Procter & Gamble Company Stock Prediction

TIME SERIES ANALYTICS

Submitted By:

Srilekha Matam - (ik1394)

Chandana kandula - (dl5702)

Kranthi Kumar Irala - (mo5225)

Adeel Nasir - (gd8408)

SUMMARY

Our time series analysis focused on forecasting the future performance of Procter & Gamble's stock using various statistical models and techniques. Leveraging historical stock data obtained from Yahoo! Finance, we applied multiple approaches to predict P&G's stock prices for the upcoming period.

The analysis began with the exploration of different time series models, including linear trend and seasonal models, Holt-Winter's exponential smoothing, ARIMA models, and auto ARIMA. Each model was evaluated based on its ability to capture the underlying patterns and trends in the stock prices and make accurate predictions.

The linear trend and seasonal model provided insights into the overall trend and seasonal fluctuations in P&G's stock prices. However, the accuracy of this model was limited, indicating the need for additional techniques to improve forecasting performance.

Holt-Winter's ZZZ model offered a more sophisticated approach by capturing both trend and seasonality while considering the data's inherent volatility. This model yielded promising results, with relatively low error measures and good predictive accuracy.

The ARIMA(2,1,2) model, incorporating autoregressive and moving average components, provided further insights into the underlying dynamics of P&G's stock prices. Despite some challenges in capturing the data's volatility, this model demonstrated reasonable forecasting performance.

Auto ARIMA, an automated approach to identifying the best-fitting ARIMA model, also produced satisfactory results, with comparable accuracy to manually selected models.

Overall, our analysis suggests that combining multiple forecasting models and techniques can enhance the robustness of stock price predictions. However, it's essential to consider the limitations and uncertainties inherent in financial forecasting and continuously refine our approaches to adapt to changing market conditions.

Through this analysis, we provide valuable insights into Procter & Gamble's stock performance, enabling investors and stakeholders to make informed decisions about their investments and market strategies.

This summary encapsulates our exploration of Procter & Gamble's stock prices and forecasting results, offering a comprehensive overview of our time series analysis and its implications for the financial markets.

INTRODUCTION

Procter & Gamble Company, a global leader in consumer goods, operates in five segments: Beauty, Grooming, Health Care, Fabric & Home Care, and Baby, Feminine & Family Care. Brands like Pantene, Gillette, Crest, Tide, and Pampers offer a wide range of products, from hair care to household essentials. Founded in 1837 and headquartered in Cincinnati, Ohio, P&G continues to provide quality products through various channels worldwide.

This project aims to forecast future sales for the next year utilizing comprehensive data on Procter & Gamble's performance in the consumer goods market. Drawing from decades of historical data, the project seeks to provide valuable insights into P&G's sales trends, market dynamics, and consumer behavior. The findings of this project can be instrumental for policymakers, economists, and investors in making informed decisions about P&G's market position, industry trends, and the broader economic landscape.

Data Source:

The data for this project will be sourced from Procter and Gamble's stock information available on Yahoo! Finance. The dataset includes various parameters such as Date, Close price, Adjusted Close price, Open, High, and Low for the year 2023. However, for this analysis, we will only focus on the Date and Adjusted Close price columns. The Adjusted Close price reflects adjustments made for stock splits, dividend distributions, or capital gains.

https://finance.yahoo.com/quote/PG/history

MAIN CHAPTER

Before converting the data into timeseries we used R libraries such as **dplyr** and **lubridate** for data manipulation and date processing. It loads stock data from a CSV file into a dataframe, converting the date column to a Date type for subsequent analysis. Then, it extracts year and month information from the date column, facilitating the calculation of monthly average adjusted closing prices. Finally, aggregated the data to present the mean adjusted closing prices both by year-month combinations and by concatenated year-month pairs, offering insights into stock price trends over time.

Time Series Data of Stocks

```r
#Create Time Series data, plot ts components, and autocorrelation.
stocks.ts <- ts(stocks_data$AdjClose,
                start = c(2000, 1), end = c(2023, 9), freq = 12)
stocks.ts
stocks.stl <- stl(stocks.ts, s.window = "periodic")
```

In the provided code snippet, a time series object `stocks.ts` is created from the adjusted closing prices (`AdjClose`) of stocks data. This time series spans from January 2000 to September 2023 with a monthly frequency.

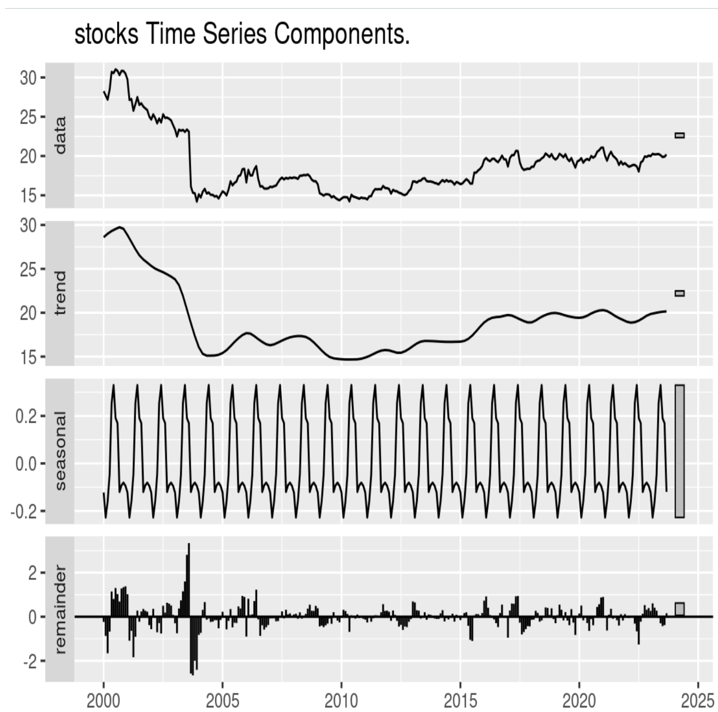
Then, the seasonal-trend decomposition procedure based on Loess (STL) is applied to `stocks.ts` using the `stl()` function with a periodic window. This procedure decomposes the time series into three components: seasonal, trend, and remainder (residuals).

