Pharmaceutical Drug Sales Analysis and Forecast

June 10, 2022

DePaul University

Time Series Analysis Forecasting

Oluwafemi Shobowale

INDIVIDUAL REPORT

This project focuses on analyzing, modeling, and forecasting the weekly pharmaceutical drug sale for all eight categories; each member divided the eight categories equally. I explored, modeled, and analyzed the forecast of sales of drugs for obstructive airway diseases and antihistamines for systemic use (R03 & R06). My analysis phase started with a profound visual dive to understand and observe the behaviors like stationarity, seasonality, drift, trendy, normality, and relationships between the two drugs. While observing the graphs, it was vital to ensure the data was valid. The Ljung-box test rejected white noise. The information from this first phase was noted and used in the second phase.

Phase two is an in-depth statistical analysis of every hypothesis I drew from my visual analysis. My hypothesis is not limited to; both series appeared to be additive. Both looked like trend stationery without no drift and no obvious seasonal pattern. How does the mean/variance change, and do drug sales depend on the previous sale? Do the series have any relationship based on time? Can a simple model capture the behavior of the series? In this stage, I plot the ACF, PACF, and EACF to check for stationarity, understand each series AR and MA components, and test for unit roots with DF and KPSS tests. Passing the series through all these tests is vital in the modeling stage. Other analyses also included detrending the series, taking the difference, observing the t-test of the difference for rw/rw+drift, and observing the ACF and PACF of the difference while also testing for white noise and all steps in the analysis.

The modeling phase this stage builds on phase two as it iterates based on the information from the previous step. I built four models, including auto Arima for obstructive airway diseases(R03) and three models for antihistamines for systemic use drugs (R06). After building all the models, it came down to evaluation and performance(forecast).

The first step in my evaluation stage occurred while building the model. I observed the sigma^2, AIC, BIC, and Ljung-Box tests on the residual, picking the model with the most significant coefficient with the highest AIC and BIC value, and all model residuals failed to reject white noise. Other performance metrics include backtesting with mean absolute percentage error (MAPE) as my preferred measure option. The vital part of my evaluation was making sure my model could forecast well beyond a reasonable doubt.

In this project, and like any other analysis project, one vital takeaway is the ability to iterate based on results. Allow the data to direct your steps and answer the question. In this project, I had to go back and forth between testing parameters and how they affect the model and using my intuition to make some decisions. I faced problems where my model had a low p-value for failing to reject white noise in the residual but had a very impressive forecast. I also noticed that the models all had the same effect on the ACF and PACF of the model residuals, making my evaluation very difficult. I used each model forecast performance as the overall model selection metric.

Another exciting takeaway is that hypothesis testing is vital. Understanding how stochastic testing can affect the model building or analysis of a time series or any data is essential. It is also crucial to note that a higher p-value in failing to reject white noise does not make the model superior.