Implementation using R to Forecast GDP

Ms. Priyanka Nandal
Assistant Professor, Dept of Computer Science and Engineering
Maharaja Surajmal Institute of Technology, GGSIPU, Delhi, India

Abstract- *A multiple regression technique is implemented using R in this work. The objective of the work is to predict the GDP (Gross Domestic Product) of a large number of countries. The top ten countries with the highest GDP are then chosen from the obtained results.*

Keywords- GDP forecasting; R; Multiple Regression

I. INTRODUCTION

For the sustainable development of any nation, GDP (Gross Domestic Product) forecasting is a vital process. GDP is the monetary value of all the finished goods and services produced within a country's borders in a specific time period. The demand for products has increased exponentially in the recent years. It has been predicted that if the current consumption pattern of energy continues, the consumption of energy in the world will increase by over fifty percent before 2030 [1]. Therefore prediction of growth rate of GDP becomes an essential process. GDP indicates the financial health of a country as a whole which is actually a hunting ground of researchers in the field of business in general and of economics in particular. The issues related to GDP have become the most concerned amongst macro economy variables. Data on GDP is regarded as an important index for assessing the national economic development and for judging the operating status of macro economy as a whole.

R is a language and environment for statistical computing and graphics [2]. R is a platform free independent language. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques, and is highly extensible. R provides an Open Source route to participation in the activity. R can be made custom-fit according to a particular application area as specialized packages can be written in it. The power of R can be utilized in forecasting GDP.

A forecast is a simulation of a model. The gross domestic product growth and its relation to energy consumption using the Data Envelopment Analysis (DEA) model was studied by Ramanathan [3]. Branch and Evans proposed a forecasting method using recursion [4]. GDP has also been forecasted using neural networks [5]. Schumacher and Breitung proposed a method of real time forecasting of GDP for monthly and quarterly data [6]. For real time GDP forecasting real time data set is used [7]. Linear regression analysis and genetic algorithms have also been used in literature for demand forecasting [8]. Bianco et al. used linear regression models to forecast GDP [9]. It has also been proposed to forecast GDP on a monthly basis for France using Bridge models [10]. Many factors models have also emerged which can be applied to large data sets that serve as an alternative for short term forecasting activity [11-17].

In this work multiple regression method is used for GDP forecasting.

Plan of paper:

This paper is organized as follows: Section II describes GDP forecasting methods. This section also includes the methodology of the used algorithm. Section III describes results and discussions. Section IV consists of conclusion and future work.

II. GDP FORECASTING METHOD

In this section the technique used for forecasting GDP is presented.

A. Multiple Regression

The general purpose of multiple regression is to learn about the relationship between several independent or predictor variables and a dependent or criterion variable.

The Regression Equation is

$$y = b0 + b1*x1 + b2*x2$$

where,
y is dependent variable, GDP in this case,
b0 is the intercept,
b1 is coefficient for first independent variable (x1), year in this case,
b2 is coefficient for second independent variable (x2) i.e., year^2.

B. Plan of work

Data is available at various resources which is in different formats. The size of data is too large. Some efficient sources of data can only be considered to obtain the dataset. The sources from where we can obtain the dataset are

a) World Bank
b) Github

In this work the dataset is collected from Github which consists of four data fields as per the requirements of the work. The four different dataset-attributes are as follows:

1) Country Name
2) Country Code
3) Year (The data is available from 1960 to 2017(for some countries only))
4) GDP value

After obtaining the data set, data fitting process is performed. In this step data is converted into a suitable form as data is available in raw format. The useful information is then extracted from this collected data.

After data fitting, the data is analyzed for studying so that economic growth of countries can be predicted for investigations such as their GDP rate, what are the factors affecting growth, etc.

The algorithm of GDP prediction used in this work is given below:

Input: data.csv file
Output: prediction.csv file and bar plot of 10 countries with maximum predicted GDP values.

Step 1: Start.
Step 2: Input data.csv file and save it to data-frame named 'data'.
Step 3: Extract unique country codes from 'data' into unique_codes.
Step 4: Set prediction period equal to 10.
Step 5: Extract current year from system date function.
Step 6: Set time_to_predict = pred_time_lapse + current_yr.
Step 7: Create an empty data-frame named 'prediction' with country name, country code and predicted GDP values as column attributes.
Step 8: Set x equal to total number of rows of unique_codes.
Step 9: For each entry in unique_codes, do steps 10-17.
Step 10: Set country_code = unique_codes$Country_code [current index].
Step 11: Extract all GDP entries from original data-frame for the corresponding country_code as country_values.
Step 12: Take year and year^2 (t2) as independent variable for multiple regression.
Step 13: Take GDP value as dependent variable for the same.
Step 14: Apply linear model function for the equation "Value ~ year + t2" on country_values using 'lm ()' function and store it into 'reg' variable, as follows:
Reg = lm(Value ~ year + t2, data = country_values).
Step 15: Calculate GDP for next 10 years as follows:
GDP = coef (reg)[1] + coef (reg)[2] * time_to _predict + coef (reg)[3] * time_to_predict^2,

Here,
coef (reg) [1] is the intercept
coef (reg) [2] is the coefficient for the first independent variable.
coef (reg) [3] is the coefficient for the second independent variable.

