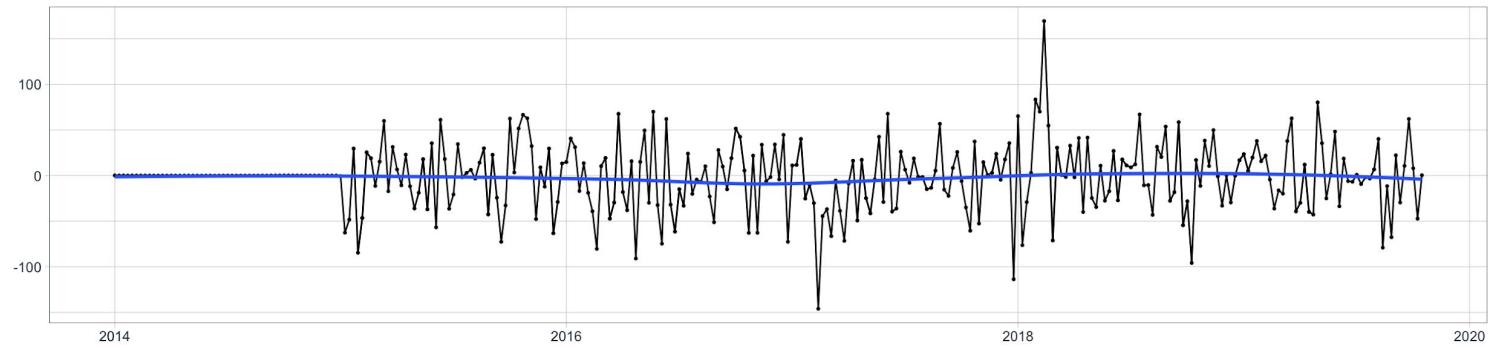
# Analysis of N02



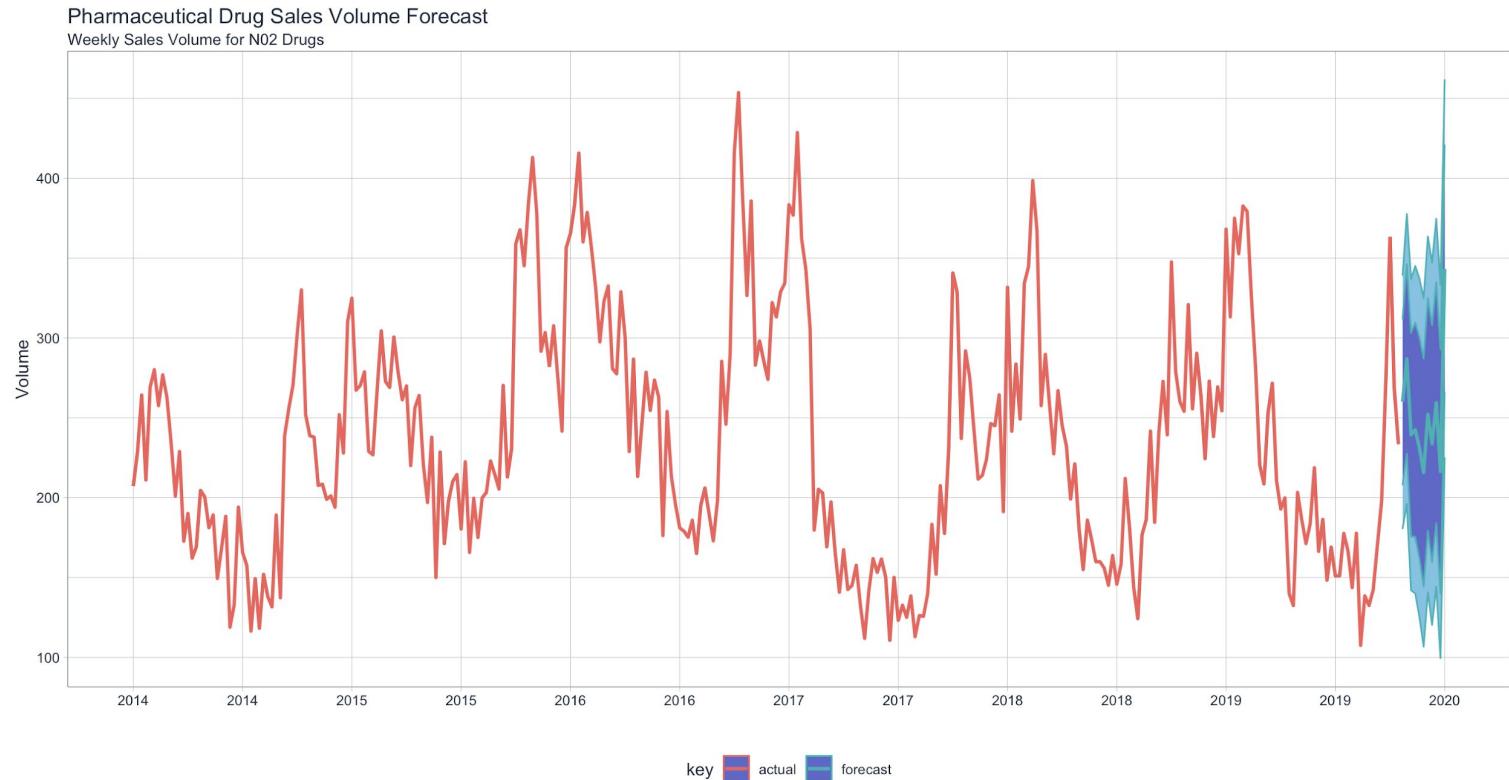
# Analysis of N02



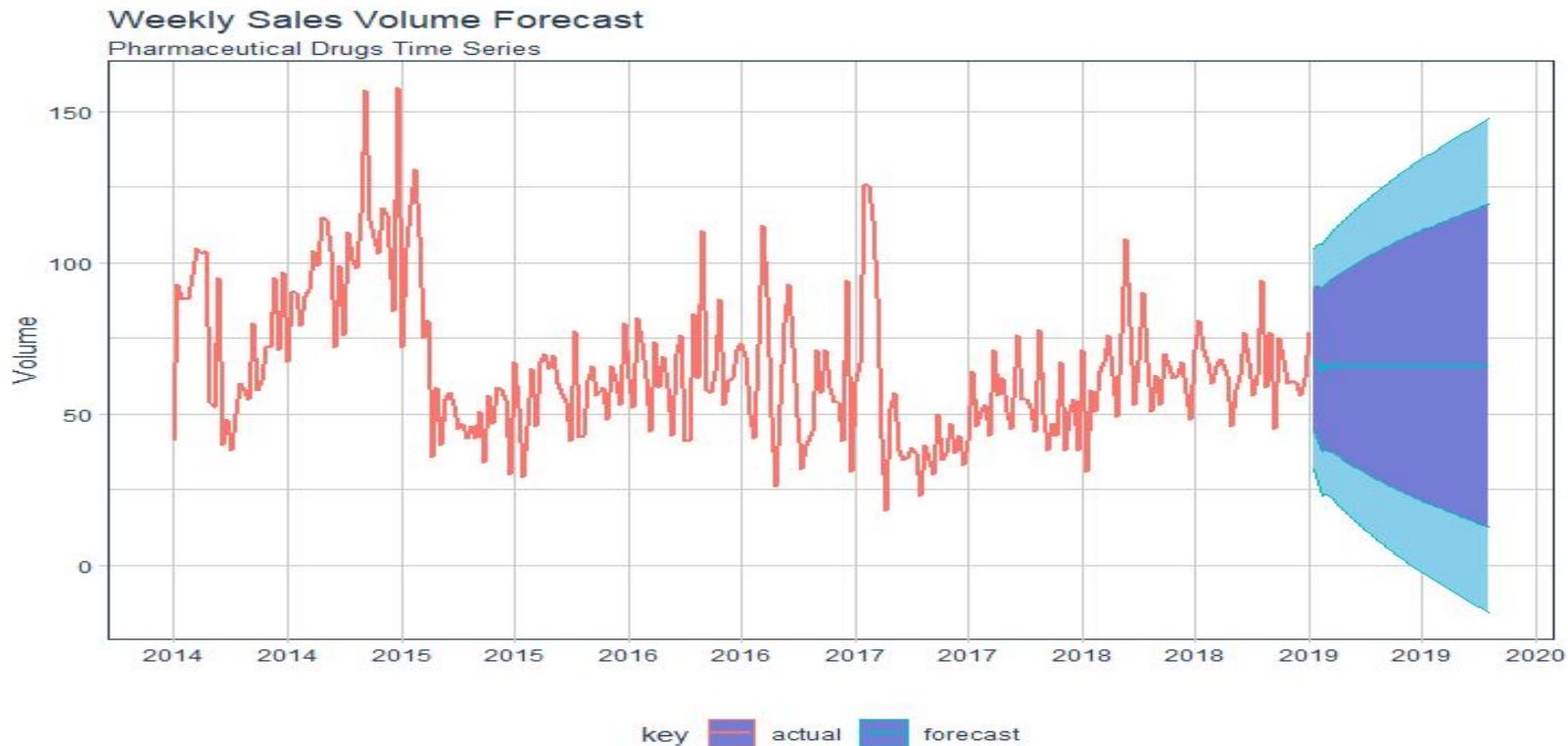
