# Forecasting using Time Series Analysis

Discussion and application of ARIMA and ETS models on stock prices

Final Examination Paper - Report
Predictive Analytics

| | |
|---|---|
| Submitted by | Natalie Schober (141354) |
| Submitted on | 31 August 2021 |
| Number of pages | 15 |
| Number of characters | 29,025 |

## Abstract

This paper analyzes the suitability of time series analysis models to forecast stock prices. For the analysis, the stock prices of Microsoft are collected from 1987, around the time when the company became listed on the stock market, to 2021. It is discussed which of the ARIMA and ETS models give a higher goodness of fit for the underlying stock price series. The most suitable models are chosen and evaluated regarding their forecasting accuracy. It is concluded that the ARIMA(0,1,1) model with drift slightly outperforms the ETS(A,A,N) model for the underlying series. Over the last decades, Microsoft's stock prices faced severe breaks, making forecasting particularly complex. To better account for such complexity, further research is suggested in combining time series analysis with Machine Learning and Deep Learning approaches.



*Keywords: Time Series Analysis, Predictive Analytics, Stock Price, ARIMA, ETS, Microsoft*


## Contents

# 1 Introduction

Due to financial crises, severe bank losses and defaults in the past, the demand for better forecasting in financial markets is rapidly increasing. Immense amounts of money had been lost due to poor forecasting models. In the last decades, many researchers have tried to apply time series analysis to build stock price forecast models. While [Devi et al., 2013] suggests ARIMA models as suitable, [Zhang et al., 2009] proposes ARCH models, as they consider different variance which better reflects the behavior of stock prices. [Sun, 2020] uses the S&P500 stock index to compare ETS and ARIMA models and finds the latter the most suitable. [Ahn and Cho, 2010] proposes a hybrid model that combines the power of ARMA models with news text mining to improve conventional forecasting techniques. Also [Panigrahi and Behera, 2017] uses a hybrid approach combining linear and nonlinear ETS models with Artificial Neural Networks. He explains that such a methodology is better able to capture different combinations of linear and/or nonlinear patterns in time series.

This paper seeks to follow up on this and to investigate

- to what extent time series analysis models are suitable for stock price forecasting.
- which model among them is best suited for stock price forecasting.

To answer these questions, the paper starts with (visually) exploring the stock price series in section 2. The original series is then transformed into log returns to stabilize the variance. The returns are tested for stationarity to avoid biased forecasts. In section 3, two time series analysis models, namely ARIMA and ETS, are extensively discussed and compared. The most suitable ARIMA and ETS models are chosen and the stock returns are trained on them. In section 4, the fit of the trained models is then analyzed to see which model is most suitable to forecast the underlying stock series. Section 5 concludes the paper and identifies limitations as well as potential improvements.

For the purpose of this paper, Microsoft's stock prices are used as the underlying time series. Since Microsoft became listed on the stock market in 1985, its stock took a fascinating development. In the last decades, Microsoft has gone through some crises (see details in section 2) but above all, accomplished huge achievements. Microsoft has greatly shaped and impacted the world through its hardware and software products. While many industries are stumbling, the technology industry is growing at a rapid pace. Microsoft is not only a part of this industry but one of its leaders. Because of Microsoft's relevance for the economy and the global advancement, the stock prices of this company are chosen.

# 2 Preprocessing

## 2.1 Acquiring and exploring stock prices

The data is obtained from Yahoo!Finance through the R package quantmod. The dataset comprises closing stock prices of Microsoft (MSFT) with daily frequency. Since monthly frequency are to be considered for this analysis, the closing stock prices are converted to monthly averages. Monthly averages instead of end-of-month values are used, as the latter would be subject to stronger fluctuations and more prone to heteroscedasticity.
The monthly data is then transformed into a time series object, starting in August 1987, around the time when Microsoft became listed on the stock market, and ending in July 2021. Thereby, the time series comprises more than 30 years and 408 observations.

To get a first an overview of the series, the output of the `summary` command is displayed in figure 1. The stock prices range from around \$0.3 to \$282 while the median value is \$27, suggesting that the price distribution is very variant and likely to be left-skewed.

```
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
 0.3283   6.1331  27.0080  37.9578  39.3650 281.5024
```

Figure 1: Summary of the price distribution

To get a better understanding of the behavior of Microsoft's stock price series, it is plotted alongside its autocorrelation function (ACF) in figure 2. When observing the plot of the original data, it seems that no clear trend emerges. However, several structural breaks can be identified. The breaks seem to occur in 2000 and 2008 and can be justified by global historical events: In the late 1990s, there was a lot of hype around investment in US technology stock, leading to rapid growth in value. In March 2000, the speculative 'dot-com' bubble burst, explaining the structural break that is visible in figure 2. Especially technology companies, so-called dot-com companies, suffered great losses. Microsoft's stock price fell from \$119.94 in 1999 to \$40.25 in late 2000. Besides the dot-com bubble, Microsoft faced another break around 2008, as can be seen in figure 2. Before the global financial crisis, the stock value of Microsoft was about \$37.50 but dropped to up to \$14.87 during the crisis in 2008 and 2009. It took Microsoft more than 14 years to get back to the stock value it was last at in 2000. Since that point in 2014, Microsoft's stock price has risen steadily, reaching \$100 in 2018 and \$200 in 2020. Microsoft owes this success in part to its strong focus on cloud technologies and the expansion of its gaming business in recent decades.

