Analyzing the Influence of Various Factors on Global Economies

Adithi Satish
Computer Science & Engineering
PES University
Bangalore, India
adithisatish@pesu.pes.edu

Raksha Ramesh
Computer Science & Engineering
PES University
Bangalore, India
raksharamesh@pesu.pes.edu

Shriya B Shankar
Computer Science & Engineering
PES University
Bangalore, India
shriya.shankar19@gmail.com

***Abstract*—The primary goal of this project is to use statistical analysis methods like multiple linear regression and ARIMAX to gain insights into the numerous and diverse factors that affect economies of all countries globally, from quite obvious factors like Trade and the Finance Sector, to other parameters like Poverty, Gender and Urban & Social Development. The data is obtained from the World Bank's list of World Development Indicators, collected over the last six decades.**

Analysis of these variables on a macroeconomic scale is imperative in order to gain insight as to how developed and developing nations fare, and to see if minor improvements in any sector would lead to monumental change and growth overall.

***Index Terms*—macroeconomic, world development indicators, time series analysis, multiple linear regression**

I. Introduction

The World Development Indicators (WDI) is the World Bank's premier compilation about relevant, credible and internationally comparable statistics to determine global development. It consists of enumerable indicators, encompassing measures of poverty, growth (in public and private sectors), education, trade and finance, economy, climate change, to list a few. The data provides a comparison on how countries and the world in general has fared over the span of almost 6 decades, from the 60s upto 2019 (as of now). It aims to provide a quantitative look at some of the issues that have been plaguing the world in recent times.

Parameters were chosen with respect to how much of an influence they may have on a country's economical growth and social development. The primary feature chosen as the target variable was GDP per capita growth. The solution approach is based on analysing which indicators have the highest influence on the GDP growth and how they vary in developing countries in comparison to developed countries.

Some of the parameters considered for initial exploaratory data analysis include: GDP per capita growth (in %), Population growth (in %), Crude birth and death rates (per 1000 people), Unemployment (modelled ISO estimate), Income share held by highest, lowest 10%, Government expenditure on education (% of total government expenditure).

II. Review of Literature

A. Are We on the Right Path to Achieve the Sustainable Development Goals?

The aim of this research paper [1], was to analyse if the world was on the right track to achieving the targets for the Sustainable Development Goals, set by the UN in 2015, for 2030. The paper tries to forecast the scenario in 2030, using the International Futures (IF) forecasting model to do the same. The targets considered by the authors include indicators for poverty, child mortality and morbidity, undernourishment, access to safe water and sanitation, education and electricity. The datasets used for analysis and forecasting differ with each target, and have been taken from UNICEF, WHO and the World Bank's WDI datasets designated for SDG analysis.
The scenario analysis was done using the SSP2, or the Shared Socioeconomic Pathways. They include five scenarios that frame potential global development trajectories and allow for cross-model collaboration.

The SSP2 scenario analysis done by the authors provided a glimpse of a moderately optimistic world in 2030 in which economies grow and convergence of low and high economic countries are prevalent. The indicator "undernourished child" showed the least improvement of only an additional 6.5% countries, as compared to primary school completion and access to water, both of which show immense improvements, with 76% and 72% countries achieving the targets respectively. It also provides a country and continent wise analysis, inferring that although the SDG targets chosen may be reached by minimum 40-50% of countries overall, African countries still fare the worst and don't achieve multiple, if any targets and are aptly termed MVCs, or Most Vulnerable Countries.

The study conclusively showed that, without considering exogenous factors, the world (i.e all 186 countries analysed) is on its way to achieve two out of the nine targets by 2030 - primary school completion and a decrease in child mortality.

In conclusion, this paper proves to be an incredibly strong starting point, as the dataset used here and in our problem statements seem to intersect. Along with this, the target variables analysed in this study might also be very influential with respect to our problem statement at hand. In addition to this, the study also is a great reference for models like

SSPs and IFs, although both have cons that would hinder their usage and application. The study also fails to consider exogenous variables, which proves to be a deterrent to the credibility of the results obtained. All in all, it can be viewed as a foundation for our analysis, giving a brief idea as to how macroeconomic and socioeconomic analysis is done, in order for us to improve, develop further and newer insights by bringing in other indicators into the foray.

B. Forecasting Egyptian GDP Using ARIMA Models

This paper[2], seeks to forecast the Egyptian GDP and perform time-series analysis on the same using well known methods like the ARIMA using the Box-Jenkins approach to do the same, as well as running diagnostic tests to check for most optimal parameters and homoscedasticity of residuals. This paper uses data from the World Bank over the last 5 decades (from 1965 to 2016) for analysis and forecasting, 52 data points to be precise.

