MTH 416A COURSE PROJECT

Analysis of Economic Development Indicators

Anupreet Porwal - 12143
Ashish Jain - 12167
Nalin Dhameliya - 12431
Shreesh Laddha - 12679

Abstract

The Gross Domestic Product (GDP) growth rate is the most important indicator of economic health. When the economy is expanding, the GDP growth rate is positive. If it's growing, so will business, jobs and personal income. In this project, we are trying to model the dependence the **GDP growth rate (gdpgr)** on a variety of other economic development indicators for 121 countries, using Multiple Linear Regression (MLR) model. In the end, four categories of countries on the basis of Human Development Index(HDI) namely - Very high, high, medium and low HDI, were used and we tried to see how this categorical variable affected our model.

1 Introduction

We have used the world Economic development data set (Source - International Monetary Fund (IMF) - International Financial Statistics, 1996) of 121 countries containing the following 12 economic development indicators which have been described below:

- **Gross National Product per capita at PPP (gnpper)** : The total value of all the goods and services produced by a country in a year determined by the relative value of different currencies. It describes the economic level of a country.

- **Gross Domestic Product (dominv)** : It consists of outlays on additions to the fixed assets of the economy plus net changes in the level of inventories. It also shows the level of investment in a country.

- **GDP Deflator (gdpdfl)** : It is a metric that accounts for inflation. The GDP deflator shows how much a change in the base year's GDP relies upon changes in the price level.

- **Agriculture Value Added (agrvlad)*** : Agriculture includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs.

- **Industry Value Added (indvlad)*** : It comprises of value added in mining, manufacturing, construction, electricity, water, and gas.

- **Export of Goods and Services (exp)*** : It represents the value of all goods and other market services provided to the rest of the world. It indicates the openness of an economy.

- **General Government Consumption (govcon)*** : It includes all government current expenditures for purchases of goods and services. It also includes most expenditures on national defense and security, but excludes government military expenditures that are part of government capital formation.

- **Resource Balance (resbl)*** : It refers to the net borrowing/lending on account of merchandise trade. Merchandise trade means imports and exports of goods across the borders.

- **Domestic Credit provided by the Banking Sector (domcrdt)*** : It refers to financial resources provided to the private sector by financial corporations, such as through loans, purchases of nonequity securities, and trade credits and other accounts receivable, that establish a claim for repayment.

- **Ratio of gross international reserves to imports (girimp)** : It denotes the strength of the foreign exchange reserves of a country.

- **Interest Spread (intsprd)** : Interest rate spread is the interest rate charged by banks on loans to private sector customers minus the interest rate paid by commercial or similar banks for demand, time, or savings deposits. It denotes the efficiency of a financial market.

- **Number of months of import cover (impcov)** : It is another measure of the foreign exchange reserves of a country.
  **Regressors are measured as a percentage of GDP*

2 Methodology

2.1 Model Fitting - I

We fit the 12 regressors mentioned above in an MLR model with the response variable as gdpgr and the following analysis were done:

- **Check for Multicollinearity** :
  It is possible that two or more regressors may linearly depend on each other, which would inflate the estimates and variance of the coefficients. Hence, a check for multicollinearity was done by analysing Variance Inflation Factors (VIFs). All the VIF's were found to be less than 5, hence no multicollinearity was detected.

- **Variable Selection** :
  A lot of regressors that we have described may represent the same information and therefore we do not need use all the regressors. For eg. agrvlad and indvlad, both describe the structure of output aspect of the economic development and therefore we do not need to add redundant information to the model. So we applied backward elimination variable selection technique to select the best 9 regressors that describe our model.

$$\textbf{gdpgr} = 0.0165 - 0.2587 * gnpper + 0.284 * dominv - 0.6219 * gdpdfl + 0.1312 * agrvlad - 0.1985 * exp + 0.2197 * resbl + 0.1994 * domcrdt + 0.1607 * girimp - 0.1557 * intsprd$$

- **Outliers Removal** :
  We used studentized residuals for the detection of outliers from our data set. A studentized residual is the quotient obtained from the division of a residual by an estimate of its standard deviation. Usually the standard deviations of residuals in a sample vary greatly from one data point to another even when the errors all have the same standard deviation, thus it makes more sense to compare residuals at different data points after studentizing them. We removed all the data points having studentized residuals with values greater than or less than ± 3. In this process we removed 6 countries namely - 'Azerbaijan', 'Angola', 'Moldova', 'Burundi', 'SierraLeone' and 'Singapore'.