The ADF confirms the observations from the time series plot, presenting strong and significant autocorrelations for all the selected lags. It shows a slowly decaying trend, and the lags start extremely close to 1, indicating that the data is non-stationary and has a unit root. Also, cyclical fluctuations around a trending pattern are assumed, as there are several slumps and upswings of different lengths present.

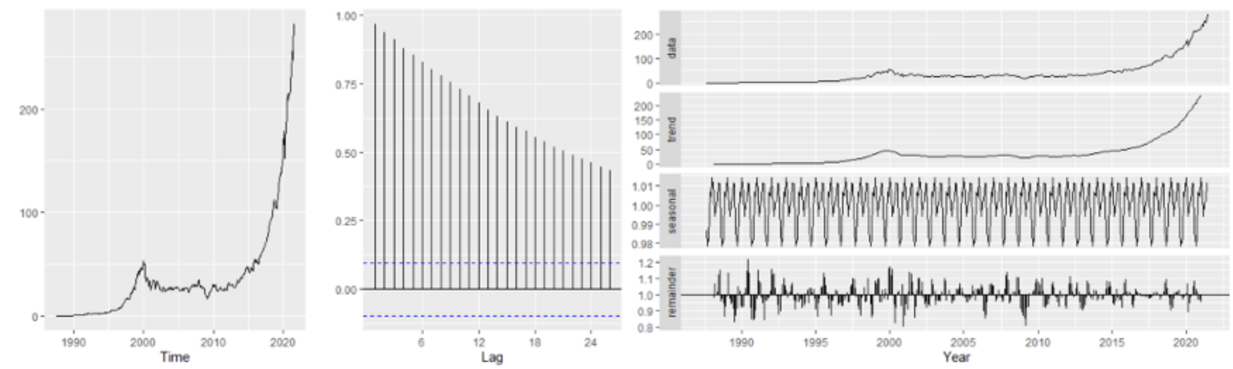



Figure 2: The original series, its autocorrelation function and decomposition

To further analyze the properties of the time series, it is decomposed into a trend, cycle, and seasonal component in figure 2. For the decomposition, the multiplicative instead of the additive type is chosen. By bare eye, it is difficult to see whether the seasonal variation increases or stays constant over time. In this case, the multiplicative decomposition shows more stable remainders than the additive one. Also, the data is in levels, suggesting that the multiplication of components (Trend * Seasonal * Random) is more suitable.

Based on the results of the multiplicative decomposition in figure 2, the data seems to show a strong upward trend starting in 2015. The fluctuation of the remainder is stronger at the beginning of the time series, especially before the year 2000. This might confirm the structural break that was discussed in the generic time series plot in figure 2. The seasonality seen in the decomposition does not become clear in the generic plot or the ACF so it is likely just a product of the decomposition function. The results obtained from the decomposition should only be interpreted as suggestive because decomposition has the caveat to assume that seasonality exists and is the same each year.

## 2.2 Computing log returns

In the next step, the closing price values P are transformed by taking the natural logarithm. The monthly log returns $y_1$, $y_2$, $y_3$, ..., $y_n$ are derived from the values $P$ to build the return distribution.

$$y_t = 100 \times ln\frac{P_t}{P_{t-1}} \tag{1}$$

$P_t$ defines the price at date t and $P_{t-1}$ defines the price at the previous trading month $t-1$. To visualize the computed returns, the autocorrelation function as well as the decomposition of the transformed data are plotted in figure 3. Compared to figure 2, the transformation to log returns has stabilized the variance and has led to a bit more linear trend. Thus, the transformed data will be used for the further analysis, as significant fluctuations were discovered in the original series and the log transformation improved that. The fluctuation of the remainder is quite strong at the beginning of the series, again indicating a break.

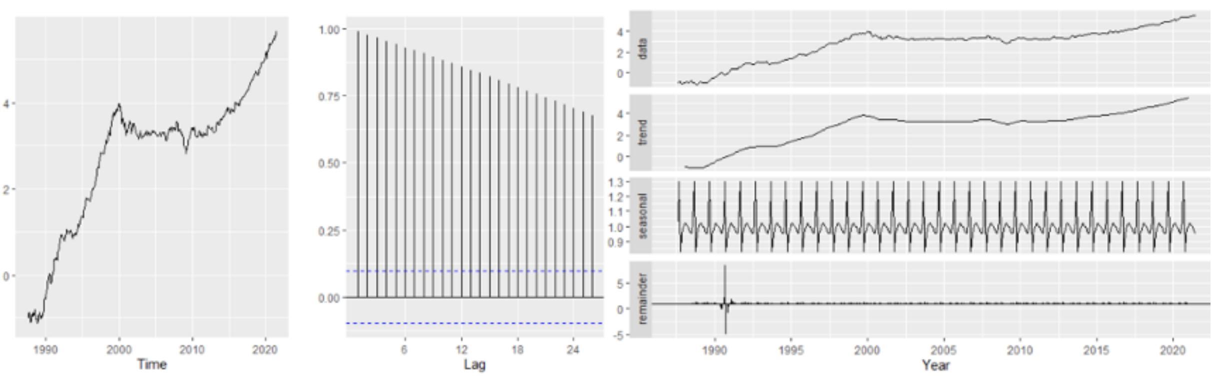



Figure 3: The log transformed series, its autocorrelation function and decomposition

## 2.3 Testing for stationarity

Before time series forecasting models (such as ARIMA) are applied to the stock series, stationarity must be ensured to avoid biased forecasts. To check for stationarity, two hypothesis tests, namely the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) and Augmented Dickey Fuller (ADF) test, are used. The following discusses which functional form of the KPSS and ADF test is most suitable for the underlying time series. Thereafter, both hypothesis tests are run and interpreted regarding (non-)stationarity.