The plot is showing the components of the time series.

Trend Component: This represents the long-term movement or direction of the time series.

Seasonality Component: Seasonality refers to patterns that repeat at fixed intervals. This component captures recurring patterns or fluctuations within the data that occur with a consistent periodicity, such as daily, weekly, or monthly cycles.

Remainder Component (Residuals): The remainder, also known as residuals, represents the random or irregular fluctuations in the data that are not accounted for by the trend or seasonality. **components**.

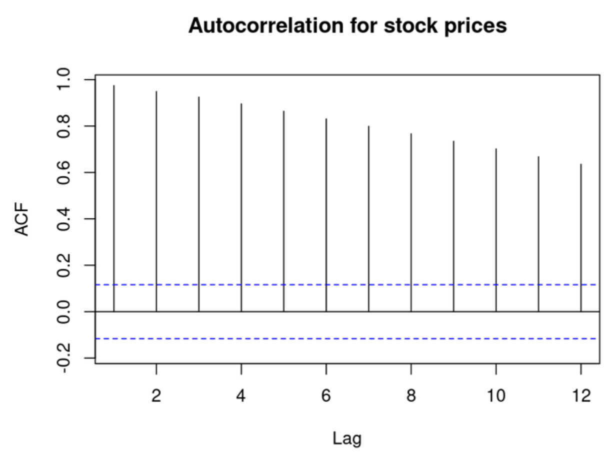


Arima Model

Predictability of the data is assessed by fitting an autoregressive model of order 1 (AR(1)) to the time series using the Arima() function.

```
> summary(stocks.ar1)
Series: stocks.ts
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1     mean
      0.9899  21.0481
s.e.  0.0080   3.1225

sigma^2 = 0.4281:  log likelihood = -284.44
AIC=574.88   AICc=574.97   BIC=585.84

Training set error measures:
                      ME      RMSE       MAE        MPE     MAPE      MASE       ACF1
Training set -0.04687621 0.6519721 0.3757228 -0.3194528 1.991744 0.2395706 0.02571765
```

The fitted AR(1) model for the stock prices time series indicates a strong positive correlation with lagged values, with an autoregressive coefficient of approximately 0.9899. The non-zero mean component suggests that the series has a mean value different from zero, estimated to be around 21.0481. Evaluation metrics such as AIC, RMSE, and ACF1 indicate good model fit and predictive performance on the training dataset.



In the autocorrelation plot, all lag bars exceeding the threshold limit suggest significant autocorrelation between the time series and its lagged values. This indicates that the current observation is highly influenced by its past values, confirming the presence of a strong temporal pattern within the data.

z-test

Z-test is applied to test the null hypothesis that the beta coefficient of the AR(1) model is equal to 1.

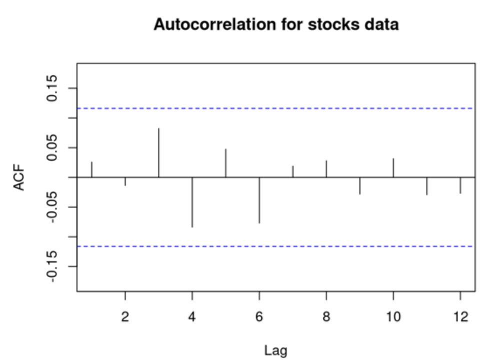
- The estimated AR(1) coefficient (ar1) is 0.9899, and the standard error is 0.0080.
- The null hypothesis is set to 1.
- A significance level (alpha) of 0.05 is chosen.
- The z-statistic is calculated as (ar1 - null_mean) / s.e., resulting in a z-statistic value.

```
> if (p.value<alpha) {
+   "Reject null hypothesis"
+ } else {
+   "Accept null hypothesis"
+ }
[1] "Accept null hypothesis"
```

Certainly, if the p-value is not less than the chosen significance level (alpha), which is 0.05 in this case, we "Accept null hypothesis." This decision indicates that there isn't sufficient statistical evidence to reject the null hypothesis. It suggests that the estimated AR(1) coefficient is not significantly different from 1 at the 5% significance level. Therefore, we conclude that the autocorrelation in the time series data does not deviate significantly from a first-order autoregressive process with a coefficient of 1.

Autocorrelation function (ACF) is computed for the residuals.

This ACF plot helps assess the adequacy of the model by examining the autocorrelation structure of the residuals.



The table displays the point forecast along with the lower and upper bounds of the prediction interval for each forecasted period.

For instance, the forecast for October 2023 is 20.19405, with both lower and upper bounds also set to 20.19405. This indicates a point forecast without prediction intervals.

```
> stocks.ar1.pred <- forecast(stocks.ar1, h = r
> stocks.ar1.pred
         Point Forecast     Lo 0     Hi 0
Oct 2023       20.19405 20.19405 20.19405
Nov 2023       20.20269 20.20269 20.20269
Dec 2023       20.21124 20.21124 20.21124
Jan 2024       20.21970 20.21970 20.21970
Feb 2024       20.22808 20.22808 20.22808
Mar 2024       20.23637 20.23637 20.23637
Apr 2024       20.24457 20.24457 20.24457
May 2024       20.25270 20.25270 20.25270
Jun 2024       20.26074 20.26074 20.26074
Jul 2024       20.26870 20.26870 20.26870
Aug 2024       20.27658 20.27658 20.27658
Sep 2024       20.28438 20.28438 20.28438
Oct 2024       20.29211 20.29211 20.29211
Nov 2024       20.29975 20.29975 20.29975
Dec 2024       20.30732 20.30732 20.30732
```

