COVID-19 and the Economy

Adam Kiehl and Oriana Meldrum

May 12, 2021

Abstract

This project took a look at the COVID-19 recession of 2020 through the lens of the American stock market and health care sector with the ultimate goal of predicting economic recovery in 2021. Time series data were obtained for the S&P 500 stock index and the Real Personal Health Care Consumption Expenditures metric and were used to train predictive ARMA models. While reasonably constructed, the models produced unsatisfactory forecasts that were hampered by difficulties in handling structural breaks present in the training data.

1 Introduction

The COVID-19 pandemic took a great toll on all facets of the economy and attempting to describe its impact in totality would be a massive endeavour. Rather, this project examined the pandemic recession from two specific perspectives: that of the stock market and that of the health care sector.

The effect of the COVID recession on the American stock market was studied through a close look at its effects on the underlying sectors that compose the market. The economy is divided into eleven sectors (information technologies, materials, e.g.) which all reacted in different ways to the exogenous price shock of a global pandemic. Understanding the intersection of sector behavior and such a structural break was the goal of modeling and was believed to be key in constructing accurate price forecasts.

The health care industry is typically thought to be recession proof [4]. However the COVID-19 pandemic has shown that while the industry might be recession proof, it is not necessarily pandemic proof. Once the WHO declared COVID-19 a pandemic and various lock down measures were introduced, all elective procedures and other non-elective procedures that could be delayed where canceled or postponed. In the first quarter of 2020, the health care industry lost 1.4 million jobs, and the real personal consumption for health care services plummeted nearly 5 percent from the 4th quarter of 2019 [7]. With the introduction of the COVID long haulers (those who have lingering symptoms and ailments from their infection with COVID-19) and the resuming of both non-elective and elective procedures, we were hoping to explore the forecasting models for how health care might recover from the COVID-19 pandemic.

1.1 Data

Stock market data was obtained from the Alpha Vantage Stock API [2] using the alphavantager CRAN package in R [11]. Alpha Vantage contains a massive repository of publicly-available worldwide asset prices, trading volumes, and company profiles. Price data, specifically, were extracted for the S&P 500 index to limit the scope of study to American companies and to only the largest and most influential of those. Additionally, the data were truncated at the beginning of 2016 to include a reasonable training sample while simultaneously maintaining consistency with recent trends as compared to trends that may be found corresponding to past recessions. Incomplete time series were discarded with company symbols AMCR, BF.B, BKR, CARR, CTVA, DXC, EVRG, FTV, FOXA, FOX, IR, LW, LIN, OTIS, UA, VTRS, and VNT. Complete time series were then aggregated according to each of the sectors of the stock market, producing eleven time series of sector-wide price averages and a single *total* time series that is an index-wide mean. This *total* time series was considered a simulator of the S&P 500 as a whole and was the eventual target of predictive modeling. The sector-specific time series can be seen below from January 2020 forward. Additionally, data from before January 1, 2021 were used as a training set while data after were saved for model validation.

