Financial Markets during The 2020 Pandemic

CSCI 5707 Principles of Database System

Group 5

Members: Miao Yang, Tiruo Yan, Darry Zhang, Linxin Li

University of Minnesota
Driven to Discover℠

Introduction **3**

Datasource **3**

2.1 Equity Market Datasource 3

2.1.1 Wiki S&P 500 members table 3

2.1.2 Yahoo Finance 4

2.3 Covid Datasource 5

Data Architecture **6**

3.1 Dataset 1 6

3.2 Dataset 2 7

3.3 Architecture 7

Financial Market Behavior under 2020 Pandemic **8**

4.1 Equity Market Behavior 8

4.2 Other Financial Assets Behaviors 10

Liquidity and assets performance pattern **11**

5.1 Model Construction 11

Conclusions **15**

Reference **15**

Appendix **15**

1. Introduction

This project explores how the financial market responds to the 2020 pandemic using databases, machine learning and financial knowledge. This report is arranged as follows. In the first part, it is introduced the datasources, approaches to acquire data, and data architecture. In the second part, a snapshot of this year's financial market behaviors through queries conducted in spark SQL is reported. The third part illustrates a vector autoregressive model to apply impulse response analysis with the aim of exploring the relationship between liquidity shock and asset performance pattern. The model is built using the MATLAB Econometrics toolbox.

2. Datasource

2.1 Equity Market Datasource

2.1.1 Wiki S&P 500 members table

Description: month-end S&P 500 provided by Wikipedia.

Schema:

Field	Data Type	Description
Company Ticker	String	Public Traded Company's ticker on exchange
Company Name	String	Company's legal name
GICS Sector	String	MSCI GICS Sector Info.
GICS Sub Sector	String	MSCI GICS Sub Sector Info
Headquarter Location	String	The location of the company's headquarter.
Date first Added	String	The first date the company is added into s&p 500
CIK	String	Company's Cusip No

Fields related to our research:

Company's ticker and Company's name: we'll examine companies that performed the best and worst in our financial market behavior analysis.

GICS Sector: We explore individual stocks performance within each sector.

To study each sector performance, 11 tickers of GICS (global industry classification standard) sector ETFs are added into the list of Wiki S&P 500 members ticker. These sector ETFs are traded the same way as individual stocks, and their performances represent the corresponding sector performances.

Data acquisition:

The information of S&P 500 was extracted from Wikipedia using Python. As the whole list of S&P 500 was provided as a table on the Wikipedia page, we were able to extract the data using the page's html link, and further render them into csv tables using pandas. After acquiring all the information about individual stocks, we can then manually add all the ETF tickers into the table, since there are only 11 ETF tickers, manual-load is feasible.

2.1.2 Yahoo Finance

Yahoo finance provides the largest amount of free financial data regarding public traded companies among other free finance websites. They offer historical daily/monthly/annually stock prices, analyst ratings and firm fundamental data.

Schema:

Field	**Data Type**	**Description**
Date	String	Month-date, year
Open	Floating Point Number	Market open price
High	Floating Point Number	The highest stock price during the date
Low	Floating Point Number	The lowest stock price during the date
Close	Floating Point Number	The market closed price
Adj Close	Floating Point Number	Adjusted stock price with dividends payment, split. etc
Volume	Floating Point Number	Trading volume

Fields related to our research:

Adj Close price: we get year to 11/11/2020 daily Adjusted close price to run this year performance research.

Data acquisition:

All the financial data were crawled from Yahoo Finance using Python and the yfinance package, which was the official data crawling API provided by Yahoo Finance. With such we were able to download the information we need, giving tickers, into csv form.

2.2 Other Financial Datasource

To explore the relationship between liquidity shock and asset performance pattern, long term quarterly economic data is used, which includes US CPI, commodity price, US housing price index, gold price, US Equity Market S&P 500 Index, Nominal money supply M2, US 3 month treasury yield, and US real GDP from 1981 1st quarter to 2020 3rd quarter.
Economic data such as M2, US 3 month treasury yield, and assets data such as commodity and gold are from the Federal Reserve Bank(FRED) website. US real GDP, CPI are from the U.S. Bureau of Labor statistics. US housing price data is from Federal Housing Finance Agency(FHFA), S&P 500 index data is from Yahoo finance.
Data downloaded from FRED, FHFA and U.S. Bureau of Labor Statistics website has the same schema listed as follows.

