

# File

*by Sanoj kumar*

---

**Submission date:** 16-Nov-2023 12:21AM (UTC+0530)

**Submission ID:** 2229237654

**File name:** g\_Multilinear\_Regression\_for\_Prediction\_of\_GDP\_of\_India-new.docx (184.1K)

**Word count:** 1823

**Character count:** 10580



# Using Multilinear Regression for Prediction of GDP of India

## **TEAM MEMBERS**

Anas Faham (2022CH71485)

Sanoj Kumar (2022CH11454)

Naina Jain (2022CH1221450)

Shriya Singh (2022CH11470)

November 15<sup>th</sup>, 2023

## Contents

<b>Abstract:</b> .....	3
<b>Introduction:</b> .....	3
<b>Problem Formulation:</b> .....	4
3.1 Regression Formulation .....	4
3.2 Variable Formulation.....	5
<b>Numerical Analysis:</b> .....	6
Data:.....	6
Expressions: .....	7
<b>Result:</b> .....	8
.....	9
<b>Conclusion:</b> .....	9
<b>Level-2:</b> .....	10
Increase of Independent Variables: .....	10
Changing from Matrix Method to Derivative Method .....	10
<b>Self-Assessment:</b> .....	11
<b>Graphs:</b> .....	12
<b>References:</b> .....	13

## Abstract:

The Gross Domestic Product (GDP) of a country is a topic of key interest for various reasons. The overall economic health of a country is strongly correlated to the GDP. In this paper, we explore how to predict future variations of the GDP based on several variables such as population, interest rates, unemployment rates, etc. In addition, some new variables are also added such as p-values to assess the importance of each variable. This review, proposes a method to fit a multiple linear regression model to determine a linear relationship between these variables.

## Introduction:

The use of multiple linear regression in predicting a nation's Gross Domestic Product (GDP) is a powerful analytical approach that allows for a comprehensive understanding of the complex interplay between various macroeconomic factors. This statistical method provides a structured framework to explore and quantify the relationships between the nation's economic output and a multitude of independent variables, such as population size, interest rates, and unemployment rates.

In the context of economic modelling, linear regression serves as a valuable tool by offering insights into how changes in independent variables relate to changes in the dependent variable, which, in this case, is the GDP. The basic premise of linear regression involves establishing a linear equation that best fits the observed data, enabling the estimation and interpretation of the impact of each independent variable on the overall economic performance.

For a single independent variable, the analysis employs simple linear regression. However, the real-world dynamics of national economies are typically influenced by a myriad of factors. To address this complexity, multiple linear regression is introduced.

Multiple linear regression extends the analysis to accommodate several independent variables, allowing for a more nuanced and realistic representation of the intricate web of economic influencers. The inclusion of multiple independent variables enables a more comprehensive understanding of the multidimensional factors impacting a nation's economic performance.

The general linear model, expressed in matrix form, further enhances the analytical sophistication. This matrix representation encapsulates the entire system of equations, providing a holistic view of the relationships between the dependent variable and multiple independent variables across various observations.

The investigation into the predictive capacity of multilinear regression in the realm of a nation's economic indicators is not merely a statistical exercise but a crucial endeavour with real-world implications. This work fosters the potential for future research and discoveries, potentially uncovering new patterns and relationships that can further refine our ability to predict and interpret a nation's economic performance. Ultimately, the use of multiple linear regression stands as a cornerstone in the empirical toolkit for understanding and forecasting complex economic systems.

## Problem Formulation:

### 3.1 Regression Formulation

For performing this multiple linear regression(mlr) to estimate GDP, we will use least square method as a base model. This method squares and sums all the vertical deviations from each observation(data points) to the line, and our aim is to minimize the error. Here we have taken  $(x_i, y_i)$  ( $i=0,1,2,...n$ ) data points. The least square method is commonly used in fitting regression line.

