Assignment No. 1

Deadline: September 30, 2024

Solutions Available: October 1, 2024

Background

Gross Domestic Product (GDP) measures the monetary value of final goods and services—those purchased by the final user—produced within a country over a specified period (such as a year). It accounts for all output generated within a country's borders, including both market-produced goods and services as well as certain non-market production, such as defense and education services provided by the government.

GDP is typically calculated by national statistical agencies, which compile data from various sources while adhering to, The System of National Accounts (SNA), which is the internationally agreed standard set of recommendations on how to compile measures of economic activity. In Pakistan, the `Pakistan Bureau of Statistics (PBS)` is responsible for calculating GDP.

There are three primary methods used to estimate GDP:

- The expenditure approach,
- The production approach, and
- The income approach

Each of these methods provides a unique perspective on economic activity, yet they all aim to capture the same overall economic output.

In theory, the three approaches to measuring GDP should produce identical values. This is because the total expenditure on goods and services must equal the value of those goods and services produced, which in turn must match the total income earned by the factors involved in their production.

In national accounts data, estimates of the true value of gross domestic product (GDP) from the expenditure approach should equal those from the value-added approach, i.e.,

$$GDP_E = GDP_{VA} \tag{1}$$

$$C + I + G + X - M = A_g + I_n + S_e + T_a \tag{2}$$

where

GDP_E = expenditure measure of GDP

GDP_{VA} = value-added measure of GDP

C = consumption spending

I = total investment

G = government spending

X = exports of goods and services

M = imports of goods and services

A_g = agriculture

I_n = industry

S_e = services

T_a = indirect taxes less subsidies

In general, the statistical discrepancy (SD) is reported on the expenditure side, such that equations (1) and (2) become:

$$\widehat{GDP_E} + SD = \widehat{GDP_{VA}} \tag{3}$$

However, since each measurement approach is conducted independently, discrepancies can arise in the resulting estimates. These differences are referred to as `statistical discrepancies (SD)`.

$$SD = \widehat{GDP_{VA}} - \widehat{GDP_E} \tag{4}$$

$$SD\_percentage = \left(\frac{\widehat{GDP_{VA}} - \widehat{GDP_E}}{\widehat{GDP_{VA}}} \right) \times 100 \tag{5}$$

Presently, no consensus exists among the profession of national accountants on how to deal with discrepancies. The discrepancies in most cases have been in a range of plus or minus 1 percent.

$$SD\_percentage = \left(\frac{\widehat{GDP_{VA}} - \widehat{GDP_E}}{\widehat{GDP_{VA}}} \right) \times 100 \approx -1\% \leq SD\_percentage \leq 1\% \tag{6}$$

Each of these two GDP_{VA} and GDP_E have Real (current basic prices in a base year) and nominal (current prices) values.

For example, in Pakistan, the PBS estimates the GDP at constant and current basic prices of base year 2015-16. For the Latest time series figures of GDP, See Table 04 — Gross Domestic Product of Pakistan (at current basic prices of 2015-16, and Table 05 — Gross Domestic Product of Pakistan (at Constant basic prices of 2015-16).

The Background section is based on Camingue et al. [2008], Grimm [2007], Bloem [1997]

Problem Setup

The Dataset **IFS_09-16-2024 16-36-10-58_timeSeries.csv** has bulk data from the International Financial Statistics (IFS).

- **Frequency:** Annual
- **Latest Update Date:** 9/16/2024
- **Sectoral Coverage:** National Accounts, Indicators of Economic Activity, Labor Markets, Prices, Government and Public Sector Finance, Financial Indicators, Balance of Payments, International Investment Position, International Reserves, Fund Accounts, External Trade, Exchange Rates, and Population.

Questions/Tasks

Question 1

Read the CSV file, and initially import all data as strings.