Field	Data Type	Description
Date	String	Month/Date/Year
Index Name	Floating Point Number	Index Price Value

Data Acquisition: These websites provides data download link, and data can be directly downloaded and saved as .CSV

2.3 Covid Datasource

Covid Datasource is from Centers for Disease Control and Prevention (CDC), can be downloaded from CDC website and saved as CSV format. Including date, total confirmed case in the US, new confirmed case in the US, 7-day moving average, and rate per 100k.

Field	Data Type	Description
Date	String	Month Date Year
Total cases	Integer	Total confirmed COVID-19 cases in the US
New cases	Integer	New confirmed COVID-19 cases in the US
7-day moving average	Integer	Average new confirmed COVID-19 cases in the US for the last 7-days
Rate per 100000	Integer	Average confirmed COVID-19 cases per 100k people according to US population

3. Data Architecture

In this report, it is first summarized the 2020 financial market behavior, and then the relationship between liquidity shock and asset performance pattern is studied. Two datasets are created for the 2 research goals.

3.1 Dataset 1

To examine the 2020 financial market behavior, Covid data and Yahoo Finance data are combined to create *Dataset1*. Both of the two sets of data are at daily frequency, and merged using the US trading dates from end of 2019 to 11/11/2020.

Schema:

Field	**Data Type**	**Description**
Date	Integer	YearMonthDate
Ticker	String	For stocks and ETFs, this field is their public traded ticker on exchange. For covid data, this field is an abbreviation for meaning of the data.. For example, the seven day moving average is SevenDmov, total cases number is totalcasesN.
PriceorValue	Floating Point Number	For stocks and ETFs, this field is their daily price. For Covid data, it is the value.
Name	String	Full name of each ticker
Type	String	For financial assets, this field is the type of security. For example, ETF and stock. For Covid data, this field is given as "CovidNo." So we can use type to distinct different data to help query.
Sector	String	For financial data, this is the sector info of each stock and ETFs. For Covid data, this field is None.

3.2 Dataset 2

To explore the relationship between liquidity shock and asset performance pattern, a vector autoregressive model is developed on a long-term dataset of time-series financial data. We create *Dataset2* that combines CPI, GDP, M2, short-term interest rate, gold price, commodity price, US house price index, and S&P 500 index price; the data variables are reported in other financial datasource above. *Dataset2* is from 1981 to 2020 on a quarterly basis.

Schema:

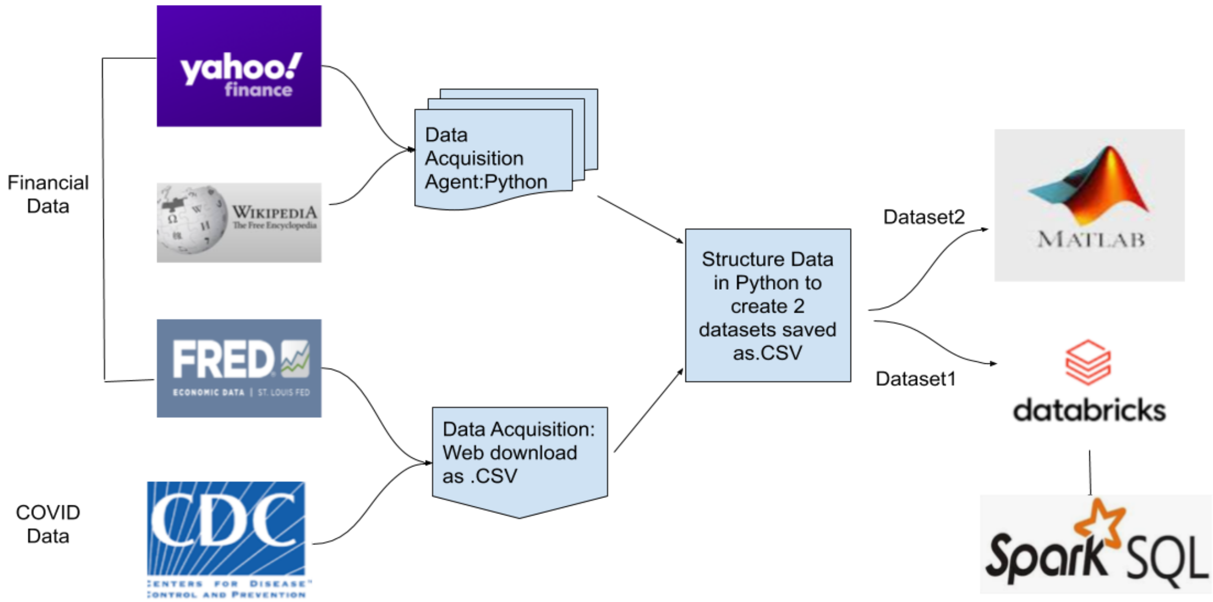
Field	Data Type	Description
Date	String	Month/Date/Year
Index Name	String	Index Name showed on its data source.
Value	Floating Point Number	Index value level

