**Thomas Parmentier**

Zypressenstrasse, 49

8004 Zürich

Thomas.parmentierlux@gmail.com

Stefan

XXX

XXX

XXX

**Data Science Project**

**The impact of macroeconomic variables on the Nasdaq composite index**

**Conceptual Design Report**

**31 October 2021**

# Abstract

The dynamic relationship among macroeconomic indicators such as gross domestic product, interest rate, unemployment, and inflation are crucial to understand stock market performance. Any unexpected or abrupt changes in economic indicators can lead to high market volatility and shift in market behavior, any actor of financial market needs to understand how they interact with each other. Several studies have been conducted to determine the relationship between the macroeconomic variable and stock prices in the past, with mixed results. The objective of this study is to discover the impact of changes in macroeconomic variables on the Nasdaq composite index. For this purpose, we have computed the OLS (Ordinary Least square) and Bidirectional Granger causality test.

# Table of Contents

[Abstract 0](#_Toc85970637)

[Table of Contents 1](#_Toc85970638)

[1 Project Objectives 2](#_Toc85970639)

[2 Methods 3](#_Toc85970640)

[3 Data 4](#_Toc85970641)

[4 Metadata 7](#_Toc85970642)

[5 Data Quality 9](#_Toc85970643)

[6 Data Flow 9](#_Toc85970644)

[7 Data Model 10](#_Toc85970645)

[8 Risks 11](#_Toc85970646)

[9 Preliminary Studies 12](#_Toc85970647)

[10 Conclusions 12](#_Toc85970648)

[Appendix A 12](#_Toc85970649)

[References and Bibliography 13](#_Toc85970650)

# 

# 1 Project Objectives

The objective of this report is to test the relationship between four macroeconomic variables, Gross Domestic Product, Inflation, Unemployment rate and interest rate, and the Nasdaq composite index. Stock market participants are looking for “trading signal” to define if they must buy or sell the markets, economic indicators are amongst the most famous signals market participants are using. The strategy based on macroeconomics has even become a whole industry amongst Hedge Fund. Global Macro Hedge Funds aims to actively managed funds from markets move caused by economic indicators and political events, this kind of strategy manages around 800 billion USD and has made famous some Hedge Funds managers like George Soros or Steve Cohen [1].

Literature has widely studied the impact of macroeconomic variables on broader stock markets such has the S&P 500 or Dow Jones, but relatively few looks specifically at the Nasdaq composite index, which focus on technology stocks in the US, which have emerged as the big winners of the COVID-19 crisis [2]. Overall, studies output has been mixed on the impact and the strength of economic variables and financial markets. For example, even it is commonly admitted that inflation is negative for stock markets, as demonstrated in 1979 by Fama and Schwert who found that stock returns were negatively correlated with the expected component of inflation rate [3], other researchers like Gjerde and Saettern [4], in 1999 found that inflation might not have a direct impact on stock performance. The same discussion can be made on other variables such as the GDP, interest rate level and unemployment rate. In 2021 a study [5] found that there is no ‎causal relation between macro factors and stock return.

With this project, our objectives are to look through the four following independent variables, which will be described in more details in the section 3.

**We will answer the following question research: Can we see a meaningful relationship between the four independent variables and how the independent variables affect the Nasdaq composite index.** To achieve this task, we will perform the following tasks:

* Extract, clean and combine information of time series on macro-variables together with stock market time series
* Visualize the data and run some inferential statistics to understand our dataset
* Look at the correlation between the independent variables and dependent variable
* Run an Augmented Dick Fuller (ADF) test to check for non-stationarity
* Run two econometric analyses: OLS and Granger causality test

# 2 Methods

**Database:** We used the Saint Louis Fred database (FRED for Federal Reserve Economic Data) to extract macroeconomic data. It is one of the most comprehensive and trustful sources of economic indicator in the US, this database is widely used and reported in the media and looked by financial market participants closely every time new data is released. The database is also publicly available. From the FRED we have extracted time series for our four independent variables, which will be described in the section 3 Date (GDP, CPI (consumer price index), US interest rates and US unemployment rate.

For our dependent variable, we used data from the Nasdaq website, which give the daily time serie of the price of the Nasdaq composite index.

**Tools:** We use Python and the Jupyter notebook environment to write a program with our dataset. The program will import, clean, consolidate, arrange, and merge the data, allowing for statistical and visualization analyses. The data are downloaded in CSV and then loaded into Jupyter notebook.

