

# Dynamic models for analysing stock market behaviour under the COVID-19 pandemic

MAARTEN PETERS, University of Amsterdam, the Netherlands

DR. CRISTIAN RODRIGUEZ RIVERO\*, University of Amsterdam, FNWI, IvI, the Netherlands

DR. JULIÁN ANTONIO PUCHETA†, Universidad Nacional de Córdoba, Argentina

During the COVID-19 pandemic [35] questions were raised on how to balance government measures ensuring population health and allowing economic development. At the start of the pandemic, Saudi-Arabia and Russia were in a pricing war over crude oil [20] which, along with speculation on the economic impact of COVID-19, led to a unprecedented negative crude oil price in the West Texas Intermediate (WTI) [5]. As the WTI serves as a benchmark for crude oil prices in North America, and a proxy for economic development [23], it is an interesting candidate to use for price forecasting [40]. The pandemic provides a unique perspective, as it introduces a new set of variables [11, 17], such as infections, deaths, vaccinations and government measures<sup>1</sup>, that might aid in predicting economic development [1, 4, 14]. Related studies generally focus on macroeconomic development, such as gross domestic product (GDP), unemployment or inflation over years or decades, rather than short-term development over days, weeks or months. This study attempts to combine data from the COVID-19 pandemic, weather, stock pricing data and machine learning techniques to determine the relationship between these variables and their value towards more accurate price forecasting. As stock prices have high variance, extreme values might indicate local or global stock market crashes, an optimal model would be able to predict these crashes. To determine the outcome of our research question, we compare the value of our data between a baseline model, linear model and two state-of-the-art (SOTA) models, the random forest regressor (RFR) and auto-regressive integrated moving average (ARIMA) model. Both SOTA models tend to perform better or similar with less features, indicating the data does not add significant value to the prediction of stock market values.<sup>2</sup>

CCS Concepts: • **Computing methodologies** → *Learning linear models*.

Additional Key Words and Phrases: COVID-19, dynamic processes, stock market index, multivariable analysis, time series forecasting

## ACM Reference Format:

Maarten Peters, Dr. Cristian Rodriguez Rivero, and Dr. Julián Antonio Pucheta. 2022. Dynamic models for analysing stock market behaviour under the COVID-19 pandemic. In *Chongqing '22: 14th International Conference on Computer Modeling and Simulation*, June 24–26, 2022, Chongqing, China. ACM, New York, NY, USA, 13 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 Introduction

At the start and during the COVID-19 pandemic, rising infection rates, deaths and widespread lockdown measures raised questions on how to limit the pandemic's impact on economic development [35]. This thesis looks into how the

\*Internal Supervisor from the University of Amsterdam as part of the master thesis programme

†External Supervisor from the National University of Córdoba

<sup>1</sup>Provided by the National Institute for Public Health and the Environment (in Dutch: Rijksinstituut voor Volksgezondheid en Milieu, RIVM), the Johns Hopkins University Coronavirus Resource Center (JHU) and the Oxford COVID-19 Government Response Tracker (OxCGRT).

<sup>2</sup>An overview of the code, documentation and data collected and/or produced during this research has been made available on Github: <https://github.com/maartenpeters/dynamic-models-for-stock-market-behaviour.git>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2022 Association for Computing Machinery.

Manuscript submitted to ACM

pandemic affects stock market values, as these are notoriously volatile in their pricing and influenced by a wide range of factors, demonstrated by the WTI value dipping in the negative at the first wave of infections [5]. The relationship between stock markets and economic factors falls under the study of macroeconomics and finds its origins in the business cycle theory, dating back to the early 1800s [8]. This theory generally states that all economic development follows a cycle or waveform, going through the phases expansion, crisis, recession and recovery. These business cycles can be measured [2] and are part of the tools that a government can use to mitigate economic downturn. Predicting the development of a business cycle is generally based on economic indicators and economic theory, such as Keynesian theory [24] and real-business cycle (RBC) theory [26]. One of the generally accepted economic indicators is the Standard and Poor's (S&P) 500 as an indicator for the United States (US) economy, stating a relationship between economic development and the top 500 US companies. One of the complexities of these relationships, between a country and a national representation of the local stock market, is that a stock market is not a local phenomenon. Stocks and other financial products are traded internationally via a stock exchange and these stock indexes, such as the S&P500, are a representation of the stock exchange itself and not (necessarily) of the country it resides in. As the pandemic's impact and government restrictions varies between countries, it is needed to isolate a region to perform a causal analysis of its impact on a stock market. When the WTI value dropped, all oil-related stocks dropped in value as a consequence, such as Royal Dutch Shell, one of the largest companies in the Netherlands. As previous research noted a relationship between pandemic data and economic development, and a relationship between economic development and stock indexes, the question can be posed whether these relationships aid in predictive modelling of stock values in a region. This research will focus on the Netherlands, with the Amsterdam Exchange index (AEX) as an analog for the S&P500, the former featuring the top 25 most traded securities on the Amsterdam Exchange. As macroeconomics looks at the large scale development of an economy, the models generally apply to years or decades of data. Daily stock market development is a lot more volatile than other economic indicators, generally displaying a non-linear patterns. Machine learning can aid in these predictions, as it can capture non-linear behaviour better than traditional linear models [41]. Still, stock market prediction is non-trivial as the efficient market hypothesis (EMH) [32] states that no investor can gain an advantage over others based on historical and current information. Some investors base their decisions on fundamental analysis [28] to gain an advantage on others. Machine learning has shown to be effective at improving prediction accuracy on stock market securities or crude oil pricing [21, 33, 40, 41], especially using recurrent- (RNN) or artificial neural networks (ANN). Having established the relationship between stock market values and economic development, the question remains whether the pandemic actually influenced stock market values. Social and economic factors have shown historical importance in stock market crashes, such as with the Tulip Mania, the Wall Street Crash of 1929 and the Great Depression [25, 39] and the global financial crisis of 2008. An important topic during the recent pandemic was whether it would influence the economy and what its impact would be. Historically pandemics have shown their effects [6, 22, 31], but the quantitative relationship between the COVID-19 pandemic and economic development is still an object of research [1, 4, 9, 14]. During the pandemic daily data was registered on its development, such as infection rates, deaths, vaccinations and government restrictions [11, 17]. As stock market crashes generally are a social phenomenon, this research investigates the relationship between these features and stock market pricing. The following research questions are posed:

RQ1 ... To what extent does the COVID-19 pandemic data influence stock market values?

