**Andrew Vo – Hiep Vo**

**Econ 342**

**Andrew Vo Project Data Sets**

* Main Data Source: All datasets below has the same starting link: <https://datahub.io/collections/economic-data#us-economic-data> which includes 8 individual sections of US economy data which will be specified below.
* The goal in the future is to select and main attributes in these dataset, try to combine the same time zones and create one unite dataset for model.
* For list of Variables info for each dataset, Readers can go to each specific link mentioned below and go to Field Information Section and there are also Data and Preparation section that has useful information about how the data is collected and interpreted

1, Gross Domestic Product (GDP) of the United States (US) both nominal and real on an annual basis:

* Bibliographic citation: DataHub. (2019). Gross Domestic Product of the United States (US GDP). [Data file]. Retrieved from <https://datahub.io/core/gdp-us>
* Date first obtained by author: 2018 – harvested from US Government’s Bureau of Economic Analysis (BEA)
* How to obtain original data file: Readers can go directly to link <https://datahub.io/core/gdp-us> then go down to Data Files section then choose the csv download option of the *year* file. Change the name of the downloaded file to year\_GDP for easier access.
* Additional Information: Gross Domestic Product (GDP) of the United States (US) both nominal and real on an annual basis being provided since 1930. Both total GDP (levels) and annualized percentage change in GDP are provided. Both levels and changes are available both in current dollars (nominal GDP) and in chained 2009 dollars (real GDP). Data is sourced from US Government’s Bureau of Economic Analysis (BEA) and provided in standardized CSV.

2, US Employment and Unemployment rates since 1940:

* Bibliographic citation: DataHub. (2019). US Employment and Unemployment rates since 1940. [Data file]. Retrieved from <https://datahub.io/core/employment-us>
* Date first obtained by author: 2018 – harvested from USA Bureau of Labor Statistics
* How to obtain original data file: Readers can go directly to link <https://datahub.io/core/employment-us> then go down to Data Files section then choose the csv download option of the *aat1* file. Change the name of the downloaded file to year\_empRate for easier access.
* Additional Information: Numbers are in thousands. US Employment and Unemployment rates since 1940 From the [USA Bureau of Labor Statistics Employment Related Data](https://www.bls.gov/)

3, US Consumer Price Index and Inflation (CPI):

* Bibliographic citation: DataHub. (2019). US Consumer Price Index and Inflation (CPI). [Data file]. Retrieved from <https://datahub.io/core/cpi-us>
* Date first obtained by author: 2018 – harvested from U.S. Department Of Labor Bureau of Labor Statistics
* How to obtain original data file: Readers can go directly to link <https://datahub.io/core/cpi-us> then go down to Data Files section then choose the csv download option of the *cpiai* file. Change the name of the downloaded file to month\_CPI for easier access.
* Consumer Price Index for All Urban Consumers (CPI-U) from U.S. Department Of Labor Bureau of Labor Statistics. This is a monthly time series from January 1913. Values are U.S. city averages for all items and 1982-84=100. Note that there are many price indices and this is only one of them (albeit the most standard and with the longest set of data).

4, 10 year US Government Bond Yields (long-term interest rate):

* Bibliographic citation: DataHub. (2020). 10 year US Government Bond Yields (long-term interest rate). [Data file]. Retrieved from <https://datahub.io/core/bond-yields-us-10y>
* Date first obtained by author: 2018 – harvested from [Release H.15 from the Federal Reserve - Selected Interest Rates Daily](http://www.federalreserve.gov/releases/h15/data.htm) specifically the [10 year US Treasury (monthly, csv)](http://www.federalreserve.gov/datadownload/Output.aspx?rel=H15&series=0809abf197c17f1ff0b2180fe7015cc3&lastObs=&from=&to=&filetype=csv&label=include&layout=seriescolumn).
* How to obtain original data file: Readers can go directly to link <https://datahub.io/core/bond-yields-us-10y> then go down to Data Files section then choose the csv download option of the *monthly* file. Change the name of the downloaded file to month\_longtermInterest for easier access.
* 10 year nominal yields on US government bonds from the Federal Reserve. The 10 year government bond yield is considered a standard indicator of long-term interest rates.

5, US House Price Index (Case-Shiller):

* Bibliographic citation: DataHub. (2019). US House Price Index (Case-Shiller). [Data file]. Retrieved from <https://datahub.io/core/house-prices-us>
* Date first obtained by author: 2018 – harvested from S&P Case-Shiller data and includes both the national index and the indices for 20 metropolitan regions.
* How to obtain original data file: Readers can go directly to link <https://datahub.io/core/house-prices-us> then go down to Data Files section then choose the csv download option of the *cities* file. Change the name of the downloaded file to month\_housePrice for easier access.
* As per the [home page for Indices on S&P website](http://www.spindices.com/index-family/real-estate/sp-case-shiller):

The S&P/Case-Shiller U.S. National Home Price Index is a composite of single-family

home price indices for the nine U.S. Census divisions and is calculated monthly. It is included in the S&P/Case-Shiller Home Price Index Series which seeks to measure changes in the total value of all existing single-family housing stock.