Partitioning of Data

Partitioning the data into training and validation sets allows for the development, evaluation, and refinement of time series forecasting models, leading to more accurate and reliable predictions.

```
length(stocks.ts)


nValid <- 75
nTrain <- length(stocks.ts) - nValid
train.ts <- window(stocks.ts, start = c(2000, 1), end = c(2000, nTrain))
valid.ts <- window(stocks.ts, start = c(2000, nTrain + 1),
                   end = c(2000, nTrain + nValid))
```

The length of the time series data (**stocks.ts**) is calculated to determine the total number of observations.

nValid is set to 75, indicating the number of observations to be included in the validation set.

nTrain is calculated as the total length of the time series minus the number of observations in the validation set, representing the number of observations in the training set

```
nTrain <- total_length - nValid

nTrain <- length(stocks.ts) - nValid
nTrain
```

Validation = 75

Training = 210

Partitioning the data into training and validation sets is a crucial step in time series forecasting for several reasons:

- Model Training
- Model Evaluation
- Prevention of Data Leakage
- Tuning Model Parameters
- Assessing Generalization

Applying Forecasting Methods

Regression model with linear trend and seasonality for training partition

This regression model aims to capture both the linear trend and seasonal patterns present in the training partition of the time series data, allowing for forecasting based on these components.

```
Call:
tslm(formula = train.ts ~ trend + season)

Residuals:
   Min     1Q Median     3Q    Max
-6.413 -2.992 -1.131  2.920  8.630

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 22.840070   0.981976  23.259   <2e-16 ***
trend       -0.040501   0.004247  -9.535   <2e-16 ***
season2     -0.119839   1.243293  -0.096    0.923
season3     -0.029039   1.243315  -0.023    0.981
season4      0.057275   1.243351   0.046    0.963
season5      0.437022   1.243402   0.351    0.726
season6      0.566782   1.243467   0.456    0.649
season7      0.183736   1.261437   0.146    0.884
season8      0.154169   1.261444   0.122    0.903
season9     -0.232927   1.261465  -0.185    0.854
season10    -0.228900   1.261501  -0.181    0.856
season11    -0.189683   1.261551  -0.150    0.881
season12    -0.277915   1.261616  -0.220    0.826
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.73 on 197 degrees of freedom
Multiple R-squared:  0.3182,    Adjusted R-squared:  0.2766
F-statistic: 7.661 on 12 and 197 DF,  p-value: 1.352e-11
```

R= 31.82%

The regression model exhibits a moderate fit, as evidenced by significant intercept and trend coefficients, suggesting a relationship between these predictors and the response variable. However, the non-significant seasonal coefficients and the relatively low R-squared value (31.82%) indicate potential inadequacies in capturing the variation in the data.

Trailing Moving Average (MA)

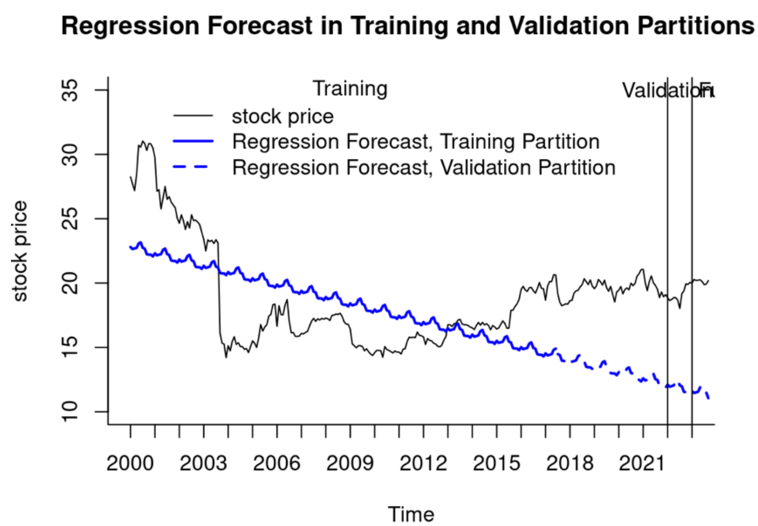
The Trailing moving average (MA) with a window width of 4 is applied to the residuals of the regression model for the training partition. The rollmean() function from the zoo package is used to compute the rolling mean.

The resulting trailing MA values are then used to forecast the residuals for the validation partition.

```
> trend.seas.res.valid
          Jan      Feb      Mar      Apr      May      Jun      Jul      Aug      Sep      Oct      Nov
2017                                                       4.695873 4.111712 4.254132 4.374483 4.442872
2018 4.585816 4.762923 5.282984 5.404925 5.116000 5.278361 5.886441 6.409432 6.618958 6.420581 6.790913
2019 5.994461 6.372881 6.825833 6.595480 6.004618 6.301183 6.137610 5.922500 6.735922 6.151723 5.565877
2020 6.363050 6.808560 6.187917 6.477608 6.272561 6.032329 6.975893 6.844662 7.456786 8.046845 8.249424
2021 8.476244 7.596530 6.925562 7.584298 7.681200 7.071915 7.277388 7.062287 6.835657 7.308277 6.931971
2022 6.755847 6.697050 6.731041 6.820076 6.480830 6.206149 5.982397 7.198733 7.825919 8.431498 8.381529
2023 8.369947 8.818880 8.690360 8.668825 8.359249 8.121652 8.313237 8.429159 9.121072
          Dec
2017 4.554830
2018 6.433167
2019 6.466262
2020 8.696884
2021 7.229268
2022 8.669385
2023
```

Additionally, a regression forecast (is generated for the validation period using the previously fitted regression model. The residuals for the validation period are calculated by subtracting the mean forecasted values from the actual validation set.