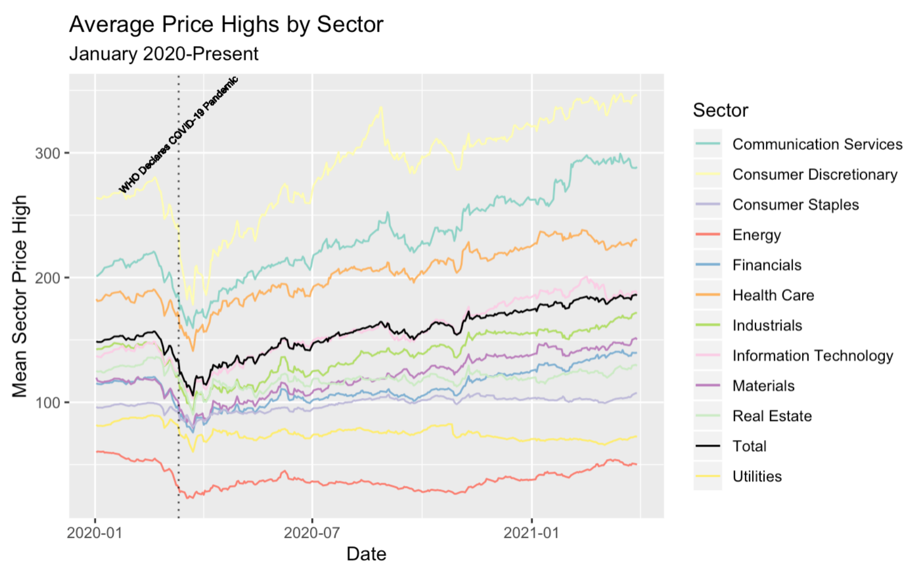

Figure 1: Time series for each sector-aggregated mean price including a *total* time series that is both the simulator of the S&P 500 and the target of predictive efforts.

Health care data were obtained from the Federal Reserve Economic Data (FRED). This is an economic time series aggregated by the Federal Reserve Bank of St. Louis center from a variety of sources. The data is updated quarterly, since 2002, in billions of chained 2021 dollars to account for inflation. This data were specifically selected as it has been seasonally adjusted, thus eliminating the need for future transforms for seasonality. The entire time frame since 2002 was selected to help train the model on the trends in Real Personal Consumption and how it is recession proof and increasing over time. The time series for this data can be seen below, encompassing the entire time frame.

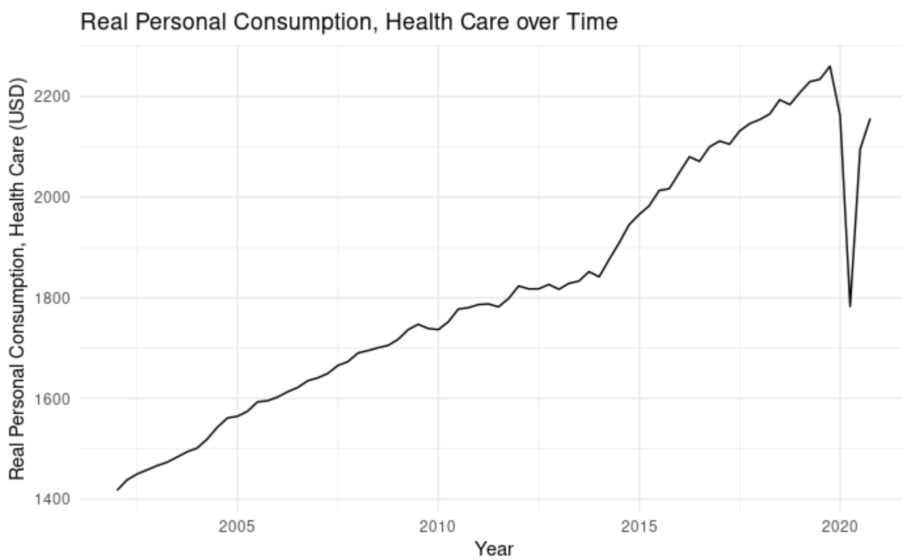

Figure 2: Time series for Health Care dating back to 2002

2 Methods

The first objective of this analysis was to pinpoint the occurrences of structural breaks that were clearly present in the S&P 500 time series for 2020. To do this, a Chow Test [5] performed using the strucchange CRAN package in R [16]. The Chow Test identifies structural breaks by comparing linear regression coefficients of different segments of time. In this case, March 11th, June 10th, and September 4th were identified as significant structural break points and were confirmed both visually and historically. The first break point aligns exactly with the World Health Organization (WHO) declaring COVID-19 a global pandemic. The second aligns closely with the coincidence of the end of several state stay-at-home orders, the reopening of Miami Beach, Disney World, and New York City, and a one-day stock plummet on March 12th [13]. The third follows a sharp decline in stock prices after the United States refused to join a WHO-led global vaccine effort and also occurs near the beginning of many school years [8]. Of the periods separated by structural breaks, the second, beginning on March 11th and ending on June 9th, is especially distinguished from the others by its low mean and high variance. For this reason, it was isolated as a particular interest of study and a dummy variable was created to represent its span.

