Forecasting Amazon Revenue/ Net Income
Time Series Forecasting Final Project – Report

Prepared by Group 1 –
Harshit Mittal
Linh Ha
Nam Dang
Neelam Arya
Vishal Doshi

Introduction

Background:
Amazon has always had a large discrepancy between its revenue and profits (Net Income). This is because Amazon invests a large chunk of their revenue on capital expenses such as Data centers, warehouses, upgrading distribution centers, etc. for 2 reasons:

- To not incur interest rates on borrowing sums of money to do the aforementioned projects.
- To write off the amount spent as expenses and reduce tax lability.

To be able to forecast with reasonable accuracy the revenue and net income for the upcoming quarter would enable amazon to plan in a better way their expenses and help them make better business decisions.

Business Problem:
To forecast Amazon finances.

Objectives:

- To build forecasting models and to forecast the revenue and profits projected for Amazon.
- To compare the forecasted figures with the real-world numbers provided by Amazon.
- To determine which forecasting model performs best.

Motivation:
Amazon is one of the largest corporations on the planet, they have hundreds of divisions that require prior resource allocation to be able to plan their activities and projects for the forthcoming quarter. We want to see how accurately we are able to forecast their finances.

Analysis Methods:
We have used 2 different iterations of 2 forecasting methods (ETS and ARIMA) to make the forecasts for Revenue forecasting as well as Net Income forecasting.

Data Exploration

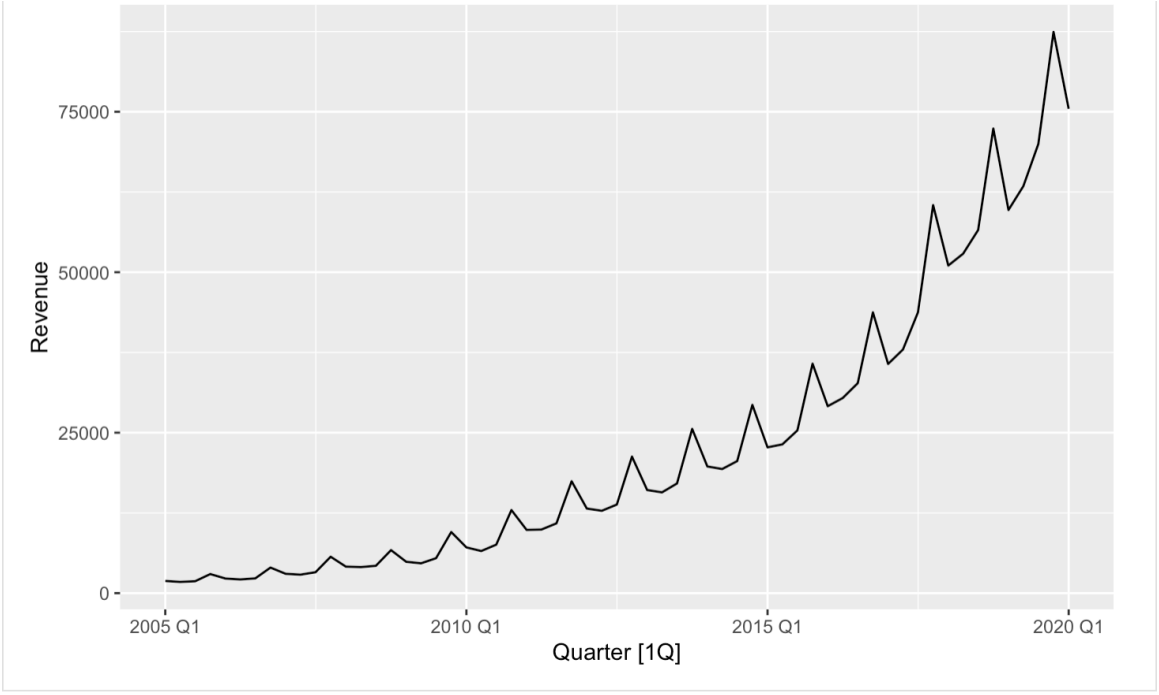
Source:
We are using a dataset of Amazon profits provided by Amazon.

