Macrosynergy Data Project

Alvaro Crespo Serrano

June 2024

1 Data process

The first step is to read the file and create a Pandas DataFrame to store all the data:

```
file_path = 'data/Data.csv'
data = pd.read_csv(file_path)
```

Once all the data is in a DataFrame, it needs to be cleaned and prepared for proper analysis. Functions such as `data.info()` and `data.head()` can provide information about the data format, columns, and the number of elements:

#	Column	Non-Null Count	Dtype
0	Date	7514	object
1	General government debt, % of GDP: net JPY	7514	float64
2	General government debt, % of GDP: net GBP	7514	float64
3	General government balance, % of GDP: overall JPY	7514	float64
4	General government balance, % of GDP: overall GBP	7514	float64
5	Main producer price index: %oya, 3mma GBP	6793	float64
6	Main producer price index: %oya, 3mma Japan	7514	float64
7	Merchandise trade balance ratio change (sa): 3M/3M Japan	6373	float64
8	Merchandise trade balance ratio change (sa): 3M/3M GBP	6361	float64
9	Generic government bond returns: 5-year maturity Japan	3063	float64
10	Generic government bond returns: 5-year maturity GBP	4302	float64

Table 1 **Summary of the initial data provided**

```
dtypes:  float64(10), object(1)
memory usage:  645.9+ KB
```

After an initial analysis, the first column (`Date`) needs to be transformed into a common Python datetime object (datetime64) to properly represent each date. This ensures that different functions applied to it (e.g., matplotlib) can interpret the data correctly:

```
data['Date'] = pd.to_datetime(data['Date'], format='%d-%b-%y')
```

Note that the rest of the values are already in the correct numeric form (`float`), so no additional transformation is needed. The next step is to separate the data into two DataFrames, one for each of the countries, as their data is independent of each other. They will only have the first column (`Date`) in common:

```
japan_data = data[['Date', ' General government debt, % of GDP: net JPY',
                   'General government balance, % of GDP: overall JPY',
                   'Main producer price index: %oya, 3mma Japan',
```

```
                               ' Merchandise trade balance ratio change (sa): 3M/3M Japan',
                               ' Generic government bond returns: 5-year maturity Japan']]

uk_data = data[['Date', 'General government debt, % of GDP: net GBP',
                ' General government balance, % of GDP: overall GBP',
                'Main producer price index: %oya, 3mma GBP',
                ' Merchandise trade balance ratio change (sa): 3M/3M GBP',
                'Generic government bond returns: 5-year maturity GBP']]
```

This separates the information from the initial data frame (`data`) with 11 columns to two dataframes with 6 columns each one, one for the UK (`uk_data`) and the other for Japan (`japan_data`).

Another step that can be performed to manage the data more comfortably is to simplify the DataFrame column names, keeping in mind that all units are presented as % of GDP:

```
# Rename columns for easier plotting
japan_data.columns = ['Date', 'Net Debt', 'Overall Balance', 'PPI',
                      'Trade Balance', 'Bond Returns']
uk_data.columns = ['Date', 'Net Debt', 'Overall Balance', 'PPI',
                   'Trade Balance', 'Bond Returns']
```

Finally, as shown in Table 1, the number of elements for the different columns is not consistent. To get an initial visualization of the data, all values are plotted with respect to time in the following Figure 1. Economic indicators are represented with respect to the left axis in % of GDP, while the bond returns are in the right axis as % of change:

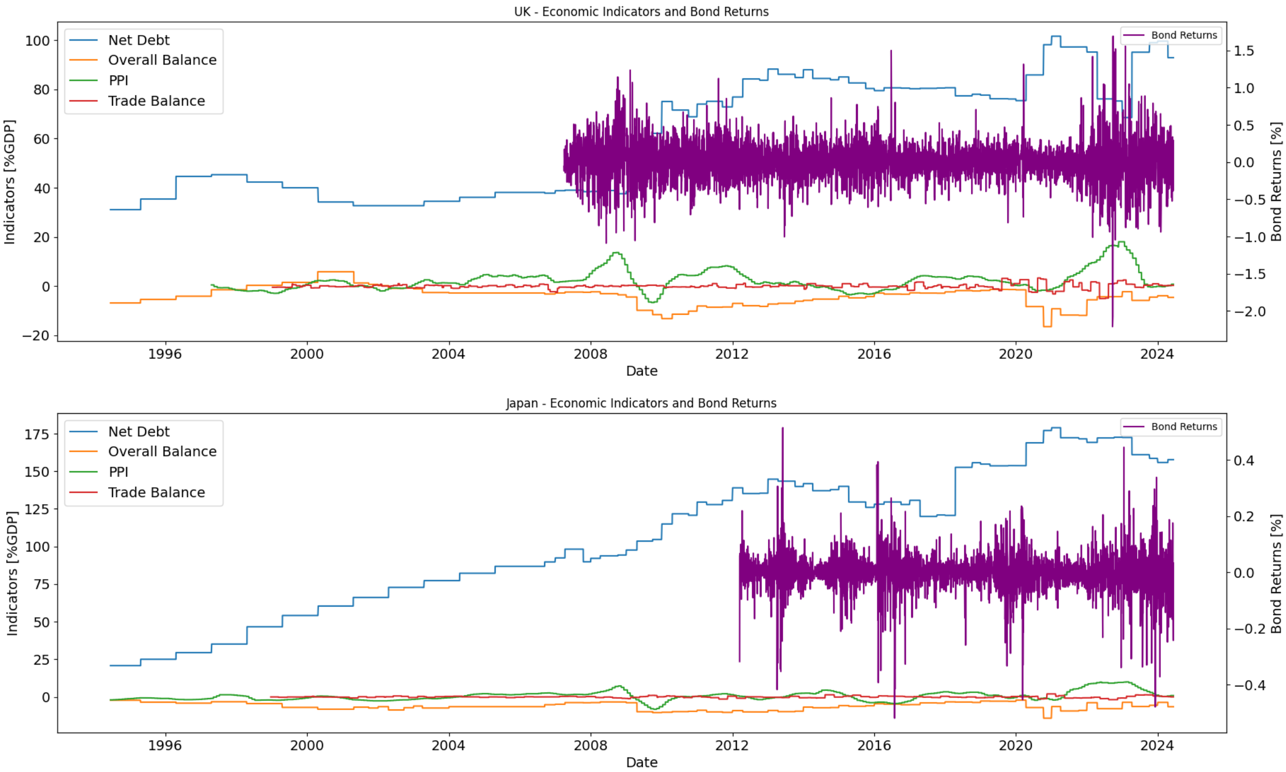

Figure 1: Evolution of economic indicators and bond returns during time for Japan and UK.

As the goal of the analysis is to see the influence of the indicators on bond returns, a time analysis is performed to determine when the data starts to be available for each column. This is accomplished with a simple loop through the dataframe columns and a dictionary:

```
# Initialize an empty dictionary to store the first non-null dates
first_non_null_dates = {}
# Iterate over each column and find the first non-null value
for column in data.columns[1:]:
    first_valid_index = data[column].first_valid_index()
    first_non_null_dates[column] = data.loc[first_valid_index, 'Date']
```

Column	**First Non-Null Date**
General government debt, % of GDP: net JPY	1994-06-13
General government debt, % of GDP: net GBP	1994-06-13
General government balance, % of GDP: overall JPY	1994-06-13
General government balance, % of GDP: overall GBP	1994-06-13
Main producer price index: %oya, 3mma GBP	1997-04-18
Main producer price index: %oya, 3mma Japan	1994-06-13
Merchandise trade balance ratio change (sa): 3M/3M Japan	1998-12-21
Merchandise trade balance ratio change (sa): 3M/3M GBP	1999-01-08
Generic government bond returns: 5-year maturity Japan	**2012-03-15**
Generic government bond returns: 5-year maturity GBP	**2007-04-02**

Table 2 **First non-null date for all data columns.**

From here on, the analysis is performed starting from 2012-03-15 for Japan and from 2007-04-02 for the UK.