Step 16: Save country_code , country_name and predicted GDP value in prediction data-frame.
Step 17: Plot actual GDP values as time-series and regression fashion of GDP values and save it as jpeg format.
Step 18: Sort the prediction data-frame in decreasing order according to predicted GDP values and save it as 'prediction.csv'.
Step 19: Plot bar graph for 10 countries having maximum predicted GDP values.
Step 20: Stop.

C. Implementation

Applying multiple regression to GDP data for all the countries, the aim is to predict the GDP for next ten years. Here it shows the regression summary for the country 'USA' which was calculated using R inbuilt linear model function:

>summary (reg)
Call:
lm (formula = Value ~ year + t2, data = c_v)
Residuals:
Min 1Q Median 3Q Max
-4.126e+11 -1.147e+11 -2.799e+10 8.430e+10 8.522e+11
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.231e+16 5.949e+14 37.50 <2e-16 ***
Year -2.277e+13 5.988e+11 -38.03 <2e-16 ***
t2 5.811e+09 1.507e+08 38.56 <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 2.518e+11 on 52 degrees of freedom

Multiple R-squared: 0.9978, Adjusted R-squared: 0.9978
F-statistic: 1.198e+04 on 2 and 52 DF, p-value: < 2.2e-16

Applying the coefficient data information to predict, using the regression formula:

GDP = 2.231*10^16 + 2027 * -2.277*10^13 + 2027^2 * 5.811 * 10^09 = 2.701*10^13
= 27010 billion dollars (For country 'USA')
GDP is calculated for all 206 countries for next ten years and stored in a CSV file named 'prediction.csv'.
Fig. 1 represents the screen shot f the Rstudio interpreter. Fig. 2 represents the regression summary for the USA country.

III. RESULTS AND DISCUSSION

In this work we have presented a model for forecasting GDP of various countries. The GDP of a total of 206 countries is forecasted and the results show that USA has the highest GDP value. It is followed by few more countries. The GDP value of India comes at tenth position as can be seen from the Figure 3.

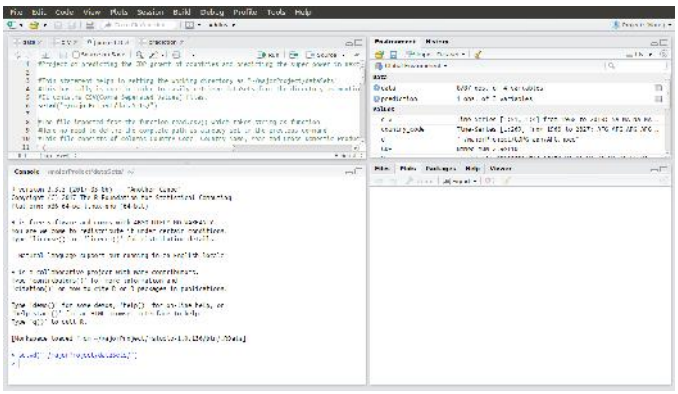

Fig. 1 Rstudio interpreter

```
Console ~/majorProject/dataSets/
> summary(reg)

Call:
lm(formula = Value ~ Year + t2, data = c_v)

Residuals:
       Min         1Q     Median         3Q        Max
-4.126e+11 -1.147e+11 -2.799e+10  8.430e+10  8.522e+11

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.231e+16  5.949e+14   37.50   <2e-16 ***
Year        -2.277e+13  5.988e+11  -38.03   <2e-16 ***
t2           5.811e+09  1.507e+08   38.56   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.518e+11 on 52 degrees of freedom
Multiple R-squared:  0.9978,    Adjusted R-squared:  0.9978
F-statistic: 1.198e+04 on 2 and 52 DF,  p-value: < 2.2e-16
```