Based on the results obtained in figure 3, an upward trend seems to be present. Hence, a random walk with drift and trend might be suitable for this stock series (type “tau” in KPSS and “trend” in ADF). For a more conservative ADF, the Akaike Information Criterion (AIC) is used to select the lags, returning the following results for the KPSS and ADF test.

```
Value of test-statistic is: 1.1398

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.119 0.146  0.176 0.216
```

```
Value of test-statistic is: -1.6332 6.4728 1.7421

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.98 -3.42 -3.13
phi2  6.15  4.71  4.05
phi3  8.34  6.30  5.36
```

Figure 4: KPSS test (type tau) and ADF test (type trend)

In the KPSS test in figure 4, the t-stat is above the significance level, which means the null of stationarity is rejected. The t-stat of the ADF-test for tau3 is below the significance level at 10%, therefore accepting the null hypothesis of a unit root. The t-stat for phi2 is slightly larger than the significance level at 1%, which rejects the null that there is a unit root without a trend and without a drift. This suggests that at least either drift or trend are different from zero. The t-stat of phi3 is below the significance level at 10%, therefore accepting the null that there is a unit root without a trend. Thus, the results of the ADF suggest that the time series is non-stationary, with a drift and without a trend.

It is tested again for a random walk with drift (type "mu" in KPSS and "drift" in ADF). "Mu" signals that the deterministic component seems to represent solely a constant (without trend).

```
Value of test-statistic is: 5.2819

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

```
Value of test-statistic is: -1.562 9.1861

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.44 -2.87 -2.57
phi1  6.47  4.61  3.79
```

Figure 5: KPSS test (type mu) and ADF test (type drift)

The KPSS test again rejects the null of stationarity. For the ADF-test, the t-stat for tau2 is below the significance level, therefore accepting the null of non-stationarity. The t-stat for phi1 rejects the null of non-stationarity and absence of drift. Thus, both tests confirm that the time series can be described by a random walk with drift. A random walk with drift $\delta$ can be denoted as follows.

$$y_t = \delta + y_{t-1} + \epsilon_t \tag{2}$$

## 2.4 Achieving stationarity

Both the plots and the hypothesis tests demonstrate that the working data is non-stationary. To make it stationary, the time series is differenced, i.e. the difference between each consecutive observation is computed. For the random walk with drift, the process of differencing looks as follows:

$$\begin{aligned} y_t &= \delta + y_{t-1} + \epsilon_t \\ y_t - y_{t-1} &= \delta + y_{t-1} - y_{t-1} + \epsilon_t \\ \Delta y_t &= \delta + \epsilon_t \end{aligned} \tag{3}$$

where $\Delta y_t$ denotes the difference between $y_t$ and its previous value $y_{t-1}$. The result looks like a white noise process.
The series is tested again by the KPSS and ADF. The t-stat of the KPSS test is below the significance level at 5%, therefore accepting the null of stationarity at the 5% level. However, it is rejected at the 10% level. Thus, differencing a second time could be considered.
For the ADF-test, the t-stat for tau2 is above the significance level, therefore, rejecting the null of non-stationarity. The t-stat for phi1 rejects the null of non-stationarity and absence of drift at the 1% level. For an accurate investigation, the ADF-test is again performed on a simple random walk (type “none”). The t-stat for tau1 again rejects the null of non-stationarity at 1%.