A final safety check is conducted to ensure that the input data provided does not contain missing values. This can be verified by using the results given by `DataFrame.isna()`. No missing days are found, as shown in Figure 2:

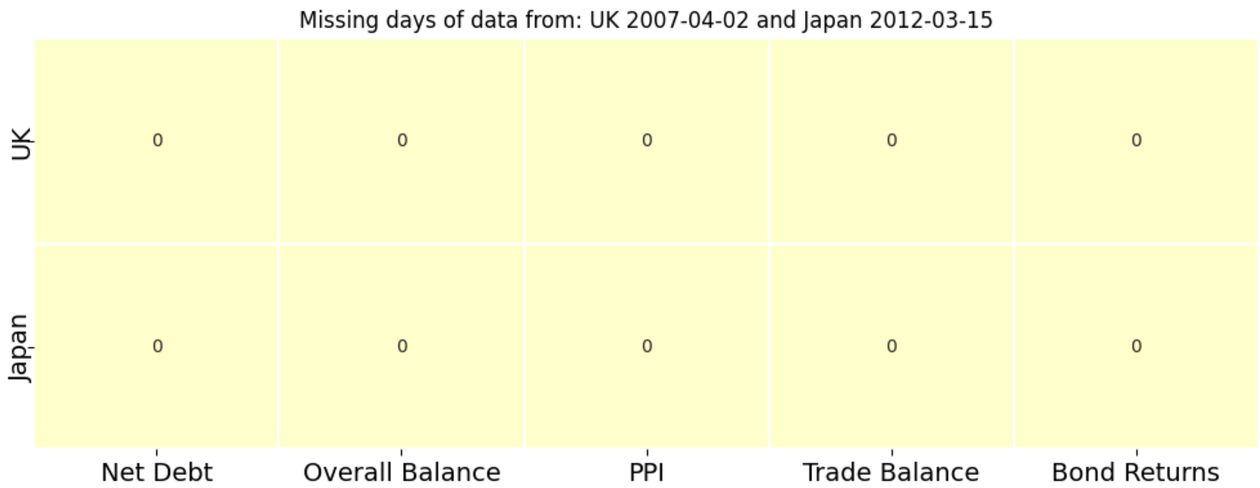

Figure 2: **Number of missing days of data during the time of analysis for UK & Japan**

2 Economic Indicators on Government Bonds

The government bonds are given as daily returns, calculated from the continuous series of zero-coupon bond yields on a daily basis for 5-year maturity. A first analysis can be done by studying the relation between this data and the different economic indicators presented as a percentage of GDP. In Figure 3, the evolution over time of all economic indicators is presented. The only noticeable relationship is during the economic recession of 2008 and the pandemic in 2019, where the different economic indicators begin to fluctuate, resulting in an increase in the variance of daily bond returns.