Documentation of the methodology can be found at:

<http://www.spindices.com/documents/methodologies/methodology-sp-cs-home-price-indices.pdf>

Key points are (excerpted from methodology):

* The indices use the “repeat sales method” of index calculation which uses data on properties that have sold at least twice, in order to capture the true appreciated value of each specific sales unit.
* The quarterly S&P/Case-Shiller U.S. National Home Price Index aggregates nine quarterly U.S. Census division repeat sales indices using a base period a nd estimates of the aggregate value of single family housing stock for those periods.
* The S&P/Case - Shiller Home Price Indices originated in the 1980s by Case Shiller Weiss’s research principals, Karl E. Case and Robert J. Shiller. At the time, Case and Shiller developed the repeat sales pricing technique. This methodology is recognized as the most reliable means to measure housing price movements and is used by other home price ind ex publishers, including the Office of Federal Housing Enterprise Oversight (OFHEO)

6, Income Limits for Each Fifth and Top 5 Percent of All Households: 1967 to 2016:

* Bibliographic citation: DataHub. (2020). Income Limits for Each Fifth and Top 5 Percent of All Households: 1967 to 2016. [Data file]. Retrieved from <https://datahub.io/core/household-income-us-historical>
* Date first obtained by author: 2018 – harvested from This dataset is acquired from U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplements. <https://www2.census.gov/>
* How to obtain original data file: Readers can go directly to link <https://datahub.io/core/household-income-us-historical> then go down to Data Files section then choose the csv download option of the *household-income-us-historical* file. Change the name of the downloaded file to year\_houseIncome for easier access.
* Upper limits of annual incomes for each fifth and lower limit of income for top 5 percent of all housholds from 1967 to last year

7, US Investor Flow of Funds into Investment Classes (Bonds, Equities etc):

* Bibliographic citation: DataHub. (2020). US Investor Flow of Funds into Investment Classes (Bonds, Equities etc). [Data file]. Retrieved from <https://datahub.io/core/investor-flow-of-funds-us>
* Date first obtained by author: 2018 – harvested from Investment Company Institute (ICI).
* How to obtain original data file: Readers can go directly to link <https://datahub.io/core/investor-flow-of-funds-us> then go down to Data Files section then choose the csv download option of the *monthly* file. Change the name of the downloaded file to month\_InvestorFlow for easier access.
* Data comes from the data provided on the [ICI Statistics pages](http://www.ici.org/research/stats), in particular:
* Summary: Estimated Long-Term Mutual Fund Flows Data (xls)
* Notes for Long-Term Mutual Fund Flows Data:
* All figures are (nominal) millions of US dollars (USD)
* Weekly cash flows are estimates based on reporting covering 98 percent of industry assets, while monthly flows are actual numbers as reported in ICI’s “Trends in Mutual Fund Investing.”

8, United States of America education budget analysis:

* Bibliographic citation: DataHub. (2019). United States of America education budget analysis. [Data file]. Retrieved from <https://datahub.io/core/usa-education-budget-analysis>
* Date first obtained by author: 2018 – harvested from Office of Management and Budget, President’s Budget from white house official website on <https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/budget/fy2018/hist05z2.xls>
* How to obtain original data file: Readers can go directly to link <https://datahub.io/core/usa-education-budget-analysis> then go down to Data Files section then choose the csv download option of the *data* file. Change the name of the downloaded file to year\_educationBudget for easier access.
* Dataset consists of useful information about BUDGET AUTHORITY BY AGENCY in the range 1976–2022.

Gross Domestic Value(GDP) comes from DataHub <http://datahub.io/core/gdp/r/gdp.csv> since it is regularly updated and includes all country codes.

Note that data in *data/budget.csv* starting 2017, the value is estimate value

9, Federal Reserve Interest Rates, 1954-Present:

* Bibliographic citation: Kaggle. (2020). Federal Reserve Interest Rates, 1954-Present. [Data file]. Retrieved from <https://www.kaggle.com/federalreserve/interest-rates>
* The interest rate data was published by the Federal Reserve Bank of St. Louis' economic data portal. The gross domestic product data was provided by the US Bureau of Economic Analysis; the unemployment and consumer price index data was provided by the US Bureau of Labor Statistics.
* How to obtain original data file: Readers can go directly to link [https://www.kaggle.com/federalreserve/interest-rates then click Download (26](https://www.kaggle.com/federalreserve/interest-rates%20then%20click%20Download%20(26) KB). Change the name of the downloaded file to month\_interestRates for easier access.
* This dataset includes data on the economic conditions in the United States on a monthly basis since 1954. The federal funds rate is the [interest rate](https://www.kaggle.com/federalreserve/interest-rates) at which depository institutions trade federal funds (balances held at Federal Reserve Banks) with each other overnight.