

# Report on data collection and cleaning

Chenning Xu

## 1 NIPA data

### 1.1 Quartelry NIPA data:

Source: NIPA <https://www.bea.gov/itable/national-gdp-and-personal-income>

Time: 1947Q1 to 2024Q2

Note: All data are seasonally adjusted at annual rate, in \$ billions.

### 1.2 Annual NIPA data:

Source: NIPA <https://www.bea.gov/itable/national-gdp-and-personal-income>

Time: 1928 to 2023

Note: All data are in \$ billions.

### 1.3 Variables and their location in NIPA tables

1. Personal saving(ps): table 5.1 line 9.
2. Net government saving (gs): table 5.1 line 10.
3. Net investment (ni): table 5.1 line 49.
4. Foreign saving (fi): table 4.1 line 33.
5. Net investment (nd): table 1.12 line 16.
6. Corporate profit with inventory and depreciation adjustment (cpwi): table 1.12 line 15.
7. Corporate profit without adjustments: table 1.12 line 45.
8. Statistical discrepancy: table 5.1 line 42

## 1.4 Data cleaning process

1. I copied the data I need from the NIPA tables, and pasted them into a new CSV file(data\_quarterly.csv and data\_annual.csv), which is in wide form without column names.
2. I imported them into Stata and transpose(xpose function) them into long form, and add variable names to the columns.
3. The processed data are saved in the processed\_data folder.

## 2 CBO data

Source: [https://github.com/US-CBO/eval-projections/blob/main/input\\_data/baselines.csv](https://github.com/US-CBO/eval-projections/blob/main/input_data/baselines.csv)

Time: 1982 to 2024, annual Note: All data are in \$ billions.

Data cleaning process: I imported data into stata. Deleted forecast for variables other than deficit. Keep only the forecast for next year since CBO forecasts 10 years. Keep only the last forecast for the next year in one year since CBO forecasts many times in one year for the next year. The processed data is saved in the processed\_data folder.

## 3 SP500 data

Source: <http://www.econ.yale.edu/~shiller/data.htm>

Time: 1871 to 2024, monthly

Data cleaning process: I cropped EPS and price data from Shiller's excel sheet (sp.xls), and saved in sp.csv. Since data are monthly, I transformed them into quarterly data. Price % change is calculated.