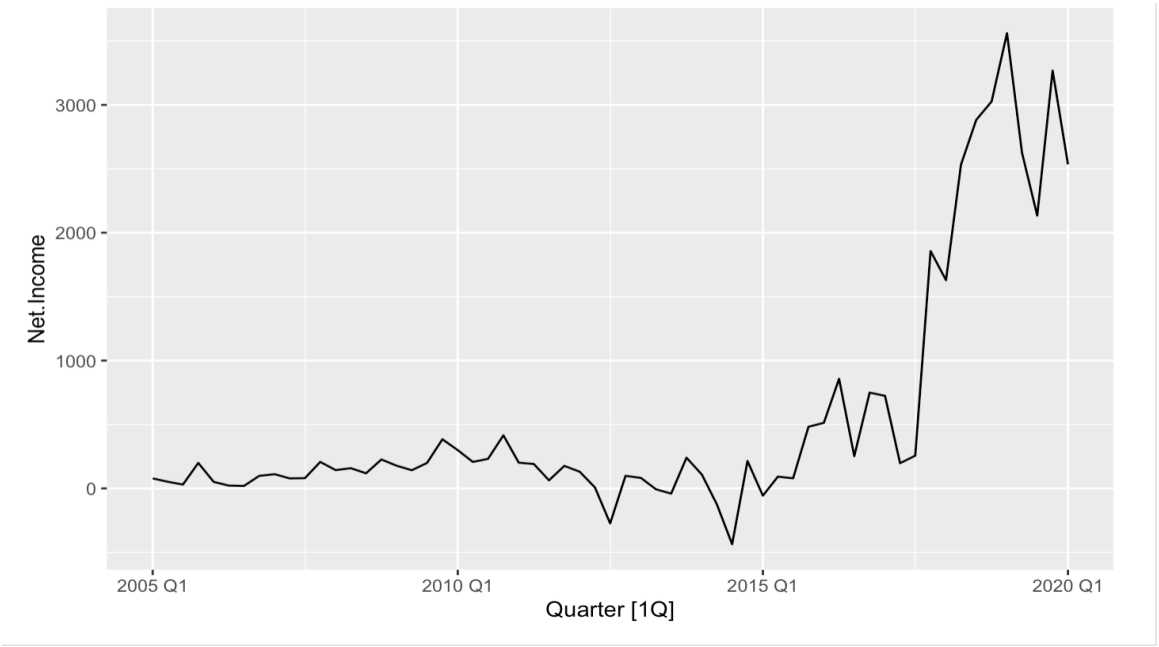
Dimensionality and description:
The dataset contains **62 rows and 3 columns**.
The data contains **quarterly Revenue and Net Income values** for Amazon from **Q4 of 2005 to Q4 of 2019**.

Data Quality:
There are **no missing values** in the dataset.

Data Visualization

In our data exploration, we wanted to see if Amazon Revenue and Net Income have a trend and/or seasonality. **We notice Revenue of Amazon is seasonal with a peak in Q4** and its revenue has a sharply **upward trending**. However, for **Net Income, there seems no clear seasonality**. Net Income of Amazon **starts flat** at the beginning of selected period, and **suddenly increases around 2018**. This is due to the revised tax plan that came into force because of president Trump.





Model Considered

1. **ETS Model**

   An 80-20 split for training and testing is used for both Revenue and Net income forecasting.

   a. **For Revenue:**
      i. R suggests that we should use a model with components **ETS(M,A,M)**
      ii. Forecasting methods of **Holt and Damped Holt's** method are used.

   b. **For Net Income:**
      i. R suggests that we should decide to use a model with components **ETS(A,N,A)**
      ii. Forecasting methods of **Holt and Damped Holt's** method are used.

2. **Arima Model**

   A 90-10 split for training and testing is used for both Revenue and Net income forecasting.

   a. **For Revenue:**

   We took one seasonal and one non-seasonal difference as the series was not stationary.

      i. R suggests that we should use the model ARIMA(3,1,2)(1,1,0)4
      ii. We used another model ARIMA(0,1,2)(0,1,1)4, because we see that the PACF lags are exponentially decaying so we checked for MA process, there were two significant non-seasonal lags in the ACF plot, so we chose q=2 and p=0, and there was one seasonal significant lag in ACF therefore we chose Q=1and P=0.