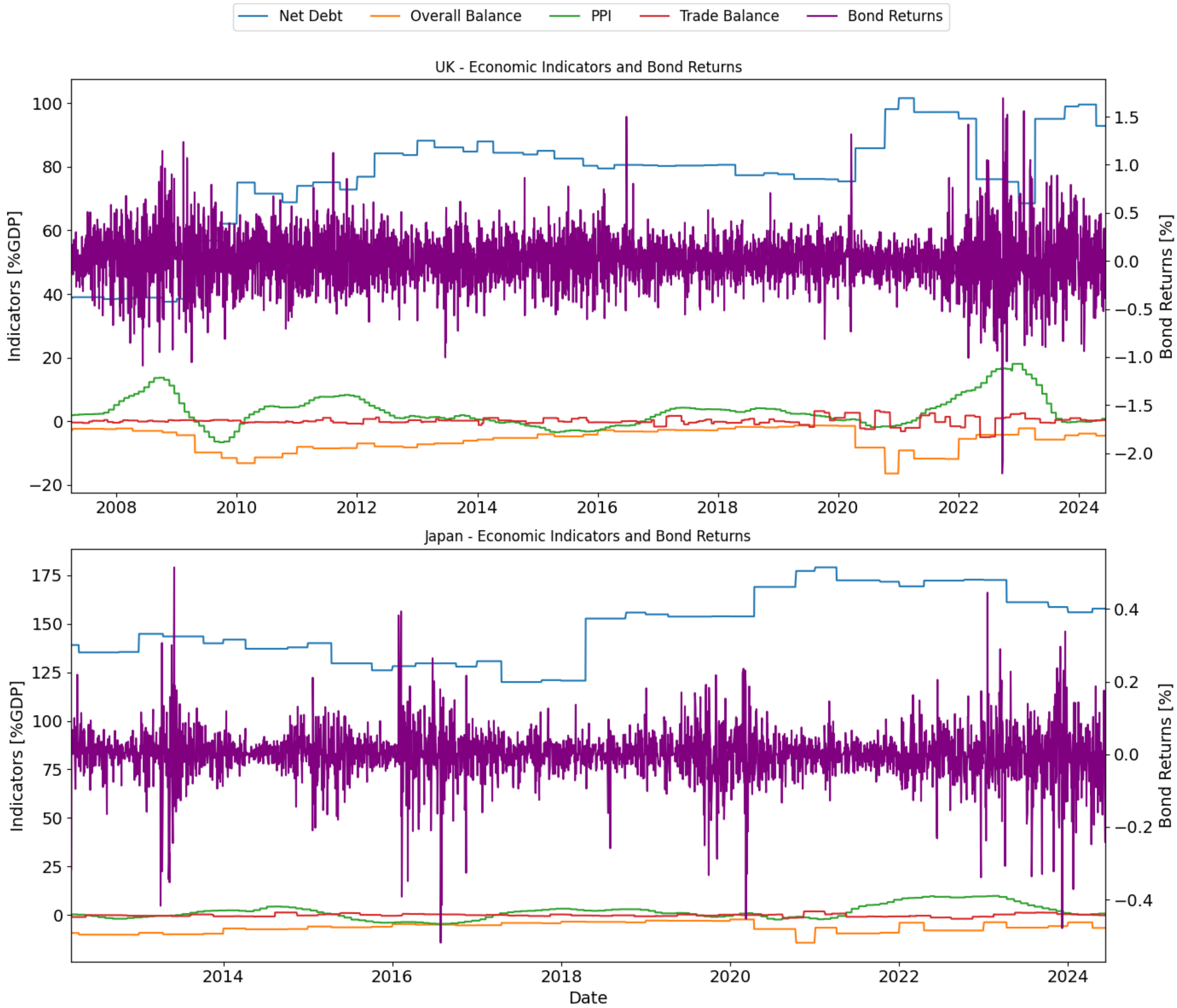

Figure 3: Evolution of indicators and bond returns during the analysis time for Japan and UK.

For example, in the UK from 2008 to 2012, the PPI and overall balance start to fluctuate because of this global economic recession, and at the same time, the daily bond returns become more unstable.

Furthermore, the cumulative bond returns can be obtained using the cumulative product of the daily bond returns. This is calculated as follows in the Python script.

```
def calculate_cumulative_returns(df, column):
    daily_returns = df[column] / 100
    cumulative_returns = (1 + daily_returns).cumprod() - 1
    return cumulative_returns * 100
```

The evolution over time is shown in Figure 4. For UK bonds, it shows stronger growth and higher returns over the period, suggesting more favorable economic conditions. In contrast, Japan bonds have lower growth and higher volatility.

