COVID-19 pandemic influences on stock market development

Maarten Peters

ABSTRACT

During the COVID-19 pandemic [VM20] 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 [Jac+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) [CGG20]. As the WTI serves as a benchmark for economic development [Kau11], it is an interesting candidate to use for price forecasting [YWL08]. The pandemic provides a unique perspective, as it introduces a new set of variables [DDG20; Hal+20], such as COVID-19 infections, deaths, vaccinations and government measures [1], that might aid in predicting economic development [Chu+20; BD20; Fer20]. 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. In this study we will attempt 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. If the model can not explain part of the variance with the added pandemic data, we can conclude that it does not play a direct role in price development. To determine the outcome of our research question, we will compare the value of our data over a set of base line (non-)linear models and machine learning models.

1 PERSONAL DETAILS

Student maarten.peters@gmail.com
Examiner 1 c.m.rodriguezrivero@uva.nl
Examiner 2 m.j.marx@uva.nl
Supervisor jpucheta@unc.edu.ar
GitHub maartenpeters/thesis-uva-dynamic-modelling

2 RESEARCH QUESTION

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 [VM20]. In this thesis we will look into how the 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 [CGG20]. Although the relationship between stock markets and economies is complex, there is a relation between stock market development and economic development, demonstrated by the 2008 stock market crash initiated by housing prices. A model that could take pandemic data into account and forecast stock market development, might be able to predict a stock market crash. If not, it could at least explain the effect of the COVID-19 pandemic on stock market pricing. With that, we propose the following research questions:

RQ1 …To what extent does the COVID-19 pandemic data influence stock market pricing?
- What is the baseline performance for standard statistical models on short term and long term?
- How does a machine learning model compare to the baseline?
- How do models generalize to different regions, time-periods and other pandemics?
- Which variables in a pandemic affect stock market pricing and to what extent? Variables such as ...
  - Pandemic (COVID-19) data incl.
    - Infections (provided by the RIVM and JHU)
    - Deaths (provided by the RIVM and JHU)
    - Vaccinations (provided by the JHU)
    - Government restrictions, such as lockdown measures (provided by the OxCGRT)
  - Environment factors incl.
    - Calendar data
    - Weather
    - Country/region

RQ2 …Can machine learning techniques aid in predicting stock market pricing?

3 RELATED LITERATURE

There have been multiple studies of the COVID-19 pandemic and its relation to economics [Chu+20; BD20; Fer20], as well as historic pandemics and their effects [Ost17; CLV18; JST20]. Most of the cited research focuses 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. This gives economists/politicians some insight into what they can prepare for in future years, but provides little value for short-term, such days, weeks or months, of decision making. As modern technology along with the COVID-19 pandemic provides a unprecedented amount of data [DDG20; Hal+20], some research has been done on the short-term development and effects of the pandemic [Zha+20; Deb+20; Car+21]. Where economic development is mostly affected on the short-term by pandemics, is businesses and stock market pricing [SAY20; Mha20]. The WTI, a benchmark for crude oil pricing, dropped into the negative values as a result of the pandemic and competition between Saudi-Arabia and Russia over oil pricing [Jac+20]. This served as a precursor of a global stock market crash, where the exact causes are still debated [CGG20], but it has served as a benchmark to market value [Kau11]. Stock market crashes are known for causing recessions, so predicting variance in crude oil pricing or stock market values could give an indication in future economic development. There has been previous research

[1] 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).

on predicting crude oil pricing using machine learning [YWL08], so the question remains whether pandemic information could aid in this prediction.

4 METHODOLOGY

After defining our research questions, we need to adhere to the basic steps of forecasting [HA18; DSS21]:

Data	: Information gathering and exploratory data analysis (EDA)
Models	: Choosing and fitting models
Evaluation	: Model evaluation methods

4.1 Data: Information gathering and exploratory data

To fulfill our data need, we will base our research on a small set of sources:

- Calendar data, such as day of the week, weeks, months, years, etc.. We will scope our research up to daily data, as that is the highest amount of detail available along all historical data sources.
- Weather data, as it generally has small but noticeable effects on human behaviour. As a classic example, people generally buy more ice-cream during warm weather, as a way of cooling down.
- Stock market data, collected from the Yahoo Finance API [Aro17], for values such as crude oil pricing, index funds and large companies.
- COVID-19 data, provided by ...
  - the Dutch institute of national well-being and environment (RIVM), providing official information on COVID-19 infections and deaths in the Netherlands.
  - the Johns Hopkins University repository for COVID-19 data [DDG20], a global dataset providing insight into infections and deaths, but also vaccinations, currently not provided by the Dutch authorities.
  - the Oxford COVID-19 Government Response Tracker [Hal+20], a global tracker of government responses to the COVID-19 pandemic, including lockdown, social distancing and business measures.