- What is the baseline performance for standard statistical models on short term and long term?
- How do different statistical and machine learning models compare given the same data?

- How do models generalize to different time-periods, such as distinctions between different infection waves?
- Which variables in a pandemic affect stock market pricing and to what extent?

RQ2 ...To what extent can machine learning techniques aid in predicting stock market pricing, given a comparison of baseline models and the addition of pandemic data?

As the pandemic influenced a wide array of human behaviours, variables are restricted to the ones mentioned above. Other variables, such as mobility, energy consumption and social contacts have shown large differences during the pandemic, but are outside of the scope of this research.

## 2 Related Work

### 2.1 Data

**2.1.1 Pandemic Data** As the COVID-19 pandemic began to increase in number of cases, world-wide spread and severity, a large part of the academic world focused their efforts on solving issues related to the pandemic or contributing to the cause in general. One of the most influential parties was the Johns Hopkins University's (JHU) with their Coronavirus Research Center (CRC) and their Center for Systems Science and Engineering (CSSE) [11]. They started one of the first and most successful data collections around the COVID-19 pandemic and provided world-wide figures into the number of infections, deceased and (in a later stage) vaccinations, for varying degrees of granularity. In the early stages of the pandemic, the focus was much more regional [13, 35, 37] and built around statistics such as reproductive numbers, health effects and possible treatments. As the effects of COVID-19 on the community became more apparent and the severity of the pandemic increased, governments became concerned on its effect on society and if their could be economic consequences. It became clear that lockdown measures were necessary to reduce the number of infections. The exact measures differed from country to country, as the effectiveness of measures were always a matter of debate. Societal measures ranged from curfews mandating people to stay at home and reduce non-essential traveling, mandatory social distancing assuring people refrain from physical contact [29] or the wearing of face masks in public spaces to reduce the spread of airborne viral particles [12]. Businesses also faced several restrictions, as the hospitality and tourism industries were forced to close, as they could become hot-spots for disease transmission [15]. This in turn also affected the air-travel industry, as people were recommended to stay home and borders were closed off for international travel. And lastly, any other business were often required to provide means for personnel to work remotely from their homes, as offices could also become hot-spots for virus transmission [10]. As these measures varied, especially when each country developed infection peaks at different times, there became a need for COVID-19 related research to have a unified way of charting these measures. The Blavatnik School of Government, part of the University of Oxford, provides a centralized dataset that collects and standardizes the measures taken by countries to reduce the effect of the pandemic [17]. They provide categories such as containment and closure policies, economic policies, health system policies, vaccination policies and miscellaneous policies over which they proposed a stringency index that indicates the severity of the measures applied.

**2.1.2 Environmental Data** As this study revolves around effects on stock market data, environmental factors can not be excluded. Although stock market data are very volatile, their variance can be explained as results from human behaviour. And as humans are affected by a variety of physical and psychological effects, these can be attributed to environmental factors, such as weather [18]. For example, without going into the principals behind causality, one might argue that there is a causal relationship between sunny weather and ice-cream sales, as people tend to buy more ice-cream during sunny (summer) days. This is an example of human behaviour being influenced by environmental factors, a phenomenon also influencing decisions of traders on the stock market. Still, human behaviour on the stock

market can be erratic as people have to deal with uncertainty [34]. Another large factor in stock market data might be stock related news [30, 34, 36, 38]. The difficulty of news articles is that they're not easily quantifiable, unless they have been labeled for the specific use case. Another factor in the research questions is the use of calendar data. As stock markets tend to close during weekends and holidays, this data might aid in identifying seasonal patterns. This research will focus on daily and weekly data, as this is the highest frequency data available for stock markets without the use of paid data providers. Pandemic data is also provided on a daily interval, so this coincides fittingly with the other datasets.

## 2.2 Forecasting models

As the COVID-19 pandemic is a recent phenomenon, studies modelling economic development after the pandemic have been performed but not yet proven [1, 4, 14]. There have been efforts on modelling historic pandemics [6, 22, 31], but most of these studies focus on long-term macroeconomic effects, such as gross domestic product (GDP), unemployment, individual/national income, consumption or inflation, looking at years or decades of development. As economic development itself is a slow process, influence by relatively short-term factors such as the stock market is a non-trivial matter. As modern technology provides an unprecedented amount of data for the pandemic [11, 17], some research has been done on the short-term development and effects of the pandemic [3, 9, 42]. A common practice in the development and selection of forecasting models is to perform exploratory data analysis (EDA) and select the model based on an error metric and/or subjective knowledge of the specific domain. Econometric and statistical models generally differ between these disciplines, as the former looks more at domain knowledge and adjusts the model accordingly to reduce error, while the latter relies on statistical methods to reduce model errors. As stock market data is quite volatile, using either of these disciplines can lead to a successful forecasting model, but the theoretical basis differs. As the discipline of data science and machine learning is in essence an evolution of statistics, its methods rely more on error metrics, even though a lot of gains can be made with problem specific model tuning. This study will look solely at the error metrics of models to determine the effectiveness of the model in an automatic manner. This study looks into an automatic method of fitting a multiple models, with linear regression and ARIMA as statistical models. The Random Forest Regressor (RFR) represents a precursor of the deep learning [27] field, implemented in a standard Python library<sup>3</sup>.