- **Residual Analysis/Model Adequacy Checking** :
  There are three prior assumptions for estimation of parameters associated with an MLR model namely - Normality of the residuals, Homoscedasticity of residuals and residuals being uncorrelated and independent.
  For normality check on residuals, we looked at the Q-Q plot of residuals. We observed the graph to be roughly at an angle of $45°$, as it was expected. Thus the residuals correctly follow normal distribution.
  For checking homoscedasticity, we looked at residuals vs predicted and also residuals vs all regressor plots to observe if there was some specific pattern/trend (open/closed funnel etc.) in the plots as shown in Figure 1. However, we didn't find any such pattern and the residuals seemed to be randomly distributed. Thus the linear model assumption was verified. Find the graphs appended below.
  We used the Durbin-Watson test statistic to check for any autocorrelation within the residuals. The statistic is defined below. We found the statistic to be equal to 2.164 , which suggested no autocorrelation among the residuals.

2.2 Model Fitting - II

Apart from the 12 existing variables, we added information about the current status of the country's development, represented in terms of Human Development Index (HDI). HDI is a tool developed by the United Nations to measure and rank countries levels of social and economic development. The HDI makes it possible to track changes in development levels over time and to compare development levels in different countries.
Data regarding the same was gathered from existing Human Development Reports, United Development Programme. Since the data was collected on a ten year basis, we used the average of data of the year 1990 and the year 2000, for the approximation of HDI in the year 1996. UN had already divided the countries into these 4 categories - Very High, High, Medium and low based on their HDI. We tried to capture this information in our model using 3 dummy variables (to manage one categorical variable with four categories). This was done to observe the effect of how HDI captured the information related to the twelve already present regressors and also express response (gdpgr) more accurately.

