Macy's

Revenue Prediction

Yiyang Duanmu

Table of Contents ★

Executive Summary

Competition in the department store industry has been more intense in recent years. Macy's, Inc. has been regarded as a leading expert in fashion by countless customers over the years. With one of the leading e-commerce platforms in the country, powered by macys.com and a mobile app, along with a countrywide network of shops, Macy's provides the most convenient and seamless shopping experience possible, delivering excellent deals in a variety of categories including clothing, home goods, cosmetics, and accessories. (Macy's About) Macy's gives customers even more options to shop and own their style by offering discounted merchandise at Macy's Backstage and a carefully chosen range of goods in a more compact store format under the name Market by Macy's. Macy's gives awe-inspiring experiences to millions of people every year via its Macy's Thanksgiving Day Parade® and Macy's 4th of July Fireworks® events. (Macy's About) Additionally, the department store assists consumers in commemorating all types of significant life events.

Our dataset consists of information about Macy's annual and quarterly revenue histories, along with their growth rates from 2008 through 2022. The revenue is the first line item on an income statement, and from that number, all other expenditures and expenses are deducted to arrive at the net income.

This project aims to determine how the pandemic has affected Macy's financial performance and to provide ideas for the department store's ongoing reaction to the pandemic. It is clear that Macy's revenue has been on a declining trend throughout the epidemic; this investigation aims to determine whether or not the pandemic is responsible for this revenue trend. We attempted a seasonal naïve model, moving average model, exponential smoothing, ARIMA models, and regression models. The regression models were of the utmost importance to us since we wanted to understand the impact that external variables had on Macy's revenue. Suppose a model that is driven entirely by the data and takes its inputs only from that data is successful. In that case, it will demonstrate that external influences are not the most significant predictors. The absence of a forecasting capacity is fascinating in and of itself. Revenue at Macy's is considerably impacted when external economic variables such as unemployment and GDP growth rates are taken into consideration. We used a time series linear regression model to examine these two factors impact on Macy's revenue. It is possible to observe the effects of the pandemic in both of these cases (the unemployment rate and the GDP growth rate), which indicates that the epidemic has significantly influenced Macy's Revenue.

Introduction and Motivation

Macy's is an American chain of high-end department stores founded in 1858 by Rowland Hussey. Macy's provides services through the Internet under two brand names: Macy's and Bloomingdale's which sells a wide range of merchandise, including men's, women's, and children's apparel and accessories, cosmetics, home furnishings, and other consumer goods. The company had 130,000 employees and earned annual revenue of $24.8 billion as of 2017.

In 2020, COVID-19 devastated the global economy, especially the industry with close contact such as hotels, sports, performing arts, restaurants, clothing stores, and airlines. COVID-19 also forced physical shops to close and consumers to stay at home, therefore, companies such as Uber eats, DoorDash, and Amazon grew well during the pandemic. Specifically, Amazon was in a prime spot to capitalize. Amazon reported a near 200-percent rise in profits, accelerated by much of North America's swift shift to exclusively online shopping. Amazon's revenue was US$96.1 billion, up 37% from 2019, with profits rising to a jaw-dropping US$6.3 billion. The pandemic hasn't only increased the company's profits but also its expansion. Amazon expanded its fulfillment infrastructure by 50% in 2020, adding more than 250,000 employees in the process. For the first time in the company's history, Amazon now employs more than one million workers around the world.

In this project, we wanted to find out how the pandemic has affected Macy's financial performance by forecasting Macy's quarterly revenue. We also wanted to provide ideas for the department store's ongoing reaction to the future pandemic by forecasting Amazon's quarterly revenue data and making comparisons to Macy's forecast.

Data Description

To make forecasts for both Macy's revenue and Amazon's revenue, we used Macy's revenue dataset and Amazon revenue dataset. Since we also incorporate external factors such as GDP Growth Rate and Unemployment Rate for our forecasts, we also used the historical GDP Growth Rate dataset and Unemployment Rate dataset.

[Macy's Revenue]
- **Source:** Macrotrends (The Premier Research Platform for Long Term Investors)
- **Frequency:** Each record represents the quarterly revenue of Macy's ranging from 2008 Q4 to 2022 Q2.

[Amazon's Revenue]
- **Source:** Macrotrends (The Premier Research Platform for Long Term Investors)

- **Frequency:** Each record represents the quarterly revenue of Amazon ranging from 2009 Q1 to 2022 Q2.

[Unemployment Rate]

- **Source:** U.S. Bureau of Labor Statistics
- **Frequency:** Each record represents the monthly unemployment rate ranging from November 2008 to July 2022. We can then calculate the quarterly unemployment rate by averaging the months.

[GDP Growth Rate]

- **Source:** U.S. Bureau of Economic Analysis
- **Frequency:** Each record represents the quarterly GDP growth rate with adjusted inflation ranging from 2008 Q4 to 2022 Q2.

Revenue Analysis

Our project mainly did two parts of analysis. We first try to better forecast Macy's revenue by deploying different models or incorporating external factors such as GDP Growth Rate and Unemployment Rate. Models we chose include Naive and Seasonal Naive Forecast, Moving Average, Exponential Smoothing, ARIMA, and Linear Regression with external factors. Secondly, we build forecast models on Amazon's revenue by deploying ARIMA models and Linear Regression with external factors. After the analysis, we want to find the difference between the model forecast accuracy between the two companies and provide recommendations to Macy's to better prepare for unexpected pandemics like COVID-19 in the future.

For the modeling, we decided to split our Macy's and Amazon's Revenue datasets into training and testing dataset. For the testing dataset, we left the last 10 months, from 2020Q1 to 2022Q2, as the testing dataset, and the previous quarters, from 2008 Q4 to 2019 Q4 as the training dataset. For the external factors, we also chose time frames that equal to the one for the training dataset.