## 3 Methodology

Given the defined research questions and related work to this study, it needs to adhere to the basic steps of forecasting [7, 19]: data 3.1, models 3.3 and evaluation 3.4. To determine the extent of influence of the pandemic on stock market pricing, all factors in the questions are isolated and answered separately. The effect of the data can be determined by using comparisons in correlation scores between features or non-parametric tests. This gives a reliable result for one-on-one relationships between explanatory and response variables, but this becomes a lot more complex in multivariate analysis. This study proposes an alternative solution by comparing forecasting performance using a dynamic model framework where feature importance is compared by applying the mean absolute percentage error (MAPE) on a selection of machine learning models, isolating features between comparisons.

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

To determine the features to be tested, a couple of choices concerning the available data surrounding the pandemic 3.1 have to be made. As the performance of the data is also dependent on the model trained with it, a collection of models will account for linear- and non-linearity. Statistical dependency can be accounted for by comparing model scores of a LR model with an ARIMA model, as the former assumes statistical dependency 3.3 and the latter does not. To evaluate

<sup>3</sup>sklearn: Machine Learning in Python: <https://scikit-learn.org/>

performance of the models and data, it is necessary to methodically isolate different parts of the research question given the MAPE metric.

### 3.1 Data: Information gathering and EDA

To fulfill the data requirements, this research is based on a small set of sources. Calendar data, such as day of the week, weeks, months, years, etc. Its scope will go up to daily data, as that is the finest granularity between all historical data sources. Weather data, as it has small but noticeable effects on human behaviour. This study will focus its efforts on the Dutch market, where the Royal Netherlands Meteorological Institute (KNMI) is the authority on Dutch weather <sup>4</sup>, with The Bilt (Utrecht) as the central point of measurements. Stock market data is collected from the Yahoo Finance API<sup>5</sup>, with the 25 companies representing the Amsterdam Stock Exchange (AEX) index as a representation of the Dutch stock market. As the AEX index is evaluated each quarter, it is assumed as a static composition of companies as of 30th June 2021. The COVID-19 data is provided by the RIVM, providing official information on COVID-19 infections and deaths. The JHU repository for data on vaccinations [11], as currently (before June 2021) this is not officially provided by the RIVM. And the Oxford COVID-19 Government Response Tracker [17] provides a global tracker of government responses to the pandemic, including lockdown, social distancing and business measures.

*3.1.1 Preprocessing of calendar data* Calendar data can be generated and as this study models time-series data, it serves as the connection for all other features. It also has its own characteristics which could aid in distinguishing patterns in the data, such as weekdays, boundaries of weeks, months or years and the distinction between workdays and weekends. During EDA it was concluded that these features are counter-productive for the models, as they are discrete variables but produce artifacts at the boundary between periods that do not represent reality. For instance, week 52 is a later than week 1, except at the transition of two years. This could be resolved by counting from an arbitrary starting point in the data and only counting up, but generally this just produces more features that are confusing to the models. The days of the week are different, as these are more categorical in nature rather than discrete variables. For instance when every Monday is counted incrementally, it would represent the number of weeks rather than every Monday. Therefore its better to one-hot encode weekdays.

*3.1.2 Preprocessing of weather data* Weather data consists of a few basic variables: air humidity and -pressure, wind speed, cloud coverage and conversely hours of sunshine, precipitation and temperature. All these features have their respective minimum and maximum values with the time-slot where that value was measured and an average or total sum depending on the type of variable. As the granularity of the data does not allow for hourly intervals and extreme values might badly influence model training, this study restricts the weather features to averages and totals of the main features mentioned.

*3.1.3 Preprocessing of stock market data* Stock market data consists of two basic variables: pricing and volume. Pricing is generally represented by a opening- and closing price, a high and a low value per day. In stock trading average values tend to be less informative, as it is the extreme differences in value between certain periods that are of value for stock trading. There are numerous metrics that exist within economics and financial theory that claim to be good indicators of the actual value of an asset, but only open, close, high and low values along with volume are agreed upon as being representative indicators. As all other constructions of these indicators are open to speculation, this study treats these values individually as predictor values. A different challenge with stock market data is that it is not always available, as

<sup>4</sup>KNMI - Royal Netherlands Meteorological Institute: <https://www.knmi.nl/home>

<sup>5</sup>Yahoo! Finance's API Python package yfinance: <https://github.com/ranaroussi/yfinance>

a stock market is not always open. Weekends, holidays and other periods might not contain data, so with preprocessing these are indicated as a variable in the dataset. As zero-filling is not representative of these variables, it is assumed that the value of prices or volume remain static during these periods. This has its dangers as the world does not 'switch off' during these periods and speculation on share prices will still occur, without it being reflected in actual trading prices. Although recording the closing days of the stock market might help models detect and predict these periods, it will inevitably introduce noise to the predictions and it might be necessary to move toward a weekly prediction interval rather than a daily interval.