**Liabraries:** We have used numerous librairies to perform our analysis, here we will list the most importants.

* **Pandas** is a library for Python allowing data manipulation and analysis. Pandas offers data structures and operations for manipulating numerical tables and time series.
* **Numpy** allows for manipulation of matrix and multidimensional tables. It also allows to perform a wide range of mathematical operations on these tables
* **Matplotlib** is a library used to make graphics and visualization of data
* **Statsmodels** is a python library that provide a wide range of statistical models to conduct statistical data exploration and statistics
* **Scikit-learn is** a machine learning library built on Numpy, Scipy and matplotlib and is used to do predictive data analysis

We also used other libraries mainly for data manipulation such as request an OS

**Statistical methods**: At first, we will use descriptive analysis and graphics to understand our data set. Secondly,when working with macro/stock timeseries it is important to check for stationarity of the data and if non-stationarity is discovered to tackle this issue. For that purpose, we will use the ADF test and first-difference (or more) if necessary. Then we will use the OLS method to test the impact of macro variables on the Nasdaq and the Granger causality test to check if we can forecast one time series using another one (if for example by using inflation time series we can forecast Nasdaq price). A last it would also be useful to check if residuals are white noise.

# 3 Data

As describe above, the data were obtained through FRED database and the Nasdaq stock exchange.

We will work with monthly data which represent 95 observations with 4 independent variables from January 2012 until December 2019:

- **GDP:** We use GGP in percent change from preceding Period, Monthly, Seasonally Adjusted Annual Rate. The GDP refers to the total amount of goods and services a country produces. It is a snapshot of the economy at a certain point in time. US GDP has moved around 4% over the time-period selected with a strong growth between 2016 and 2018 following a looser regulation and tax environment.

Chart

Description automatically generated

- **CPI (consumer price index):** Growth Rate Same Period Previous Year, Monthly, Not Seasonally Adjusted. The CPI gives the current price of a selected basket of goods and services that is updated periodically. The CPI has moved around 2% to 2.5% over the time-period selected.

Chart, line chart

Description automatically generated

- **US Interest rates:** Effective Federal Funds Rate, Percent, Monthly, Not Seasonally Adjusted. The Fed fund rates have been extremely low following the global financial crisis, up to 2016 where the FED decided to tighten its monetary policy. The easing of the monetary policy in 2019 was due to the market crash of December 2018 following the trade war between China and the US and the slowdown of the US economy.

Chart, line chart

Description automatically generated

- **US Unemployment rate** in Percent, Monthly, Seasonally Adjusted. The [unemployment rate](https://www.investopedia.com/terms/u/unemploymentrate.asp) shows how many people from the available pool of labor are not able to find a job. US unemployment rate constantly went down following the global financial crisis to reach a low point of 3.8%, end 2019.

Chart, scatter chart

Description automatically generated

As a first observation we can already see that some of the series are not stationary. CPI looks stationary as the mean and the standard deviation look constant and there is no clear trend, while for the unemployment rate we can clearly see that the mean is not constant. We will confirm all of it in the statistical analysis when we test for unit root through the Augmented Dickey Fuller test.

Our **independent variable** is the Nasdaq composite index. The Nasdaq composite index had an impressive performance over the period, with a drop in December 2018 and a catch back beginning 2019 following the decrease of the Fed Fund rate and the rebound of the US economy.

Chart, line chart

Description automatically generated

To put everything on the same scale we have converted the daily Nasdaq composite price in monthly return, which will be the name of our new independent variable by using the following code:

Graphical user interface, text, application

Description automatically generated

Below is the dataset we will work with to perform our statistical analysis. Data quality will be asset in the section 5 of this report.