```
Value of test-statistic is: 0.4434

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

```
Value of test-statistic is: -13.7144 94.0425

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.44 -2.87 -2.57
phi1  6.47  4.61  3.79
```

```
Value of test-statistic is: -12.8454

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Figure 6: KPSS test (type mu), ADF test (type drift), ADF test (type none)

# 3 Modelling

To forecast the stock time series, ARIMA and ETS models are used. This section aims to identify and discuss the most suitable ARIMA and ETS processes for the underlying series.

## 3.1 Autoregressive Integrated Moving Average (ARIMA)

Autoregressive moving average models are a combination of past values of the variable. The order of each ARIMA model consists of three parameters: p – the order of the autoregressive part (AR), q - the order of the moving average (MA), and d- the order of differencing [Hyndman and Athanasopoulos, 2018]. An ARIMA model (p,d,q) can be denoted as follows

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - ... - \theta_p \epsilon_{t-p} \tag{4}$$

where $y_t$ represents a stationary stochastic process, $\phi_0$ the constant coefficient, $\epsilon_t$ the error term, $\phi_1$ and $\phi_2$ autoregressive coefficients and $\theta_1$ and $\theta_2$ represent the moving average coefficients [Sun, 2020].

After it has been confirmed that the data is stationary, the parameters of the ARIMA(p,d,q) model can be estimated observing the ACF and PACF plots. When looking at the plots, it is important to consider the significant spikes not only in the ACF but also in the PACF. This is because the significant spikes in the ACF $y_t$ and $y_{t-1}$ can be correlated because they are both correlated to $y_{t-2}$ and not because $y_{t-2}$ is relevant in predicting $y_t$ [Hyndman and Athanasopoulos, 2018].

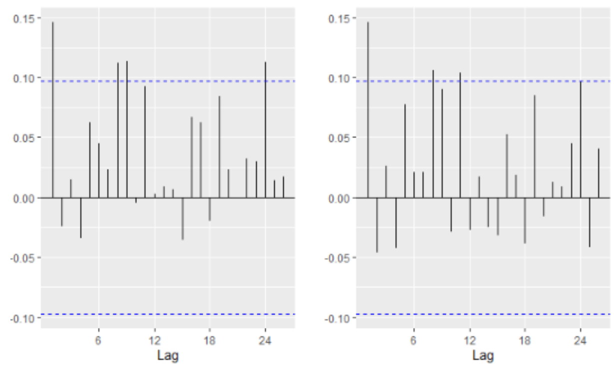



Figure 7: ADF and PACF of the differenced time series

In figure 7, both the ACF and the PACF show a significant spike at lag 1, with the subsequent spikes being not significant. The significance of lag 1 in the PACF indicates that the number of AR terms might be 1. The lag at which the ACF cuts off might indicate the number of MA terms being 1.
In addition to lag 1, also lag 8, 9, 11 and 24 are significant (especially in the ACF) in figure 7. All significant lags are positive. If lag 12, 24, etc. were significant, that could indicate a seasonal pattern. Perhaps a weakened and blurred form of seasonality is presented here. To account for seasonality, a seasonal MA(1) component could be included. Summing up, an ARIMA(1,1,1)(0,0,1) is suggested.

Of course, other models could also be suitable, so another one is proposed here: As [Hyndman and Athanasopoulos, 2018] describe, ACF and PACF only help in determining the order of the process if either p=0 or q=0. In general, for ARIMA (p,d,0), the ACF shows sinusoidal/exponentially decaying behaviour and the PACF truncates after lag p. An ARIMA(0,d,q) shows the opposite behaviour: The ACF truncates after lag q and the PACF shows sinusoidal/exponentially decaying behaviour. In Figure 13, the lags of both ACF and PACF do not exponentially decay, but truncate after lag 1. Especially the PACF shows a bit of scalloped positive patterns, followed by a little scalloped negative patterns. Thus, an ARIMA(0,d,q) model might be appropriate. The results of the ARIMA(0,1,1)(0,0,1) are displayed in figure 8 and 9.

```
ARIMA(1,1,1)(0,0,1)[12]

Coefficients:
         ar1      ma1     sma1
      0.1201  -0.9742  -0.0327
s.e.  0.0512   0.0133   0.0482

sigma^2 estimated as 0.004583:  log likelihood=517.25
AIC=-1026.51   AICc=-1026.41   BIC=-1010.48

Training set error measures:
                       ME       RMSE        MAE      MPE     MAPE      MASE
ACF1
Training set 0.0003655639 0.06736406 0.05034001 79.71354 133.5409 0.7183074
0.01209415

        Ljung-Box test

data:  Residuals from ARIMA(1,1,1)(0,0,1)[12]
Q* = 24.167, df = 21, p-value = 0.285

Model df: 3.   Total lags used: 24
```

```
ARIMA(0,1,1)(0,0,1)[12]

Coefficients:
          ma1     sma1
      -0.9652  -0.0322
s.e.   0.0162   0.0484

sigma^2 estimated as 0.004634:  log likelihood=514.52
AIC=-1023.05   AICc=-1022.99   BIC=-1011.03

Training set error measures:
                       ME       RMSE        MAE      MPE    MAPE      MASE      ACF1
Training set 0.0006721967 0.06782393 0.05086293 76.78228 135.763 0.7257691 0.1149054

        Ljung-Box test

data:  Residuals from ARIMA(0,1,1)(0,0,1)[12]
Q* = 26.846, df = 22, p-value = 0.2171