**3.1.4 Preprocessing of pandemic data** COVID-19 data features four variables: infected, deceased and vaccinated people, along with government measures. All these can either be viewed as daily recorded values or cumulative values, with the exception infected people, as the duration of COVID-19 infection varies, generally showing a two week average duration. Data is provided by two sources, the RIVM and the JHU, where the former can be viewed as reliable and the latter can be changeable, as the data is crowd-sourced. Both suffer from the flaw that the recording of a value might differ from the actual value at that time. Infected people might not always know they are infected and their infection status is only recorded if they decide to be officially tested. A similar phenomenon occurs with the number of deceased people, as the cause of death may not always be determined (to a single cause), people are not required to notify the government and the deceased are recorded by this notification date and not time of death. Vaccination data features these same inconsistencies, but also has the deficit of being crowd-sourced from various sources at the start of recording, as the RIVM did not officially record these statistics at the start of this study. Therefore this study will rely on the vetting mechanism that open-source data applies, accepting any errors this may introduce. Some gaps in the data may occur and with cumulative values a 'steady state' is assumed, meaning no change is expected in the state unless recorded. For daily recorded statistics this means data is zero-filled if no value is known. This differs from the approach with stock market data, as vaccinations could be halted due to political or medical re-evaluations.<sup>6</sup> Although it can be assumed infections and deaths maintain relatively stable figures, this study is aware that this data is curated and historically corrected, removing the necessity for gap-filling. All data is prepared in a similar fashion before model training and evaluation. Because it can not be assumed that features are homogeneous, min-max-normalization is applied on the data between all transformations so that each feature is scaled identically without modifying their distribution. As for transforming the distribution of non-normally distributed data, it was concluded from EDA that applying a log-transformation, a second re-scaling along with a Yeo-Johnson transformation, yields the best results for data distribution, while maintaining distribution characteristics for normally distributed data. Finally the data is re-scaled once more before model training so each feature will be weighted similarly. As the response variable this study chose to look at opening prices of stocks, as a single variable allows for easier error metrics. From EDA it was concluded that there is little difference in either opening, closing, high or low pricing values.

## 3.2 Feature selection

As the principal of 'simple is better' generally applies to time-series forecasting, it can not be assumed that all features combined as the full training set is best. To remove these assumptions, recursive feature elimination is implemented during model training based on the mean squared error (MSE) score of a prediction with five fold cross validation. However, this only applies for the LR and RFR models as the used ARIMA model implementation does not allow

<sup>6</sup>The AstraZeneca and Janssen vaccines were both found to cause embolic/thrombotic events in patients and were suspended until safety was reassessed: <https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2021-01/02-COVID-Villafana.pdf>  
<https://www.fda.gov/news-events/press-announcements/joint-cdc-and-fda-statement-johnson-johnson-covid-19-vaccine>

for automatic feature selection.<sup>7</sup> For the ARIMA model a Fourier Featurizer is implemented to fit seasonal as well as non-seasonal features effectively. And lastly our baseline model does not necessitate any feature selection as it disregards all input features other than the last value of the response variable, which results as a constant for each training cycle.

### 3.3 Models: Choosing and fitting models

To accommodate a valid conclusion to the research questions, a variety of models must be considered and compared in their individual performances to the larger research questions. Only then can this study conclude if it is the model that accounted for a gain in performance.

**3.3.1 Baseline model** For our baseline model, it is necessary to ensure the performance is consistent irrespective of the input data. Effectively this means the baseline model should predict a constant value from the last known value in a time-series. Assuming the error of the predicted values is normally distributed, the MAPE metric should be relatively consistent over multiple predictions and inputs. For every time-step  $t$  and feature  $i$ , we define the predictor  $y^{t+1}$  as the last time-step  $t$  of input value  $x$ :

$$\hat{Y}_{T+h|T} = Y_T$$

**3.3.2 Linear regression** As the standard for most econometric prediction models [16], a model based on multiple linear regression (LR) is implemented. As the model expects constant variance, independence of errors and linearity of the data, it will only perform adequately if these conditions are met. With specific transformations applied to the data to improve model fitting, it serves as a good comparison to the other models as its value has been proven in many fields and disciplines.

**3.3.3 ARIMA** As a more advanced regression model, ARIMA has the potential to fit the data quite well. As an autoregressive (AR) model, it does not assume independent data and serves as a good comparison to LR on this feature of the model. The moving average (MA) accounts for the errors in the AR part of the model, improving the model fitting. Still, this ARMA model does not work on non-stationary data, which is where the (I) integrated part of the model comes in. This accounts for the difference in time-steps ( $Y_t + 1 - Y_t$ ) in the data and transforms the series so it does become stationary for the ARMA part to work. The key takeaway being that the three parameters  $p$ ,  $d$  and  $q$ , respectively representing AR, I and MA, need to be tuned for a correct model fitting. This can be performed automatically by performing differencing tests for the  $p$  parameter and iterating step-wise through the  $d$  and  $q$  parameters. This could lead to over-fitting the model, but with multiple samples of the data this study should get representative model performance.

**3.3.4 Random Forest Regressor** As a state-of-the-art (SOTA) model a random forest regressor (RFR) is implemented. It can fit non-linear data, but is a 'black-box' model, so compared to LR and ARIMA its inner workings are not fully intuitive. But where it lacks in explainability it gains in performance. Related literature also points out that RNN/LSTM networks can perform well on stock market time-series data [40], but these networks require quite large amounts of data compared to an RFR model. Therefore it is the best machine learning model we could implement on our data.

### 3.4 Evaluation: Model evaluation methods

To evaluate model performance it is compared between different segments of the research questions using the MAPE metric. In these questions the following components can be isolated: model performance, the time period of the sample, feature performance. Using these cross-sections of the research questions a statistical hypothesis test can be performed

<sup>7</sup>pmdarima: ARIMA estimators for Python <https://alkaline-ml.com/pmdarima/index.html>



on the MAPE metric to draw a valid conclusion. When a model will not significantly outperform the baseline model, this study has to acknowledge that the research question does not have a conclusive answer given the data and models.

#### 4 Results

Model training is performed for all 25 companies in the AEX index, with five randomly sampled time periods for each of the models. This ensures there is enough data to answer the research questions. Although all parameters are kept consistent between model training, some errors do inevitably creep into the results necessitating result selection. MAPE scores equal or above 1.0, i.e. a +100% error rate, are discarded, as these are obvious errors where the models did not perform as expected. With identical model training cycles, random samples seeded with identical values and identical training data, identical results were expected, but the reality is that results have to be manually selected to ensure quality. To assure each model is represented equally, 50 results were sampled from each model and plot our results.