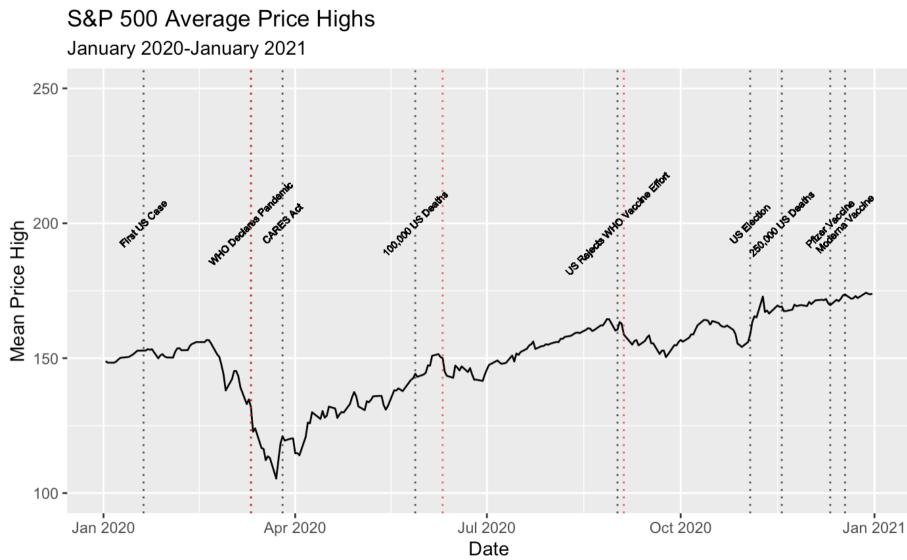

Figure 3: Structural breaks identified using a Chow Test in the 2020 time series.

Unlike the stock market data, the health care data from FRED was already accounting for seasonal variations, achieving stationarity was not necessary. The FRED removed the seasonality of the data prior to publishing. After much exploration [6], the astsa package written by David Stoffer was used as it seemed the most simple and straight forward of all time series packages available in CRAN for estimating linear time series models [3]. Data must be continuous and regularly spaced, and the FRED data fits this criteria as it has been recorded quarterly since 2002.

2.1 Achieving Stationarity

Before any modeling could be performed, the stock market sector time series had to be made stationary by removing the inherent trend and seasonality present in the data. This task was accomplished by several different methods. The first involved applying a seasonal differencing transform

$$\mathbf{Y}_{\nabla 252}(t) = \mathbf{Y}(t) - \mathbf{Y}(t - 252) \tag{1}$$

followed by a first-differencing transform to remove linear trend.

$$\mathbf{Y}_{\nabla}(t) = \mathbf{Y}(t) - \mathbf{Y}(t - 1) \tag{2}$$

The second applied only a first-differencing transform and used dummy variables to represent each season for modeling. The third used a twice-differencing transform applying consecutive first-differencing transforms to remove non-linear trends [10]. All three methods produced a satisfyingly stationary time series like the one below. Stationarity was further confirmed by the decay feature of each time series' auto-correlation function (ACF). Following the Box-Jenkins method, the shape of each ACF plot also suggested that a full Auto-Regressive Moving Average (ARMA) was the best choice for modelling [9].