In this plot stock price is increasing while the regression forecasts for both the training and validation partitions are decreasing, it suggests that the regression model may not adequately capture the upward trend present in the stock price data.



Two-Level Forecast For Validation Period

In the provided code, a two-level forecast for the validation period is developed by combining the regression forecast with the trailing MA forecast for residuals. This combination is achieved by adding the forecasted values from both components together.

```
> round(accuracy(fst.2level, valid.ts), 3)
            ME  RMSE   MAE   MPE  MAPE  ACF1 Theil's U
Test set 0.546 1.305 1.037 2.645 5.262 0.893     3.267
```

The two-level forecast demonstrates reasonable performance, with measures such as RMSE (1.305) and MAE (1.037) indicating moderate accuracy. However, a positive ME (0.546) suggests a slight tendency to overestimate.

Exponential Smoothing

Simple exponential smoothing is applied to the training data. The model parameter is set to "ANN", which specifies that the model comprises additive error (A), no trend (N), and no seasonality (N). The smoothing parameter (alpha) is set to 0.2, controlling the weight assigned to the most recent observation in the smoothing process.

```
ETS(A,N,N)

Call:
 ets(y = train.ts, model = "ANN", alpha = 0.2)

  Smoothing parameters:
    alpha = 0.2

  Initial states:
    l = 28.8103

  sigma:  1.2664

     AIC     AICc      BIC
1224.073 1224.131 1230.767
```

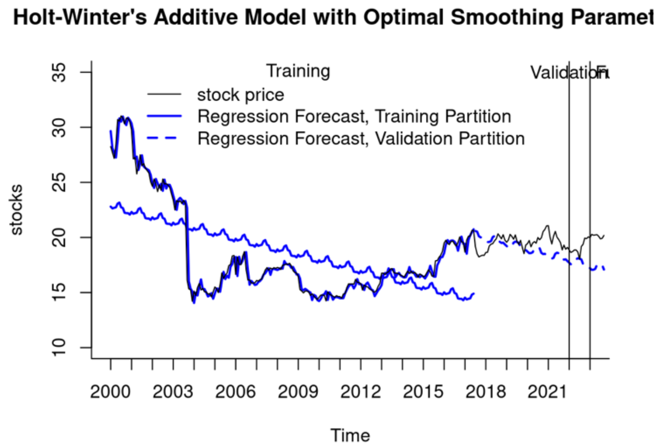
The ETS(A,N,N) model with a smoothing parameter (alpha) of 0.2 is applied to the training data. The model estimates an initial level (l) of approximately 28.8103 and a standard deviation (sigma) of the error term around 1.2664. Model selection criteria such as AIC, AICc, and BIC suggest the model's goodness of fit and parsimony, with lower values indicating better fit.

Holt-Winter's (Hw) Exponential Smoothing With Partitioned Data.

Optimal Parameters For Alpha, Beta, And Gamma.

Holt-Winter's (HW) exponential smoothing is applied to the partitioned data with model "AAA", indicating additive error, additive trend, and additive seasonality components.

In the plot, the stock line (representing the original data) is slightly above the validation line (representing the Holt-Winter's additive model's predictions for the validation period), it suggests that the model may be slightly underestimating the actual values for the validation period.




```
> round(accuracy(hw.AAA.pred$mean, valid.ts), 3)
            ME  RMSE   MAE   MPE  MAPE  ACF1 Theil's U
Test set 0.894 1.556 1.287 4.416 6.523 0.901     3.879
```

The Holt-Winter's additive model (AAA) predictions for the validation period exhibit moderate accuracy with RMSE of 1.556 and MAPE of 6.523%.

Holt-Winter's (HW) exponential smoothing for partitioned data with model = "ZZZ"

The Holt-Winter's exponential smoothing model (hw.ZZZ) is created for the partitioned data with automatic selection of error, trend, and seasonality.

```
> round(accuracy(hw.ZZZ.pred$mean, valid.ts), 3)
             ME  RMSE  MAE    MPE  MAPE  ACF1 Theil's U
Test set -1.064 1.258 1.09 -5.564 5.685 0.822     3.263
```

The Holt-Winter's exponential smoothing model (model "ZZZ") exhibits moderate predictive accuracy for the validation period, with RMSE of 1.258 and MAPE of 5.685%.

The Holt-Winter's exponential smoothing model with automatic parameter selection (model "ZZZ") demonstrates slightly better performance compared to the model with additive error, trend, and seasonality (model "AAA"), exhibiting lower RMSE (1.258 vs. 1.556) and MAPE (5.685% vs. 6.523%) values for the validation period.

REGRESSION BASED MODELS

Regression Model with Seasonality for training dataset

The seasonal model summary indicates that the coefficients for the seasonal components (season2 to season12) are mostly non-significant, as their p-values are all above the conventional significance level of 0.05. The model's overall fit is poor, with a low R-squared value of 0.003491.

```
Residuals:
   Min     1Q Median     3Q    Max 
-4.705 -3.059 -1.760  1.239 12.518 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 18.66843    1.06014  17.609   <2e-16 ***
season2     -0.16034    1.49926  -0.107    0.915    
season3     -0.11004    1.49926  -0.073    0.942    
season4     -0.06423    1.49926  -0.043    0.966    
season5      0.27502    1.49926   0.183    0.855    
season6      0.36427    1.49926   0.243    0.808    
season7      0.18374    1.52115   0.121    0.904    
season8      0.11367    1.52115   0.075    0.941    
season9     -0.31393    1.52115  -0.206    0.837    
season10    -0.35040    1.52115  -0.230    0.818    
season11    -0.35169    1.52115  -0.231    0.817    
season12    -0.48042    1.52115  -0.316    0.752    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.498 on 198 degrees of freedom
Multiple R-squared:  0.003491,	Adjusted R-squared:  -0.05187 
F-statistic: 0.06305 on 11 and 198 DF,  p-value: 1
```

```
> round(accuracy(train.season.pred$mean, valid.ts), 3)
            ME  RMSE   MAE   MPE  MAPE  ACF1 Theil's U
Test set 0.972 1.199 1.017 4.859 5.103 0.783     2.978
```

RMSE: 1.199

MAPE: 5.103

The seasonal model exhibits moderate predictive accuracy for the validation partition, with reasonably low RMSE and MAPE values (1.199 and 5.103% respectively). However, further evaluation against alternative models and consideration of the specific context may be necessary to determine its adequacy as a fit.