$$hdev(x) = \begin{cases} 1 & : x \in A \\ 0 & : x \notin A \end{cases}$$

$$mdev(x) = \begin{cases} 1 & : x \in B \\ 0 & : x \notin B \end{cases}$$

$$ldev(x) = \begin{cases} 1 & : x \in C \\ 0 & : x \notin C \end{cases}$$

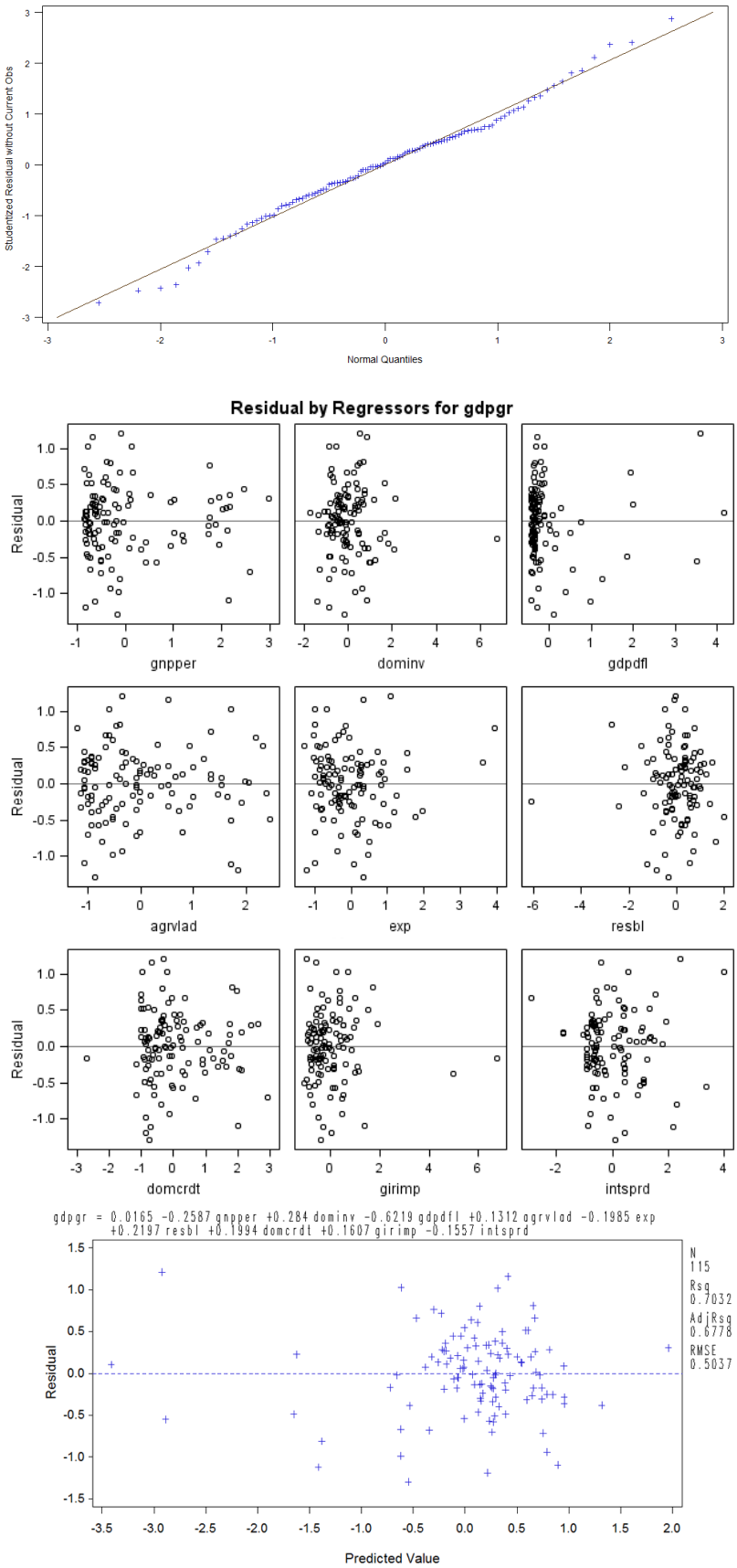

Figure 1: Q-Q Plot, Plots of residuals vs regressors and residual vs predicted values corresponsing to Model I

Where $A = \{$ Countries with very high HDI value$\}$
$B = \{$ Countries with high HDI value$\}$
$C = \{$ Countries with medium HDI value$\}$
Thus for a country with low HDI value all the three dummy variables described above will be zero. Thus preventing the problem of linear dependency with the intercept column.

An MLR model was fit on this data and the following analysis was performed:

- **Check for Multicollinearity** :
  A check for multicollinearity was done by analysing Variance Inflation Factors (VIFs). All the VIF's were found to be less than 5, hence no multicollinearity was detected.

- **Variable Selection** :
  As described above we applied Backward elimination variable selection technique to select the best 9 (out of 15) regressors that describe our model.

  $$\textbf{gdpgr} = 0.07441 + 0.2382 * dominv - 0.7252 * gdpdfl - 0.1129 * indvlad - 0.1440 * govcon + 0.1426 * resbl + 0.2643 * impcov - 0.1658 * intsprd - 0.3479 * hdev + 0.2036 * ldev$$

- **Outliers Removal** :
  We used studentized residuals for the detection of outliers from our data set. We removed all the data points having high studentized residuals. In this process we removed 7 countries namely - 'Azerbaijan','Angola','Moldova','Burundi', 'SierraLeone','Botswana' and 'Belarus'.

- **Residual Analysis/Model Adequacy checking** :
  For normality check on residuals, we looked at the Q-Q plot of residuals. We observed the graph to be roughly at an angle of 45°, as it was expected. Thus the residuals correctly follow normal distribution.
  For checking homoscedasticity, we looked at residuals vs predicted and also residuals vs all regressor plots to observe if there was some specific pattern/trend (open/closed funnel etc.) in the plots as shown in Figure 2. However, we didn't find any such pattern and the residuals seemed to be randomly distributed. Thus the linear model assumption was verified.

  We used the Durbin-Watson test statistic to check for any autocorrelation within the residuals. The statistic is defined below. We found the statistic to be equal to 2.370, which suggested a very low (insignificant) autocorrelation among the residuals.