Forecasting Macy's Revenue

Naive and Seasonal Naive Forecast

To analyze Macy's Revenue from the 2008 third quarter to the 2022 second quarter, our team first performed a Naive forecast and a seasonal naive forecast. As the Naive or seasonal naive forecast model only uses the last period observation or the last identical season as the next period forecast without actually predicting, it is better to use it for comparison with more complex models. However, comparing the accuracy of the naive forecast and seasonal naive forecast model, it is more obvious that the seasonal naive forecast model has better accuracy as it has lower MAPE.

Naive & Seasonal Naive Forecast
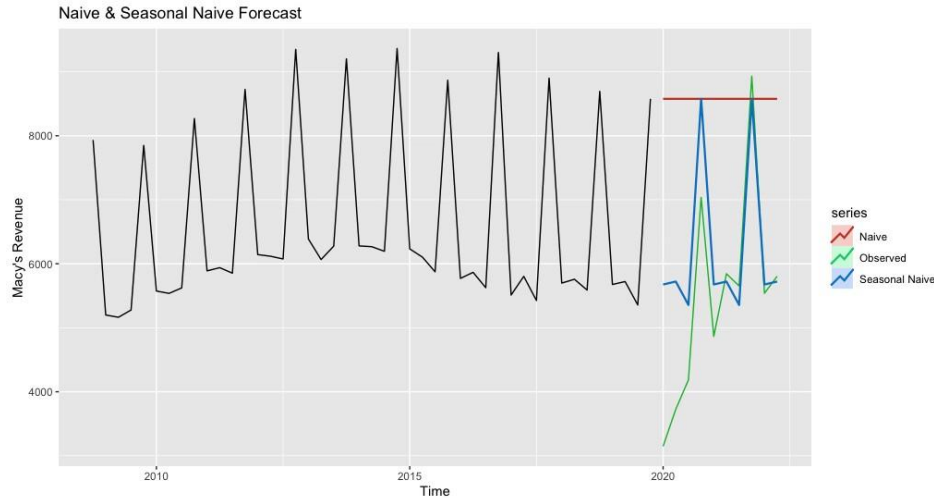
Naive Forecast	RMSE	MAPE
Training set	2142.21	22.47
Test set	3485.68	71.08

| Seasonal Naive Forecast	RMSE	MAPE
Training set	279.06	3.60
Test set	1225.75	21.55

Moving Average

Then we performed the Moving Average smoothing method with trailing right and centered. After applying rollmean, rollmax, rollmedian with trailing right and checking the centered moving average, we identified that moving mean and moving median are more reasonable in presenting the observations in the graph, so we used a one-step ahead rolling forecast of these two methods and compared the accuracy of these two. Finally, in the moving average method, we concluded that using roll median is more accurate than using roll mean because it generated fewer errors. However, as they still have high MAPE, we want to apply more forecast methods to find the most accurate model of all.

Moving Average
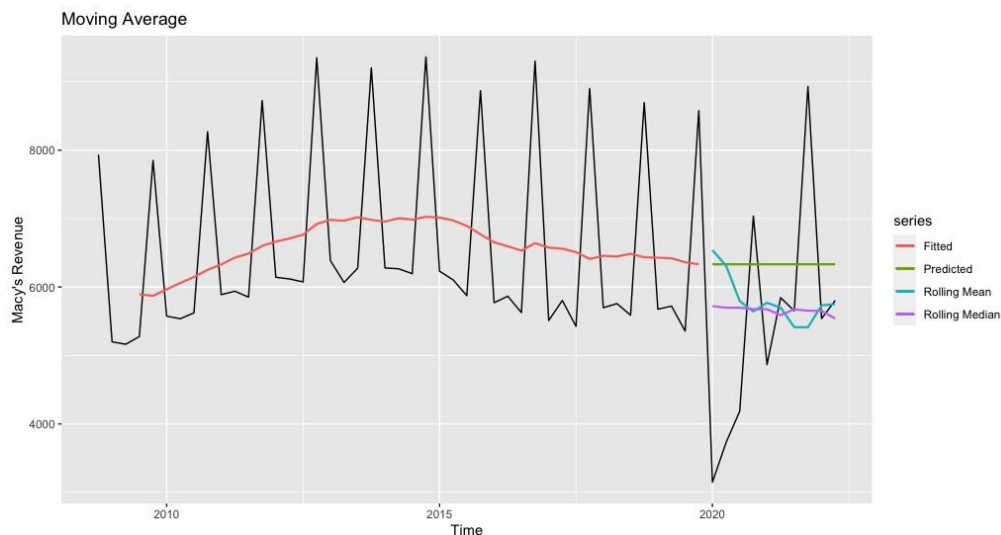
| Test Set	RMSE	MAPE
Roll Mean	1898.38	30.45
Roll Median	1618.55	25.49

Exponential Smoothing

Apply the Holt-Winter's exponential smoothing (with additive trend and seasonality). First, we plot the observed data during the training set, and the forecasted revenue during the validation periods. After comparing all models, we believe model ZZZ(MNA) performs the best. It has the lowest RMSE in both the training and test sets.