3.3 Architecture

Two separate datasets are created in this project. *Dataset1* is constructed to study the 2020 financial behavior. It is uploaded to Databrick, and Spark SQL is used to assess the market behavior. *Dataset2* is built to a regression model to analyze the relationship between liquidity shock and asset performance pattern. This dataset is loaded into Matlab to conduct our economic analysis using Matlab Econometrics toolbox.

We prefer Spark SQL to other SQL platforms such as MySQL for the following reasons.
The one biggest advantage of Spark SQL is that if we execute the same queries in MySQL and Spark SQL, queries execution time of Spark SQL is much faster than MySQL' s. This is because Spark automatically uses all CPU cores and executes complexes in parallels but MySQL can only use one CPU core to execute one query. Although in MySQL, creating an index in a query would speed up execution time, we cannot create an index for each ad-hoc query while Spark is capable of processing a variety of queries.
The second advantage is that in general, Spark can handle extract, transform and load（ETL) processes easily. ETL process allows us to gather data from different sources (such as Yahoo!, Wikipedia, and so on) and different types of data and consolidate it into a single table, which is what we need in our data analysis.

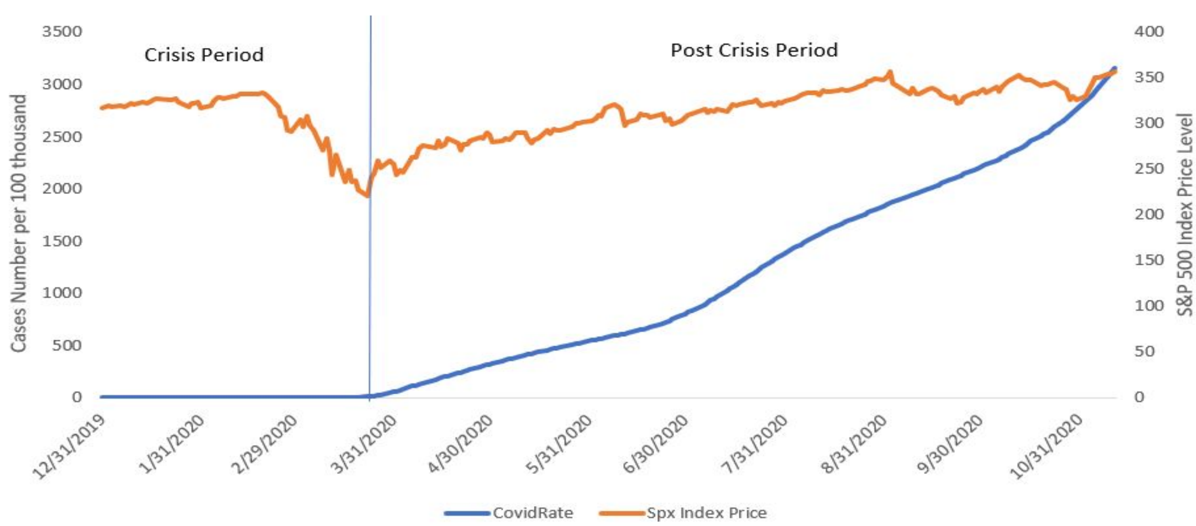
Architecture structure is shown below.



4. Financial Market Behavior under 2020 Pandemic

4.1 Equity Market Behavior

Since the pandemic outbreak globally in the beginning of 2020, the US equity market entered correction territory in February. However, it started to recover from the beginning of the second quarter regardless of the exponentially increasing Covid-19 new cases. According to the equity market reaction, we call the first quarter crisis period, and rest of the year post crisis or recovery period.



During the crisis period, the best performing sector is information technology, while the worst is the financial sector, followed by the energy sector.

The most popular equity index weighting scheme in the financial industry is market cap weighting, which means larger companies (larger market cap stocks) receive larger index weights. As a result, mega companies are dominating in almost all popular broad equity indices and products, such as S&P 500 index and sector ETFs, which is known as the concentration feature of market cap weighted indices.