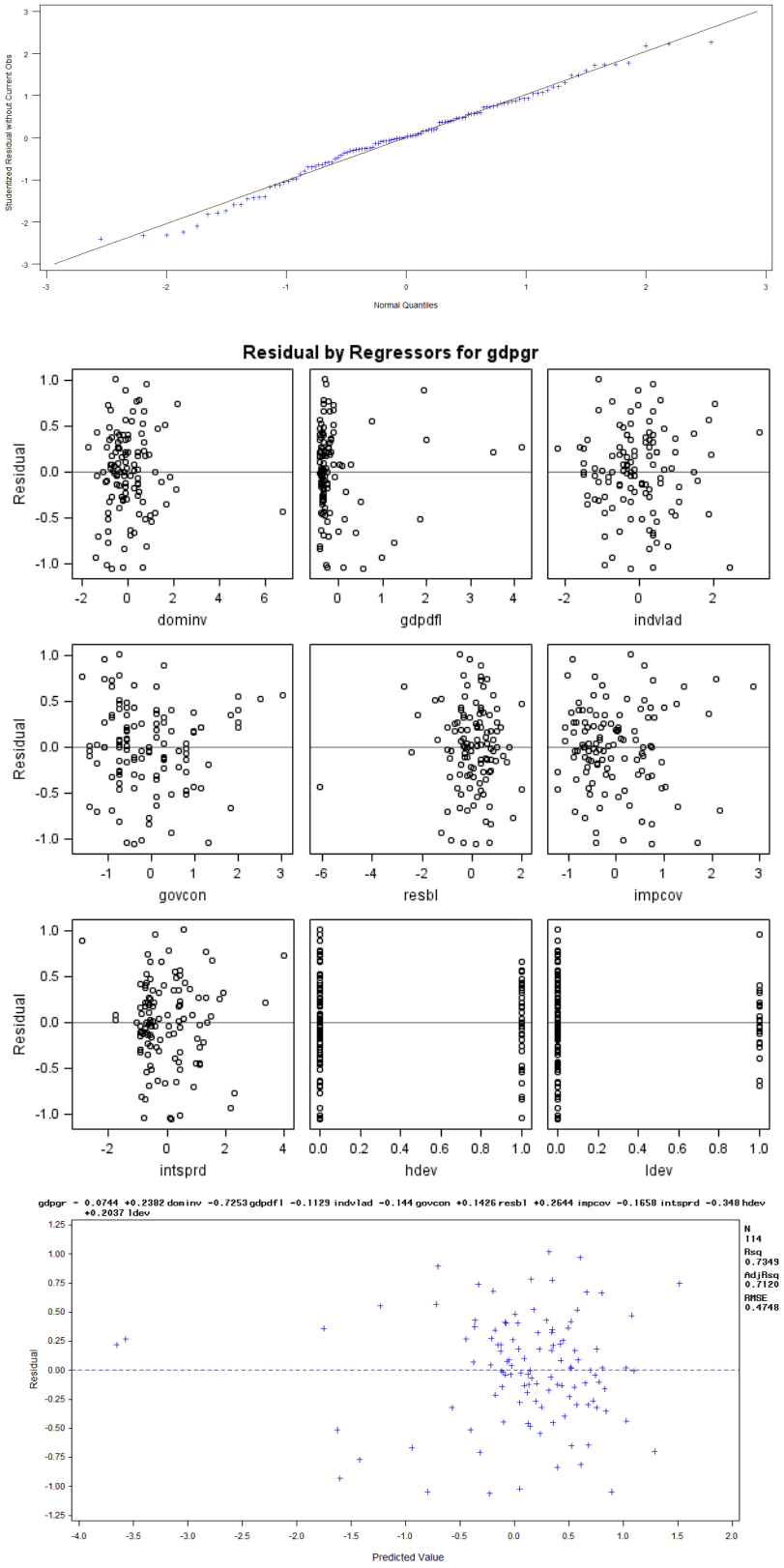

Figure 2: Q-Q Plot, Plots of residuals vs regressors and residual vs predicted values corresponsing to Model II (Own Computation)

The REG Procedure
Model: MODEL1
Dependent Variable: gdpgr gdpgr

Number of Observations Read	115
Number of Observations Used	115

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	63.12584	7.01398	27.64	<.0001
Error	105	26.64298	0.25374		
Corrected Total	114	89.76883			