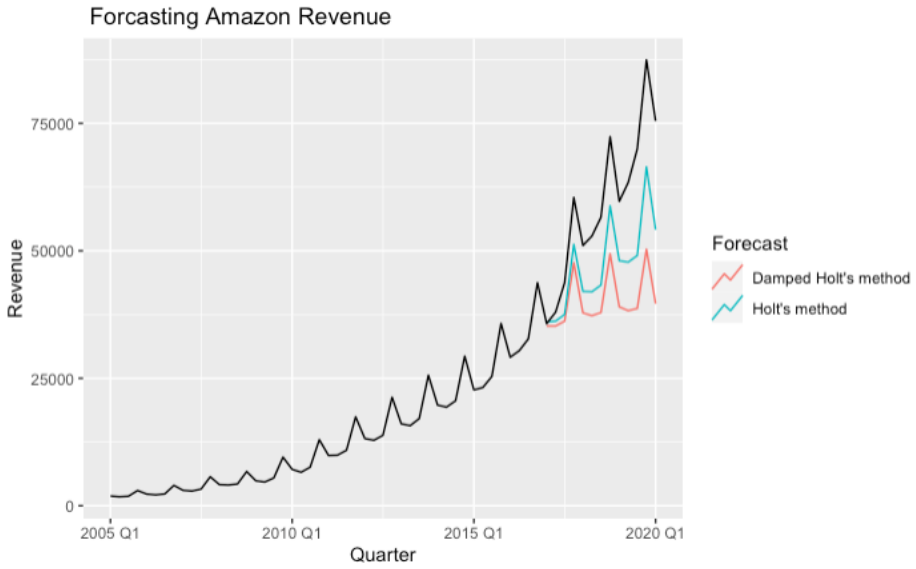
   b. **For Net-Income:**

   We took one non-seasonal difference as the series did not have constant variation.

      i. R suggests that we should use the model ARIMA(0,1,0)(2,0,0)4
      ii. We used another model ARIMA(4,1,0) because we see that the ACF lags are exponentially decaying so we checked for AR process, there were 4 significant non-seasonal lags in the PACF plot, so we chose p=4 and q=0.

Model Forecasts and Accuracy

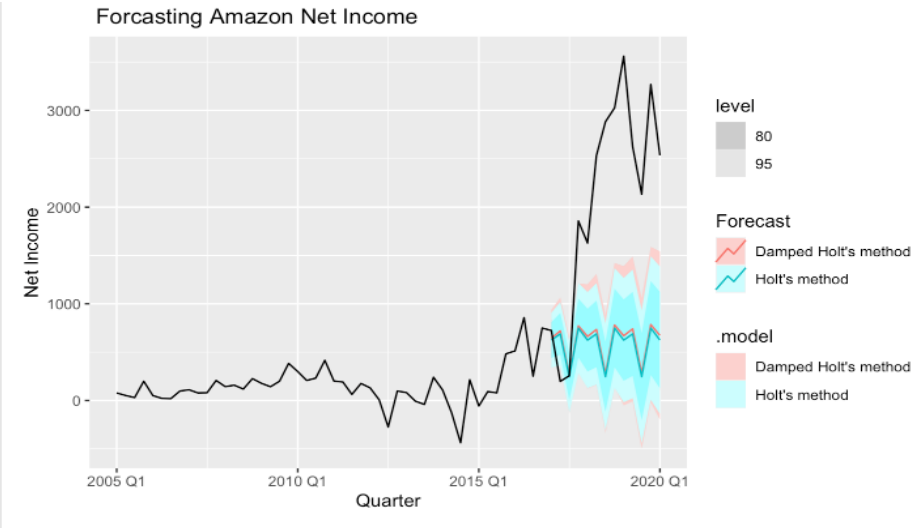
1. **ETS Model**
    a. **For Revenue:**



.model <chr>	.type <chr>	ME <dbl>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>	MASE <dbl>	RMSSE <dbl>	ACF1 <dbl>
Damped Holt's method	Test	18779.44	21885.69	18779.44	28.92978	28.92978	6.480902	6.266761	0.7520005
Holt's method	Test	11853.10	13585.96	11907.70	18.45265	18.60552	4.109420	3.890212	0.7130938