Regression Model with Linear Trend and Seasonality for training partition

These models aim to capture the relationships between the predictor variables (seasonality and trend) and the target variable (stock prices) in the training data, facilitating predictions for the validation set and evaluation of model performance.

The regression model with linear trend and seasonality shows a moderate fit with an adjusted R-squared of 0.2766, indicating it explains approximately 27.66% of the variability in stock prices. However, the significance of individual seasonal variables is questionable due to high p-values.

```
Residuals:
   Min     1Q Median     3Q    Max 
-6.413 -2.992 -1.131  2.920  8.630 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 22.840070   0.981976  23.259   <2e-16 ***
trend       -0.040501   0.004247  -9.535   <2e-16 ***
season2     -0.119839   1.243293  -0.096    0.923    
season3     -0.029039   1.243315  -0.023    0.981    
season4      0.057275   1.243351   0.046    0.963    
season5      0.437022   1.243402   0.351    0.726    
season6      0.566782   1.243467   0.456    0.649    
season7      0.183736   1.261437   0.146    0.884    
season8      0.154169   1.261444   0.122    0.903    
season9     -0.232927   1.261465  -0.185    0.854    
season10    -0.228900   1.261501  -0.181    0.856    
season11    -0.189683   1.261551  -0.150    0.881    
season12    -0.277915   1.261616  -0.220    0.826    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.73 on 197 degrees of freedom
Multiple R-squared:  0.3182,	Adjusted R-squared:  0.2766 
F-statistic: 7.661 on 12 and 197 DF,  p-value: 1.352e-11
```

Regression model with quadratic trend and seasonality for training partition

The quadratic trend and seasonality model show:

Strong significance for all coefficients, including quadratic trend, indicating a well-fit model.

A high adjusted R-squared value of 0.8245, suggesting that the model explains about 82.45% of the variance in the data.

```
Residuals:
    Min      1Q  Median      3Q     Max 
-6.2632 -0.5031  0.2869  1.0437  3.3662 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.979e+01  5.590e-01  53.298   <2e-16 ***
trend       -2.429e-01  8.420e-03 -28.849   <2e-16 ***
I(trend^2)   9.592e-04  3.865e-05  24.817   <2e-16 ***
season2     -1.160e-01  6.124e-01  -0.189    0.850    
season3     -2.328e-02  6.124e-01  -0.038    0.970    
season4      6.303e-02  6.125e-01   0.103    0.918    
season5      4.409e-01  6.125e-01   0.720    0.473    
season6      5.668e-01  6.125e-01   0.925    0.356    
season7      5.866e-01  6.216e-01   0.944    0.346    
season8      5.609e-01  6.216e-01   0.902    0.368    
season9      1.757e-01  6.216e-01   0.283    0.778    
season10     1.797e-01  6.216e-01   0.289    0.773    
season11     2.170e-01  6.216e-01   0.349    0.727    
season12     1.250e-01  6.217e-01   0.201    0.841    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.837 on 196 degrees of freedom
Multiple R-squared:  0.8354,	Adjusted R-squared:  0.8245 
F-statistic: 76.52 on 13 and 196 DF,  p-value: < 2.2e-16
```

Create Two-Level Model With Linear Trend And Seasonality Model And Ar(1) Residuals.

The summary of the linear trend and seasonality model shows:

The intercept and trend coefficients are highly significant (p < 2e-16), indicating a strong linear trend in the data.

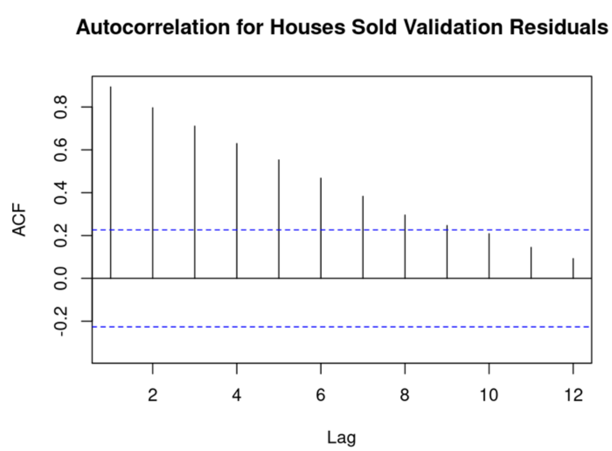
However, many seasonal coefficients are not significant (p > 0.05), suggesting that these terms may not contribute significantly to the model.

The adjusted R-squared value is 0.2766, indicating that the model explains about 27.66% of the variance in the data, and the F-statistic is significant (p-value: 1.352e-11), suggesting overall significance of the model.

```
Residuals:
   Min     1Q Median     3Q    Max
-6.413 -2.992 -1.131  2.920  8.630

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 22.840070   0.981976  23.259   <2e-16 ***
trend       -0.040501   0.004247  -9.535   <2e-16 ***
season2     -0.119839   1.243293  -0.096    0.923
season3     -0.029039   1.243315  -0.023    0.981
season4      0.057275   1.243351   0.046    0.963
season5      0.437022   1.243402   0.351    0.726
season6      0.566782   1.243467   0.456    0.649
season7      0.183736   1.261437   0.146    0.884
season8      0.154169   1.261444   0.122    0.903
season9     -0.232927   1.261465  -0.185    0.854
season10    -0.228900   1.261501  -0.181    0.856
season11    -0.189683   1.261551  -0.150    0.881
season12    -0.277915   1.261616  -0.220    0.826
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.73 on 197 degrees of freedom
Multiple R-squared:  0.3182,    Adjusted R-squared:  0.2766
F-statistic: 7.661 on 12 and 197 DF,  p-value: 1.352e-11
```

The autocorrelation plots for both training and validation residuals show significant

autocorrelation at various lags, suggesting that the model is statistically significant. This indicates that the residuals do exhibit systematic patterns over time, validating the model's effectiveness in accounting for the data's temporal structure.