This feature is the most significant in the information technology sector as the largest 5 underlying companies add up to almost 50% dollar value of the whole sector. Such mega tech companies (e.g. Apple and Microsoft) have mature businesses with less reliance on the financial market for capital. Therefore, they exhibited low volatility during the crisis period, and drove the whole information technology sector to be the best performer.

Interest rate cuts and lack of liquidity hurt the financial sector, especially profits of big banks, while the gloomy economy and stay at home order hurt the energy sector the most.

Sector Name	**Crisis Period Performance (%)**
Info Tech	-11.869
Health Care	-12.615
Consumer Staples	-13.017
Utilities	-13.394
Communication Services	-17.276
Real Estate	-19.234
Equity Index	-19.448
Consumer Discretionary	-21.419
Materials	-26.179
Industrials	-27.006
Financials	-31.785
Energy	-50.516

During the post crisis period, the best performer is consumer discretionary, while the worst is still the energy sector. Because of alleviated liquidity stress under a series of stimulus, hope of vaccines, and attractive valuation caused by the sold off in 2020Q1, companies in consumer discretionary, such as entertainment and restaurants, started to be back in favor. In contrast, the energy sector experienced an anomaly not seen in modern finance that negative crude oil price in April 2020 resulting from critical oil storage facilities filling to capacity as production vastly exceeds demand.

Sector Name	Post Crisis Period Performance (%)
Consumer Discretionary	56.7429
Materials	55.4371
Info Tech	51.6281
Industrials	45.0231
Communication Services	44.1976
EquityIndex	39.5463
Financials	31.4225
Health Care	26.4896
Consumer Staples	24.3093
Utilities	22.0756
Real Estate	21.1609
Energy	19.6698

Besides sector performance, we also assess the winners and losers, as well as the performance discrepancy between them in each sector.

Crisis period

Sector	Best Performer Name	Best performer Return	Worst Performer Name	Worst Performer Return	Discrepancy (BstPerfor Ret-WstPerfor Ret)
Consumer Discretionary	DPZ	0.105697	NCLH	-0.81236	0.91805783
Consumer Staples	CLX	0.135898	SYY	-0.46374	0.5996418
Energy	COG	-0.0062	APA	-0.8354	0.829199635
Financials	MSCI	0.121516	CMA	-0.58364	0.70515074
Health Care	REGN	0.300442	HCA	-0.39007	0.69050783
Industrials	ROL	0.093203	UAL	-0.64184	0.735046434
Information Technology	CTXS	0.280445	DXC	-0.64597	0.92641933
Materials	NEM	0.045059	MOS	-0.49853	0.543585307
Communication Services	NFLX	0.160491	VIAC	-0.66115	0.82164222
Real Estate	DLR	0.169233	SPG	-0.62607	0.7953037
Utilities	NEE	-0.00117	CNP	-0.42731	0.426140361

Post Crisis period

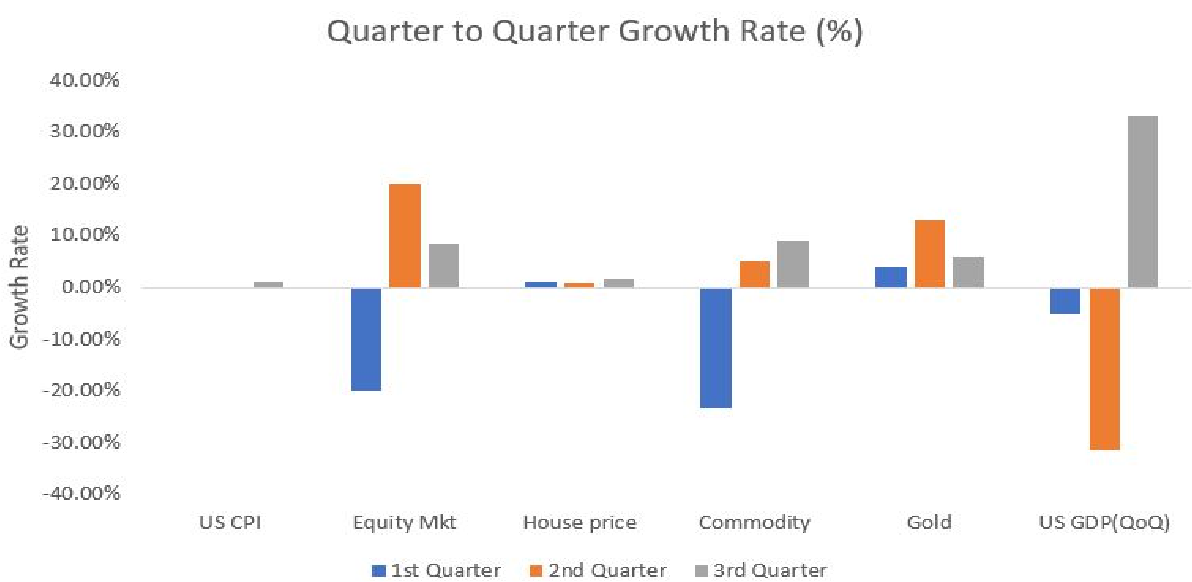
Sector	Best Performer Name	Best performer Return	Worst Performer Name	Worst Performer Return	Discrepancy
Consumer Discretionary	ETSY	2.397243	TIF	0.024722	2.372520334
Consumer Staples	EL	0.561327	WBA	-0.07341	0.63473577
Energy	APA	1.398056	HFC	-0.10882	1.506873686
Financials	SIVB	1.176463	WFC	-0.13251	1.30897611