The system of equation for the given data points  $(X_i, Y_i)$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

can be written as:

$$[y_1 : y_n] = [1 x_{11} \dots x_{1k} : \vdots : 1 x_{n1} \dots x_{nk}] [\beta_0 : \beta_n] + [e_1 : e_n]$$

We can write this as  $b = Ax$ :

$$\text{Where } b = [y_1 : y_n], A = [1 x_{11} \dots x_{1k} : \vdots : 1 x_{n1} \dots x_{nk}], x = [\beta_0 : \beta_n]$$

Here:  $1 x_{11} \dots x_{1k} : \vdots : 1 x_{n1} \dots x_{nk}$  represent the different variables on which the GDP is dependent on.

On implementing least square method we will get  $\hat{\beta}$  which is

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad \text{here } X^T \text{ is transpose of matrix } X.$$

Thus we know the values of  $\beta_i (i=0,1,\dots,n)$

Hence, the predicted value of  $Y$  is

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y.$$

### 3.2 Variable Formulation

Here, the  $Y$  matrix is the GDP which is the dependent variable. The  $X$  matrix contains all the independent variables.

<sup>1</sup>  
Dependent Variable: GDP (Annual Percent) – GDP

Independent Variables:

- Population
- Disposable Income
- Unemployment Rate

## Numerical Analysis:

In our model, we use independent variables to establish a linear regression between the overall GDP and various macroeconomic variables. There are several variables which affect the GDP, such as population, labour force, unemployment rate, interest rate, maturity rate, disposable income, etc. The most important factors were determined using the corresponding p-values (provided by M. Samiyu):

Input Variable	P-Value
Population	**
Interest Rate	
Maturity Rate	**
Disposable Income	***
Unemployment Rate	***
House Price Index	***

**Key:** \*\*\* - High Significance    \*\* - Moderate Significance    No Significance (Empty)

In our paper, Interest Rate & Maturity Rate was neglected due to low P-Values, and due to lack of information, House Price Index was not used.

## Data:

Year	Unemployment Rate	Disposable Income	Population in Crores	GDP (Predicted)	(GDP Actual)
1990	0.0566	4.797	87.04522	220.2769	320.98
1991	0.0674	5.478	88.89418	180.69	270.11
1992	0.0682	6.385	90.7574	222.0731	288.21
1993	0.068	7.394	92.63513	272.5186	279.3
1994	0.0683	8.678	94.5262	321.6899	327.28
1995	0.0701	9.917	96.42791	358.8932	360.28
1996	0.0718	11.837	98.32812	402.7354	392.9
1997	0.0728	13.047	100.2335	446.064	415.87
1998	0.0749	15.231	102.1435	489.2466	421.35
1999	0.0771	16.711	104.05	525.4871	458.82
2000	0.0777	18.314	105.9634	575.5299	468.39
2001	0.0796	20.171	107.8971	617.9702	485.44
2002	0.081	21.357	109.8313	658.559	514.94
2003	0.0836	23.564	111.7415	698.0107	607.7
2004	0.0853	32.559	113.6265	802.6553	709.15
2005	0.087	37.145	115.4639	868.1891	820.38





## Result:

Macroeconomics variable	coefficients	estimate	P-value
Intercept	B <sub>0</sub>	-1237.04	
Interest rate	B <sub>1</sub>	-7871.21	***
Disposal income	B <sub>2</sub>	8.6401	***
Population	B <sub>3</sub>	21.384	**

Model:

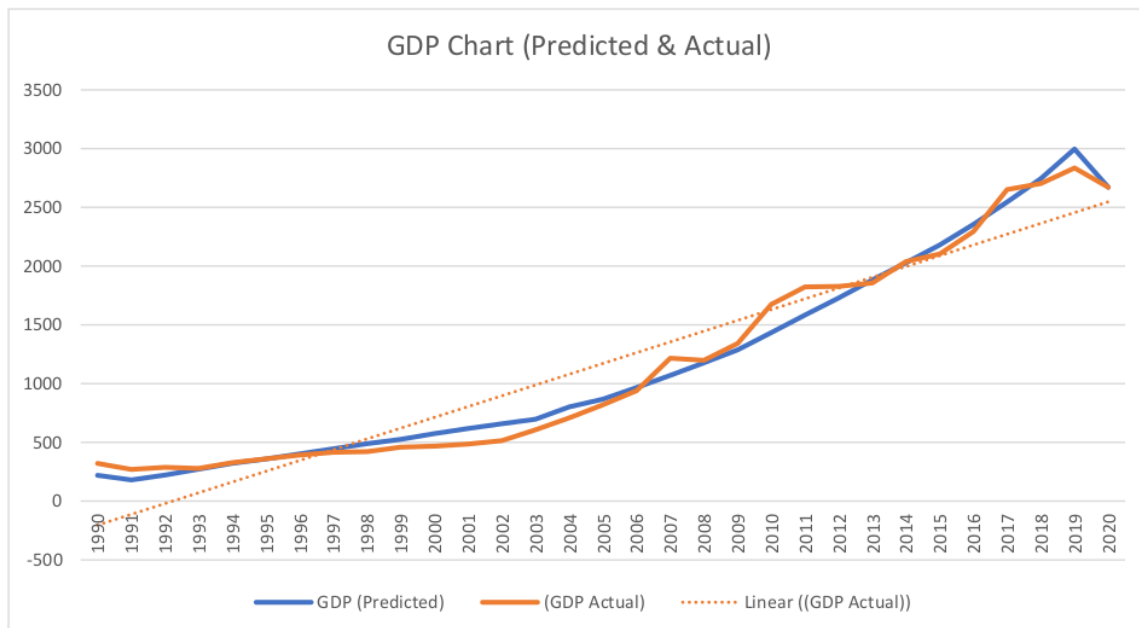
$$GDP = -1237.04 + (-7871.21) \times R + (8.6401) \times D + 21.384 \times P$$

We use data from 1990 to 2020 to train our model to predict the GDP of the Indian economy.

Now we predict the GDP of 2021 and 2022 using our model.

Year	Unemployment rate	Disposable income	Population	Predicted GDP(Billion\$)	Actual GDP
2021	7.71%	238.573	140.756	3227.32	3150.31
2022	7.33%	274.133	141.717	3585.03	3385.09

As the prediction shows, the predicted and actual GDP are within range of error, hence demonstrating the validity of the above parameters.



*Graph of GDP vs Year – Predicted & Actual (Linear Trendline Provided)*

As you can see from the above graph, the multilinear regression is quite close to the actual GDP variation. For reference, a trendline is provided showing single-variable linear regression and we can observe the variation.

## Conclusion:

In our term paper, we looked into predicting GDP by studying different economic factors. We used a method called multiple linear regression to do this. At first, we made a simple model, and then we added more factors to see if it would work better. We wanted to make sure these extra factors were actually making a difference, so we checked things like VIFs and p-values at a 5% level of significance.

It turns out that our multiple linear regression model shows a clear connection between the main features (like the original factors we considered) and the extra ones we added. This helps us understand how these different economic factors work together to influence GDP.

## Level-2:

To advance our term paper, we aim to reach Level 2 by implementing additional details. This extension will offer a more comprehensive understanding of our subject matter, enriching the depth and scope of our study.

## Increase of Independent Variables:

We know that GDP depends on many macroeconomic parameters but in our model we only used three variables so we can add the following variables to get more accurate results.

1. House price index(HPI)- It is statistics that measure the average change in the sales price of residential housing price. if the HPI is increasing, it will lead to an increase in housing investment, which will boost GDP growth of a economy
2. Corporate Profit- it is the money a firm has left over after paying all its expenses. When corporate profit is high it means the firms are doing well and hence, they produce more goods and services. This leads to a positive impact on GDP.

The addition of these parameters can have a significant impact on our model.