Forecast Macy Revenue (Exponential Smoothing)
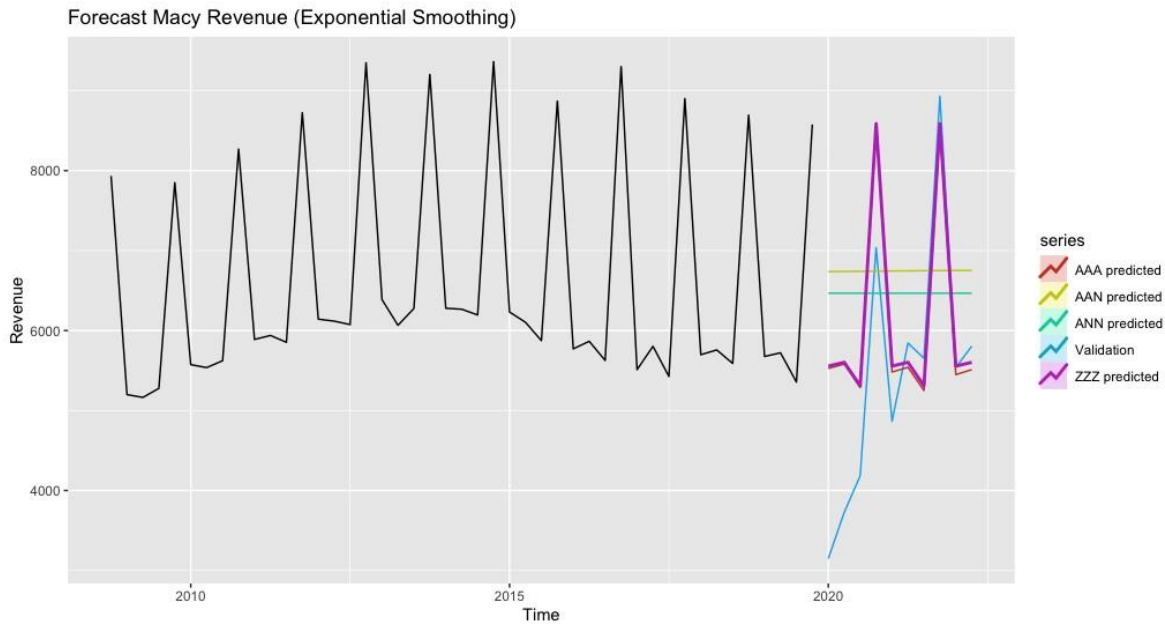
| ZZZ Model (MNA)	RMSE	MAPE
Training set	211.30	2.46
Test set	1174.49	20.77

Seasonal ARIMA

ARIMA model is using the combination of Autoregression and Moving Average approaches and makes forecasts. To choose the model parameters, we observed the training dataset after detrending or deseasonality or removing both trends and seasonality. From the plots, we found that detrend provides a more obvious pattern, with tailing off ACF and cutting of PACF. What's more, because there is a seasonality pattern shown in the ACF, we decided to incorporate a seasonal component in our ARIMA model. After, we tried several combinations of parameters and decided on the best one with the lowest test MAPE.
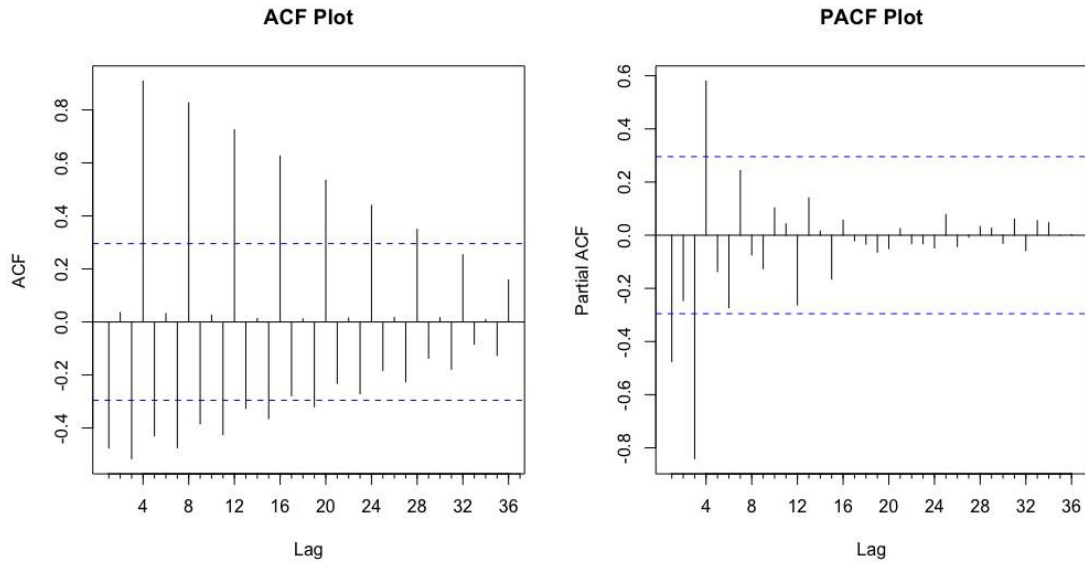
ACF Plot

PACF Plot

We first tried ARIMA(1,1,1)(1,1,0)[4] according to the ACF and PACF plots. Because PACF is showing significant value when lag = 1, and it is also showing a little tailing off pattern. Therefore, we chose (1,1,1) for the first part. For the seasonality part, we choose to apply both autoregressive and a constant trend into it. The model fitted successfully, and the residuals met the assumptions of the ARIMA model.

	RMSE	MAPE
Training set	201.89	2.23
Test set	1170.49	21.71

Residuals from ARIMA(1,1,1)(1,1,0)[4]
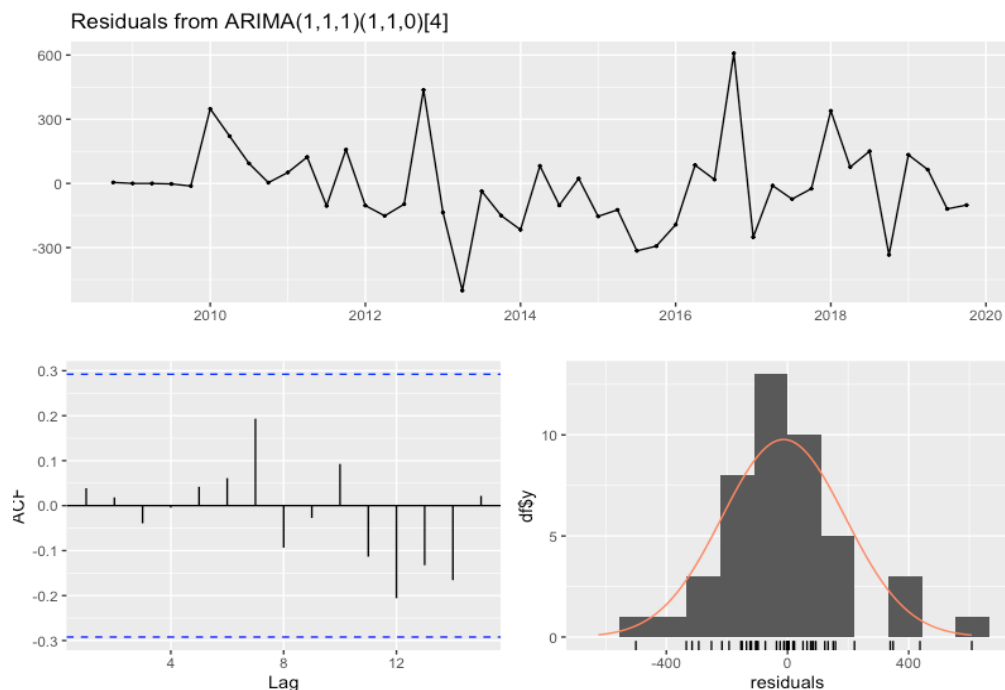
After, we also tried several combinations of parameters and found ARIMA(2,0,1)(1,0,0)[4] is performing well with the lowest MAPE value. We think this is caused by the instability of the ARIMA model itself, but since it is leading the lowest MAPE value, we decided to pick this model as the final ARIMA model we chose. The residuals also met the assumptions of the ARIMA model.