Health Care	ALGN	1.654671	BIIB	-0.22868	1.88335171
Industrials	FDX	1.225505	HII	-0.12242	1.34792013
Information Technology	QCOM	1.211829	CTXS	-0.18421	1.39604382
Materials	FCX	1.893333	IFF	0.083211	1.81012254
Communication Services	VIAC	1.173418	LUMN	0.022199	1.151219265
Real Estate	VTR	0.849064	EQR	-0.01325	0.862315008
Utilities	AES	0.607823	FE	-0.22893	0.83675219

4.2 Other Financial Assets Behaviors

The data of other financial assets is on a quarterly basis. We show their performances below.

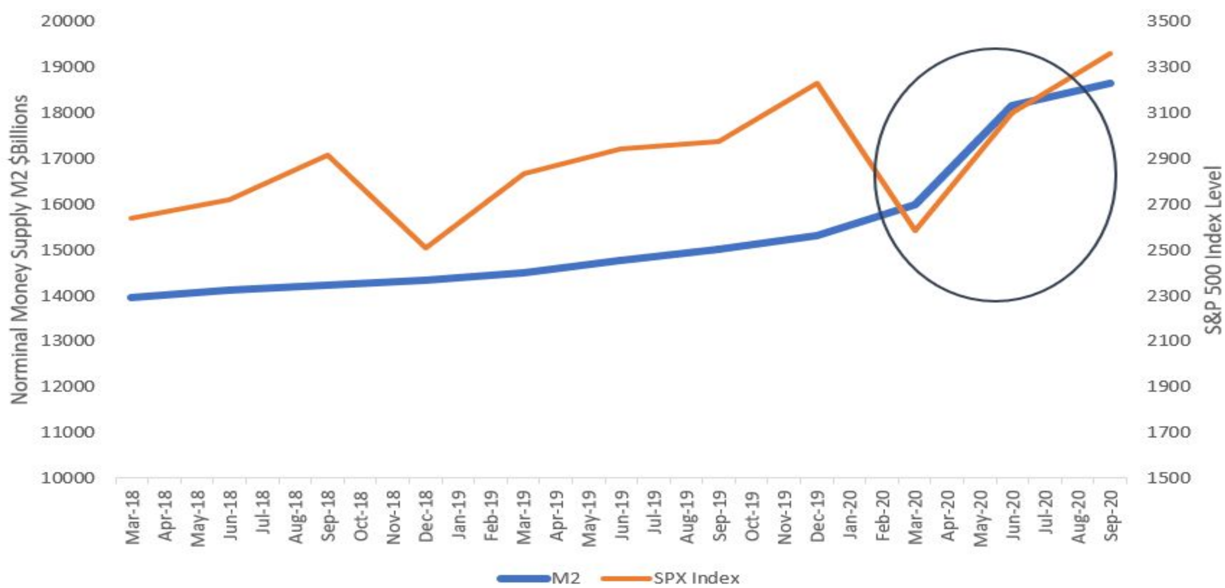
The US CPI was quite flat in the first 2 quarters, and slightly up in the third quarter. The commodity performed poorly in the first quarter, but quickly picked up in the second and third quarter. House price and gold performed strongly. The US GDP was improving in the third quarter after the sharp drop in the previous quarter. In the next section, we analyze the relationship between M2 (nominal money supply) and these assets.