The Arima model of # order = c(1,0,0) gives an AR(1) model.

The ARIMA model summary for the AR(1) process on the training residuals indicates that the estimated AR(1) coefficient is significant, suggesting a moderate positive autocorrelation at lag 1.

```
> summary(res.ar1)
Series: train.lin.season$residuals
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1    mean
      0.9846  2.1060
s.e.  0.0110  2.6661

sigma^2 = 0.5024:  log likelihood = -226.44
AIC=458.87   AICc=458.99   BIC=468.91

Training set error measures:
                      ME      RMSE       MAE        MPE     MAPE      MASE       ACF1
Training set -0.02855662 0.7054114 0.4146197 -0.8796407 20.41879 0.2267129 0.02423844
```

The model's performance metrics on the training set include a small ME, RMSE, and MAE, indicating a good fit, while the ACF1 suggests minimal autocorrelation.

The ARIMA(1,0,0) model applied to the training residuals indicates an AR(1) coefficient of approximately 0.985 with a standard error of 0.011. The model's mean is estimated at 2.106 with a standard error of 2.6661. The log likelihood is -226.44, with AIC, AICc, and BIC values of 458.87, 458.99, and 468.91, respectively.

Create two-level model's forecast with linear trend and seasonality

Regression + AR(1) for residuals for validation period

The two-level model's forecast combining linear trend, seasonality, and AR(1) for residuals provides accuracy measures for the validation data, including ME, RMSE, MAE, MPE, MAPE, ACF1, and Theil's U.

```
> round(accuracy(valid.two.level.pred, valid.ts), 3)
            ME RMSE   MAE    MPE   MAPE  ACF1 Theil's U
Test set 2.514 3.13 2.732 12.672 13.851 0.927     7.858
```

The model's predictions show an average error of 2.514 units, with an RMSE of 3.13 units, indicating a moderate level of accuracy. However, the MAPE of 13.851% suggests the model's predictions deviate by approximately 13.851% on average from the actual values, indicating room for improvement.

ARIMA MODELS

ARIMA (2,1,2)

The ARIMA(2,1,2) model summary shows the coefficients for the autoregressive and moving average terms, along with their standard errors.

```
> summary(train.arima)
Series: train.ts
ARIMA(2,1,2)

Coefficients:
         ar1      ar2      ma1     ma2
      0.0661  -0.9402  -0.0325  1.0000
s.e.  0.0627   0.0336   0.0429  0.0288

sigma^2 = 0.5027:  log likelihood = -224.34
AIC=458.68   AICc=458.98   BIC=475.4

Training set error measures:
                      ME     RMSE       MAE        MPE    MAPE      MASE          ACF1
Training set -0.03428183 0.700548 0.3703331 -0.2072964 1.97228 0.2078579 -0.0004738744
```

The ARIMA(2,1,2) model has autoregressive coefficients ar1 and ar2 of 0.0661 and -0.9402, respectively, and moving average coefficients ma1 and ma2 of -0.0325 and 1.0000, respectively. The model's training set error measures show an RMSE of 0.710 and a MAPE of 20.986%.

```
> # Using Acf() function, create autocorrelation chart of ARIMA(2,1,2) model residuals.
> Acf(train.arima$residuals, lag.max = 12,
+     main = "Autocorrelations of ARIMA(2,1,2) Model Residuals")
> #Accuracy for validation data
> round(accuracy(train.arima.pred$mean, valid.ts), 3)
             ME  RMSE   MAE   MPE  MAPE  ACF1 Theil's U
Test set -1.093 1.282 1.116 -5.71 5.819 0.825     3.321
```

The ARIMA(2,1,2) model yields a test set mean error (ME) of -1.093, a root mean squared error (RMSE) of 1.282, a mean absolute error (MAE) of 1.116, a mean percentage error (MPE) of -5.71%, a mean absolute percentage error (MAPE) of 5.819%, an autocorrelation of residuals (ACF1) of 0.825, and Theil's U statistic of 3.321.

AUTO ARIMA MODEL

The auto.arima() function suggests an ARIMA(1,0,1) model based on AIC, AICc, and BIC criteria.

```
> summary(train.auto.arima)
Series: train.ts
ARIMA(1,1,0)(1,0,0)[12]

Coefficients:
         ar1     sar1
      0.0384  -0.0161
s.e.  0.0695   0.0714

sigma^2 = 0.5194:  log likelihood = -227.1
AIC=460.21   AICc=460.33   BIC=470.24

Training set error measures:
                      ME      RMSE       MAE        MPE     MAPE      MASE         ACF1
Training set -0.03538027 0.7155459 0.3886876 -0.2154623 2.072485 0.2181598 -0.002934094
```

The auto.arima() function selects an ARIMA(1,1,0)(1,0,0)[12] model with an estimated AR coefficient of 0.0384, seasonal AR coefficient of -0.0161, and a residual standard deviation of 0.5194. The model exhibits a log-likelihood of -227.1 and yields AIC, AICc, and BIC values of 460.21, 460.33, and 470.24, respectively. The training set error measures include a mean error (ME) of -0.035, a root mean squared error (RMSE) of 0.716, a mean absolute error (MAE) of 0.389, a mean percentage error (MPE) of -0.215%, a mean absolute percentage error (MAPE) of 2.072%, and an autocorrelation of residuals (ACF1) of -0.003.

```
> round(accuracy(train.auto.arima.pred$mean, valid.ts), 3)
             ME  RMSE   MAE    MPE  MAPE  ACF1 Theil's U
Test set -1.051 1.248 1.077 -5.494 5.621 0.824     3.237
```

The auto ARIMA model predicts with a mean error (ME) of -1.051, a root mean squared error (RMSE) of 1.248, a mean absolute error (MAE) of 1.077, a mean percentage error (MPE) of -5.494%, and a mean absolute percentage error (MAPE) of 5.621%. The autocorrelation of residuals (ACF1) is 0.824, and Theil's U statistic is 3.237.