## Changing from Matrix Method to Derivative Method

These can be implemented by increasing the rank of the matrix. In the original code, Brute Force method is used to solve a matrix of rank 4. However, if variables are increased, the rank of the matrix also increases making the inverse of the matrix increasingly ill conditioned and difficult to find (via. Gauss Elimination or other methods).

Hence to achieve conditioned matrix, we use the derivative method which is detailed below. Here  $S_r$  represents the error which needs to be minimized:

$$\begin{aligned} S_r &= \sum_{i=1}^N e_i^2 = \sum_{i=1}^N (y_{i,measured} - y_{i,modeled})^2 \\ &= \sum_{i=1}^N (y_i - a_0 - a_1x_1 - a_2x_2 \dots - a_nx_n)^2 \end{aligned}$$

$$S_r = \sum_{i=1}^N (y_i - a_0 - a_1 x_1 - a_2 x_2 \dots - a_n x_n)^2$$

$$S_r = S_r(a_0, a_1, \dots, a_n)$$

$$\Rightarrow S_r|_{\min} \Rightarrow \frac{\partial S_r}{\partial a_0} = 0, \frac{\partial S_r}{\partial a_1} = 0, \dots, \frac{\partial S_r}{\partial a_n} = 0$$

Upon solving the series of equations for  $s_r$ , we can get  $n$  different equations which we can then subsequently solve using Gauss Elimination. Here, we have avoided trying to find the inverse of  $n$ -order matrix, hence avoiding ill-conditioned matrices and errors.

### Self-Assessment:

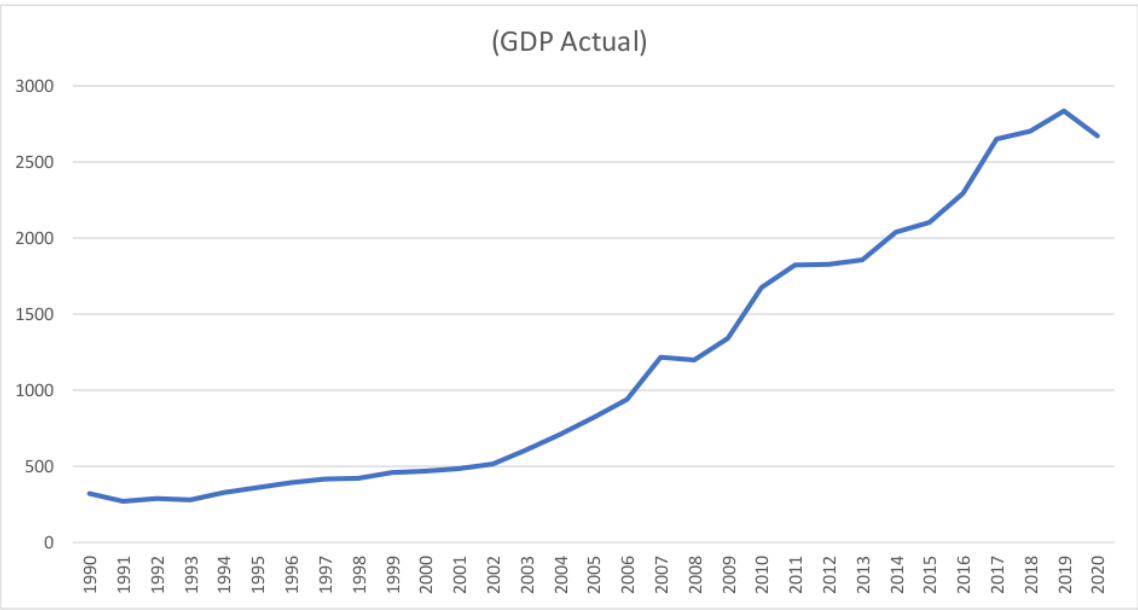
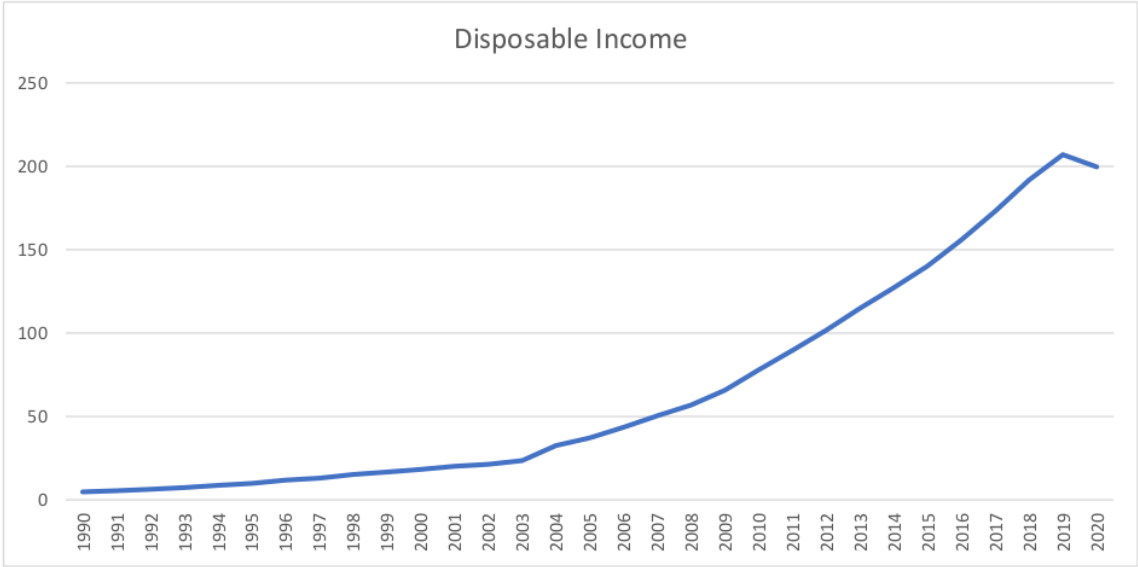
Our term paper served as a tremendously beneficial learning experience for us. We delved into the realm of predicting a country's GDP through the application of multiple linear regression. As we endeavoured to replicate the original results, we grappled with various challenges associated with preparing the data, particularly in coding for population rates.