Model df: 2.   Total lags used: 24
```

Figure 8: ARIMA(1,1,1)(0,0,1) and ARIMA(0,1,1)(0,0,1)

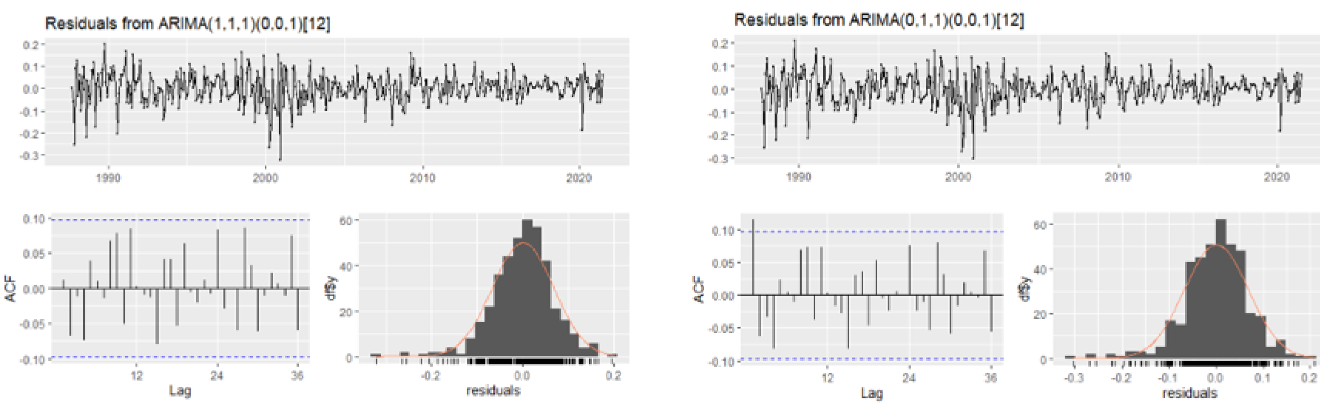



Figure 9: Residuals from ARIMA(1,1,1)(0,0,1) and ARIMA(0,1,1)(0,0,1)

In figure 8, the Ljung-Box test suggests that the residuals of ARIMA (1,1,1)(0,0,1) and ARIMA(0,1,1)(0,0,1) are independent with p-values of 0.285 and 0.2171, respectively. The AICs of -1026.71 and -1023.05 suspect that the ARIMA (1,1,1)(0,0,1) fits slightly better. In figure 9, the ARIMA model, where an AR component is included, shows no significant spikes. Also, the residuals from the ARIMA(1,1,1)(0,0,1) fluctuate a bit more around zero mean and share more similarity with a normal distribution shape than those from the ARIMA(0,1,1)(0,0,1). All these observations suggest that the ARIMA (1,1,1)(0,0,1) is slightly preferred and an AR component should be included in the model. This, however, would not make a substantial difference, as the only minimally different AICs of the two models show.

Besides manually searching for the optimum ARIMA order, it can also be done automatically using the function `auto.arima(x)`. To get more accurate results, the parameter `stepwise` and `approximation` are both set to `FALSE`. This ensures that many different combinations of ARIMA models are considered, instead of only approximating an ARIMA model. The function `auto.arima(x, seasonal = FALSE)` declares an ARIMA(0,1,1) with drift as optimal for the underlying time series. When setting `seasonal = TRUE`, an ARIMA(0,1,1)(0,0,2) with drift is selected.

Both selected ARIMA models have a MA(1) component and a drift included. The only difference between the models is that one includes seasonality, while the other does not. According to the results of the Ljung Box test in figure 10, no significant autocorrelation has been detected, with both p-values 0.1108 and 0.1635 being just above the 10% significance level. The AICs of the two models are with -1034.18 and -1036.2 very similar. The ARIMA (0,1,1)(0,0,2) with drift only has a slightly smaller AIC and therefore shows a better fit. According to [Hyndman and Athanasopoulos, 2018], a difference in AIC values of 2 or less is not regarded as substantial. Thus, the simpler but non-optimal model can be chosen following the principle of parsimony. In this case, the **ARIMA (0,1,1) model with drift** is chosen for the forecasting in the next section. This simple exponential smoothing model can be denoted as follows

$$y_t = y_{t-1} - \theta_1 \epsilon_{t-1} + \delta \tag{5}$$

where $\theta$ represents the moving average coefficient and $\epsilon_{t-1}$ is the error term at t-1. As seen in figure 10, the coefficients are estimated as

$$y_t = y_{t-1} - 0.1594 \epsilon_{t-1} + 0.0164 \tag{6}$$

```
ARIMA(0,1,1) with drift

Coefficients:
         ma1   drift
      0.1594  0.0164
s.e.  0.0510  0.0039

sigma^2 estimated as 0.004568:  log likelihood=520.09
AIC=-1034.18   AICc=-1034.12   BIC=-1022.15

Training set error measures:
                        ME       RMSE        MAE       MPE    MAPE      MASE
ACF1
Training set -3.437515e-05 0.06733604 0.05003777 -0.250163 5.60881 0.1819049
-0.005433134

        Ljung-Box test

data:  Residuals from ARIMA(0,1,1) with drift
Q* = 31.509, df = 22, p-value = 0.08617

Model df: 2.   Total lags used: 24
```

```
ARIMA(0,1,1)(0,0,2)[12] with drift

Coefficients:
         ma1    sma1    sma2   drift
      0.1624  0.0028  0.1294  0.0163
s.e.  0.0519  0.0516  0.0522  0.0043

sigma^2 estimated as 0.004519:  log likelihood=523.1
AIC=-1036.2   AICc=-1036.05   BIC=-1016.15

Training set error measures:
                       ME       RMSE        MAE        MPE     MAPE      MASE
ACF1
Training set 7.983939e-05 0.06680672 0.04965503 -0.4784408 5.457295 0.1805136
-0.00739865

        Ljung-Box test

data:  Residuals from ARIMA(0,1,1)(0,0,2)[12] with drift
Q* = 26.071, df = 20, p-value = 0.1635