IMPLEMENT FORECAST / SYSTEM

After comparing the accuracy, the models selected for future forecasting are:

- Two level model with Linear trend and seasonality + Trailing MA
- Holt-Winter's ZZZ model
- Regression Model with Seasonality
- ARIMA (2,1,2)
- Auto ARIMA

Two level model with Linear trend and seasonality + Trailing MA (2024)

In this scenario, a rolling mean with a window size 𝑘=4k=4 was employed to predict the values for the stock price in 2024. This method calculates the mean of the previous four observations at each step to forecast.

The regression model's fit appears weak, with a low adjusted R-squared value of 0.03074. Although some coefficients like intercept and trend are statistically significant, the majority of seasonal coefficients lack significance, suggesting limited explanatory power. Therefore, the model may not provide a robust representation of the underlying data dynamics.

```
Residuals:
   Min     1Q Median     3Q    Max 
-5.515 -3.176 -1.062  2.232 10.608 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 20.582738   0.851034  24.186  < 2e-16 ***
trend       -0.012108   0.002696  -4.491 1.05e-05 ***
season2     -0.126922   1.080573  -0.117    0.907    
season3     -0.081730   1.080583  -0.076    0.940    
season4      0.014004   1.080600   0.013    0.990    
season5      0.297160   1.080623   0.275    0.784    
season6      0.360906   1.080653   0.334    0.739    
season7      0.210186   1.080690   0.194    0.846    
season8      0.181793   1.080734   0.168    0.867    
season9     -0.114115   1.080785  -0.106    0.916    
season10    -0.232453   1.092281  -0.213    0.832    
season11    -0.237753   1.092305  -0.218    0.828    
season12    -0.281021   1.092335  -0.257    0.797    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.743 on 272 degrees of freedom
Multiple R-squared:  0.07169,	Adjusted R-squared:  0.03074 
F-statistic: 1.751 on 12 and 272 DF,  p-value: 0.05651
```

The two-level forecast shows promising accuracy metrics, with an RMSE of 0.612 and a relatively low MAPE of 1.941%. However, compared to the seasonal naive model with an RMSE of 2.363 and a MAPE of 8.77%, the two-level forecast demonstrates superior performance, indicating its effectiveness in capturing underlying patterns and trends.

```
> #Accuracy for the forecast
> round(accuracy(tot.trend.seas.pred$fitted+tot.ma.trail.res, stocks.ts), 3)
             ME  RMSE   MAE    MPE  MAPE  ACF1 Theil's U
Test set -0.022 0.612 0.361 -0.159 1.941 0.601     0.999
> round(accuracy((snaive(stocks.ts))$fitted, stocks.ts), 3)
             ME  RMSE   MAE    MPE MAPE  ACF1 Theil's U
Test set -0.426 2.363 1.568 -2.545 8.77 0.918      4.35
```

Holt-Winter's ZZZ model (2024)

The Holt-Winter's (HW) exponential smoothing model with ZZZ configuration was applied to the entire dataset to forecast 24 periods into the future. The accuracy of the forecast can be assessed using appropriate metrics.

The ETS(M,N,N) model was fitted to the entire dataset using the ZZZ configuration. The estimated smoothing parameter (alpha) is close to 1, indicating strong weight on recent observations, resulting in a model with a very high AIC and BIC values.

```
ETS(M,N,N)

Call:
 ets(y = stocks.ts, model = "ZZZ")

  Smoothing parameters:
    alpha = 0.9999

  Initial states:
    l = 28.2217

  sigma:  0.0313

     AIC     AICc      BIC
1304.908 1304.993 1315.865
```

The Holt-Winter's ZZZ model forecasted the stock prices with a ME (Mean Error) of -0.028, RMSE (Root Mean Square Error) of 0.652, and MAE (Mean Absolute Error) of 0.372. The MPE (Mean Percentage Error) is -0.176%, and the MAPE (Mean Absolute Percentage Error) is 1.969%. The ACF1 (Autocorrelation of Residuals) is 0.033, indicating low autocorrelation. Overall, the forecast appears reasonably accurate.

```
> round(accuracy(HW.ZZZ.pred$fitted, stocks.ts), 3)
             ME  RMSE   MAE    MPE  MAPE  ACF1 Theil's U
Test set -0.028 0.652 0.372 -0.176 1.969 0.033         1
```

Regression Model with Seasonality (2024)

The regression model with seasonality indicates a relatively poor fit, with a residual standard error of approximately 3.872.