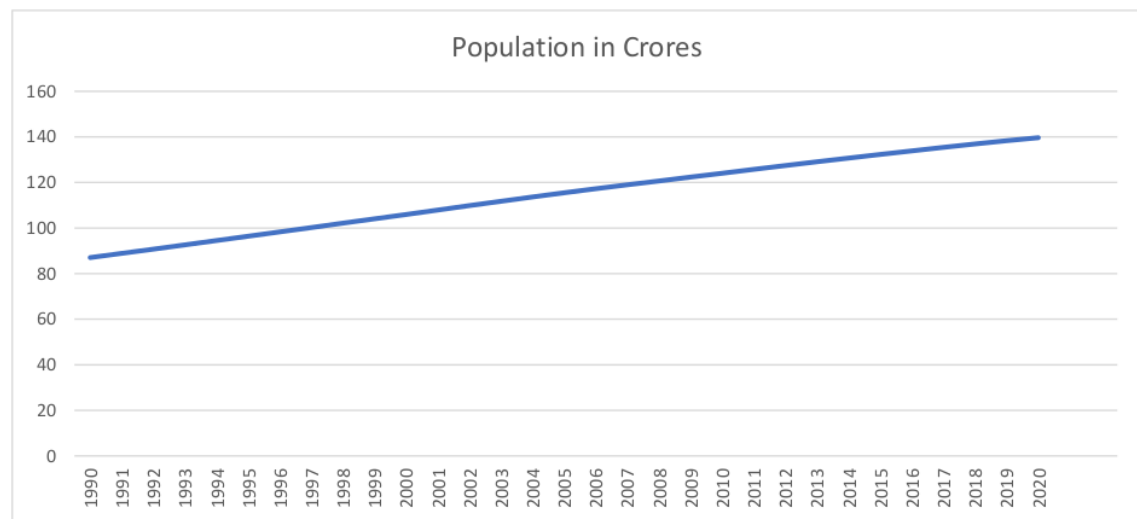
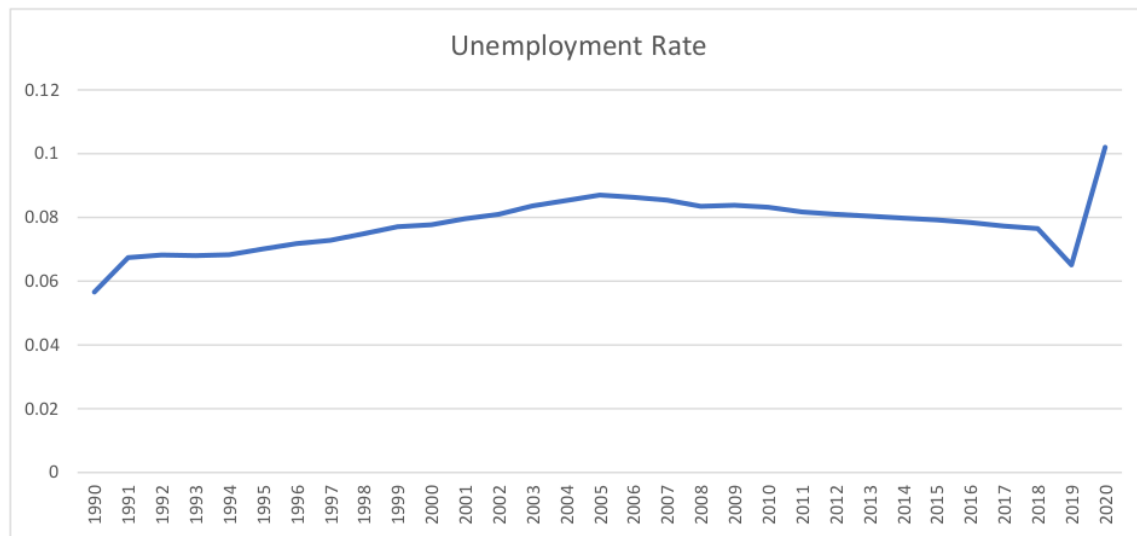
The task of handling diverse datasets introduced its own set of hurdles, primarily stemming from missing information. This complexity rendered the process quite demanding and time-consuming. However, despite these challenges, our attention to detail played a pivotal role in enhancing our proficiency in the tasks of organizing and cleaning the data.