5. Liquidity and assets performance pattern

As the global pandemic prompted deep worry of a gloomy economy and caused the lack of liquidity concern (banks are not lending, businesses are not investing and consumers are not buying), the US equity market experienced a sharp drawdown and entered into correction territory in February 2020. Responding to the unsettling market, the Fed announced 100 bps interest rate cut to zero and commitment to $700 billion bond buying at the beginning of March 2020, and surprisingly enacted unlimited QE (quantitative easing) at the end of March to alleviate building liquidity stress in the treasury and interbanks. Following Fed actions, the US equity market started to recover from the beginning of the second quarter. At the same time, the value of other financial assets, such as commodities, real estate and gold, were soaring while US inflation was flat as we discussed in Section 4.2.

Under FED unlimited QE, nominal money supply M2 has sharply increased, which is shown in the chart below. Many wall street analysts believe that it helps the equity market recover but drives up US inflation risk. However, we have not spotted high inflation yet since CPI index is quite flat in the second and third quarter. In addition, the fact that gold and equity both performed strongly in the second and third quarter condratics the common view in the investment industry that gold and equity exhibit negative correlations overtime. In the following section, we build a vector autoregressive model and conduct impulse response analysis to explore asset performance patterns after M2 got unexpectedly shocked.



5.1 Model Construction

Vector autoregressive model with P lags with no time trend or exogenous variables is specified as follows.

$$y_t = c + \sum_{j}^{p} \Phi_j y_{t-j} + \varepsilon_t ,$$

where $y_t, c \ and \ \varepsilon_t$ are k-dimensional vector, Φ_j is k*k matrix.

According to the model, impulse analysis function of y_t shocked by an impulse to variable j by one standard deviation of its innovation m periods into the future can be written as follows.

$$\psi_j(m) = C_m e_j, \ C_m = \Omega_m K,$$

where K is lower triangular factor in Cholesky, and Ω_m is the lag m coefficient of $\Omega(L)$.

The data we use to conduct impulse response analysis is *Dataset2* discussed in Section 3.2. To avoid forward-looking bias, the time period for this research is from 1st quarter 1981 to 1st quarter 2020, before when FED unlimited QE started in 2020.

A constant is added into the model, and all variables are taken to log transformation except the interest rate. Model specification is given by the following.

$$y_t = (GDP, CPI, \ Commodity, \ House\ Price, \ Gold, \ M2, \ Short\ Term\ Interest, \ Stock)_t$$

According to (Favero, 2001), monetary variables are expected to respond faster to the real economy, and thus they should be ordered last in Cholesky decomposition.

In our test, the lag length of 1 gives the model the best Akaike Information Criteria (AIC), and we use lag length of 1 in the model.

5.2 Empirical Findings

Equity price (S&P 500 index) does not show a positive response to a monetary impulse, which is consistent with Belke et al. (2010) but contradicts with that strong performance of the equity market after M2 got shocked in 2020. Many researchers found that the relationship between money and stock price is less pronounced than for the other assets (Fischer, et al., 2008) and equity price is usually modeled as random walks in academics (Agwuegbo, 2010).

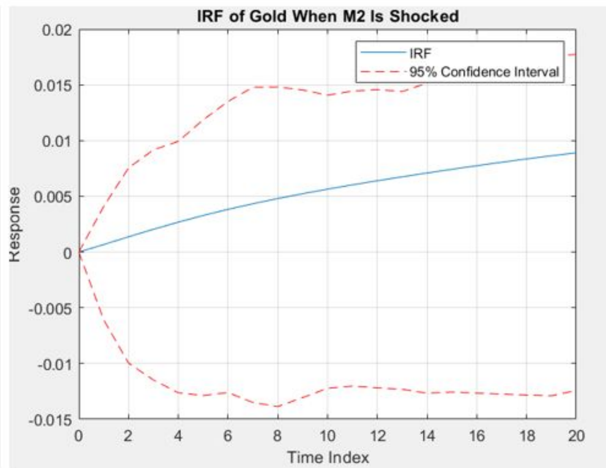
Increased liquidity tends to increase household assets. If equity expected return is high, excess assets will be invested into the stock market. It would be interesting to add equity valuation in our VAR model because lower valuation indicates higher expected return (Shiller. 2013).