Model df: 4.   Total lags used: 24
```

Figure 10: ARIMA(0,1,1) with drift and ARIMA(0,1,1)(0,0,2) with drift

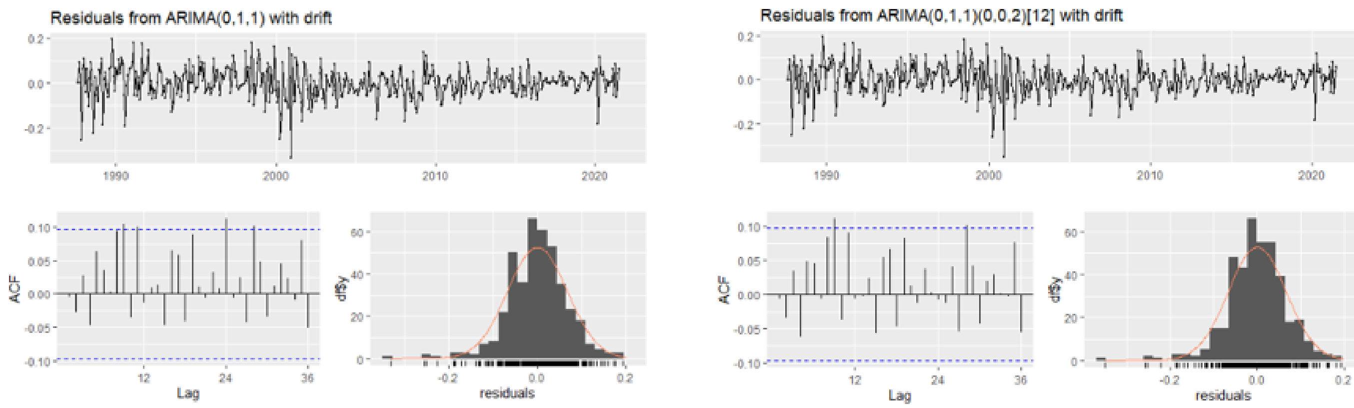



Figure 11: Residuals from ARIMA(0,1,1) with drift and ARIMA(0,1,1)(0,0,2) with drift

## 3.2 Exponential Smoothing (ETS)

After discussing which ARIMA model would best fit the underlying stock price series, another forecasting model, the ETS model, is considered. The three letters in ETS represent the following: The first letter denotes the error type, the second one the trend type and the third one the season type. The letters can take the value of "N" standing for none, "A" for additive or "M" for multiplicative.

In this case, as the time series has been log transformed, all three components (error, trend and seasonal) are suggested to be additive, making up an ETS(A,A,A) model. Some level of seasonality might be present, as shown in the decomposed plots and correlograms before. The suggested ETS(A,A,A) is also called additive Holt-Winters' method with additive errors. The `ets()` command is then used to estimate the state-space version of the ETS(A,A,A). Its residuals are discussed regarding distribution, variance, mean and autocorrelation.

```
ets(y = l.price, model = "AAA")

  Smoothing parameters:
    alpha = 0.8678
    beta  = 0.0228
    gamma = 0.0336

  Initial states:
    l = -1.2279
    b = 0.0259
    s = 0.0542 0.0055 -0.0122 0.0156 0.0065 0.0026
           -0.0092 -0.03 -0.002 -0.0247 -0.0017 -0.0047

  sigma:  0.0725

     AIC     AICc      BIC
329.1703 330.7395 397.3618

Training set error measures:
                        ME       RMSE        MAE        MPE     MAPE      MASE
ACF1
Training set -0.0001861392 0.07108313 0.05347893 -0.7983799 6.590885 0.1944148
0.2248067

        Ljung-Box test

data:  Residuals from ETS(A,A,A)
Q* = 40.945, df = 8, p-value = 2.135e-06
```

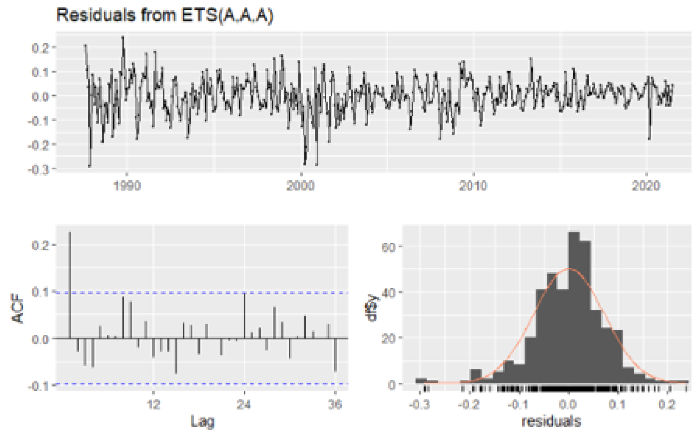



Figure 12: ETS(A,A,A)

From Figure 13, it becomes obvious that the residuals are correlated. The p-value of the Ljiung-Box is very low, therefore clearly rejecting the null of no joint autocorrelation. In addition, significant spikes can be observed in the correlogram. The residuals do not appear to be normally distributed and they do not fluctuate around zero mean. Collectively, there is considerable information remaining in the residuals which has not been captured with the ETS(A,A,A) model.

Contrary to the initial thought, perhaps there is no level of seasonality present, so that the ETS(A,A,N) model would be better suited. The `ets(x, model(ZZZ))` command, which automatically runs over different ETS models and selects the most suitable one, confirms this. It identifies the ETS(A,A,N) model as most suitable. The ETS(A,A,N) is also called Holt's linear method with additive errors. In state space form, it can be denoted as follows:

$$\begin{aligned} y_t &= l_{t-1} + b_{t-1} + \epsilon_t \\ l_t &= l_{t-1} + b_{t-1} + \alpha\epsilon_T \\ b_t &= b_{t-1} + \beta\epsilon_t \end{aligned} \tag{7}$$

Then, the forecast equations are

$$\begin{aligned} y_{T+1} &= l_T + b_T + \epsilon_{T+1} \Rightarrow \hat{y}_{T+1|T} = l_T + b_T \\ y_{T+2} &= l_{T+1} + b_{T+1} + \epsilon_{T+2} \Rightarrow \hat{y}_{T+2} = l_T + 2b_T \end{aligned} \tag{8}$$

where $l_T$ stands for level, $b_T$ for trend, $\alpha$ and $\beta$ for smoothing parameters and $\epsilon_t$ for the Gaussian white noise disturbance.

Compared to the ETS(A,A,A) model suggested before, the ETS(A,A,N) does not include a seasonal component. For the ETS(A,A,N) model, the Ljung-Box test returns a p-value that is slightly above the

10% significance level, therefore rejecting the null hypothesis that there is significant autocorrelation in the residuals. The correlogram confirms this by showing that no lags (except slightly the first lag) are significant. Also, the AIC of the ETS (A,A,N) is lower than it was for the ETS(A,A,A), Collectively, this means that the **ETS (A,A,N) model** seems the most suitable among the ETS models for this stock series.