# Analysis of N02



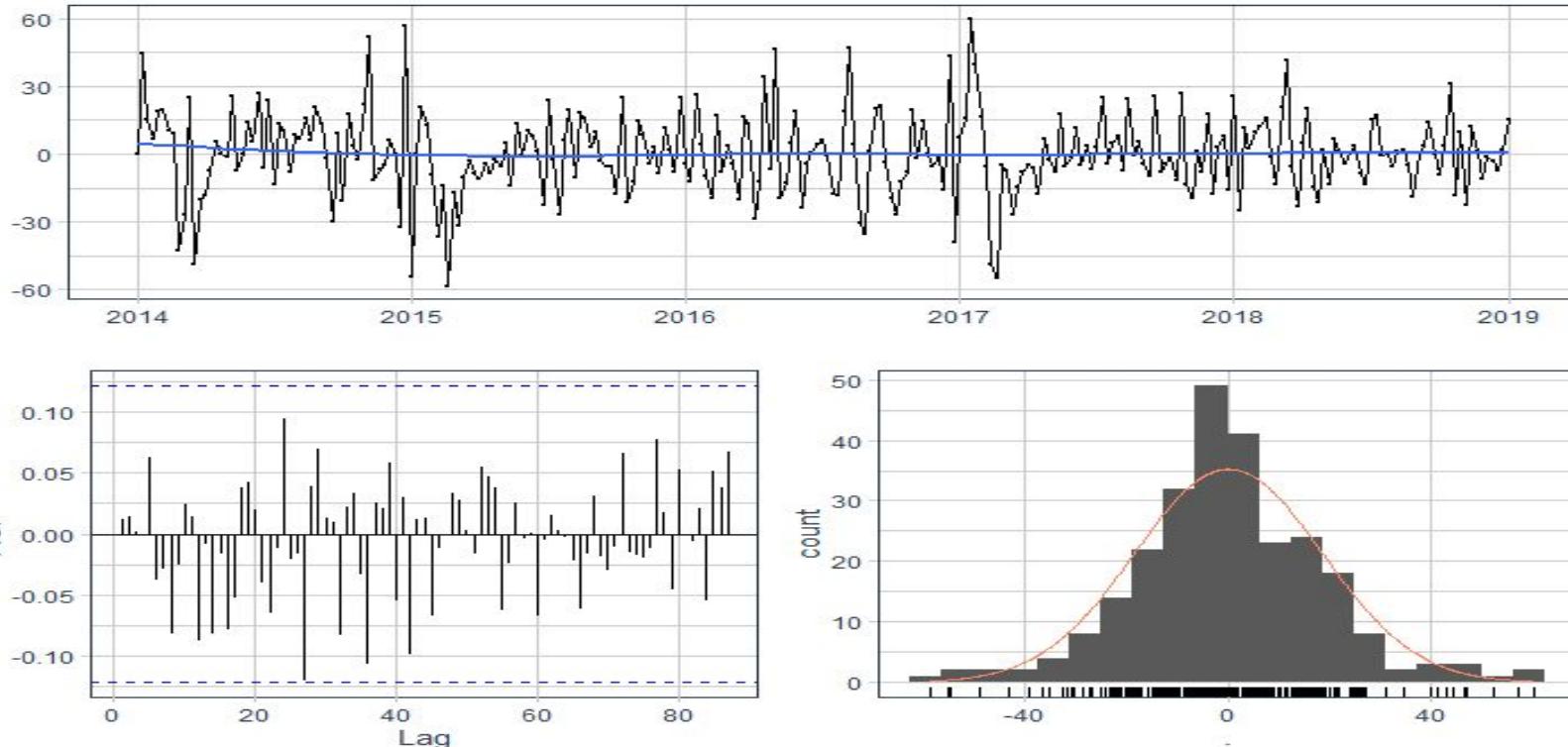
# Analysis of N02



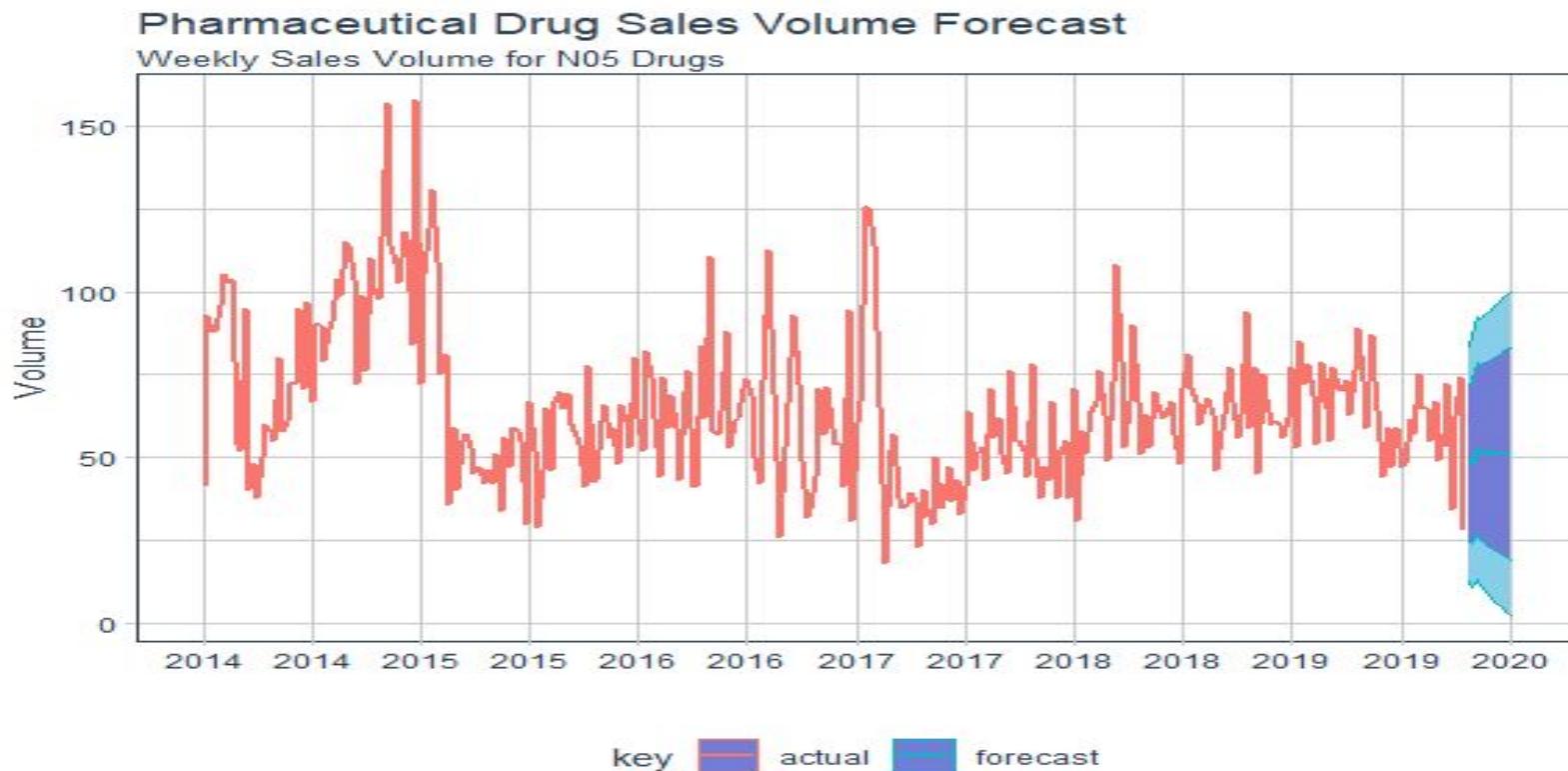
# Analysis of N05 (Psycholeptic