Question 2

Explore the data:

1. Print the data
2. Get the dimension with and without NAN values
3. Get the names of the columns
4. Get count of unique countries, based on column `"Country Name"`
5. Get count of unique countries, based on column `"Country Code"`
6. Get count of unique indicators, based on column `"Indicator Code"`

Question 3

Subset the data :

1. Select years from 2010 till 2023
2. Filter data for the indicators given in Table 1. Use the column `"Indicator Code"`

Table 1: National Accounts Data

Abbr	Indicator Name	Indicator Code
	National Accounts, Current Prices	
N	Gross Domestic Product, Nominal, Domestic Currency	NGDP_XDC
N1	Household Consumption Expenditure, incl. NPISHs, Nominal, Domestic Currency	NCP_XDC
N2	Government Consumption Expenditure, Nominal, Domestic Currency	NCGG_XDC
N3	Gross Fixed Capital Formation, Nominal, Domestic Currency	NFI_XDC
N4	Change in Inventories, Nominal, Domestic Currency	NINV_XDC
N5	Exports of Goods and Services, Nominal, Domestic Currency	NX_XDC
N6	Imports of Goods and Services, Nominal, Domestic Currency	NM_XDC
	National Accounts, Constant Prices	
R	Gross Domestic Product, Real, Domestic Currency	NGDP_R_XDC
R1	Household Consumption Expenditure, incl. NPISHs, Real, Domestic Currency	NCP_R_XDC
R2	Government Consumption Expenditure, Real, Domestic Currency	NCGG_R_XDC
R3	Gross Fixed Capital Formation, Real, Domestic Currency	NFI_R_XDC
R4	Change in Inventories, Real, Domestic Currency	NINV_R_XDC
R5	Exports of Goods and Services, Real, Domestic Currency	NX_R_XDC
R6	Imports of Goods and Services, Real, Domestic Currency	NM_R_XDC
	National Accounts, Statistical Discrepancies	
N7	Statistical Discrepancy, Nominal, Domestic Currency	NSDGDP_XDC
R7	Statistical Discrepancy, Real, Domestic Currency	NSDGDP_R_XDC

Question 4

Make sure the index runs from 0 to n for each of the following tasks, also ensure it does not affect your dataframe in any way:

1. Count the total number of entries for each indicator in the `"Indicator Code"` column.
2. Count the total number of entries for each country in the `"Country Name"` column, and sort the results by count in descending order.
3. Count the total number of entries for each country in the `"Country Name"` column, then filter the results to show only countries with 14 or more entries. Sort the results by `"Country Name"` in descending order.
4. Perform the task in (c) but print the complete list of results.
5. Count the total number of entries for each attribute in the `"Attribute"` column. Did you find only one attribute 'Value'?

Question 5

1. Create a new DataFrame called "indicator_dict" that contains unique pairs of `"Country Name"` and `"Indicator Code"`.
2. The index in above should run from 0 to n
3. This new dataframe is expected to have 16 rows and two columns, `"Country Name"` and `"Indicator Code"`.
4. Initialize an empty list called data.
5. Iterate through each row of the indicator_dict DataFrame.
6. For each row, append a list containing the row index (starting from 1 instead of 0), `"Indicator Code"`, and `"Country Name"` to the data list.

Question 6

1. Filter the dataset to remove all entries where the `"Country Name"` is 'Euro Area'.
2. Filter the dataset to include only countries with a count of 14 or more entries.
3. Remove the columns `"Attribute"`, `"count"`, and `"Indicator Name"` from the dataset.

Question 7

It is now preferable to reshape the data from wide to long form so that:

1. The years are organized under a new column named `"Year"`.
2. The values are organized under a new column named `"Value"`.
3. Your DataFrame will now have the following four columns: `"Country Name"`, `"Indicator Code"`, `"Year"`, and `"Value"`.
4.
5. Convert the `"Value"` into float64 data type
6. print data types of all four columns and confirm

Question 8

Now again reshape the data from long to wide so that:

1. Each unique indicators from column `"Indicator Code"` become separate columns.
2. The values are organized under these columns.
3. Your DataFrame will now have the following 18 columns: `"Country Name"`, , `"Year"`, and `N, N1, N2, N3, N4, N5, N6, N7, R, R1, R2, R3, R4, R5, R6, R7`. "Refer to Table 1 for details about the abbreviations."
4. Order the columns as written in 3 above.