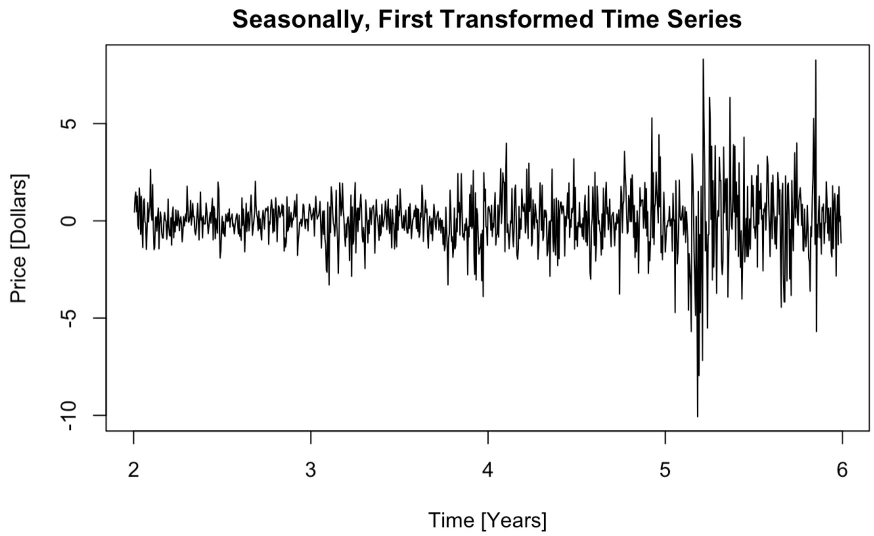

Figure 4: Stationary time series achieved through a seasonal differencing transform followed by a first-differencing transform. The structural break is still visually present early in year five.

It was further considered that some sectors could be modelled together given that they exhibited similar patterns of fluctuation over time. The stationary series were then compared using a correlation matrix that found high degrees of correlation between different groups of sectors. Ultimately, it was decided that Group A should consist of Communication Services, Consumer Discretionary, and Information Technology, Group B should consist of Financials, Industrials, and Materials, Group C should consist of Real Estate and Utilities, and Consumer Staples, Energy, and Health Care should be modelled independently. A factor analysis confirmed these groupings with separate factors dominated by each group's underlying sectors.

2.2 Modeling

ARMA(p, d, q) models were used across the board for stock market time series modeling. To tune the parameters of each model, an optimization algorithm was employed to minimize the Akaike Information Criterion (AIC)

$$AIC = 2k - 2ln(\hat{L}) \tag{3}$$

where k is the number of estimated parameters and L is the maximum value of the likelihood function for the model [1]. For the first two methods of achieving stationarity with d=0 differences, the AIC was minimized at p=7 auto-regressive parameters and q=10 moving average parameters. For the third method with d=2 differences, the AIC was minimized at p=9, q=3. Six models (one for each sector/sector grouping) were fit with ARMA(7, 0, 10) and seasonally and first differenced training data. Six more were fit with ARMA(7, 0, 10) and first-differenced training data. A final model was fit with ARMA(9, 2, 3) and raw training data. All models included a dummy variable representing the identified structural break period of interest. The Ljung-Box Test was used to test for serial auto-correlation with the standard practice of choosing a lag equal to twice the supposed period (504, in this case) [14]. The first class of models was largely rejected by the Ljung-Box test with all but one p-value less than 0.05. The second and third class of models passed with flying colors, boasting p-values greater than 0.80.