Drugs)



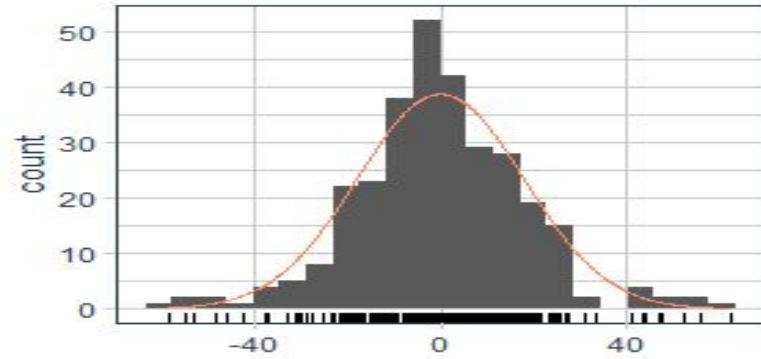
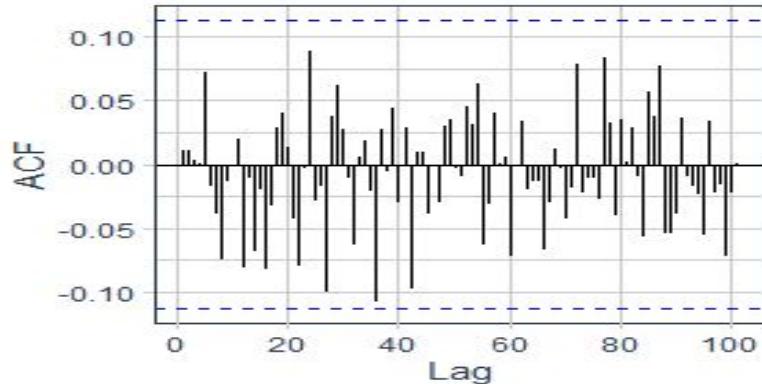
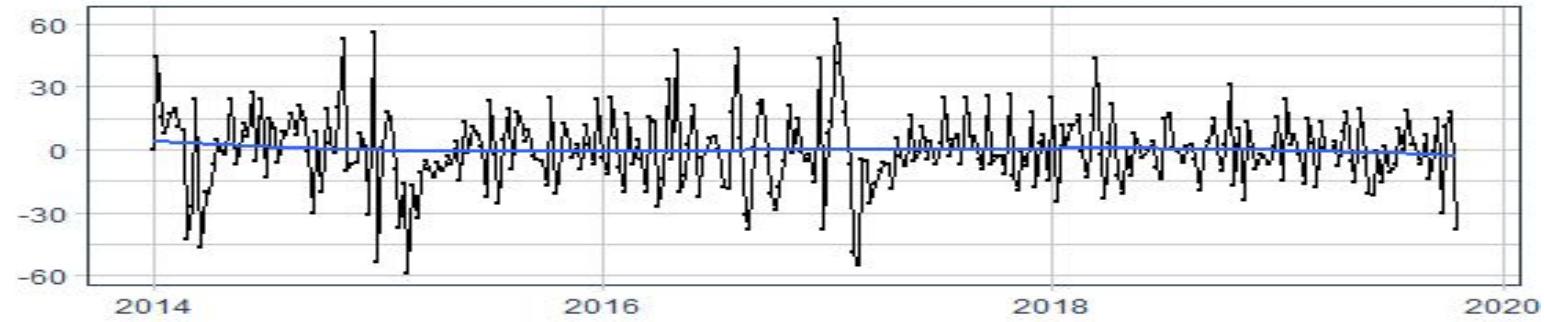
# Residual Analysis of N05