```
ets(y = l.price, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.9996
    beta  = 0.0294

  Initial states:
    l = -1.0278
    b = 0.0239

  sigma:  0.0681

     AIC     AICc      BIC
266.2996 266.4489 286.3560

Training set error measures:
                       ME       RMSE        MAE        MPE     MAPE      MASE
ACF1
Training set 0.0001628932 0.06777642 0.05062775 -0.3246314 5.724781 0.1840497
0.1122102

        Ljung-Box test

data:  Residuals from ETS(A,A,N)
Q* = 27.978, df = 20, p-value = 0.1099

Model df: 4.   Total lags used: 24
```

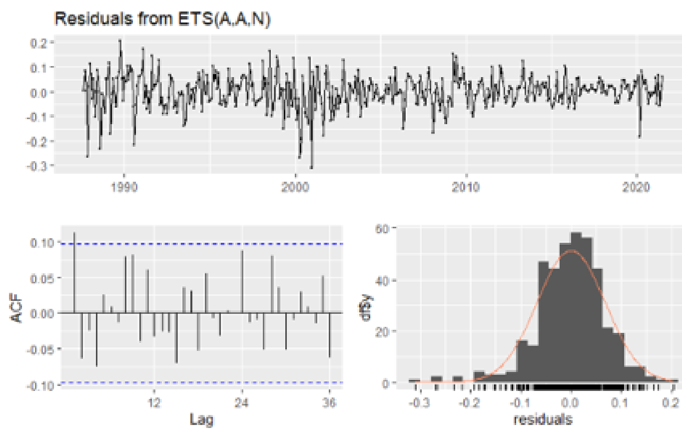



Figure 13: ETS(A,A,N)

### 3.3 Comparing ARIMA and ETS

In the preceding, both ARIMA and ETS models were applied and it may have seemed that ARIMA and ETS are very similar. Actually, until the late 1990s, many economists had only used ARIMA models, as ETS models were considered inferior. Since then, however, much research was conducted in this field, concluding that ARIMA and ETS share some features but also show differences and are both worth considering. For example, a disadvantage of ARIMA models can be the selection of the order, as it is sometimes quite complex and vague [Hyndman and Athanasopoulos, 2018].

All ETS models are non-stationary, while some ARIMA models are stationary. ARIMA models are used in the case of autocorrelation, i.e. when the past describes the present data well. In the case of a trend or seasonality in the data, ETS models are applied as they specifically consider these components.

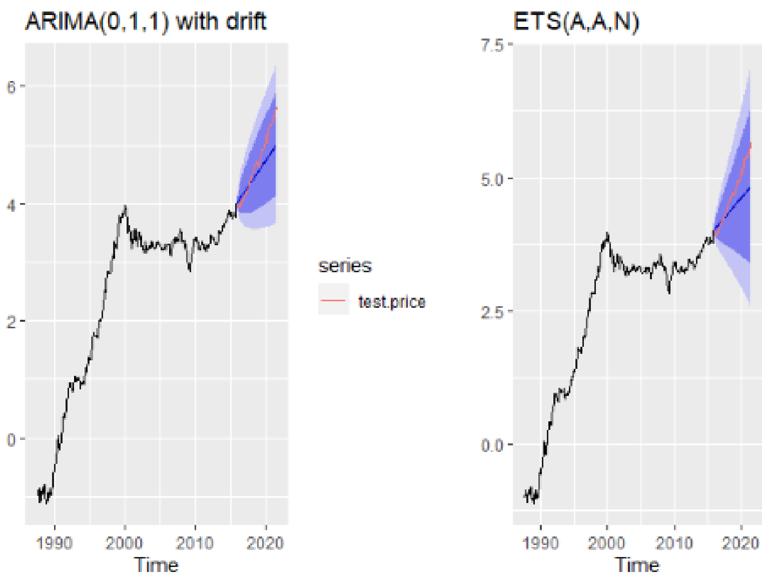
Still, ETS and ARIMA models are closely connected, as all linear ETS models have equivalent ARIMA counterparts. The previously as most suitable selected ARIMA(0,1,1) is equivalent to ETS (A,N,N), while the selected ETS(A,A,N) is equivalent to ARIMA(0,2,2). As [Svetunkov, 2021] describes, an important difference between ARIMA(0,1,1) and ETS(A,N,N) is that the boundary values for parameters are not possible for ARIMA(0,1,1), i.e. $\alpha = 0$ is possible in ETS, but the respective $\theta_1 = -1$ is not in ARIMA. A similar situation applies to the case of ARIMA(0,2,2) vs ETS(A,A,N).

Collectively, ETS models are worth considering as they have their own features and advantages.

# 4 Forecasting and evaluating performance

After the data has been splitted into training and test set, forecasts are performed using the models ARIMA (0,1,1) with drift and ETS (A,A,N), which were identified as most suitable in section 3. Based on the visual forecasts and the accuracy measures, it is discussed which model is best suited for predicting the stock price series.

The data is split 80:20 into training and test sets. The training set goes to the end of 2014 and the test set starts at the beginning of 2015. In the figure on the right, both ARIMA and ETS model predict that stock prices will increase approximately constantly from 2015. However, they actually increase sharply due to the hype around (cloud) technologies, as described in section 2. Thus, both ARIMA and ETS fail to predict this strong upward trend in stock prices.




Critically, it can be noted here that this train-test split made forecasting difficult. The train-test split was made at a point where a break occurred: While stock prices remained roughly constant for several years, they sharply increased from this point on. Complex factors led to this break, which the time series analysis processes, that are used here, are only able to model partially.

In terms of performance measures, the ARIMA model shows the best fit to the test set. From figure 14, it becomes clear that the ME, RMSE, MAE, MPE, MAPE, MASE and ACF1 of ARIMA have smaller error values than those obtained from the ETS model. For example, the RMSE (root mean squared error) and MAE (mean absolute error) of the ARIMA are around 0.29 and 0.23, respectively, while they are 0.38 and 0.30 in case of the ETS. That means, the ARIMA is better suitable for forecasting the underlying stock series than the ETS.

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 | Theil's U |
|---|---|---|---|---|---|---|---|---|
| Training set | -0.0000385368 | 0.07115262 | 0.05385186 | -0.2971407 | 6.591899 | 0.1977310 | -0.006277809 | NA |