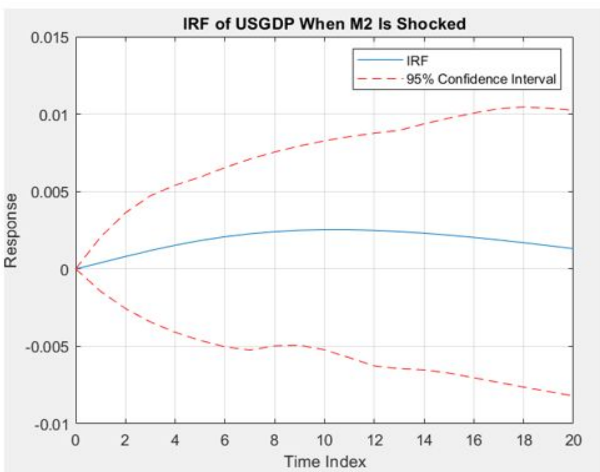
In contrast to equity price, when an unexpected shock to liquidity (M2) occurs, house, gold and commodity show a quick positive response, which is also observed in the following period. Instead, CPI index reacts slowly and moves upwards later. The positive responses are inline with economic theory, The Law of Supply and Demand. According to this theory, the expansionary monetary policy drives up the demand on assets. For assets characterized by restriction on supply, such as gold and house, the additional demand triggers immediate price rise. However, for assets which have high price elasticity of supply, such as consumer goods, the price reaction to higher demand is more subdued. As the 2020 unlimited QE started, we spotted this phenomenon that gold, commodity and house price were rising while CPI index was flat as we showed in Section 4.2, indicating that these assets behavior is inline with the model's indication in 2020.

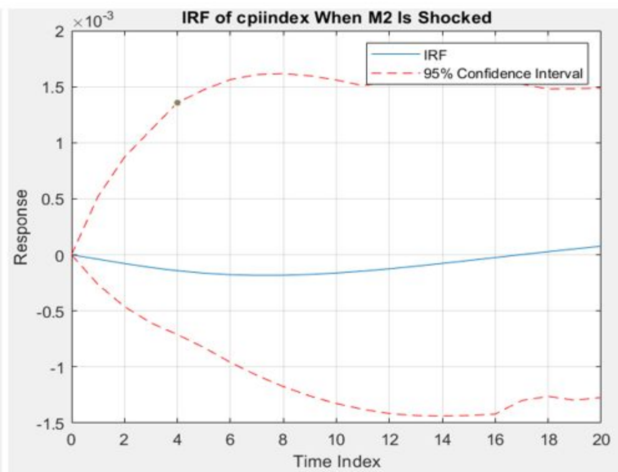
US GDP gives a positive response to liquidity as well. The fact that GDP goes up temporarily but not permanently due to the liquidity shock is in line with the theoretical consideration that money is neutral for the real economy in the long run. After M2 was shocked in 2020, GDP is improving in the third quarter, but the impact from the liquidity shock will eventually disappear according to the model's indication

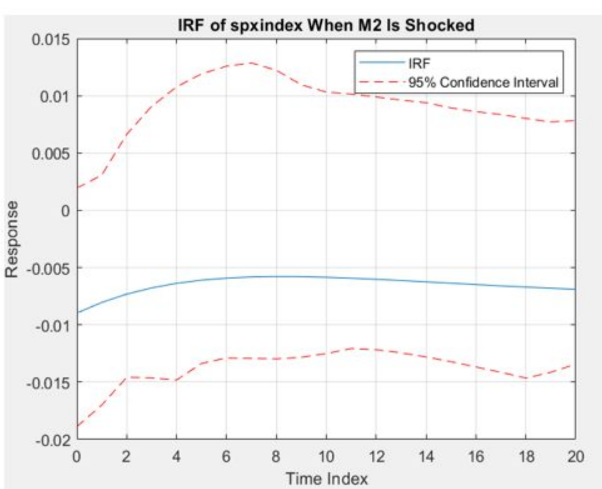
However, none of the responses we discussed above are statistically significant, which is inconsistent with Belke et al. (2010).