Fig. 2 Input File:data.csv

IV. CONCLUSION AND FUTURE WORK

A multiple regression method is used to predict the GDP of large number of countries. Forecast is becoming the sign of survival and the language of business. All requirements of the business sector need the technique of accurate and practical reading into the future. Forecasts are, therefore, very essential requirement for the survival of business. No forecast can be fully correct. The more realistic the forecasts, the more effective decisions can be taken for tomorrow. There are following fields in which forecasting of GDP is going wider in near future:

- Trade
- Currency
- Living standards
- Prices, wages, jobs
- Budget
- Interest rates
- Investment
- Housing
- Share market

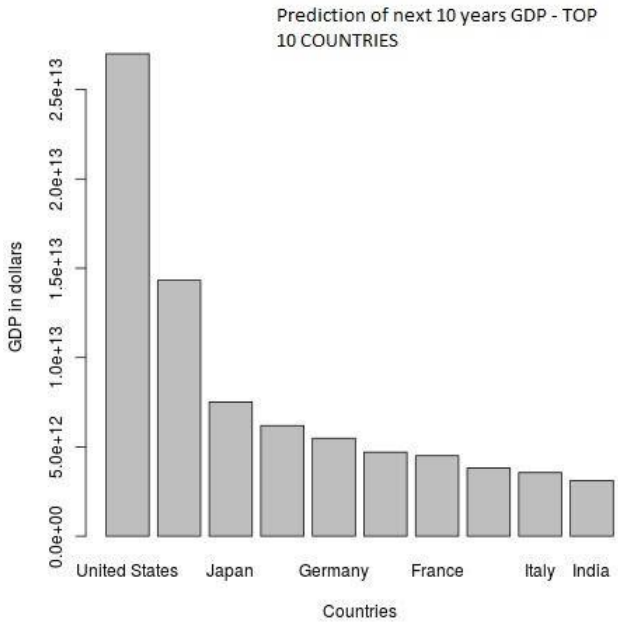

Fig.3 Bar Graph plot

REFERENCES

[1] M. Smith, K. Hargroves, P. Stasinopoulos, R. Stephens and C. Desha, *Energy transformed: sustainable energy solutions for climate change mitigation*, 2007, pp. 6.

[2] Ross and R. Gentleman. *R: a language for data analysis and graphics*, Journal of computational and graphical statistics, 1996, pp. 229-314.
[3] R. Ramanathan, *A multi-factor efficiency perspective to the relationships among world GDP, energy consumption and carbon dioxide emissions*, Technological Forecasting and Social Change, 2006, pp. 483-494.
[4] W.A. Branch and G. W. Evans, *A simple recursive forecasting model*, Economics Letters, 2006, pp. 158-166.
[5] G. Tkacz, *Neural network forecasting of Canadian GDP growth*, International Journal of Forecasting, 2001, pp. 57-69.
[6] C. Schumacher and J. Breitung, *Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data*. International Journal of Forecasting, 2008, pp. 386-398.
[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
[8] R. Golinelli and G. Parigi, *Real-time squared: A real-time data set for real-time GDP forecasting*, International Journal of Forecasting, 2008, pp. 368-385.
[9] V. Bianco, O. Manca, and S. Nardini, *Electricity consumption forecasting in Italy using linear regression models*, Energy, 2009, pp. 1413-1421.
[10] K. Barhoumi, O. Darné, L. Ferrara, and B. Pluyaud, *Monthly GDP forecasting using bridge models: Application for the French economy*, Bulletin of Economic Research, 2012.
[11] J. H. Stock and M. W. Watson, *Macroeconomic forecasting using diffusion indexes*, Journal of Business and Economic Statistics, 2002, pp. 147–62.
[12] J. Stock and M. Watson, *Forecasting with many predictors*, Handbook of Economic Forecasting, 2006.
[13] M. Forni, M. Hallin, M. Lippi and L. Reichlin, *Do financial variables help forecasting inflation and real activity in the euro area?*, Journal of Monetary Economics, 2003, pp. 1243–55.
[14] C. Schumacher, *Forecasting German GDP using alternative factor models based on large datasets*, Journal of Forecasting, 2007, pp. 271–302.
[15] C. Frale, T. Proietti, M. Marcellino and G. L. Mazzi, *A monthly indicator of the Euro area GDP*, Discussion Paper No. 7007, CEPR, 2008.
[16] E. Angelini, G. Camba-Mendez, D. Giannone, L. Reichlin and G. Runstler, *Short term forecasts of euro area GDP growth*, Discussion Paper No. 6746, CEPR, 2008.
[17] K. Barhoumi, O. Darne and L. Ferrara, *Are disaggregate data useful for forecasting French GDP with dynamic factor models?*, Journal of Forecasting, 2010, pp. 132–44.