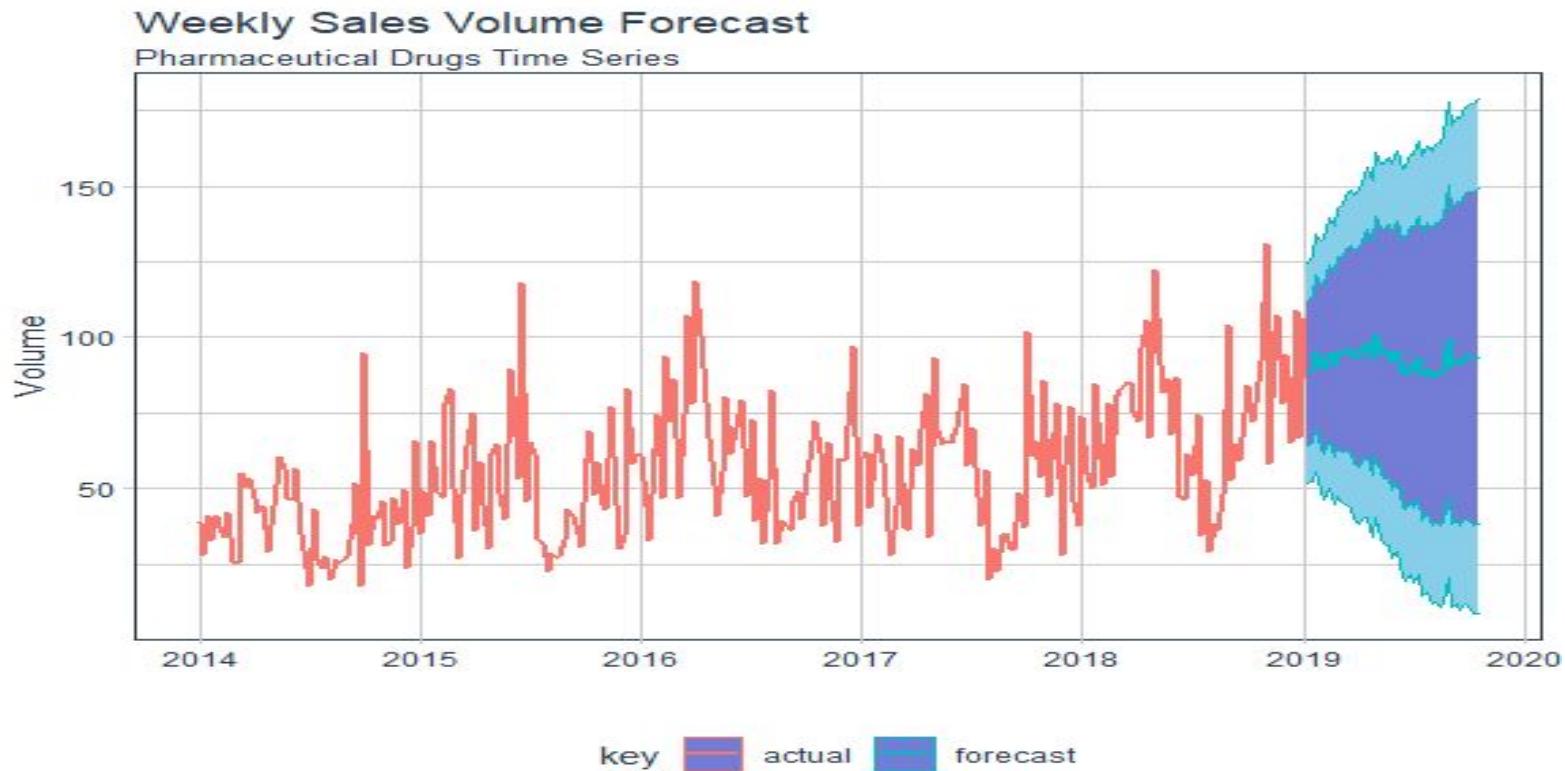
# Analysis of N05



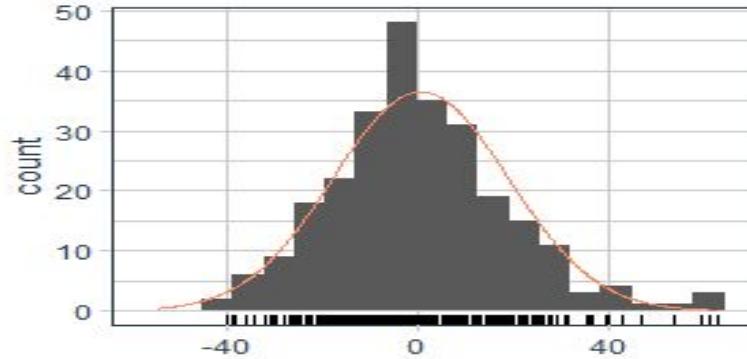
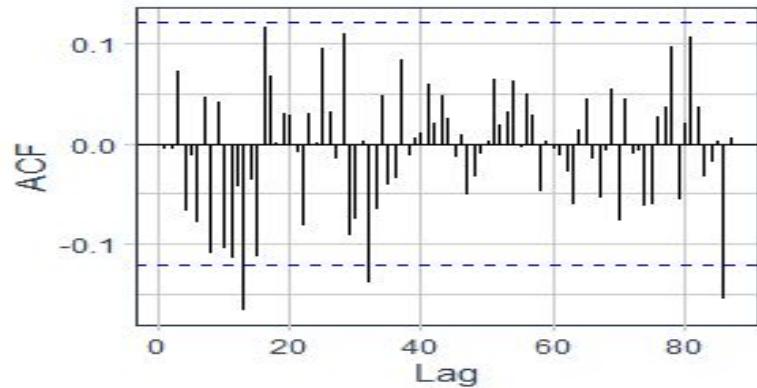
# N05 Residual Analysis for the rest of 2019



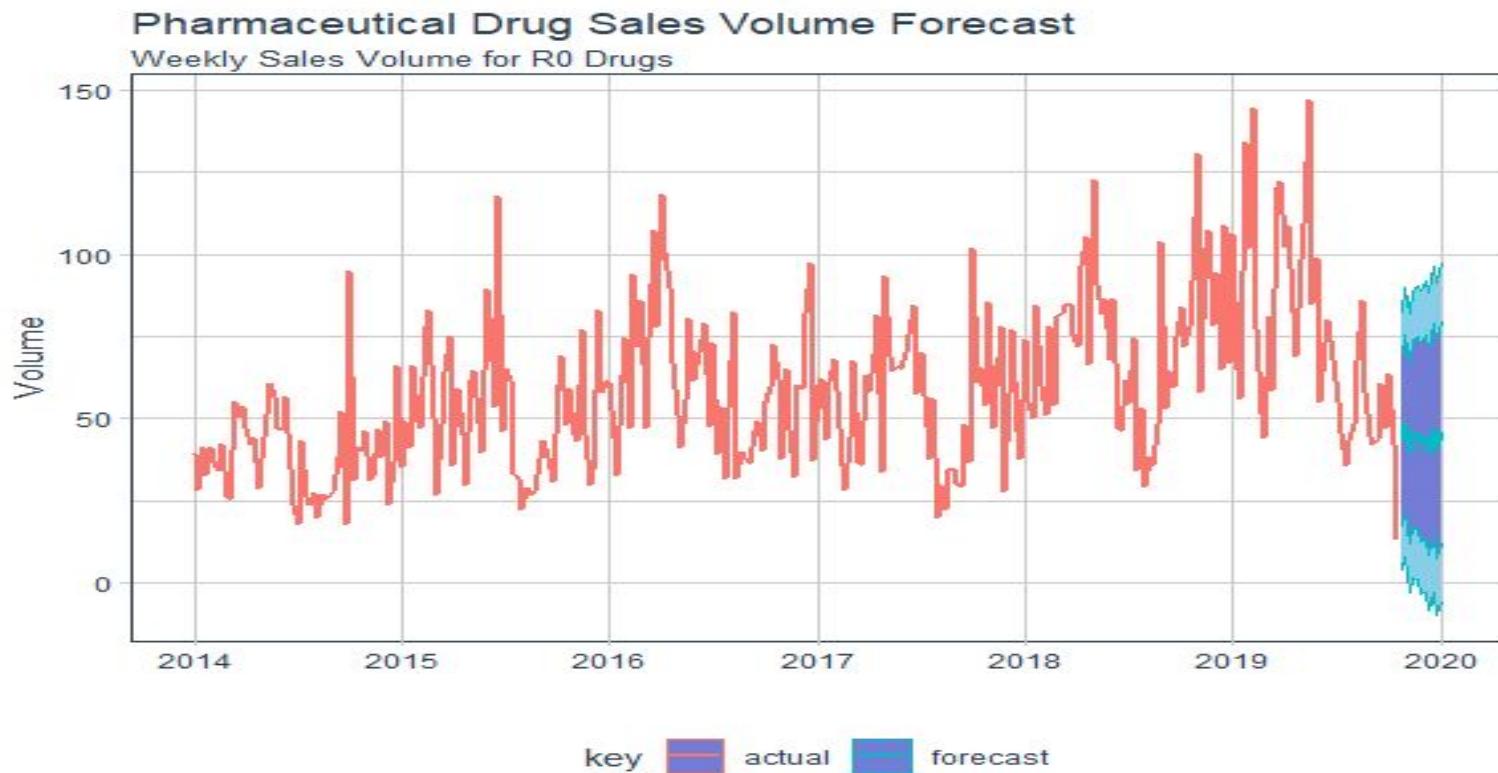
# Analysis of R0 (Antihistamines Drugs)