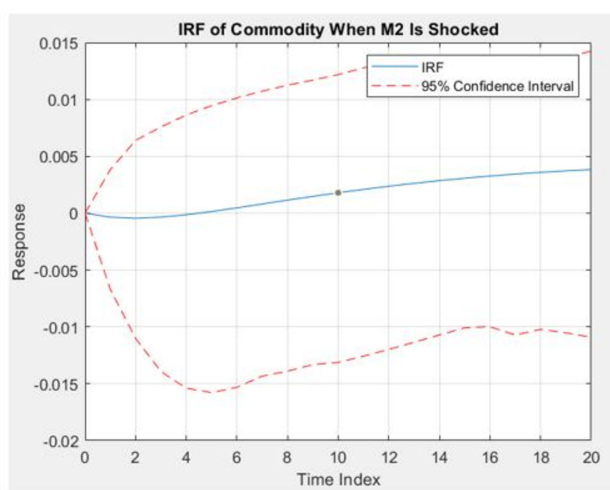
This discrepancy in statistical significance could be caused by different time periods of the data used. Belke et al. (2010) draw conclusions based on a dataset from 1984 to 2006, while ours is from 1981 to 2020. A non-negligible difference is that our sample covers the 2008 financial crisis and the 2020 pandemic. Studies suggest that a crisis can push the economy and markets into a new regime that was never observed before, for example Tan and Cheong (2016). A model that doesn't treat data of crises and post-crisis periods differently cannot capture the regime shift, and thus tend to lose statistically significance. Future study could be to incorporate the regime change into the current model. One approach is to apply Markov regime-switching VAR presented in Guo et al. (2011).











6. Conclusions

In this project, we use Spark SQL to study financial assets performance during the 2020 pandemic. We next build a VAR model to conduct impulse response analysis using Matlab Econometric toolbox, aiming at understating the relationship between liquidity and financial assets behaviors. We found that during 2020, behaviors of commodity, gold, house price, inflation and GDP are inline with our model indication, but equity market behavior is exceptional. It would be interesting to add equity market valuation into the VAR model for further research.

Reference

Agwuegbo, S. O. N., Adewole, A. P., & Maduegbuna, A. N. (2010). A random walk model for stock market prices.

Belke, A., Orth, W., & Setzer, R. (2010). Liquidity and the dynamic pattern of asset price adjustment: A global view. *Journal of Banking & Finance*, *34*(8), 1933-1945.

Favero, C. A. (2001). *Applied macroeconometrics*. Oxford University Press on Demand.

Guo, F., Chen, C. R., & Huang, Y. S. (2011). Markets contagion during financial crisis: A regime-switching approach. *International Review of Economics & Finance*, *20*(1), 95-109.

Fischer, B., Lenza, M., & Pill, H. Reichlin (2008): Money and monetary policy. The ECB experience 1999-2006. *The role of money: Money and monetary policy in the 21st century, European Central Bank*, 102-175.

Shiller, R. J. (2014). Speculative asset prices. *American Economic Review*, *104*(6), 1486-1517.

Tan, J., & Cheong, S. A. (2016). The regime shift associated with the 2004–2008 US housing market bubble. *PloS one*, *11*(9), e0162140.

Appendix

1.Data Acquisition: example given as get Apple(AAPL) from Yahoo Finance using finance:

```
import yfinance as yf
import pandas as pd
output_file =  pd.DataFrame()
stock = yf.Ticker("AAPL")
pd1 = stock.history(start="2019-12-31",end = "2020-01-01", actions=False)
output_file = output_file.append(pd1, index=False)
output_file.to_csv(address, index=False)
```

2.Spark SQL Query: gives code on how to sort crisis period sectors (to show we query on ETFs only)

```
select tem1.Ticker, tem1.Type, tem1.Sector, (tem2.PriceOrCasesN/tem1.PriceOrCasesN - 1)
as crisis_period_return
from (select Ticker, PriceOrCasesN, Type, Sector

from dataset1

where 20191231 = Date and Type = "ETF") as tem1,

(select Ticker, PriceOrCasesN, Type, Sector

from dataset1

where 20200331 = Date and Type = "ETF") as tem2

where tem1.Ticker = tem2.Ticker
```

3. Spark SQL Query on search for winner and loser companies inside each sector: crisis return rate tables are tables that we created based on calculating the return rate as measuring the percentage difference between different time points.

```
Select *

From return_crisis

Order by return_crisis desc limit1.

Select *

From return_crisis

Order by return_crisis asc limit1.
```

4. Matlab VAR & Impulse response_simplified code:

```
load dataset2                          %load dataset2 into Matlab.
TT= table2timetable(dataset2);        % transfer date into matlab time table.
Mdl.Trend = NaN;                   % do not add time trend into VAR model
EstMdl = estimate(Mdl,TT.Variables,'Display',"full");        % VAR(1) model established
nnum = 21;                           % time length for impulse response is 21 quarters
[Response,Lower,Upper] = irf(EstMdl,"Method","orthogonalized","NumObs",nnum);   %
impulse response function established
```