	RMSE	MAPE
Training set	222.37	2.64
Test set	1182.42	20.91

Residuals from ARIMA(2,0,1)(1,0,0)[4] with non-zero mean
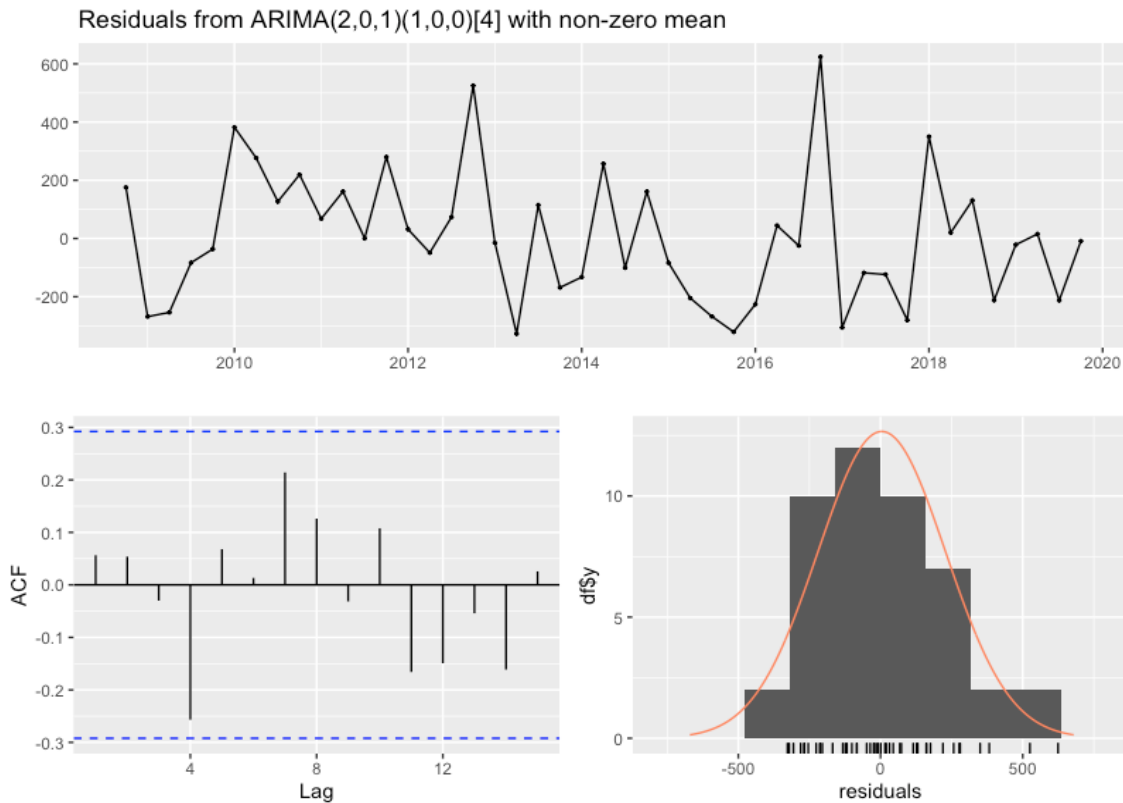
External Factors Analysis – Linear Regression & ARIMAX

Two main external factors that we think can reflect the impact of COVID-19 are GDP Growth Rate and Unemployment Rate. According to the time series plots of these two factors, we can see that there is a significant fluctuation at the beginning of the COVID-19 pandemic. Therefore, we decided to study the impact of the Unemployment Rate and GDP Growth Rate (Inflation-adjusted) on Macy's Revenue.