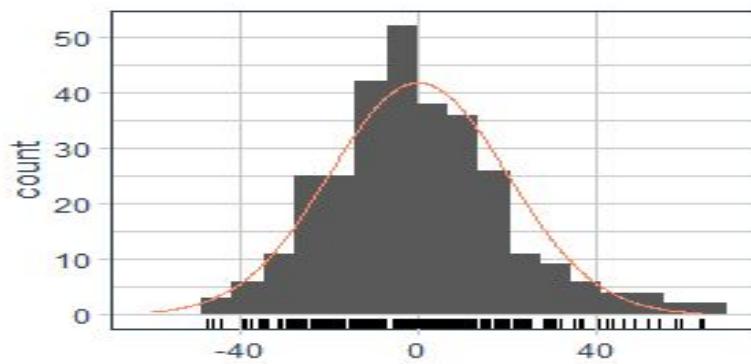
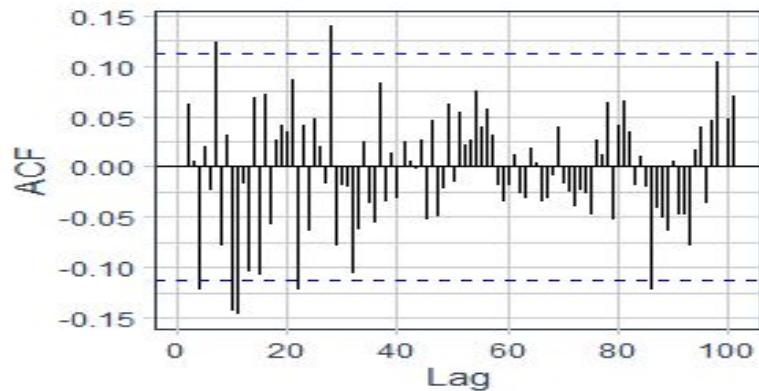
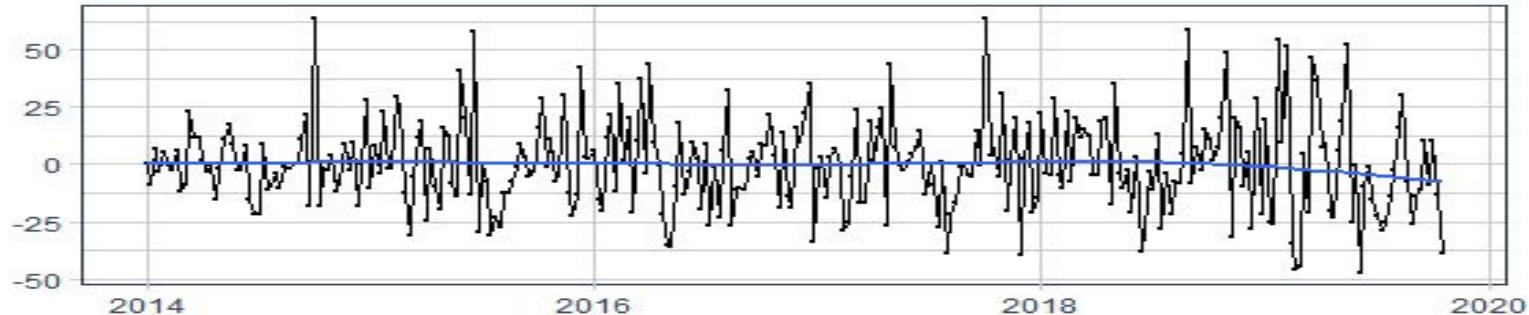
# Residual Analysis of R0