By plotting forecast result, we come up our first conclusion that **Holt's method performs better than Damped Holt's** one as the forecast values of Holt's method are closer to the actual values. Also, the **accuracy of Holt is also higher showing in its much lower RMSE and MAPE compared to the other method.** However, both models seem to have significantly **high forecast errors, so exponential smoothing models do not appear to be the most accurate model** to forecast Amazon's Revenue.

    b. **For Net Income:**



.model <chr>	.type <chr>	ME <dbl>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>	MASE <dbl>	RMSSE <dbl>	ACF1 <dbl>
Damped Holt's method	Test	1479.264	1807.300	1560.874	37.3396	78.63599	10.90134	8.425002	0.7277493
Holt's method	Test	1513.881	1839.076	1589.547	40.4657	78.87479	11.10160	8.573130	0.7304817

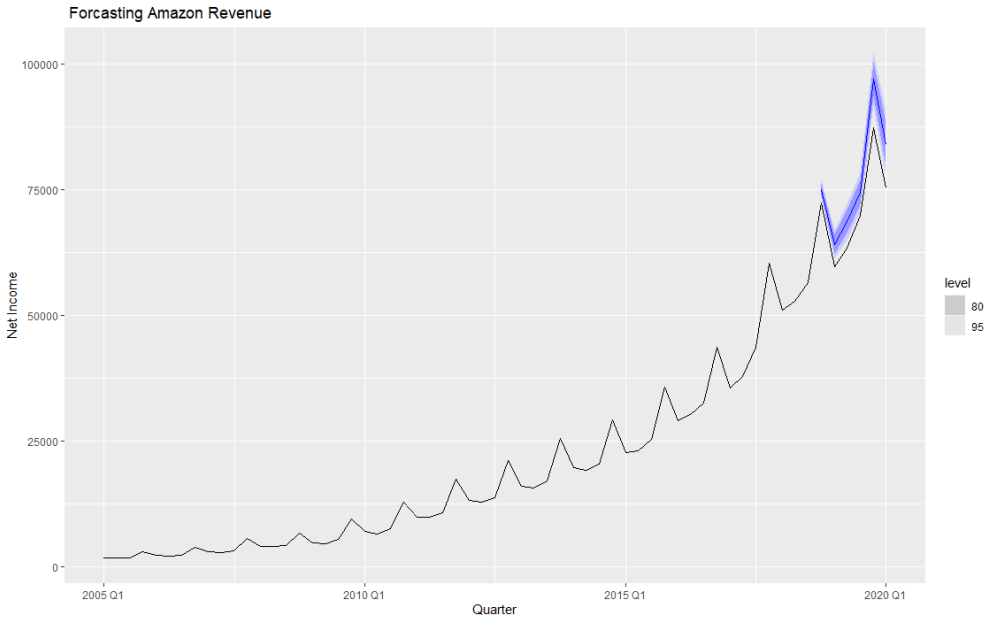
By plotting forecast result, we observe that both of the methods perform poorly on the testing dataset. **None of those models is able to forecast a sharp increase in 2018**. In fact, Amazon Net Income goes up from $1.8B to $2.5B in 2018, up 35%. Amazon had extensively expanded its footprint in E-Commerce industry by opening the second headquarter, starting Amazon Fresh, launching agreements with multiple large retail brands. As Amazon became the only popular E-commerce platform in the US, it made Amazon.com a go-to-retailer for holiday shopping. The cloud service business segment of Amazon continued to have a successful quarter, contributing 50% to the bottom lines of the company.

With strategy deployments, our **forecast models are not able to accurately predict a jump in the Net Income** which explain why the **accuracy is relatively low for both methods**. We would not recommend using this exponential smoothing models for forecasting Amazon's Net Income. We think forecasting should be more practical if we try to apply forecasting methods on the dataset from Q1'2018 until YTD.