Table

Description automatically generated

We have also use botplox to understand the spread of our dataset and if we have outliers. As we can see on the below

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

# 4 Metadata

What metadata is required for reproducing your analysis. We use the FRED database which is the database from the Federal Reserve of Saint Louis. The database use filters that have been parsed out to their most finit terms. These metadata allowed us to sort through the information of the database quickly. For example, there is 78,211 series in FRED that are associated with GDP. The FRED allow to break down the data in 8 sections:

- Concepts

- Geography

- Frequencies

- Sources

- Releases

- Seasonal Adjustements

- Citation & copyright

Below we describe the filters used for the four independent variables:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metadata filters** | **GDP** | **CPI** | **Interest rate** | **Unemployment rate** |
| Concepts | Gross Domestic Product | Consumer price inflation | FED fund rate | Unemployment rate |
| Geography types | Country | Country | Country | Country |
| Geographies | US | US | US | US |
| Frequencies | Monthly | Monthly | Monthly | Monthly |
| Sources | n.a. | n.a. | n.a. | n.a. |
| Releases | Bureau of Economic Analysis | Bureau of Economic Analysis | Bureau of Economic Analysis | Bureau of Economic Analysis |
| Seasonal Adjustements | Seasonally adjusted | Seasonally adjusted | No | Seasonally adjusted |

**Where do you store the metadata, how can people access it?**

We have saved the metadata in xls files available on our github, alternatively the metadata are available on the website of the FRED of St Louis by using the below process:

Step 1: reach out to research.stlouisfed.org and select FRED  
Graphical user interface, website

Description automatically generated  
Step 2: Input the name of the independent variable

Graphical user interface, text, application

Description automatically generated

Step 3: Refine your search using the above table

Graphical user interface, application

Description automatically generated

# 5 Data Quality

The FRED and Nasdaq exhibit a relative clear dataset. As we can see below, using the following code: df\_merged\_clean.info(verbose = True) (df\_merged\_clean being our dataframe).

There is no null variables and no missing values. All the values have the dtypes float64 which we would expect. Running the data on a daily or weekly basis would be relatively difficult as macro-economic data are usually published monthly, we would have to close multiple gaps if we would like to do it. That’s why we converted the Nasdaq price index on monthly return.

|  |  |  |
| --- | --- | --- |
| Metric | Definition | Answer |
| Ratio of Data to Errors | How many errors do you have relative to the size of your data set? | none |
| Number of Empty Values | Empty values indicate information is missing from a data set. | none |
| Data Transformation Error Rates | How many errors arise as you convert information into a different format? | none |
| Amounts of Dark Data | How much information is unusable due to data quality problems? | none |
| Data Storage Costs | How much does it cost to store your data? | none |

Text

Description automatically generated with low confidence

In order to assess our data quality we have use the following framework from the Precisely [10]

Based on the above framework, we can conclude that the quality of our dataset is good enough to proceed with our statistical analysis

# 6 Data Flow

Our data flow can be summarized as below:  
**Step 1:** collect data on the FRED and NASDAQ website and download the file in csv or xlsx

**Step 2:** Load the data in jupyter notebook and convert the data into a dataframe

**Step 3:** Clean and merge the dataset. Check the quality of the data (e.g. missing values, ect) and merge the different datasets all together. Perfom a first data analysis using the code below:

round(df\_merged\_clean.describe(),2)

**Step 4:** Run the statistical analysis:

* Test for stationarity and then modify the data if necessary (use first/second difference)
* Run the OLS model

**Simplified data flow**

FRED Saint-Louis timeseries for independent variables

Nasdaq time series for dependent variable

Loading the data into Jupyter (csv or api)

Test for stationarity then first / second difference

OLS

# 7 Data Model

**7.1. Conceptual model**

**7.2. Logical (what columns/features will you use/need)**

**7.3 Physical data model**

# 8 Risks

Financial markets are affected by a wide range of factors that go beyond what we have included in our model. GDP, inflation, Fed fund rate and unemployment rate are four important metrics but far from being the only one relevant when one wants to understand market moves. At macroeconomic level, geopolitics (the trade war between US and China), weather (the earthquake in 1906 that trigger a market crash at wall street in 1907 or the impact of hurricanes on oil production), fiscal policy or presidential elections are factors that also affect the market. At a microeconomic level, companies financial results are also a relevant indicator of market performance. Finally market psychological behavior play also a lot in the performance of financial markets (e.g. herding spirit). Markets can also exhibit cycles that we have taken into account in our analysis [6].

A second important element that matter is the methodology of calculation for the difference macroeconomic data. They are not all “perfect” and subject to emotional discussions amongst economists, they also change over time.