The exploratory analysis done shows that the time-series in this case is non-stationary, and hence, differencing has to be applied to the dataset (the ARIMA model is used). The study shows that a difference factor of 2 works best in this case, converting the time-series into a stationary one. Following this, the p and q parameters are estimated using the ACF and PACF plots to each be equal to 1. The equation for the ARIMA(1,2,1) model is found to be

$$X_t = 0.0005 + 0.1081X_{t-1} + 1.0478\epsilon_{t-1} + \epsilon_t$$

The authors performed diagnostic checks on the model, to check the normality and stationarity of the residuals obtained, in line with the Box-Jenkins approach.

On passing all diagnostic checks and analysing the out-of-sample forecasts, the authors further infered that the Egyptian GDP is predicted to rise over the next ten years, also noting that this model is just a prediction and cannot be expected to accommodate the complex and dynamic nature of the economy.

This paper provides a crisp analysis of one of the most common forecasting methods, ARIMA(p,d,q) used in the industry, and also highlights the interdisciplinary applications of statistical and data analysis models in real-world scenarios. The dataset used for analysis and forecasting is taken from the World Bank's WDI GDP indicator, which reflects the data chosen for our problem statement. The methodology followed is very thorough and all parameters are taken care of, with the appropriate diagnostic checks done. However, further valuable insights could be drawn from taking other features that underline a nation's economy into consideration and finding out the correlation between those attributes and the GDP. In conclusion, this paper provides a very clear and concise procedure that can be followed while performing time-series analysis using the ARIMA and Box-Jenkins approach.

*C. **ADD THE OTHER TWO PAPERS HERE!!!!***

III. Dataset

IV. Initial Insights

Before deciding on how to go about forecasting and making predictions, exploratory data analysis and cleaning was performed on all the datasets in order to understand the nature of the data.

Given the large amount of data, dating way back to the 60s, it was correctly assumed that the data representing most of the indicators was incredibly sparse. For indicators like Government expenditure on education and Unemployment, there was barely any or no data recorded until the late 90s or 2000s. In addition to this, using recent data (the last two decades or so) would be able to provide much more relevant insights than data from half a century ago. In order to perform dimensionality reduction, feature selection was used to drop the attributes with more than 70% of the data missing. This was done for almost all the indicators, except for population growth and birth and death rates.

Along with this, a lot of countries had almost no data recorded as well. When looked at from a bird's eye view, it seems that most of these countries with no data are not big players on the international platform, but rather place like Gibraltar, St. Martin, San Marino, etc. It is important to note that most of our exploratory data analysis revolves around the World at large, or drawing comparisons between India and other global heavyweights like the United States, China, Germany, etc. As this is the case, only these relevant rows were considered for analysis, effectively getting rid of most of the unwanted data.

In the dataset "% of government expenditure on education", dimensionality reduction was used to drop attributes since they had 80% missing values. This was the time period ranging from 1960 - 2000. For the subsequent years, there were quite a few missing values in the dataset. If the number of missing values was ¿ 8 (2001-2016 is a time period of 16 years), the country was discarded from making estimations. If above 8, to account for this, the missing value was appended with the mean value of the country's expenditure between 2001 - 2016. In the Unemployment indicator, China had two missing values: for the years 2015 and 2016. As these are values that are already accounted for by various other agencies, forecasting them would be redundant. So, the appropriate values were filled in by cross checking with other data online.

V. Proposed Solution

VI. Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections VI-A–VI-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads—LaTeX will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as "3.5-inch disk drive".
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: "Wb/m^2" or "webers per square meter", not "webers/m^2". Spell out units when they appear in text: ". . . a few henries", not ". . . a few H".
- Use a zero before decimal points: "0.25", not ".25". Use "cm^3", not "cc".)

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \tag{1}$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

D. LaTeX-Specific Advice

Please use "soft" (e.g., `\eqref{Eq}`) cross references instead of "hard" references (e.g., `(1)`). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don't use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in LaTeX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you've discovered a new method of counting.

BibTeX does not work by magic. It doesn't get the bibliographic data from thin air but from .bib files. If you use BibTeX to produce a bibliography you must send the .bib files.

LaTeX can't read your mind. If you assign the same label to a subsubsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

LaTeX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it's supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won't be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively".
- In your paper title, if the words "that uses" can accurately replace the word "using", capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect", "complement" and "compliment", "discreet" and "discrete", "principal" and "principle".
- Do not confuse "imply" and "infer".
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al.".
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [**b7**].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left

to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

Fig. 1. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

[1] Jonathan D. Moyer and Steve Hedden. "Are we on the right path to achieve the sustainable development goals?" In: *World Development* 127 (2020), p. 104749. ISSN: 0305-750X. DOI: https://doi.org/10.1016/j.worlddev.2019.104749. URL: http://www.sciencedirect.com/science/article/pii/S0305750X19303985.

[2] Mohamed R. Abonazel and Ahmed Ibrahim. "Forecasting Egyptian GDP using ARIMA models". In: *Reports on Economics and Finance* 5 (Jan. 2019), pp. 35–47. DOI: 10.12988/ref.2019.81023.