Moreover, we extended our research endeavours by taking a step further and delving into level 2, extending the number of independent variables the GDP should rely on and changing the method for solving the system of equations to obtain relation.

In summary, our term paper, despite the data-related challenges encountered, not only equipped us with essential insights but also facilitated substantial growth in our ability to predict a nation's GDP by incorporating population rates within the framework of multiple linear regression.

Graphs:





## References:

1. Samiyu, Mutiu, Multiple Regression Model for Predicting GDP Using Macroeconomic Variables (Part 1) (July 28, 2021).
2. Jayati Sarkar. (n.d.-a). NUMERICAL METHODS IN CHEMICAL ENGINEERING. [www.aps.org](http://www.aps.org).

3. <https://www.macrotrends.net/countries/IND/india/unemployment-rate#:~:text=India%20unemployment%20rate%20for%202022,a%202.48%25%20decline%20from%202020.>
4. <https://www.fxempire.com/macro/india/disposable-personal-income>
5. <https://www.worldometers.info/world-population/india-population/>

# File

## ORIGINALITY REPORT

5%

SIMILARITY INDEX

5%

INTERNET SOURCES

1%

PUBLICATIONS

3%

STUDENT PAPERS

## PRIMARY SOURCES

1

github.com

Internet Source

3%

2

dokumen.tips

Internet Source

1%

3

Submitted to University of Glamorgan

Student Paper

1%

4

www.coursehero.com

Internet Source

1%

5

udspace.udel.edu

Internet Source

<1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On