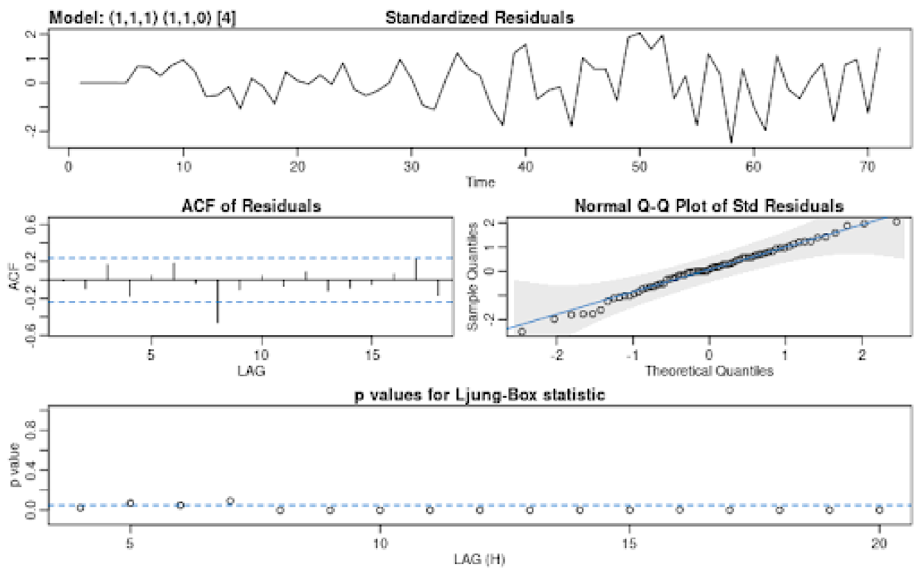

Figure 5: Health Care ARIMA(0,1,1)_4 Model Diagnostics

Due to the unique features of the asta package (the astsa package automatically produces diagnostic plots), model diagnostics are as simple as running the code and interpreting the resulting graphs. The standardized residuals plot confirms the data is stationary and is not need of any transformations. After running various models, the ARIMA(0,1,1)_4 forcasting model was selected as it had the smallest AIC.

3 Results

The forecast CRAN package was used to generate predictions for each model 60 trading days into 2021 [12]. This study used the most extreme of the upper and lower confidence bounds of each prediction to better simulate the 'fuzziness' of the market that was not otherwise captured by these models. The values from each sector model were then averaged to simulate the S&P 500 *total* series. Following standard practice, the 'differenced' predicted values were then converted into real dollar values by cumulatively adding them to the last known un-transformed training observation. For a single first-differencing transform, as with the second class of models, this was performed simply with the cumsum() function in R. However, to deal with the seasonally and first transformed class of models, some adjustments had to be made to the cumulative sum method.

Let $\mathbf{Y}$(t) be a time series of price data:

$$\mathbf{Y}(t) = y_1, ..., y_n$$

Apply a seasonal differencing transform:

$$\mathbf{Y}_{\nabla 252}(t) = \mathbf{Y}(t) - \mathbf{Y}(t-252)$$

Apply a first differencing transform:

$$\begin{aligned}
\mathbf{Y}_{\nabla\nabla 252}(t) &= \mathbf{Y}_{\nabla 252}(t) - \mathbf{Y}_{\nabla 252}(t-1) \\
\mathbf{Y}_{\nabla\nabla 252}(t) &= [\mathbf{Y}(t) - \mathbf{Y}(t-252)] - [\mathbf{Y}(t-1) - \mathbf{Y}(t-253)] \\
\mathbf{Y}_{\nabla\nabla 252}(t) &= \mathbf{Y}(t) - \mathbf{Y}(t-252) + \mathbf{Y}(t-253) - \mathbf{Y}(t-1)
\end{aligned}$$

Solve for $\mathbf{Y}$(t) to obtain inverse differencing transform:

$$\mathbf{Y}(t) = \mathbf{Y}_{\nabla\nabla 252}(t) + \mathbf{Y}(t-1) + \mathbf{Y}(t-252) - \mathbf{Y}(t-253)$$

Ultimately, three predicted time series were built and compared with the true 2021 validation series using Mean Squared Error. The twice-differenced model performed the best, achieving an MSE of 32.04. The seasonally and first transformed model appeared to perform well initially but simulated almost exactly the sudden price spike of March 2020 and so did not produce satisfyingly original predictions. A forecast combination [15] was also performed between all three series to smooth the effect of different model fits, but it did not outperform the models in terms of MSE. Visualizations can be seen on the next page.