# Analysis of R0

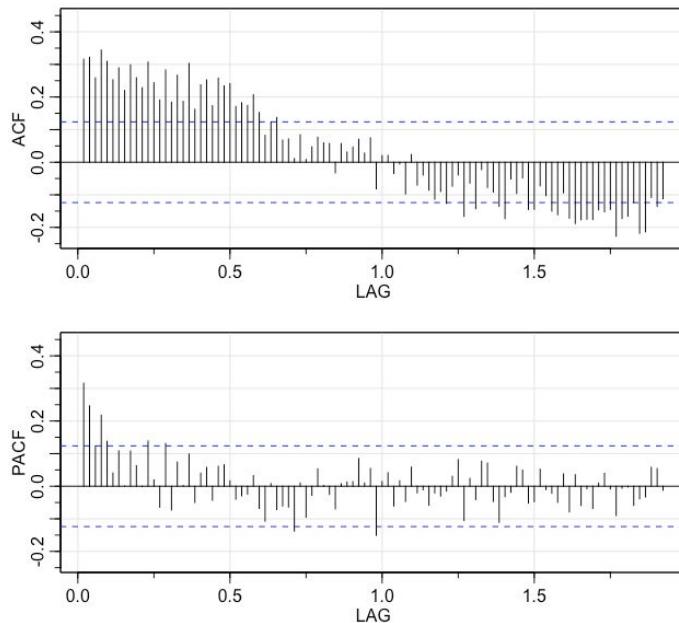


# R<sub>0</sub> Residual Analysis for the rest of 2019

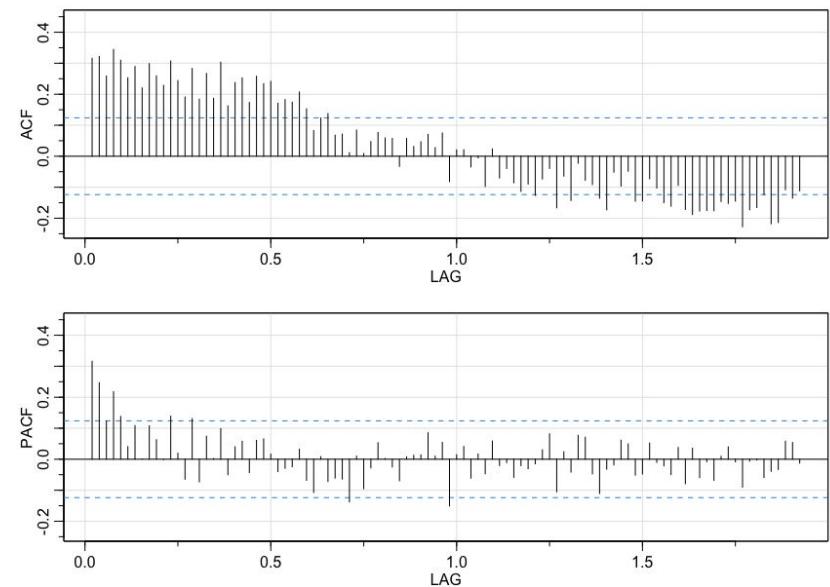


# Answers to Research Questions

# Research Question 1 Results

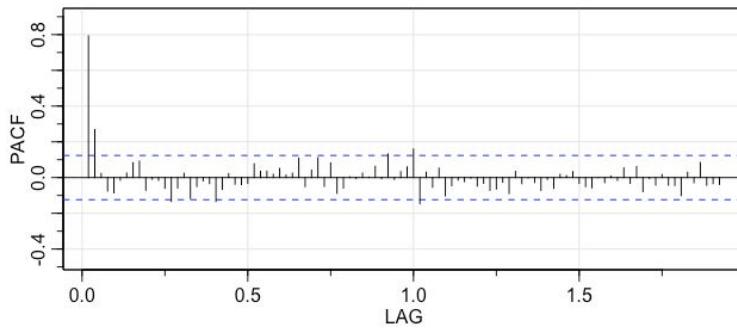
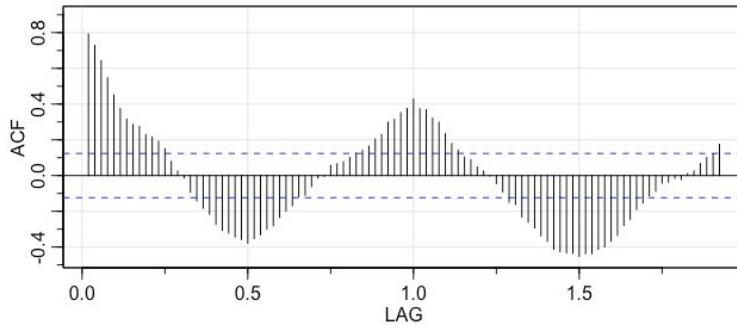


M01 without STL  
Decomposition

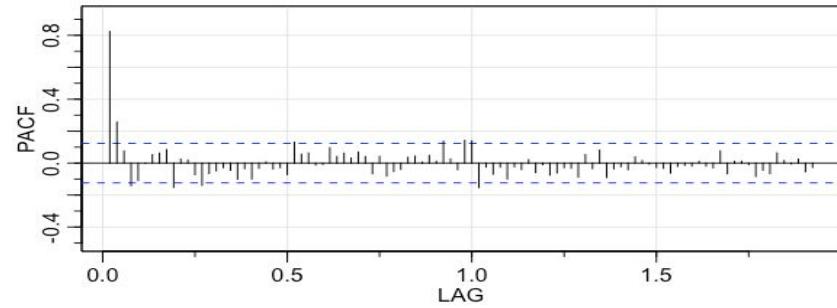
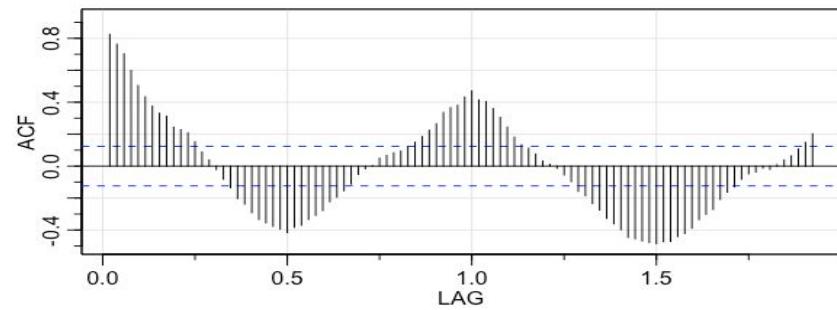


M01 with STL  
Decomposition

# Research Question 1 Results

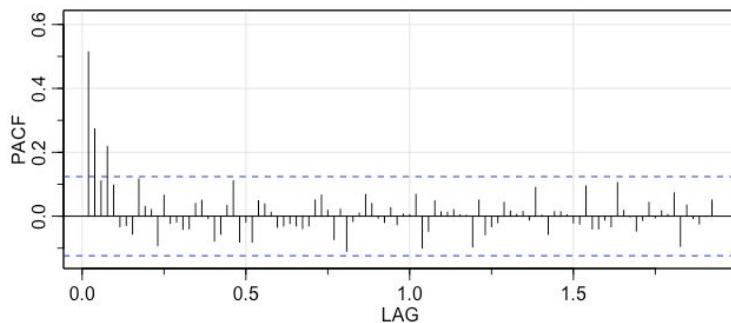
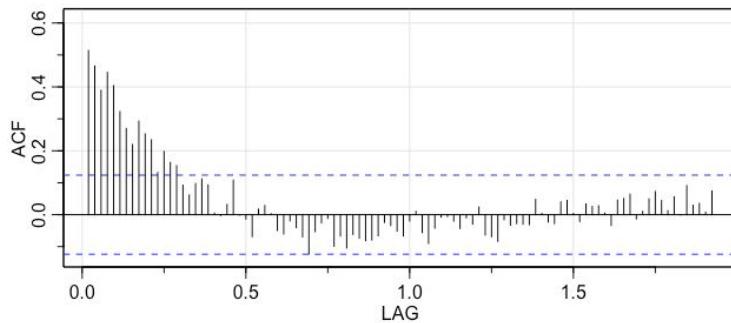


N02 without STL  
Decomposition

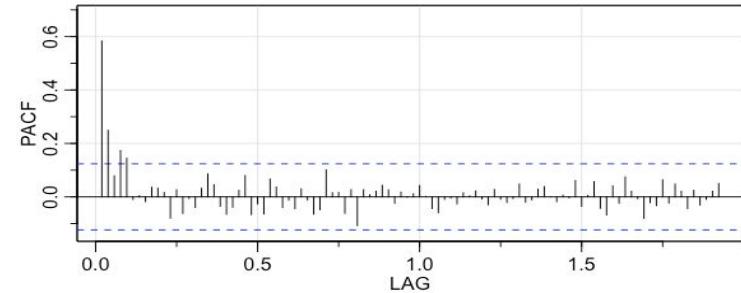
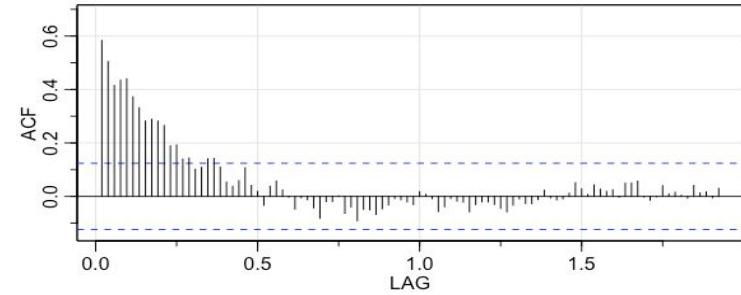