For our four independent variables with have the following pitfalls [7],[8]:

* **GDP**: The one issue of this metric is that data has to be collected within a specific time period, a figure for the GDP now would have to be an estimation.
* **CPI**: There is nothing less that 11 metrics of calculation to determine consumer price inflation (some do not include the cost of healthcare for example), but the CPI index is one of the most well known and use by the Federal Reserve to dictate its monetary policy. For example a strong increate in healthcare cost could be missed in this metrics which would lead to an understatement of inflation
* **US interest rates:** US interest rates are set by the monetary authority, issue of properly evaluating the state of the economy could lead to a mistake in the monetary policy. For example an increase in interest rates too early following the COVID crisis could lead to an economic slowdown as the cost of borrowing would go up.
* **Unemployment rate:** This metric only includes at people looking actively for a new job, others would be excluded, which could lead to an understatement of the “real” unemployment rate

Last pitfall, most of these metrics are backward looking and could already miss the current state of the economy.

**When this and that goes wrong, what counter measures do you have?**

In order to solve our first issue, we could include more macroeconomic indicators in the model. Some macro-models are including pools prediction for a presidential election in order to forecast what could be the fiscal policy after an election and the impact on the market.

Below a list of macroeconomic indicators [9], we could test our model with:

- Government Debt to GDP

- Balance of Trade

- Credit Rating

- Minimum Wages

- Non-Farm Payrolls

- Loan Growth

- Money Supply M0

- Consumer Confidence

- Consumer Credit

- Consumer Spending

- Composite PMI

- Corporate Profits

- Personal Spending

- Private Sector Credit

- Retail Sales MoM

- Retail Sales YoY

The second issue on the methodology calculation is tricky to update. The metrics we are selected are the most used by macroeconomics. We could cross-check the selected metrics by the ones using an other methodology.

Regarding the last issue, one can use alternative data in order to assess if one of the metrics is currently over-or under-estimating the current state of the economy. Some of these metrics are credit card issuance, traffic on the road, satellite image of activity…

The project could suffer some additional costs as collecting alternative data is usually expensive. The computing power would also need to increase in order to digest these data.

# 9 Preliminary Studies

Plots and numbers (from Module 2).

# 10 Conclusions

….

# Appendix A

Link to github: https://github.com/parment1/Project-M1-M2

# References and Bibliography

Please number any information source you used in the report with corresponding links here [1]:

[1] https://hedgelists.com/top-25-global-macro-hedge-funds-2021/

[2] <https://www.schroders.com/en/insights/economics/covid-19-why-the-tech-giants-have-emerged-as-winners/>

[3] Asset returns and inflation, Eugene F.Fama and G.William Schwert, November 1977, Journal of Financial Economics , volume 5, issue 2

[4] Causal relations among stock returns and macroeconomic variables in a small, open economy [Øystein Gjerde](https://econpapers.repec.org/RAS/pgj6.htm) and [Frode Sættem](https://econpapers.repec.org/RAS/pst120.htm) [Journal of International Financial Markets, Institutions and Money](https://econpapers.repec.org/article/eeeintfin/), 1999, vol. 9, issue 1, 61-74

[5] The impact of macroeconomic variables on Stock ‎market in United Kingdom NEIFAR, MALIKA and Dhouib, Salma ‎ and Bouhamed, Jihen ‎ and Ben Abdallah, Fatma ‎ and Arous, Islem ‎and Ben Braiek, Fatma ‎ and Mrabet, Donia ‎ (2021)

[6] Finance and Macroeconomics: The Preponderance of the Financial Cycle [Michel Aglietta](https://www.cairn.info/publications-de-Michel-Aglietta--6668.htm), [Revue de l'OFCE](https://www.cairn.info/revue-de-l-ofce.htm) [2018/3 (N° 157)](https://www.cairn.info/revue-de-l-ofce-2018-3.htm), 197-224

[7] BIS Working Papers No 677 Macroeconomic implications of financial imperfections: a survey by Stijn Claessens and M Ayhan Kose Monetary and Economic Department November 2017

[8] https://www.investopedia.com/articles/investing/031413/economic-indicatiors-affect-us-stock-market.asp

­[9] https://tradingeconomics.com/indicators

[10] https://www.precisely.com/blog/data-quality/how-to-measure-data-quality-7-metrics

[11] <https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>

[12] <https://machinelearningmastery.com/time-series-data-stationary-python/>

[13] <https://machinelearningmastery.com/statistical-hypothesis-tests-in-python-cheat-sheet/>

[14] <https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>

[15] <https://pandas.pydata.org/>

[16] https://www.statsmodels.org/stable/index.html