Unfortunately, the asta package did not do a good job forecasting as it tried to mimic the last four quarters entered. This of course is a problem because the last four quarters included the 2020 COVID-19 pandemic shock. While the model did produce a smaller shock, it is still not accurate. In addition, a test

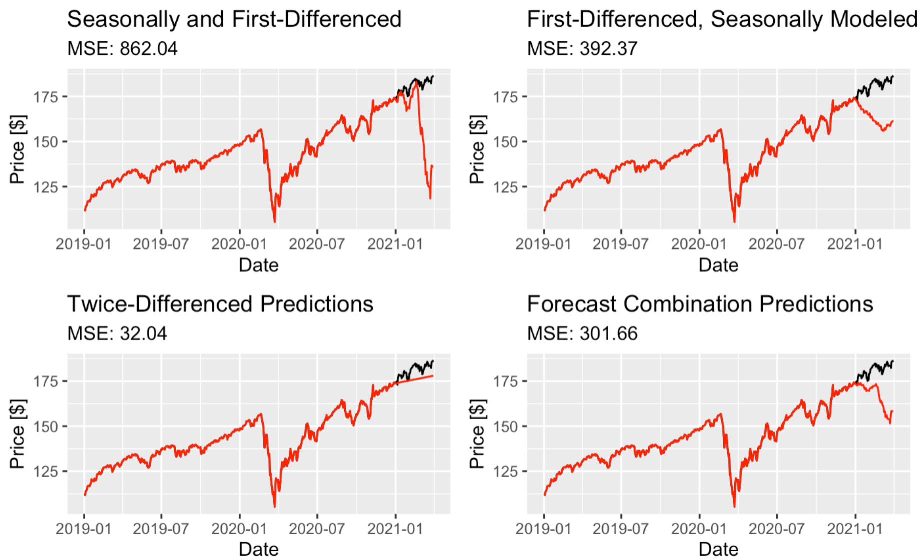

Figure 6: Three constructed model predictions along with a forecast combination. The first seems to imitate closely the downwards price spike of March 2020.

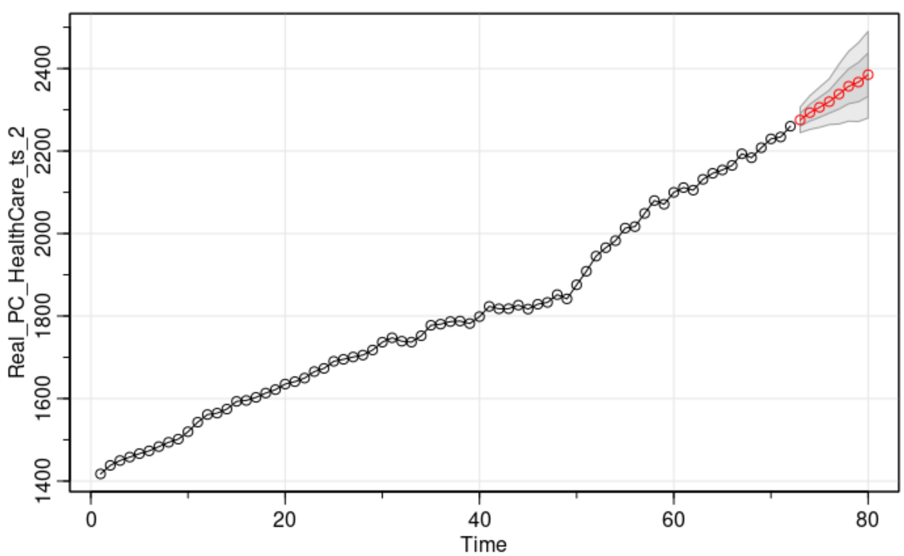

Figure 7: Health Care Forecasting from quarter four of 2019

was performed going back to the last quarter of 2019 (as the last quarter of 2019 did see some slowing of growth due to the introduction of an outbreak of a then unknown virus), however this model was similarly unsatisfactory as it did not capture the shock to the system in early 2020, and just predicted further steady growth.