Question 9

Replace all NA values with 0 in the 16 columns of data type float64 in the dataset.

$$Nominal\_GDP = (N1 + N2 + N3 + N4) + (N5 - N6) + N7 \tag{7}$$

$$Nominal\_GDP\_DIFF = N - Nominal\_GDP \tag{8}$$

$$Nominal\_GDP\_PERC\_SD = \left(\frac{Nominal\_GDP\_DIFF}{N}\right) \times 100 \tag{9}$$

$$Real\_GDP = (R1 + R2 + R3 + R4) + (R5 - R6) + R7 \tag{10}$$

$$Real\_GDP\_DIFF = R - Real\_GDP \tag{11}$$

$$Real\_GDP\_PERC\_SD = \left(\frac{Real\_GDP\_DIFF}{R}\right) \times 100 \tag{12}$$

Question 10

"Refer to Table 1 for details about the abbreviations." and

1. Calculate Equation 7,Equation 8, and Equation 9
2. Calculate Equation 10,Equation 11, and Equation 12

Question 11

1. Drop NAs from dataframe
2. Filter if both `"Real_GDP_PERC_SD"` and `"Nominal_GDP_PERC_SD"` are not in range as specified in Equation 6 i.e. in between -1% to 1% inclusive.

Question 12

Prepare a balanced panel such that

1. Each country must have data for years 2010 till 2023.

Question 13

1. Read the CSV file again, import all data as strings. Named your Dataframe "exchange_rate_df"
2. Select Years (2010 to 2023)
3. Filter data where column `"Attribute"` contains only the attribute *Value*
4. Filter data where column `"Indicator Code"` contains only "ENDA_XDC_USD_RATE" which is *Exchange Rates, Domestic Currency per U.S. Dollar, Period Average, Rate.*
5. Remove all columns except `"Country Name"`, and Years (2010 to 2023)
6. Reshape the dataframe from wide to long The years are organized under a new column named "Year" , and he values are organized under a new column named `"ENDA_XDC_USD_RATE"`
7. Join this dataframe with the data at you prepared at the end of Question 12. Make sure the result drop extra countries from Dataframe "exchange_rate_df" which are not in your main dataframe.

$$Real\_GDP\_USD = \left(\frac{R}{ENDA\_XDC\_USD\_RATE} \right) \tag{13}$$

$$Nominal\_GDP\_USD = \left(\frac{N}{ENDA\_XDC\_USD\_RATE} \right) \tag{14}$$

$$\tag{15}$$

Question 14

Calculate

1. `"Real_GDP_USD"` using Equation 13
2. `"Nominal_GDP_USD"` using Equation 14
3. Mean `"Real_GDP_USD"` and mean `"Nominal_GDP_USD"` across countries. Order the result in ascending order of mean`"Real_GDP_USD"`

References

Adriaan M. Bloem. Discrepancies between quarterly gdp estimates. *IMF Working Papers*, 1997(123): A001, 1997. doi: 10.5089/9781451854572.001.A001. URL https://www.elibrary.imf.org/view/journals/001/1997/123/article-A001-en.xml.

Shiela Camingue, Gemma Estrada, Juan Paolo Hernando, Edith Laviña, Nedelyn Magtibay-Ramos, Pilipinas Quising, and Lea Sumulong. A note on statistical discrepancies in the national income accounts of selected asian economies. In *Asian Development Outlook 2008: Workers in Asia*, pages 259–261. Asian Development Bank, Manila, Philippines, 2008. URL https://www.adb.org/sites/default/files/publication/27707/ado2008.pdf.

Bruce T. Grimm. The Statistical Discrepancy. BEA Papers 0071, Bureau of Economic Analysis, May 2007. URL https://ideas.repec.org/p/bea/papers/0071.html.

Notice: Plagiarism or copying of assignments will result in a grade of zero.

Please submit an HTML copy that includes all outputs and results.

Use Markdown cells for any descriptions or remarks.