As we are dealing with stock market data, which is notoriously volatile, there are multiple data sources that could aid predictions. Speculation on news publications, especially related to business development of publicly traded stocks, is fairly common in stock exchanges. This data is highly unstructured and requires a lot of manipulation for it to be of value to our models. Therefore we do not include this data, as it complicates model development heavily.

4.2 Models: Choosing and fitting models

To accommodate a valid conclusion to our research question, we must consider a variety of models and compare their individual performances to the larger research questions. Only then can we conclude if it was the model that accounted for a gain in performance.

We will answer our research questions with a set of baseline models. Such a model should be explainable and be comparable in their error evaluation to other models. For instance, a simple- or multi-variable- linear regression model should output a prediction in the same manner as a neural network, such that they can be compared by their shared error function, for instance (symmetric) mean absolute percentage error (sMAPE).

For our machine learning models we will look into related research [YWL08] to develop an optimal model. Because supervised machine learning models are a function of their model and data, we have to carefully consider its parameters. We can use a grid-search algorithm to determine the right parameters and use feature selection for our data.

4.3 Evaluation: Model evaluation methods

To evaluate model performance, we have to define an error metric. Our models will output a predicted response variable based on explanatory variables. We will consider our models prediction versus the known outcome, using sMAPE as our error function. But we can also compare our model with pandemic data to a similar model without this data. This would quantify the value of this new information and assess whether it assists in prediction accuracy. We can use statistical testing on the mean error of both models to assess their significance.

To measure the extent of influence of pandemic data on stock market pricing, we will look at Pearson correlation scores to determine the statistical significance of a pair of explanatory and response variables. For multivariate statistics we can apply ANOVA to perform statistical tests, which allows us to test for categorical for variables.

To test for generalization we will perform the same statistical tests between different regions and time windows, comparing MSRE over n different model fittings.

For feature selection we can also utilize Pearson correlation, assuming normality. The features still might need to be transformed or normalized to ensure an optimal result.

5 RISK ASSESSMENT

A regular thesis project follows a twelve week schedule, assuming preparation with exploratory data analysis and model experimentation, containing model development and training, and gathering results along with writing the paper itself. Although there are numerous data sources left unused, our current selection is reasonably well structured and is rich in features for a regression study. The data regarding government measures [Hal+20] for the pandemic can be regarded as the source for most news articles during the pandemic and is well structured, so it should suffice compared to raw news articles.

A possible backup plan could be to restrict the number of features and models to include in our research. This could help in a time crunch and should not affect the validity of the research as long as the methodology is left untouched. If for some reason the methodology is too time-consuming then we could reduce the number of model fitting sessions, but this should be the last resort as it could influence the power of the research.

6 PROJECT PLAN

Given the circumstances of this thesis starting outside the regular schedule and being a part-time student, we're planning to use the

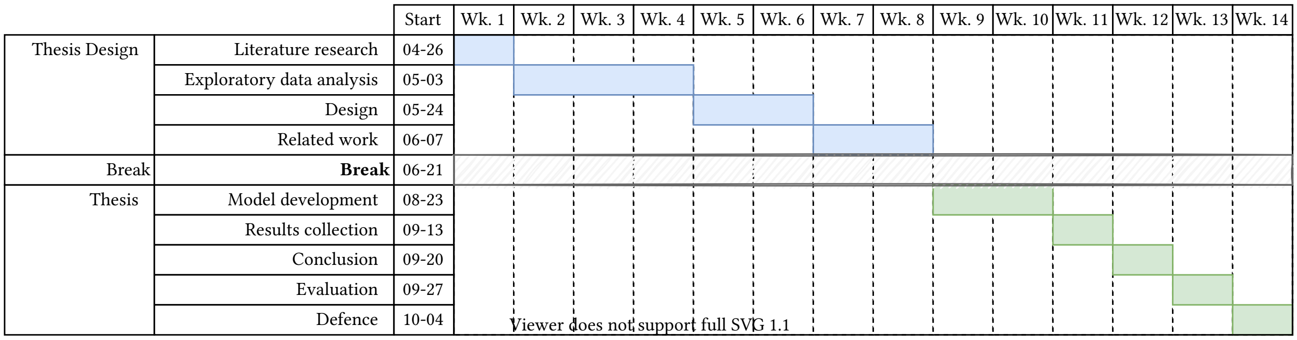

Figure 1: The gantt planning for the thesis

full twelve weeks, with a possible extension to fourteen weeks. As this thesis design had to be written within that twelve week period, we have to account for this delay (see Figure 1). As an update on this document I've communicated with my guiding supervisor and head of the master track and shifted my final weeks of planning, because of personal issues. The intended end of my thesis would be the end September 2021.