##### 4.1 RQ1: The influence of pandemic features on stock pricing

To fully answer the extent of this research question, first the performance of our models must be assessed in comparison to the baseline. During EDA and the results collection a sample of these results was taken from the training cycles and pair-plotted for quick comparison 2. During EDA this study experimented with different sized ratios of training and testing data and determined a ten to one ratio of training vs. testing data should be sufficient. The target prediction length was seven days of values for the daily interval and three weeks of values for weekly interval. From the pair-plot model fittings some prediction errors can clearly be seen, but these are expected given the variance in the training data. It can also clearly be seen that the baseline model (labeled as "dummy") functions as expected, ignoring all input features except the last known value of the response variable.

*4.1.1 Model performance* Given access to all features while implementing automatic feature selection, model performance on the MAPE metric is compared for daily and weekly data (see Figure 1b). Weekly data generally shows higher error scores, especially for the ARIMA model. An unusual result is that none of the models show any noteworthy performance differences when compared to the baseline model. The baseline should average out to a small error score, but the expectation was that with sufficient training data and feature selection, the other models should've outperformed the baseline. In our sample fits a possible cause can be seen, as some results show extreme errors in predictions.

*4.1.2 Time period* As a difference between daily and weekly data was established for predictions, the results can also distinguish the used time period for training and testing. As the pandemic started at the end of 2019, it can be seen that the error scores increase in 2020 and consequently decrease in 2021. This could support the hypothesis that the pandemic caused an upset in 2020 and was restoring in 2021. An indication that this is not just an artifact of the pandemic data, there is little difference to be seen in the baseline model between 2019 and 2021, even though the pandemic is still widespread in the latter year. The same behaviour can also be seen on a weekly basis, indicating that this also generalizes over longer prediction periods.

*4.1.3 Feature performance* To isolate the impact of data features on our error metric, the same predictions were ran given different selections of features. The results above were given full access to the data, but the following variables were isolated: Weather features, COVID-19- infections, -deaths, -partial/-full vaccinations, stringency of regulations [17] and no additional features. Looking at a comparison of the MAPE scores between weather data only and the full dataset, there is no notable difference in performance (see Figure 1c). This could indicate that either the weather features are always selected during features selection, or none of the features are of any value to the models. The latter could be



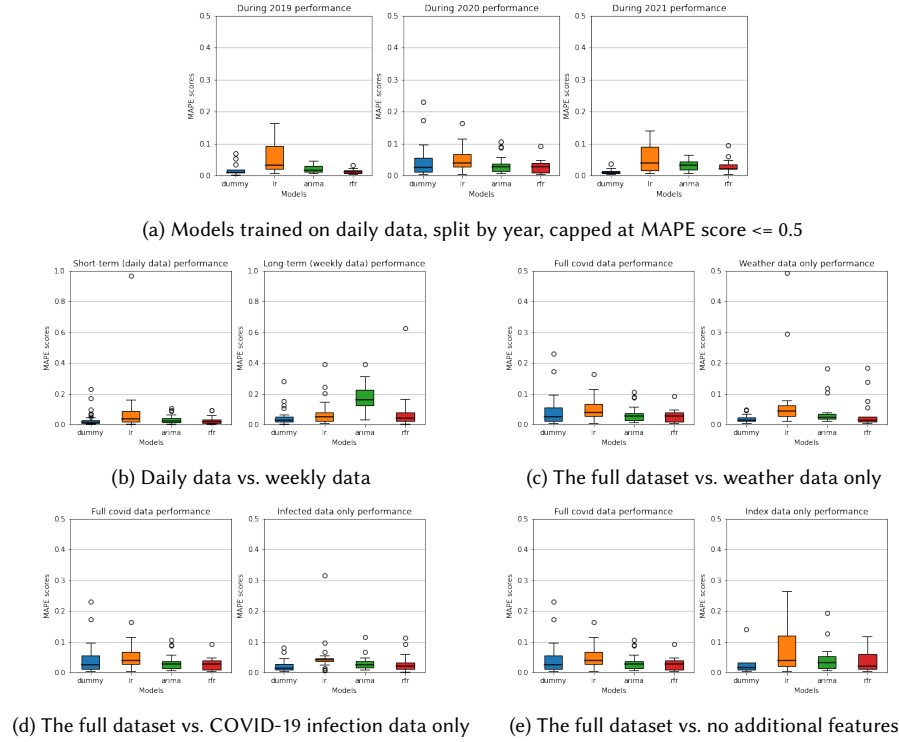


Fig. 1. Comparisons of model scores

true given none of the models show a notable performance increase from the baseline, but the feature of infections shows that pandemic data does have some influence (see Figure 1d), at least for the ARIMA and RFR models. The same can be said for deaths, vaccination data and the stringency index, where the ARIMA and RFR model (almost) outperform the baseline model. Although the differences are subtle, they are noticeable, especially compared to LR which shows consistent high error scores compared to the baseline. Weather features are the richest of all data and features can be seen as individual features. The only test left is one where the models are given no additional input features and only the response variable defines the output of the models. As a consequence all models increase in performance compared to the baseline as the number of data features decrease, posing the question whether any additional features is helpful at all (see Figure 1e).