2. ARIMA Model

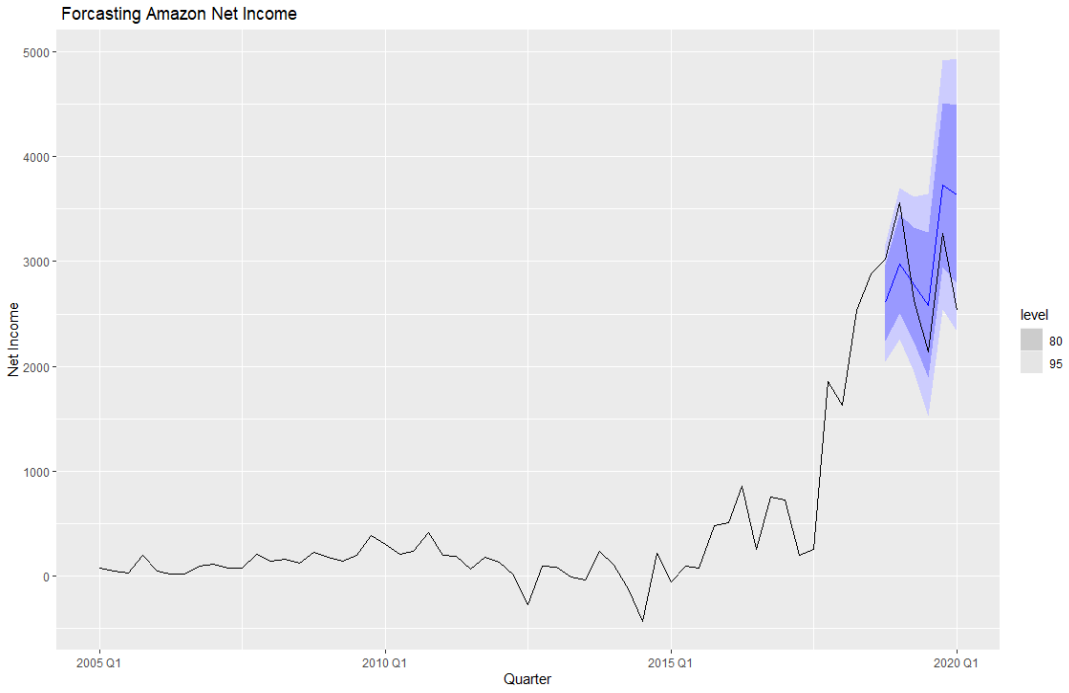
a. For Revenue:

We had same order of differencing for both the models that we tried for revenue, hence, we can make a comparison based on AICc, we find that the model that was suggested by R had the least AICc measure, so we chose the auto model ARIMA(3,1,2)(1,1,0)4 as better of the two and used it for forecasting.



b. **For Net Income**

Wen did not have same order of differencing for both the models that we tried for net income. The auto model had seasonal ARIMA whereas we chose a model without the seasonal component hence we made our comparison based on the accuracy measure RMSE, we found that ARIMA(4,1,0) had had smaller RMSE measure than the auto model, so we chose the model ARIMA(4,1,0) as better of the two and used it for forecasting.



Conclusions and Recommendations

1. Based on the Accuracy measures, error measures and graphical validation due to closeness between actual and predicted values, we came to the conclusion the **ARIMA(3,1,2)(1,1,0)4** is the better in **forecasting Revenue** compared to the **Holt's Method.**

.model <chr>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>
Damped Holt's method	21885.69	18779.44	28.92978	28.92978
Holt's method	13585.96	11907.70	18.45265	18.60552

A tibble: 1 x 5

.model <chr>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>
ARIMA(`Revenue (US $M)`)	6385.	5904.	-8.12	8.12

2. **None of the methods** can be said to accurately **forecast Net Income data** due to there being a sharp rise in 2018 due to external tax policy decisions and cannot recommend using either of them. But **ARIMA(4,1,0)** has the best Accuracy measures and predictions intervals.

.model <chr>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>
Damped Holt's method	1807.300	1560.874	37.3396	78.63599
Holt's method	1839.076	1589.547	40.4657	78.87479

A tibble: 1 x 5

.model <chr>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>
ARIMA(`Net Income (US $M)` ~ pdq(4, 1, 0))	718.	645.	-19.7	24.3

3. Based on the above conclusions, we recommend that:

    a. **ARIMA(3,1,2)(1,1,0)4** Model be used for forecasting Revenue data in the future.
    b. **ARIMA(4,1,0)** Model be used for forecasting Net Income data in the future.