Week	Start Date	Achievement
Week 1	2021-04-26	Literature research
Week 2	2021-05-03	Exploratory data analysis
Week 3	2021-05-10	Exploratory data analysis
Week 4	2021-05-17	Exploratory data analysis
Week 5	2021-05-24	Thesis design
Week 6	2021-05-31	Thesis design
Week 7	2021-06-07	Thesis related work
Week 8	2021-06-14	Thesis related work
Week -	2021-06-21	Break
Week 9	2021-08-23	Model development
Week 10	2021-09-06	Model development
Week 11	2021-09-13	Results collection
Week 12	2021-09-20	Thesis conclusion
Week 13	2021-09-27	Thesis evaluation & rework
Week 14	2021-10-04	Thesis defence

REFERENCES

[CLV18] Sergio Correia, Stephan Luck, and Emil Verner. "Pandemics depress the economy, public health interventions do not: Evidence from the 1918 flu". In: *Public Health Interventions do not: Evidence from the* (1918).

[YWL08] Lean Yu, Shouyang Wang, and Kin Keung Lai. "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm". In: *Energy Economics* 30.5 (2008), pp. 2623–2635. ISSN: 0140-9883. DOI: https://doi.org/10.1016/j.eneco.2008.05.003. URL: https://www.sciencedirect.com/science/article/pii/S0140988308000765.

[Kau11] Robert K Kaufmann. "The role of market fundamentals and speculation in recent price changes for crude oil". In: *Energy policy* 39.1 (2011), pp. 105–115.

[Aro17] Ran Aroussi. *ranaroussi/yfinance*. Mar. 2017. URL: https://github.com/ranaroussi/yfinance.

[Ost17] Michael T Osterholm. *Preparing for the next pandemic*. Routledge, 2017.

[HA18] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[BD20] Richard Baldwin and B Weder Di Mauro. "Economics in the time of COVID-19: A new eBook". In: *VOX CEPR Policy Portal* (2020).

[Chu+20] Alexander Chudik et al. "Economic consequences of Covid-19: A counterfactual multi-country analysis". In: *VoxEU. Org* (2020).

[CGG20] Shaen Corbet, John W. Goodell, and Samet Günay. "Co-movements and spillovers of oil and renewable firms under extreme conditions: New evidence from negative WTI prices during COVID-19". In: *Energy Economics* 92 (2020), p. 104978. ISSN: 0140-9883. DOI: https://doi.org/10.1016/j.eneco.2020.104978. URL: https://www.sciencedirect.com/science/article/pii/S0140988320303182.

[Deb+20] Pragyan Deb et al. "The economic effects of Covid-19 containment measures". In: (2020).

[DDG20] Ensheng Dong, Hongru Du, and Lauren Gardner. "An interactive web-based dashboard to track COVID-19 in real time". In: *The Lancet infectious diseases* 20.5 (2020), pp. 533–534.

[Fer20] Nuno Fernandes. "Economic effects of coronavirus outbreak (COVID-19) on the world economy". In: *Available at SSRN 3557504* (2020).

[Hal+20] Thomas Hale et al. "Variation in government responses to COVID-19". In: *Blavatnik school of government working paper* 31 (2020), pp. 2020–11.

[Jac+20] Trent Jacobs et al. "OPEC+ Moves To End Price War With 9.7 Million B/D Cut". In: *Journal of Petroleum Technology* 72.05 (2020), pp. 38–40.

[JST20] Òscar Jordà, Sanjay R Singh, and Alan M Taylor. *Longer-run economic consequences of pandemics*. Tech. rep. National Bureau of economic research, 2020.

[Mha20] Majdouline Mhalla. "The impact of novel coronavirus (COVID-19) on the global oil and aviation markets". In: *Journal of Asian Scientific Research* 10.2 (2020), p. 96.

[SAY20] Arshian Sharif, Chaker Aloui, and Larisa Yarovaya. "COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: Fresh evidence from the wavelet-based approach". In: *International Review of Financial Analysis* 70 (2020), p. 101496. ISSN: 1057-5219. DOI: https://doi.org/10.1016/j.irfa.2020.101496. URL: https://www.sciencedirect.com/science/article/pii/S105752192030140X.

[VM20] Thirumalaisamy P Velavan and Christian G Meyer. "The COVID-19 epidemic". In: *Tropical medicine & international health* 25.3 (2020), p. 278.

[Zha+20] Shi Zhao et al. "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". In: *International journal of infectious diseases* 92 (2020), pp. 214–217.

[Car+21] Tiziana Carpi et al. "Twitter Subjective Well-Being Indicator During COVID-19 Pandemic: A Cross-Country Comparative Study". In: *arXiv preprint arXiv:2101.07695* (2021).

[DSS21] Andrey Davydenko, Cuong Sai, and Maxim Shcherbakov. "Forecast Evaluation Techniques for I4. 0 Systems". In: *Cyber-Physical Systems: Modelling and Intelligent Control* (2021), pp. 79–102.