Features	Model	MAPE	Baseline compare	
			t-stat	p
All	ARIMA	0.0302	-0.3735	0.7096
	RFR	0.0234	0.6711	0.5037
None	ARIMA	0.0315	-1.7365	0.0856
	RFR	0.0242	-0.5310	0.5966

Table 1. Model MAPE score comparison to baseline, for all dataset features and no additional features

Model	Features: All		None		Comparison	
	MAPE		MAPE		t-stat	p
ARIMA	0.0341		0.0434		-0.8300	0.4114
RFR	0.0300		0.0408		-0.7865	0.4380
Baseline	0.0475		0.253		1.3612	0.1824

Table 2. Data feature MAPE score comparison for all the ARIMA, RFR and baseline models

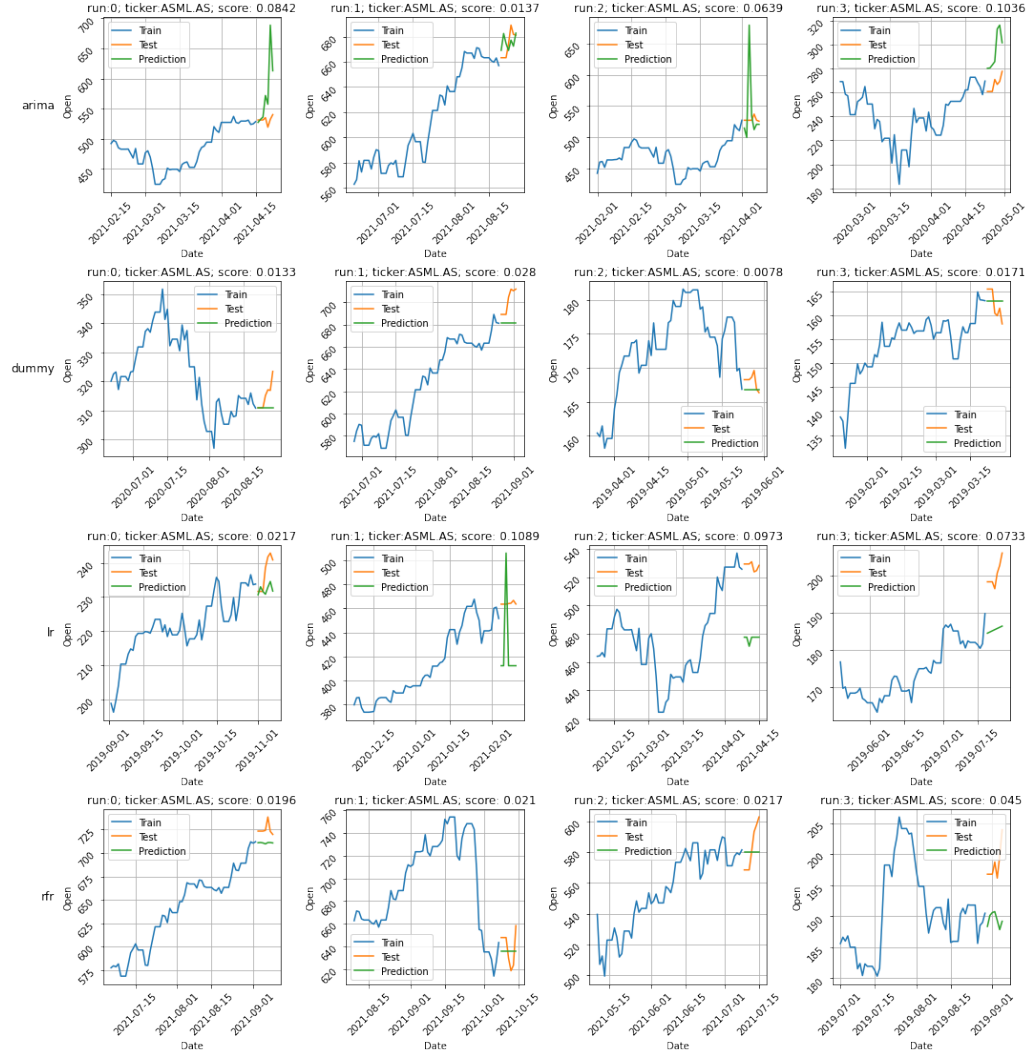


Fig. 2. Side-by-side model fittings and resulting daily predictions for ASML

#### 4.2 RQ2: Performance of machine learning techniques in stock market prediction

As the RFR model is considered as part of the machine learning techniques, it should be compared to the baseline and evaluated on its relative performance. The differences in performance are minute and necessitate an insight into the significance tests to give a reliable measure of difference in performance. For the sake of simplicity, this study has only done this comparison for the daily data, as the weekly data shows unreliable performance. Given a confidence level of 95% ( $\alpha = 0.05$ ), a T-test on the MAPE scores for the full dataset and the dataset containing no additional features results in the following table 1. With these results it can be said that there is no significant performance difference in using machine learning techniques such as RFR for stock market prediction. The same can be said for the ARIMA

model, with the notion that ARIMA does seem to show better performance with reduced features. As each row in table 1 refers to an individual T-test, the dimensions can also be flipped to test whether the features in the data influence the models in table 2. This research shows there is no significant difference in the mean MAPE error scores between all or no additional features in the data, for none of the models.