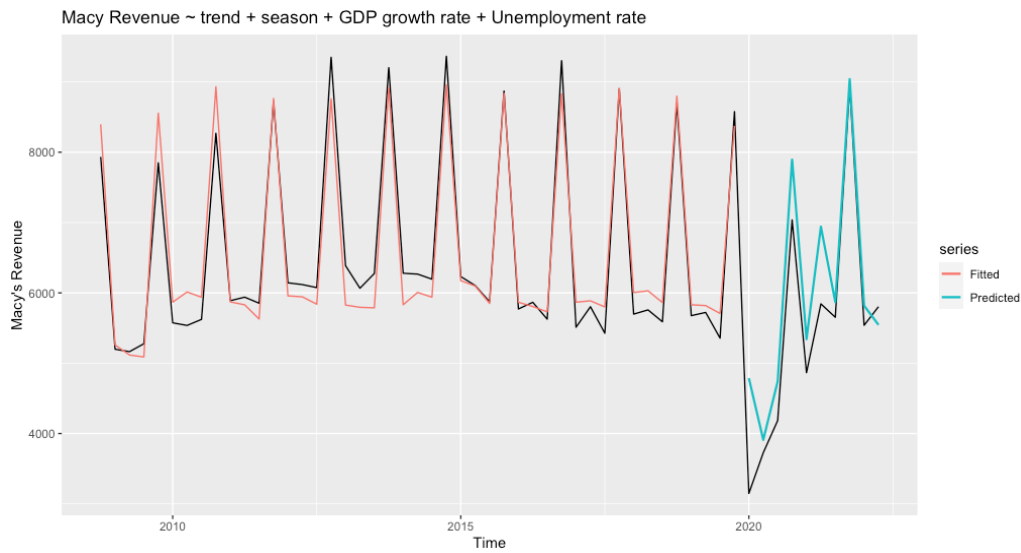**Unemployment Rate**   **GDP Growth Rate (Inflation-adjusted)**
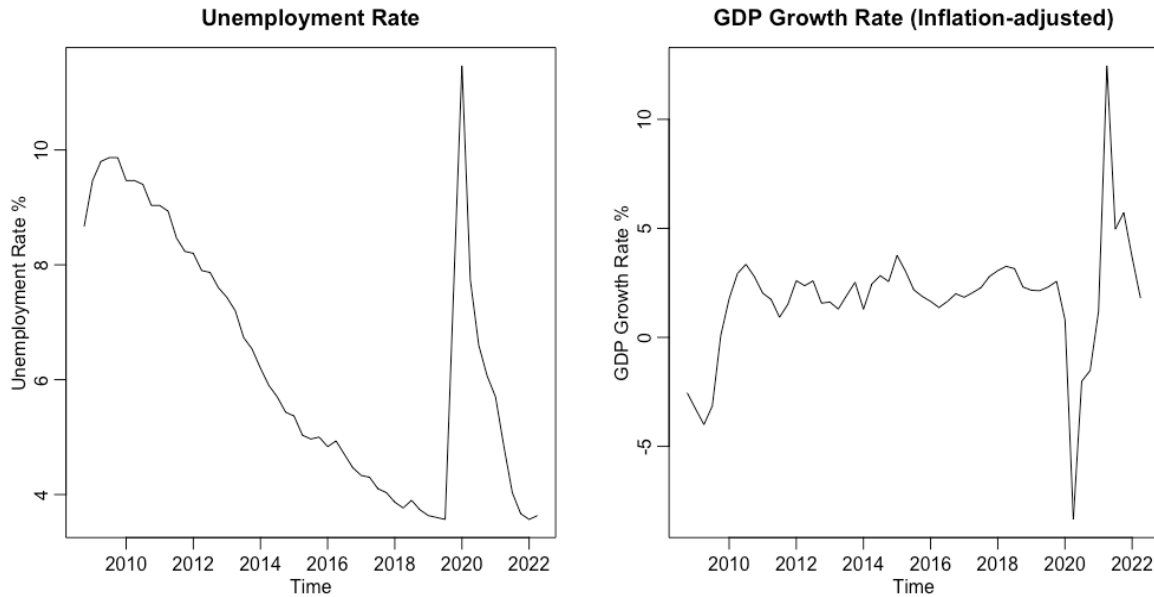
After creating a seasonally adjusted series of Macy's Revenue by rollmean, we calculated the correlation coefficient values between these two variables and Macy's Revenue and found that GDP Growth Rate (Inflation-adjusted) does have correlation with the revenue and Unemployment Rate may have a weak correlation or even have no correlation with the revenue.

COR(Unemployment Rate, Revenue)	COR(GDP Growth Rate, Revenue)
- 0.187	0.283

Then, we established several time series linear regression models which consider the impact of unemployment rate, GDP growth rate, trend, and seasonality of the revenue after splitting the data set into two parts: training set and test set.

| Model	Variables	R Square	MAPE
LM1	trend, season, gdpg	0.9443	Training: 3.88 Test: 19.90
LM2	trend, season, ur	0.9275	Training: 4.81 Test: 19.75
LM3	**trend, season, gdpg, ur**	**0.9478**	**Training: 3.78 Test: 12.56**

According to the MAPE and R Square, Model LM3 which considers the impact of both Unemployment Rate and GDP Growth Rate performs best. Although Unemployment rate has a low correlation coefficient value and its P value is not significant, we still consider it as an essential factor, because it does help improve the accuracy of forecasting (MAPE: from 19.90 to 12.56) and the Adjusted R Square of Model 3 (from 0.9371 to 0.9395).

Macy Revenue ~ trend + season + GDP growth rate + Unemployment rate

	RMSE	MAPE
Training set	309.61	3.78
Test set	734.80	12.56

Then, we also tried seasonal ARIMAX models to study the impact of GDP Growth Rate and Unemployment Rate on Macy's Revenue.

ARIMAX(2,0,1)(1,1,0)[4] Model	External Variables	MAPE
AXM1	gdpg	Training: 2.19 Test: 18.52
AXM2	ur	Training: 2.19 Test: 18.65
AXM3	**gdpg, ur**	**Training: 2.12 Test: 16.38**

ARIMAX(2,0,1)(1,1,0)[4](GDP growth rate,Unemployment rate)
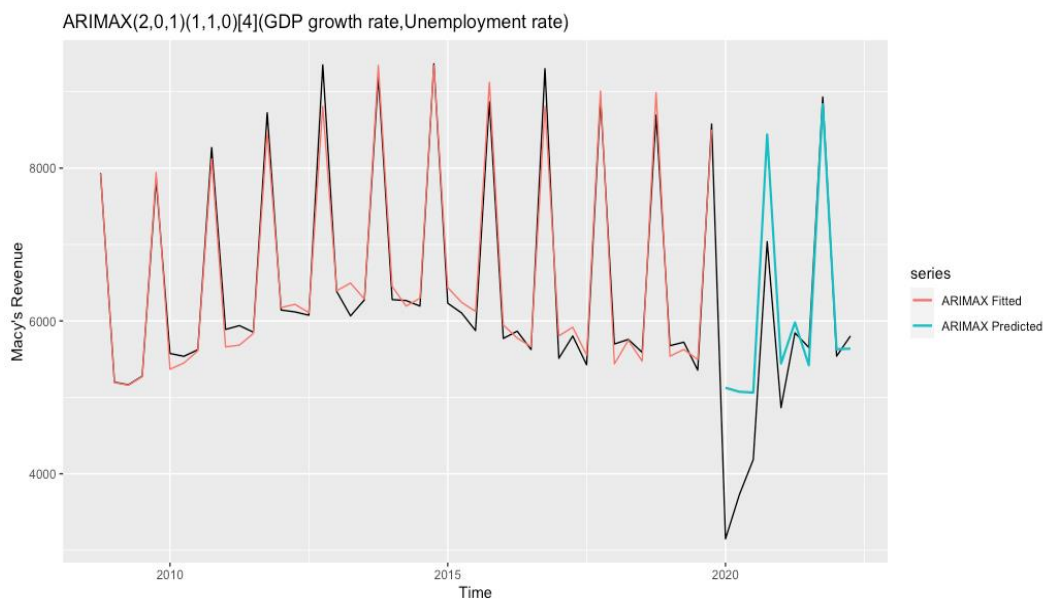
|	RMSE	MAPE
Training set	190.93	2.12
Test set	944.05	16.38