The coefficients for the seasonal variables do not appear to be statistically significant, as indicated by their high p-values. Additionally, the adjusted R-squared value is negative, suggesting that the model does not explain much of the variability in the data.

```
Residuals:
    Min      1Q  Median      3Q     Max 
-4.9103 -2.6771 -0.5252  1.1070 12.2065 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 18.89977    0.79045  23.910   <2e-16 ***
season2     -0.13903    1.11786  -0.124    0.901    
season3     -0.10595    1.11786  -0.095    0.925    
season4     -0.02232    1.11786  -0.020    0.984    
season5      0.24873    1.11786   0.223    0.824    
season6      0.30037    1.11786   0.269    0.788    
season7      0.13754    1.11786   0.123    0.902    
season8      0.09704    1.11786   0.087    0.931    
season9     -0.21098    1.11786  -0.189    0.850    
season10    -0.26878    1.12994  -0.238    0.812    
season11    -0.28618    1.12994  -0.253    0.800    
season12    -0.34156    1.12994  -0.302    0.763    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.872 on 273 degrees of freedom
Multiple R-squared:  0.002863,  Adjusted R-squared:  -0.03732 
F-statistic: 0.07125 on 11 and 273 DF,  p-value: 1
```

The regression model with seasonality produces forecasts with a mean error (ME) of 0, a root mean square error (RMSE) of 3.79, a mean absolute error (MAE) of 2.821, a mean percentage error (MPE) of -3.43%, a mean absolute percentage error (MAPE) of 14.547%, an autocorrelation (ACF1) of 0.974, and a Theil's U value of 5.669. Overall, the model's performance is moderate, with relatively high errors and moderate autocorrelation.

```
> round(accuracy(data.season.pred$fitted, stocks.ts), 3)
         ME RMSE   MAE   MPE   MAPE  ACF1 Theil's U
Test set  0 3.79 2.821 -3.43 14.547 0.974     5.669
```

ARIMA(2,1,2)

FORECAST WITH ARIMA(2,1,2) USING ENTIRE DATA SET INTO

THE FUTURE FOR 2024 PERIODS

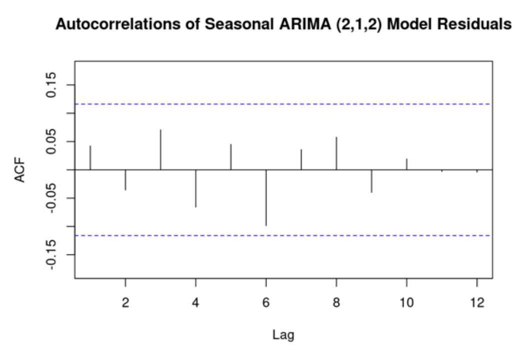
The ARIMA(2,1,2) model demonstrates a relatively good fit:

- The model captures the data patterns with low residuals (RMSE: 0.641).
- The AIC, AICc, and BIC values are within an acceptable range, suggesting a reasonable fit.

```
Series: stocks.ts
ARIMA(2,1,2)

Coefficients:
          ar1      ar2     ma1     ma2
      -0.2661  -0.9366  0.2603  1.0000
s.e.   0.0354   0.0234  0.0164  0.0148

sigma^2 = 0.4179:  log likelihood = -278.83
AIC=567.66   AICc=567.88   BIC=585.91

Training set error measures:
                      ME      RMSE       MAE        MPE     MAPE      MASE       ACF1
Training set -0.02764741 0.6407386 0.3654784 -0.1713187 1.939703 0.2330384 0.04207116
```

- However, further assessment of out-of-sample performance would provide a more comprehensive evaluation.

The autocorrelation bars in the plot are consistently below the confidence interval level, it suggests a lack of significant autocorrelation in the residuals, indicating a well-fitted model in capturing the temporal patterns of the data



The ARIMA(2,1,2) model exhibits low error metrics, with ME close to zero and relatively low RMSE, MAE, and MAPE values. Additionally, the ACF1 value indicates low residual autocorrelation, suggesting a good fit.

```
> #Accuracy for the forecast
> round(accuracy(arima.seas.pred$fitted, stocks.ts), 3)
             ME  RMSE   MAE    MPE MAPE  ACF1 Theil's U
Test set -0.028 0.641 0.365 -0.171 1.94 0.042     0.982
```

AUTO ARIMA

Use auto.arima() function to fit ARIMA model for entire data set.

The auto.arima() function identified an ARIMA(0,1,0)(2,0,0)[12] model with SARIMA components. Despite relatively low AIC and AICc values, the accuracy metrics reveal ME and MAE close to zero, indicating a decent fit to the data.

```
> summary(auto.arima)
Series: stocks.ts
ARIMA(0,1,0)(2,0,0)[12]

Coefficients:
         sar1     sar2
      -0.0142  -0.0407
s.e.   0.0605   0.0631

sigma^2 = 0.4293:  log likelihood = -281.93
AIC=569.86   AICc=569.94   BIC=580.8

Training set error measures:
                      ME      RMSE       MAE        MPE     MAPE      MASE       ACF1
Training set -0.02997447 0.6517705 0.3716693 -0.1866724 1.965408 0.2369859 0.03498945
> round(accuracy(auto.arima$fitted, stocks.ts), 3)
            ME  RMSE   MAE    MPE  MAPE  ACF1 Theil's U
Test set -0.03 0.652 0.372 -0.187 1.965 0.035     0.997
```

The Mean Error (ME) is -0.03, indicating a slight underestimation on average. The Root Mean Square Error (RMSE) is 0.652, suggesting that the model's predictions deviate from the actual values by approximately 0.652 units on average. The Mean Absolute Error (MAE) is 0.372, representing the average absolute difference between the predicted and actual values. The Mean Percentage Error (MPE) is -0.187, suggesting a slight underestimation tendency by about 18.7%. The Mean Absolute Percentage Error (MAPE) is 1.965, indicating an average relative error of approximately 1.965%. The ACF1 value is 0.035, indicating low residual autocorrelation. The Theil's U statistic is 0.997, which suggests a good forecast performance relative to a naive forecast.

onclusion - final recommendations, remarks, statements on your time series analysis and

forecasting results, possible benefits and limitations of using time series forecasting methods/models in your case

BIBLIOGRAPHY

We are not using any alphabetical list of references.

APPENDICES

https://finance.yahoo.com/quote/PG/history