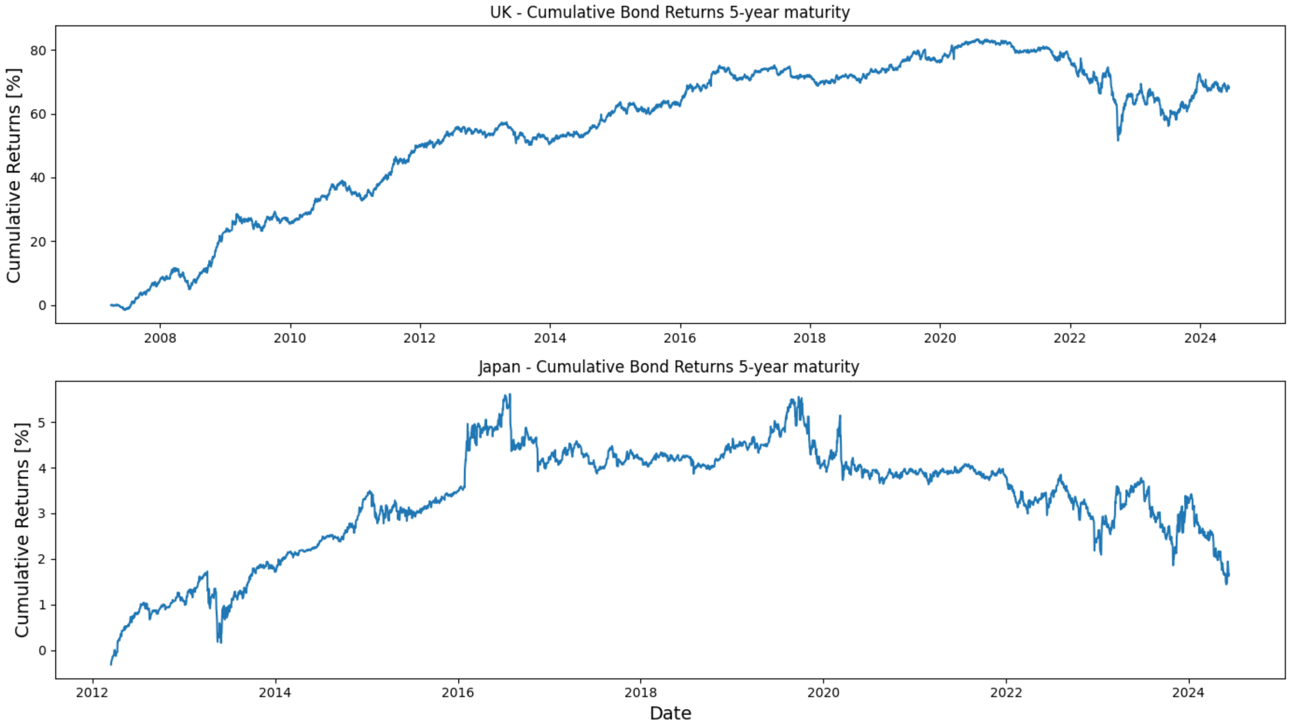

Figure 4: Evolution of the cumulative bond returns for Japan and UK.

The volatility can be calculated by computing the standard deviation of the daily bond returns and multiplying it by the square root of the number of trading days (252). In Python, this is done with the following function, which gives an annualized volatility for UK bond returns of 4.14% and 1.08% for Japan.

```
def calculate_volatility(df, column):
    df[f'{column}_Decimal'] = df[column] / 100  # Convert to decimal
    daily_volatility = df[f'{column}_Decimal'].std()
    annual_volatility = daily_volatility * np.sqrt(252)
    return annual_volatility
```

Finally, the correlation between the economic indicators and the bond returns can be computed for all elements using the `DataFrame.corr()` method. The matrix in Figure 5 shows the values obtained for the UK and Japan.

Figure 5: Correlation matrix of all elements and bond returns for UK and Japan.

In the case of the UK analysis, there is a strong positive correlation between the general government net debt and the cumulative bond returns (0.82), suggesting that fiscal policies that cause an increase in the net debt significantly influence the cumulative bond returns.

For Japan, the strongest positive correlation occurs between the general government overall balance and cumulative bond returns (0.61), indicating that a good balance is associated with better cumulative bond returns.

Daily bond returns seem to be independent of other economic indicators for both countries, as the correlation is nearly 0.