## 5 Discussion

When viewing the results compared to the baseline model, better performance was to be expected from the ARIMA and RFR models. These models should have been able to capture the variance of the non-linear time-series, when properly tuned and fitted. The additional features presented by the COVID-19 pandemic do seem to sporadically add accuracy, however the significance tests show that these are most likely a random or coincidental occurrence rather than a relationship in the data. Linear regression actually consistently performed worse than the baseline, which indicates that the data is non-linear. In light of related work, this could be due to the methodology of automated model fitting and tuning. Previous studies of forecasting models generally approach these tasks as an iterative process, where a single model is developed on very specific data with large amounts of manual tuning. In these results large parts of the test scores already needed to be discarded due to obvious fitting errors caused by noise in the data, preprocessing or failures of the code. Where EDA showed there was potential in the data and models, when implemented on a larger scale it shows that without manual intervention, these models do not generalize. This was somewhat expected, as other related work already showed that stock market prediction relies on domain knowledge or technical analysis. As the COVID-19 pandemic was new information at the end of 2019, there was a consequential dip in oil futures and related assets, which set the precedent of the pandemic becoming old information. Also it was clear that this dip in oil futures was a result of compounding events rather than just an effect of the pandemic. Another possible explanation for the insignificant results is that there is no causal relationship between our features. Although publicly available data could've influenced the behavioural psychology of traders, this is a relationship primarily determined by the traders' psychology and only has a secondary relation to this study's data. As businesses moved to working remotely from their employees homes, lockdown effects were overcome for most industries and the effect on the economy was consequently negated. Along with quick developments in vaccines, this could've led traders to the conclusion that the pandemic was of little importance after the first wave of infections. These developments were mostly covered in news stories, which are difficult to quantify, but are generally the main source of information for traders. Any relationship with pandemic information is probably only visible on long term analysis, where events on the stock market accumulate over time and variance decreases. Although technology has come a long way, and SOTA machine learning models have shown promise in varying fields of expertise, it is clear that these models do not learn 'automatically'. A lot of expertise is needed to properly fit and tune a model to very specific data and only then can it be considered as a viable solution for the specific problem in mind. Generalization of models is acquired by training with large amounts of data, but only when the model in question can do the same on very specific data. For instance, models that perform character recognition in images can generalize to other tasks, such as object classification, however this comes with complex engineering of the models and large amounts of prepared data. In this case, the EDA showed promise on prediction models combined with pandemic data. And with large amounts of fitting, tuning and engineering of a single model it might still be possible to achieve a significant result. Still, generalization of such models comes with large volumes of data and preparation, and only when a precedent is set research can work towards a more general solution. In this research experiments were performed with different models, such as Xgboost, but from sample fittings it was clear it would not outperform RFR without additional data or proper manual tuning. When reviewing our selection of control, or baseline model and other prediction models, the research also admits that there might have been more appropriate

models available at the time of writing. If this study would be revisited, one would do right to reconsider the model selection along with possible different error metrics.

## 6 Conclusions

Summarizing the results and discussion, this study can not confirm there is any added benefit of including COVID-19 pandemic data as a explanatory variable for the purpose of stock market predictions. Considering the individual contributions of environmental factors, such as weather, this study can not establish a significant relationship. The same is applicable for any of the individual COVID-19 pandemic features, where it is concluded that any relation is either coincidental or inconclusive as no significant performance difference can be measured. This does not mean there is no added benefit of applying SOTA machine learning techniques or models to improve prediction accuracy, as indicated by related work. This study could not establish a significant relationship between pandemic data and stock market values with ARIMA or RFR models, especially compared to the performance of the baseline model. Forming a basis for more complex models, such as LSTM/RNN, is therefore unsustainable as this study can not provide the data these models require to achieve significant results. There remains a possibility for pre-trained models or transfer learning to improve results, however this would require further research to verify. Financial theories debate on whether an algorithm can consistently beat (the majority of) the stock market and as a consequence one can imagine these models are not within the grasp of the general public. Within the scope of this research it is concluded that it is not possible to achieve such a result and leave this as a springboard for further research. Although this study did not establish a statistical significant result, there is anecdotal evidence of an impact during the first infection wave of the COVID-19 pandemic. With purpose built models and data there is a possibility of constructing accurate time-series forecasts, but this comes at the cost of generalization.

## References

- [1] Richard Baldwin and B Weder Di Mauro. 2020. *Economics in the time of COVID-19: A new eBook*. VOX CEPR Policy Portal, 2nd Floor, 33 Great Sutton Street, London EC1V 0DX, United Kingdom.
- [2] Marianne Baxter and Robert G King. 1999. Measuring business cycles: approximate band-pass filters for economic time series. *Review of economics and statistics* 81, 4 (1999), 575–593.
- [3] Tiziana Carpi, Airo Hino, Stefano Maria Iacus, and Giuseppe Porro. 2021. Twitter Subjective Well-Being Indicator During COVID-19 Pandemic: A Cross-Country Comparative Study. arXiv:2101.07695 [econ.GN]
- [4] Alexander Chudik, Kamiar Mohaddes, M Hashem Pesaran, Mehdi Raissi, and Alessandro Rebucci. 2020. Economic consequences of Covid-19: A counterfactual multi-country analysis.
- [5] Shaen Corbet, John W. Goodell, and Samet Günay. 2020. Co-movements and spillovers of oil and renewable firms under extreme conditions: New evidence from negative WTI prices during COVID-19. *Energy Economics* 92 (2020), 104978. <https://doi.org/10.1016/j.eneco.2020.104978>
- [6] Sergio Correia, Stephan Luck, and Emil Verner. 2020. Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu. *SSRN Electronic Journal* (Mar 2020). <https://doi.org/10.2139/ssrn.3561560>
- [7] Andrey Davydenko, Cuong Sai, and Maxim Shcherbakov. 2021. *Forecast Evaluation Techniques for I4.0 Systems*. Springer International Publishing, Cham, 79–102. [https://doi.org/10.1007/978-3-030-66077-2\\_7](https://doi.org/10.1007/978-3-030-66077-2_7)
- [8] Jean Charles Leonard Simonde De Sismondi. 1827. *Nouveaux principes d'économie politique, ou de la richesse dans ses rapports avec la population*. Vol. 2.
- [9] Pragyan Deb, Davide Furceri, Jonathan D. Ostry, and Nour Tawk. 2021. The Economic Effects of COVID-19 Containment Measures. *Open Economies Review* (14 Dec 2021). <https://doi.org/10.1007/s11079-021-09638-2>
- [10] Jonathan I Dingel and Brent Neiman. 2020. How many jobs can be done at home? *Journal of Public Economics* 189 (2020), 104235. <https://doi.org/10.1016/j.jpubeco.2020.104235>
- [11] Ensheng Dong, Hongru Du, and Lauren Gardner. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* 20, 5 (2020), 533–534.
- [12] Steffen E. Eikenberry, Marina Mancuso, Enahoro Iboi, Tin Phan, Keenan Eikenberry, Yang Kuang, Eric Kostelich, and Abba B. Gumel. 2020. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infectious Disease Modelling* 5 (2020), 293–308. <https://doi.org/10.1016/j.idm.2020.04.001>
- [13] Chaolin Huang et. al. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 395, 10223 (2020), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)