N02 with STL  
Decomposition

# Research Question 1 Results

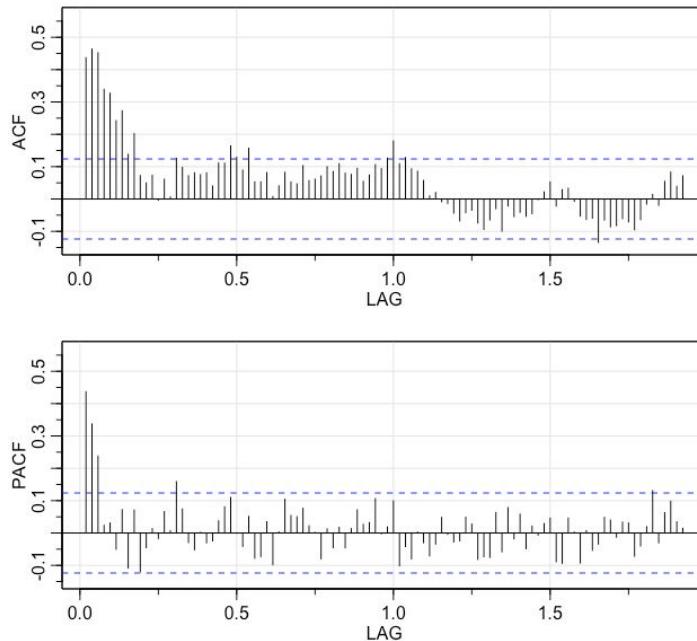


N05 without STL  
Decomposition

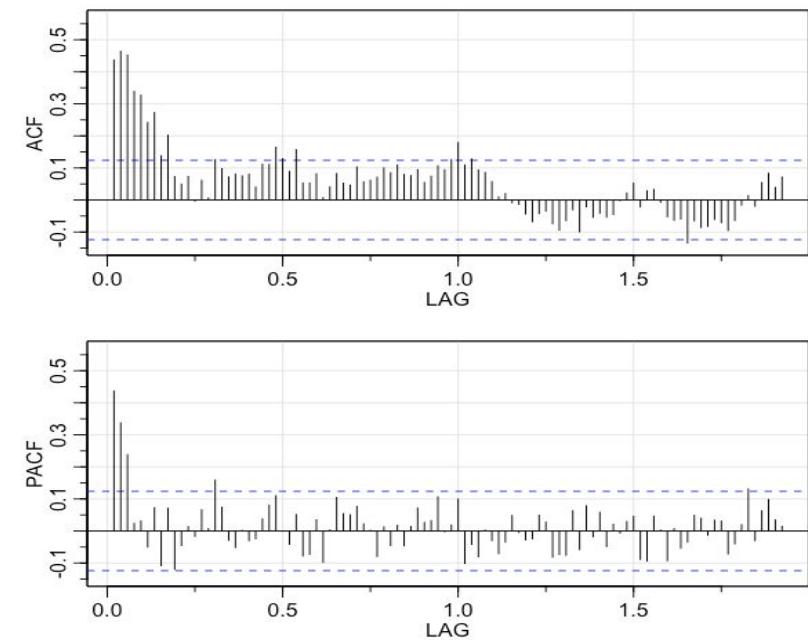


N05 with STL  
Decomposition

# Research Question 1 Results

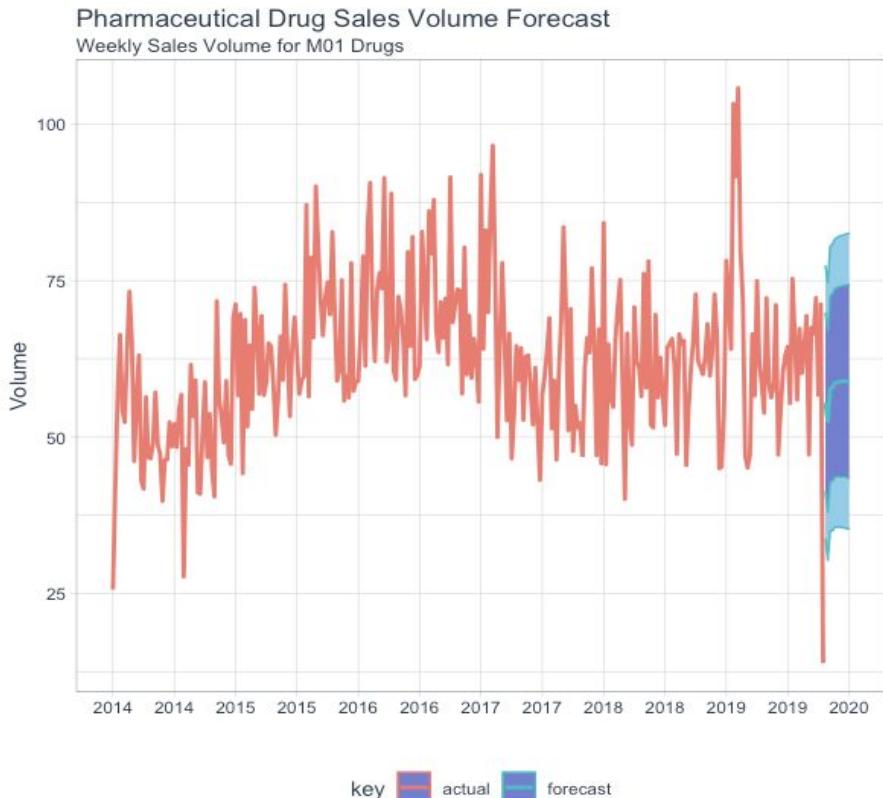


R0 without STL  
Decomposition



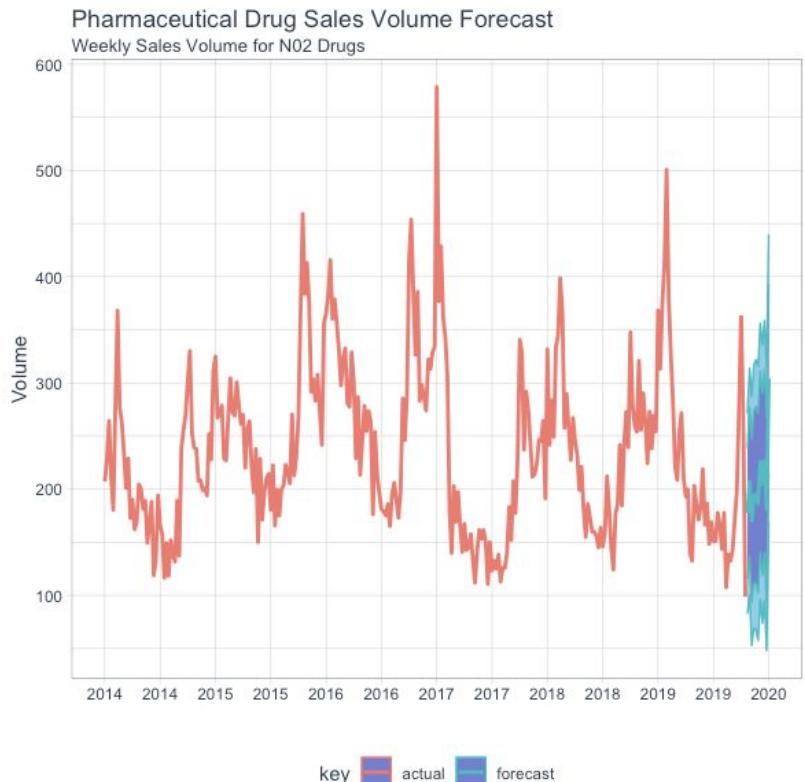
R0 with STL  
Decomposition

# Research Question 2 Results



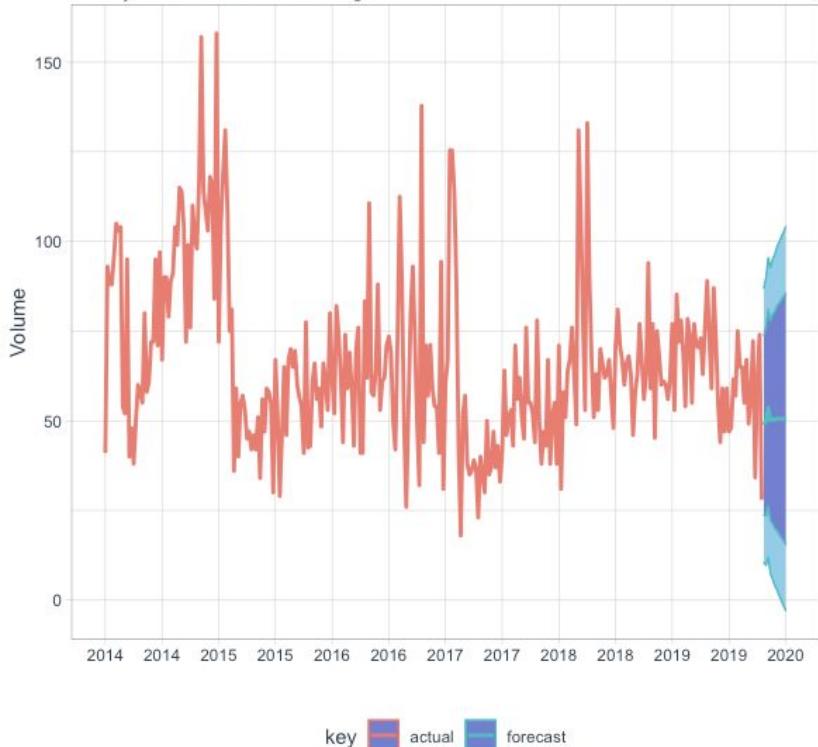
- Time Series for M01 has small prediction interval.
- Possible reasons include: decline in usage, other treatments becoming more popular

# Research Question 2 Results



- Prediction interval indicates a decrease for the forecast
- Legalization of marijuana, natural alternatives

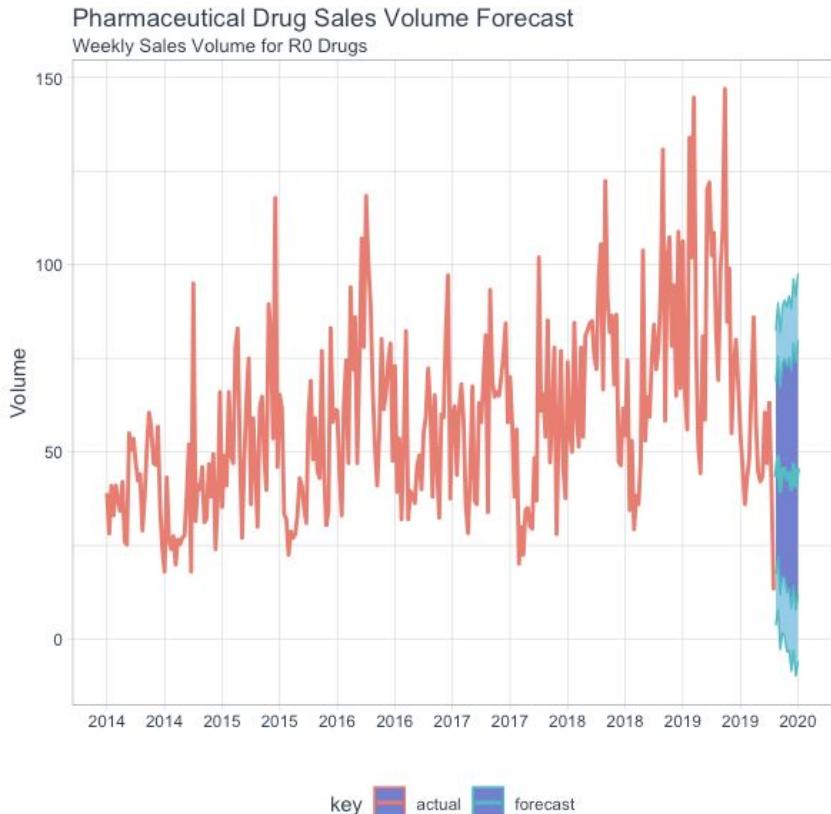
Pharmaceutical Drug Sales Volume Forecast  
Weekly Sales Volume for N05 Drugs



## Research Question 2 Results

- Time series of N05 stays constant
- Increasing behavior initially, but gradually declines and stays constant from 2015 and onward.