Summary

Based on the performances of above models, we can find the best model to forecast Macy's Revenue is the time series linear regression model LM3 which considers the impact of both Unemployment Rate and GDP Growth Rate. The MAPE of the test set of LM3 is 12.56 which is the lowest one. In addition, we can find LM3 does a really good job in predicting Macy's Revenue during the early period of COVID-19 and this reflects that the two external factors (GDP Growth Rate and Unemployment Rate) do have a strong impact on Macy's Revenue especially in the early period of COVID-19. In other words, COVID-19 does have a strong negative impact on Macy's Revenue which cannot be ignored when predicting Macy's future revenue.

Competitor Analysis - Amazon

For comparison, we chose Amazon to investigate if COVID-19 also impacted its revenue severely and if we can still make a good forecast on its revenue under the COVID-19 pandemic. If we can make a good prediction, it means Amazon's revenue was not affected by the pandemic and Macy's can learn from Amazon's operation. What's more, we also want to investigate Amazon's revenue with external factors including GDP Growth Rate (Inflation-adjusted) and the unemployment rate. The model we chose for Amazon is the seasonal ARIMA model, linear regression models and seasonal ARIMAX models.
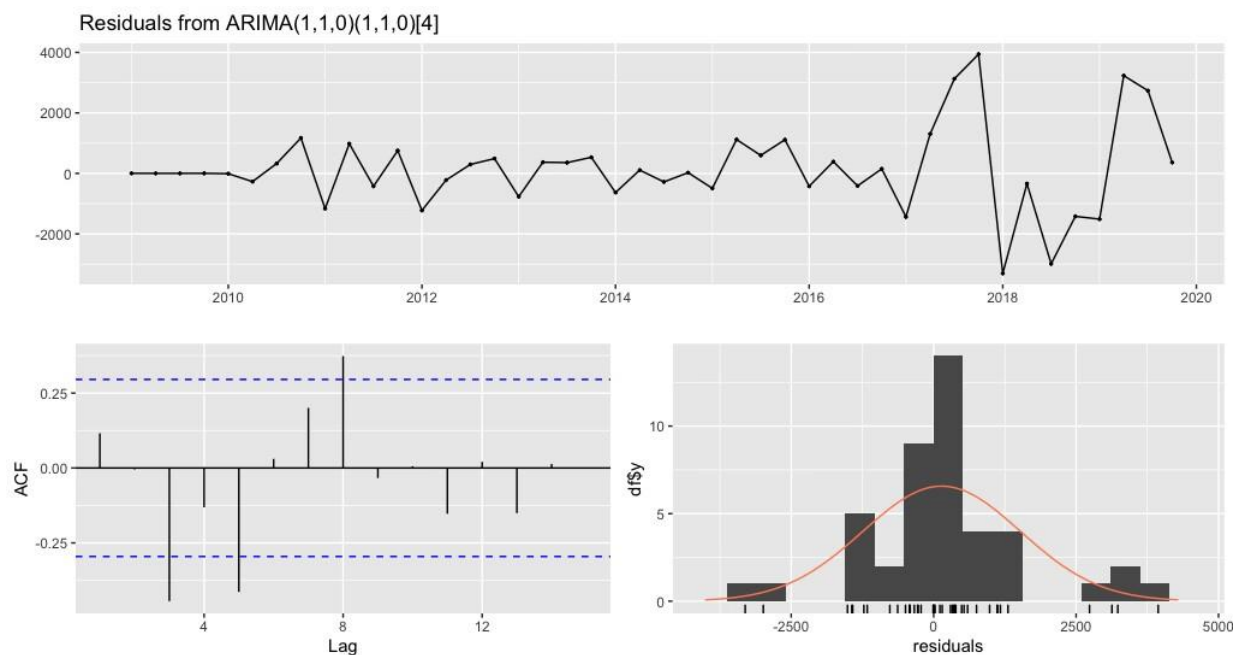
Seasonal ARIMA