- [14] Nuno Fernandes. 2020. Economic Effects of Coronavirus Outbreak (COVID-19) on the World Economy. *SSRN Electronic Journal* (Mar 2020). <https://doi.org/10.2139/ssrn.3557504>
- [15] Stefan Gössling, Daniel Scott, and C Michael Hall. 2020. Pandemics, tourism and global change: a rapid assessment of COVID-19. *Journal of Sustainable Tourism* 29, 1 (2020), 1–20. <https://doi.org/10.1080/09669582.2020.1758708> arXiv:<https://doi.org/10.1080/09669582.2020.1758708>
- [16] William H Greene. 2003. *Econometric analysis*. Pearson Education India.
- [17] Thomas Hale, Anna Petherick, Toby Phillips, and Samuel Webster. 2020. Variation in government responses to COVID-19. *Blavatnik school of government working paper* 31 (2020), 2020–11.
- [18] David Hirshleifer and Tyler Shumway. 2003. Good Day Sunshine: Stock Returns and the Weather. *The Journal of Finance* 58, 3 (2003), 1009–1032. <https://doi.org/10.1111/1540-6261.00556> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/1540-6261.00556>
- [19] Rob J Hyndman and George Athanasopoulos. 2018. *Forecasting: principles and practice*. OTexts.
- [20] Trent Jacobs et al. 2020. OPEC+ Moves To End Price War With 9.7 Million B/D Cut. *Journal of Petroleum Technology* 72, 05 (2020), 38–40.
- [21] A.K. Jain, Jianchang Mao, and K.M. Mohiuddin. 1996. Artificial neural networks: a tutorial. *Computer* 29, 3 (1996), 31–44. <https://doi.org/10.1109/2.485891>
- [22] Óscar Jordà, Sanjay R Singh, and Alan M Taylor. 2020. *Longer-run economic consequences of pandemics*. Technical Report. National Bureau of economic research.
- [23] Robert K Kaufmann. 2011. The role of market fundamentals and speculation in recent price changes for crude oil. *Energy policy* 39, 1 (2011), 105–115.
- [24] John Maynard Keynes. 2018. *The general theory of employment, interest, and money*. Springer.
- [25] Charles Poor Kindleberger et al. 1986. *The world in depression, 1929-1939*. Vol. 4. Univ of California Press.
- [26] Finn E Kydland and Edward C Prescott. 1982. Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society* (1982), 1345–1370.
- [27] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.
- [28] Baruch Lev and S Ramu Thiagarajan. 1993. Fundamental information analysis. *Journal of Accounting research* 31, 2 (1993), 190–215.
- [29] Joseph A Lewnard and Nathan C Lo. 2020. Scientific and ethical basis for social-distancing interventions against COVID-19. *The Lancet infectious diseases* 20, 6 (2020), 631–633. [https://doi.org/10.1016/S1473-3099\(20\)30190-0](https://doi.org/10.1016/S1473-3099(20)30190-0)
- [30] Chanwit Onsumran, Sotarat Thammaboosadee, and Supaporn Kiattisin. 2015. *Gold price volatility prediction by text mining in economic indicators news*. Ph.D. Dissertation. Mahidol University.
- [31] Michael T Osterholm. 2017. *Preparing for the next pandemic*. Routledge.
- [32] Edgar E Peters. 1996. *Chaos and order in the capital markets: a new view of cycles, prices, and market volatility*. John Wiley & Sons.
- [33] Suraphan Thawornwong and David Enke. 2004. Forecasting stock returns with artificial neural networks. *Neural Networks in Business Forecasting* (2004), 47–79.
- [34] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (1974), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124> arXiv:<https://www.science.org/doi/pdf/10.1126/science.185.4157.1124>
- [35] Thirumalaisamy P Velavan and Christian G Meyer. 2020. The COVID-19 epidemic. *Tropical medicine & international health* 25, 3 (2020), 278.
- [36] Pietro Veronesi. 2015. Stock market overreactions to bad news in good times: a rational expectations equilibrium model. *The Review of Financial Studies* 12, 5 (06 2015), 975–1007. <https://doi.org/10.1093/rfs/12.5.975> arXiv:<https://academic.oup.com/rfs/article-pdf/12/5/975/24420052/120975.pdf>
- [37] Chen Wang, Peter W Horby, Frederick G Hayden, and George F Gao. 2020. A novel coronavirus outbreak of global health concern. *The lancet* 395, 10223 (2020), 470–473.
- [38] Jinxiao Wang, Jiaxin Shi, Dexin Han, and Xiaoyu Zhao. 2019. Internet Financial News and Prediction for Stock Market: An Empirical Analysis of Tourism Plate Based on LDA and SVM. *Journal of Advances in Information Technology Vol* 10, 3 (2019).
- [39] Eugene N. White. 1990. The Stock Market Boom and Crash of 1929 Revisited. *Journal of Economic Perspectives* 4, 2 (June 1990), 67–83. <https://doi.org/10.1257/jep.4.2.67>
- [40] Lean Yu, Shouyang Wang, and Kin Keung Lai. 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics* 30, 5 (2008), 2623–2635. <https://doi.org/10.1016/j.eneco.2008.05.003>
- [41] Yudong Zhang and Lenan Wu. 2009. Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert systems with applications* 36, 5 (2009), 8849–8854.
- [42] Shi Zhao, Qianyin Lin, Jinjun Ran, Salihu S Musa, Guangpu Yang, Weiming Wang, Yijun Lou, Daozhou Gao, Lin Yang, Daihai He, et al. 2020. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International journal of infectious diseases* 92 (2020), 214–217.