| Test set | 0.1794098239 | 0.28748512 | 0.22663807 | 3.3527485 | 4.521685 | 0.8321601 | 0.943350174 | 5.676243 |

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 | Theil's U |
|---|---|---|---|---|---|---|---|---|
| Training set | -0.00135399 | 0.07179688 | 0.05454838 | -0.3878265 | 6.712826 | 0.2002884 | 0.1284925 | NA |
| Test set | 0.26419365 | 0.37896696 | 0.29774452 | 5.0676449 | 5.905840 | 1.0932458 | 0.9478188 | 7.484921 |

Figure 14: Performance measures of ARIMA(0,1,1) with drift (above) and ETS(A,A,N) (below)

# 5 Conclusion

This paper analyzed the suitability of the time series analysis models ARIMA and ETS to forecast stock prices. After having intensively discussed the choice of time series analysis models, it is concluded that the ARIMA(0,1,1) with drift shows the best fit for the Microsoft stock price series from 1987 to 2021. The ARIMA(0,1,1) with drift outperformed the ETS(A,A,N) model for the underlying series.
The paper has shown that time series analysis provides a solid foundation to forecast in which direction the stock prices develop. However, especially under extreme market conditions, as seen in section 4, the time series analysis processes could only partially model and forecast the series accurately.

One limitation is with regards to the underlying time series. The series contains stock prices from 1987 to 2021, i.e. it includes several major historical events that heavily (and suddenly) affected the stock prices. During these events, stocks were subject to high volatility and economic risk, which makes predicting stock prices particularly difficult. Also, the underlying time series only considers closing stock prices for forecasting. Additional features providing information external to the stocks themselves could potentially account for more of the variation.
Another limitation is with regards to the models. Both the ARIMA and ETS models assume that the log returns have zero mean and the residuals are normally distributed. In reality, however, stock series often have different variances. The time series analysis models only take into account the characteristics of the time series, without considering that the stock price movement itself is influenced by many unpredictable and complex patterns. To account for such complexity, hybrid approaches could be useful. ARIMA and ETS models could be combined, for example, with Machine Learning and Deep Learning models. Text mining of news articles and company announcements can also help in the prediction of stock prices.

Collectively, although there are certain limitations of the models and high precision is required for true stock prediction, it is believed that the analysis provides a foundation for future optimization, especially with regards to hybrid approaches.

# References


[Ahn and Cho, 2010] Ahn, S.-W. and Cho, S.-B. (2010). Stock prediction using news text mining and time series analysis. In *Proceedings of the Korean Information Science Society Conference*, pages 364–369. Korean Institute of Information Scientists and Engineers.

[Devi et al., 2013] Devi, B. U., Sundar, D., and Alli, P. (2013). An effective time series analysis for stock trend prediction using arima model for nifty midcap-50. *International Journal of Data Mining & Knowledge Management Process*, 3(1):65.

[Gujarati, 2021] Gujarati, D. N. (2021). *Essentials of econometrics*. SAGE Publications.

[Hendry et al., 2014] Hendry, D., Doornik, J. A., et al. (2014). Statistical model selection with'big data'. Technical report.

[Hyndman and Athanasopoulos, 2018] Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

[Panigrahi and Behera, 2017] Panigrahi, S. and Behera, H. S. (2017). A hybrid ets–ann model for time series forecasting. *Engineering applications of artificial intelligence*, 66:49–59.

[Sun, 2020] Sun, Z. (2020). Comparison of trend forecast using arima and ets models for s&p500 close price. In *2020 The 4th International Conference on E-Business and Internet*, pages 57–60.

[Svetunkov, 2021] Svetunkov, I. (2021). Forecasting and analytics with adam. OpenForecast. (version: [current date]).

[Zhang et al., 2009] Zhang, J., Shan, R., and Su, W. (2009). Applying time series analysis builds stock price forecast model. *Modern Applied Science*, 3(5):152–157.