Similar to the procedure we performed on Macy's revenue dataset, we detrended the dataset and observed the ACF and PACF of Amazon's revenue. It also has a tailing-off pattern for ACF and a cutting-off pattern for PACF.
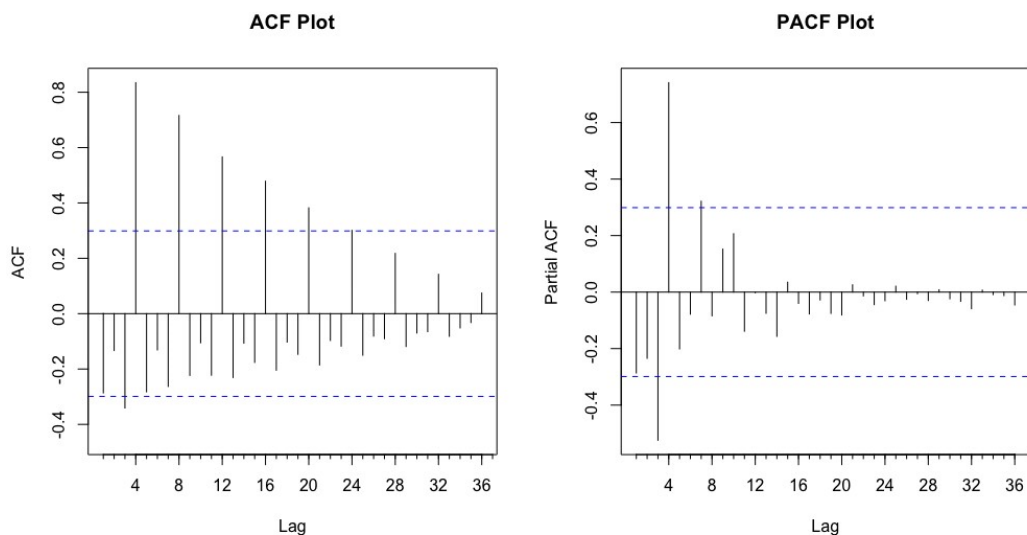
According to these two plots, we decided to incorporate AR in the model. After trying several combinations of parameters, we chose ARIMA(1,1,0)(1,1,0)[4] as our best model for Amazon. Although the residuals still have some correlation with their value when lag equals 3 or 5, we do not see a seasonality effect that we need to remove.
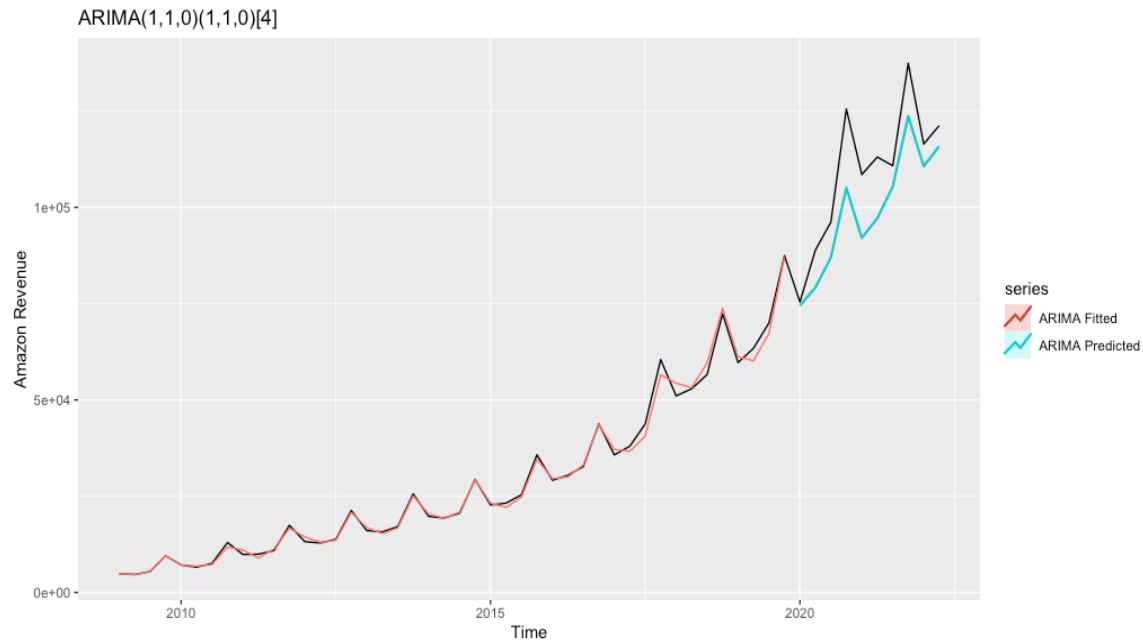

	RMSE	MAPE
Training set	1374.98	3.27
Test set	11842.37	9.16

## *External Factors Analysis – Linear Regression & ARIMAX*

We also analyzed the impact of GDP Growth Rate and Unemployment Rate on Amazon's Revenue and established several linear regression models and the seasonal ARIMAX models.

Model	Variables	R Square	MAPE
ALM1	trend, season, gdpg	0.9075	Training: 28.62 Test: 33.15
ALM2	trend, season, ur	0.9485	Training: 24.59 Test: 24.56
ALM3	trend, season, gdpg, ur	0.9575	Training: 18.39 Test: 26.56
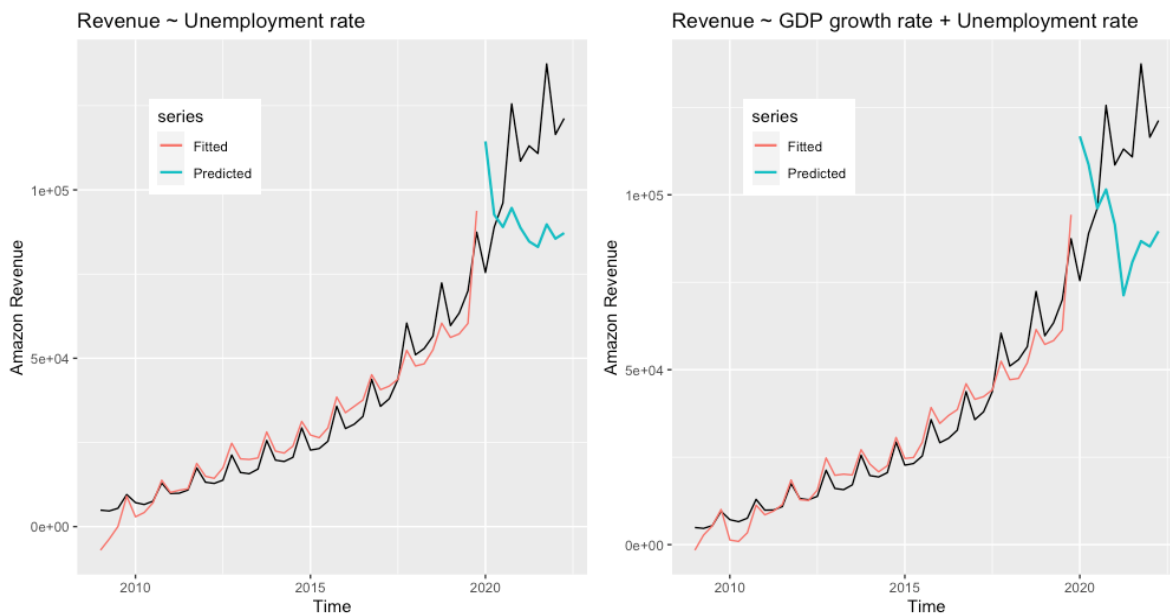
According to the performance of linear regression models, we found these two external factors will mislead the linear prediction of Amazon's Revenue especially in the period of COVID-19.