4 Discussion

While certainly interesting, this analysis of the stock market failed to produce any satisfying forecasts for 2021. Sectors were modeled independently or in small groups with the idea that each sector/sector group reacted somewhat differently to the pandemic recession of 2020 and that separate modeling would reveal those discrepancies. This may be true; however, many of the models also failed to recognize the structural break and simulated a similar price spike to the one in March 2020 in March 2021. The inverse transform of the seasonally and first transformed models, specifically, utilized values from around a year ago in its cumulative sum so the 2020 spike is problematically mirrored no matter how the model performs. Somewhat paradoxically, the only model that did not account fully for seasonality within the data performed the best according to MSE. However, the predictions from that twice-differenced model were also the most 'linear' and so it was hypothesized that the transform did not correctly remove the linear trend of the training time series and so that trend was carried over neatly into the model's predictions. Future research into this area of study should include larger amounts of training data and more greatly weight the effect of a structural break period so it is more visible to the model. It would also be interesting to experiment with interaction effects and different kinds of models including an LSTM model.

It is difficult to know how accurate the health care forecasting model is, as the results for the first quarter of 2021 have not been released by FRED, meaning we could not compare the predicted data to the real world data. However, based on the trends observed since 2002, the steady increase in spending should be accurate. Provided there are not more devastating shocks to the system, it would be reasonable to believe the health care forecasting model is accurate. Given that FRED releases data quarterly, one needs only to occasionally check the FRED for updates to see how the real personal consumption changes and if the projections contained here are accurate.

Due to time constrains, no further model predictions were explored, though it was hoped that the intricacies of this pandemic could have been further studied. Training models to recognize and acceptably handle significant structural breaks like the price shock present in spring 2020 is difficult. Some models resemble too closely the values from a year ago. Some models fail to capture the variability and 'fuzziness' of the market. Model validation is sometimes difficult with the structural break present as well. The biggest lesson we learned during this project was that it's easy to get carried away with a project and its scope. To be successful, one must have a simple, well-defined research question and one must stick to a structured plan that is based on extensive background research.

Code

Summary PDFs generated from RMarkdown scripts will be attached to this paper. A full repository of our code and data is publicly available at https://github.com/akiehl2000/STATCapstoneProject.

Acknowledgements

This paper was written under the mentorship and assistance of Dr. Aaron Nielsen, Gray Stanton, and Youngseok Song for Statistics 472 capstone class at Colorado State University.

References

[1] Autoregressive moving average arma(p, q) models for time series analysis.

[2] Alpha vantage. 2017.

[3] Asta package, Dec 2020.

[4] Is healthcare recession-proof? 2020.

[5] Structural breaks, Dec 2020.

[6] Cran task view: Time series analysis, April 2021.

[7] Real personal consumption expenditures: Services: Health care, April 2021.

[8] A timeline of covid-19 developments in 2020, Jan 2021.

[9] George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Wiley, 5 edition, 2015.

[10] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2 edition, 2002.

[11] Matt Dancho and Davis Vaughan. Alphavantager: Lightweight r interface to the alpha vantage api, Mar 2020.

[12] Rob Hyndman, George Athanasopoulos, and Christoph Bergmeir. Forecast: Forecasting functions for time series and linear models, Mar 2021.

[13] Chase Porter. Covid-19 calendar timeline data for analysts, Aug 2020.

[14] Kyle Rich. Time series in r: Stationairty testing, 2017.

[15] Allan Timmermann. Forecast combinations. Nov 2004.

[16] Achim Zeileis, Friedrich Leisch, Kurt Hornik, Christian Kleiber, Bruce Hansen, and Edgar Merkle. Strucchange: Testing, monitoring, and dating structural changes, Oct 2019.