**Thomas Parmentier**

Zypressenstrasse, 49

8004 Zürich

Thomas.parmentierlux@gmail.com

**Data Science Project**

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

**Conceptual Design Report**

**X 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](#_Toc83971363)

[Table of Contents 1](#_Toc83971364)

[1 Project Objectives 2](#_Toc83971365)

[2 Methods 3](#_Toc83971366)

[3 Data 3](#_Toc83971367)

[4 Metadata 4](#_Toc83971368)

[5 Data Quality 4](#_Toc83971369)

[6 Data Flow 4](#_Toc83971370)

[7 Data Model 4](#_Toc83971371)

[8 Risks 4](#_Toc83971372)

[9 Preliminary Studies 5](#_Toc83971373)

[10 Conclusions 5](#_Toc83971374)

[Acknowledgements 5](#_Toc83971375)

[Appendix X 5](#_Toc83971376)

[References and Bibliography 5](#_Toc83971377)

# 

# 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, 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.

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 try to answer the following question: 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:

- **GDP:** We use GGP in percent change from preceding Period, quarterly, Seasonally Adjusted Annual Rate.

- **CPI (consumer price inflation):** Growth Rate Same Period Previous Year, Monthly, Not Seasonally Adjusted.

- **US Interest rates:** Effective Federal Funds Rate, Percent, Monthly, Not Seasonally Adjusted.

- **US Unemployment rate** in Percent, Monthly, Seasonally Adjusted.

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, and we will start by running statistical analysis.

Chart, line chart

Description automatically generated

Table

Description automatically generated

A picture containing table

Description automatically generated

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

# 4 Metadata

What metadata is required for reproducing your analysis?

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

# 5 Data Quality

What are the quality requirements you have to reach your analysis goal (precision ...)?

Are they met? If not, do you expect a significant impact on your results,

Any measures to improve the quality?

# 6 Data Flow

Explain with a figure and words how the data flow of your project will be, from the data source to the final plots and numbers.

# 7 Data Model

Draw and explain your data model at the conceptual level, the logical level and the physical level. .

Conceptual

Logical (what columns/features will you use/need)

Physical (infrastructure needs)

# 8 Risks

What can go wrong?

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

What will be the impact on the quality of the aimed output, project time schedule, project cost ?

# 9 Preliminary Studies

Plots and numbers (from Module 2).

# 10 Conclusions

….

# Acknowledgements

Acknowledge persons or institutions that helped you with the CDR here.

# Appendix X

If you have something to attach to your report, do it here.

# 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