ARIMAX(1,1,0)(1,1,0)[4] Model	External Variables	MAPE
AAXM1	**gdpg**	**Training: 3.25 Test: 9.02**
AAXM2	ur	Training: 3.25 Test: 10.30
AAXM3	gdpg, ur	Training: 3.2 Test: 10.51

ARIMAX(1,1,0)(1,1,0)[4](GDP Growth Rate)
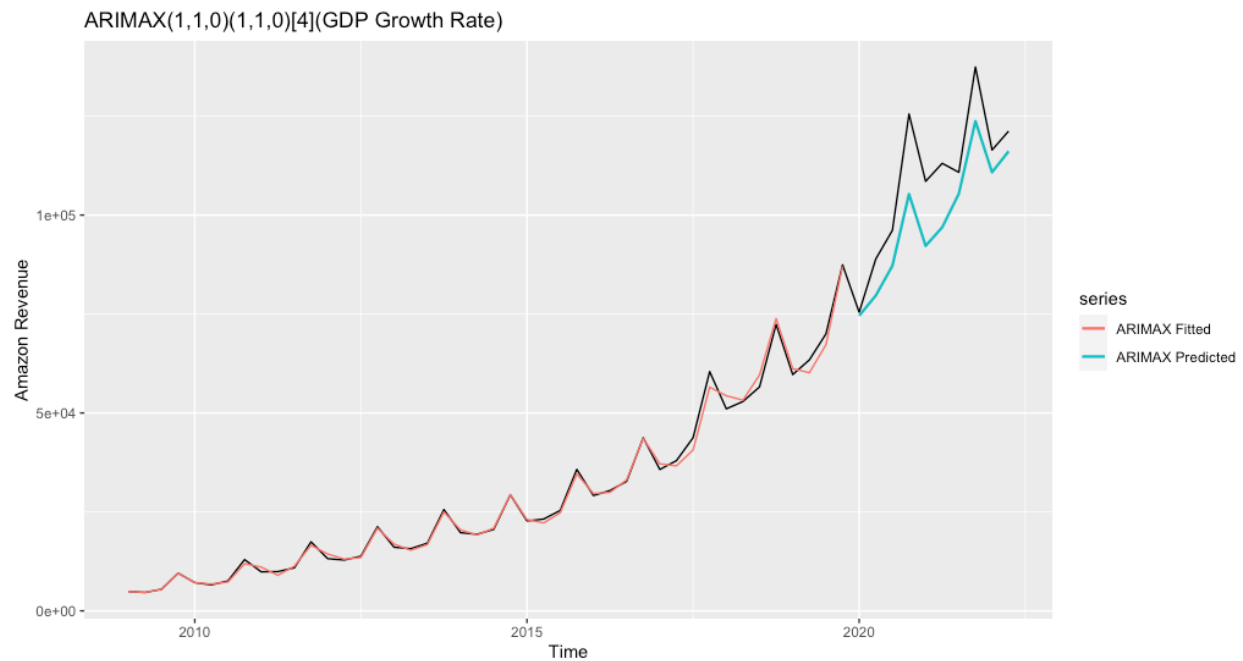
According to the performance of ARIMAX models, we found the best ARIMAX model is AAXM1 which performed slightly better than the common ARIMA model and other ARIMAX models all performed worse. This means GDP Growth Rate and Unemployment Rate may have a weak impact or even no impact on Amazon's Revenue.

Summary

Based on the performances of each model of predicting Amazon's Revenue, we can find the best model is AAXM1 which is a seasonal ARIMAX model which considers the impact of GDP Growth Rate. However, the performance of AAXM1 is slightly better than the common seasonal ARIMA model and we can find that the predicted lines of these two models are almost the same. This means that we can nearly ignore the impact of external factors when predicting Amazon's Revenue. In other words, Amazon's revenue is barely impacted by COVID-19. Therefore, we can obtain business insights and provide recommendations by comparing the difference between Macy's and Amazon.

Conclusion & Recommendations

According to the analysis above, we mainly studied the impact of COVID-19 on Macy's revenue and Amazon's revenue via focusing on the revenue itself and the impact of inflation-adjusted GDP Growth Rate and Unemployment Rate. We think COVID-19 does have strong negative impact on Macy's revenue while Amazon's revenue is barely impacted by COVID-19. Therefore, we can then provide recommendations based on the analysis.

As our revenue source mainly comes from sales, our recommendation for Macy's will mostly focus on sales. According to our revenue analysis, Macy's competitor, Amazon, didn't lose much revenue and had a constantly increasing trend during the pandemic. It calls us to think

more strategically about how Macy's should make improvements on its operating strategy to meet any potential crisis in the future.

As our analysis indicates, Macy's should and is now transferring to more digital led than focusing on off-line shopping. For the fourth quarter of 2021, Macy's online revenue occupies 58% of all revenue, so it should pay more attention to online retail now as the pandemic has changed people's shopping habits more from off-line to online.

Our team would recommend Macy's to improve the online shopping experience for customers to shop online in a simple way just like they did for offline shopping. It can achieve this by adjusting the website interface to shorten the user's shopping path, simplifying the price to make it more clear for customers, offering a more personalized shopping experience for customers, and putting more budget on online promotion to attract more customers to encourage sales online.