- Wide prediction interval

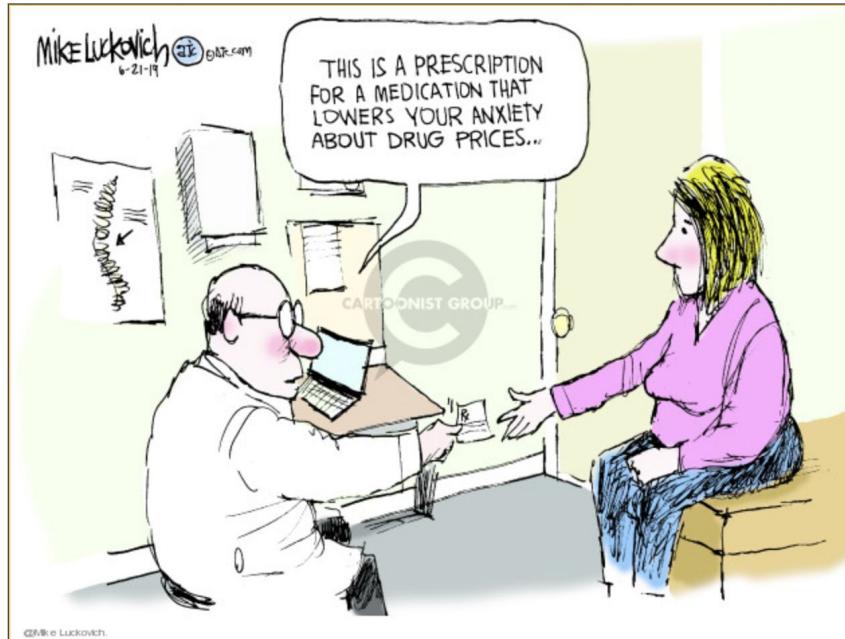
# Research Question 2 Results



- Prediction Interval of R0 indicates a decrease.
- Since 2014, the volume increases but decreases from 2018 and onward.
- This could have been attributed to other forms of medication
- Allergies are seasonal (widest prediction interval)

# Research Question 3 Results

- **M01:** (some opioids could be made illegal)
- **N02:** (legalization of marijuana)
- **N05:** (legalization of marijuana, therapy, microdosing psychedelics)
- **R0:** (increase in vaping could lead to higher usage of R0 drugs)
- New forms of treatment (homeopathic medicine, etc.)
- Covid-19



# Conclusion & Additional Work

- Time Series for drug classes are somewhat inconsistent
- R0 is only drug with trend
- No significant outliers
- Look into a time series on the total quantity of pharmaceutical drugs
- Pharma sales in other countries
- Observe trends of drug data over a longer period of time