Root MSE	0.50373	R-Square	0.7032
Dependent Mean	0.07455	Adj R-Sq	0.6778
Coeff Var	675.66514		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	Intercept	1	0.01649	0.04740	0.35	0.7287	0
gnpper	gnpper	1	-0.25869	0.08052	-3.21	0.0017	-0.28730
dominv	dominv	1	0.28397	0.06702	4.24	<.0001	0.31454
gdpdfl	gdpdfl	1	-0.62187	0.07027	-8.85	<.0001	-0.54108
agrvlad	agrvlad	1	0.13117	0.07404	1.77	0.0794	0.14513
exp	exp	1	-0.19852	0.06810	-2.92	0.0043	-0.19120
resbl	resbl	1	0.21967	0.07087	3.10	0.0025	0.23752
domcrdt	domcrdt	1	0.19937	0.06729	2.96	0.0038	0.22684
girimp	girimp	1	0.16068	0.04979	3.23	0.0017	0.18408
intsprd	intsprd	1	-0.15570	0.05407	-2.88	0.0048	-0.17824

The REG Procedure
Model: MODEL1
Dependent Variable: gdpgr gdpgr

Number of Observations Read	114
Number of Observations Used	114

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	64.99875	7.22208	32.04	<.0001
Error	104	23.44416	0.22542		
Corrected Total	113	88.44291			

Root MSE	0.47479	R-Square	0.7349
Dependent Mean	0.09883	Adj R-Sq	0.7120
Coeff Var	480.41427		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.07441	0.06762	1.10	0.2737
dominv	dominv	1	0.23824	0.05657	4.21	<.0001
gdpdfl	gdpdfl	1	-0.72528	0.07226	-10.04	<.0001
indvlad	indvlad	1	-0.11291	0.05906	-1.91	0.0587
govcon	govcon	1	-0.14404	0.06001	-2.40	0.0182
resbl	resbl	1	0.14260	0.06076	2.35	0.0208
impcov	impcov	1	0.26435	0.06778	3.90	0.0002
intsprd	intsprd	1	-0.16582	0.05155	-3.22	0.0017
hdev	hdev	1	-0.34797	0.12944	-2.69	0.0084
ldev	ldev	1	0.20366	0.12515	1.63	0.1067

Figure 3: Summary Statistics for Model I and II respectively

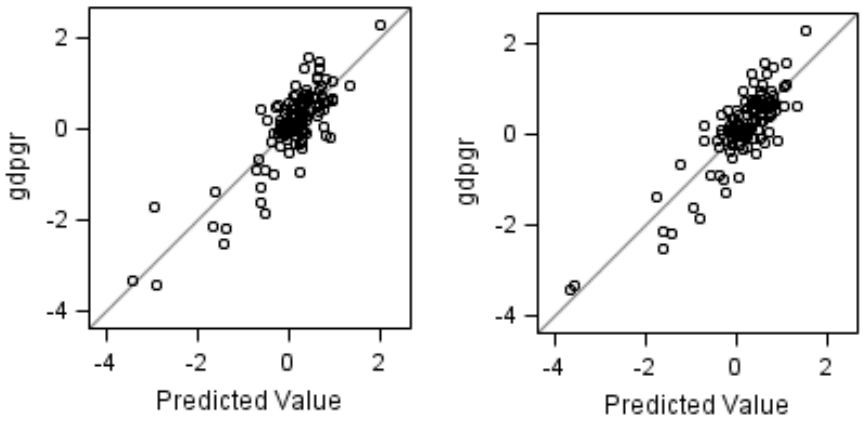

Figure 4: Plot of Observed response value and predicted response value for Model I and II respectively

3 Conclusion

The purpose of the project was to illustrate how to apply and interpret the linear multiple regression analysis to analyze the effect of economic development indicators on Gross Domestic Product. The model was strengthened by giving more categorical information in form of countries classified on the basis of Human Development Index.

After applying a set of regression specific procedures (using the statistical software SAS v.9.2), nine of the 12 regressors were chosen to represent the model I and for model II, nine out of 15 regressors were chosen to represent it, two of which were dummy variables.

The developed regression models were linear models, without signs of multicollinearity within the regressors and the residuals nearly followed a normal distribution for both the models. The computed residual values for the models presented constant variance at every level of the predictors (homoscedasticity) and did not correlate within each other (nonautocorrelation).

The value of adjR^2 from model I is 0.67 and from Model II is 0.71 suggest that the inclusion of categorical variables in the regression model helped to improve the model. However, the value of adjR^2 also suggests that GDP cannot be explained as a function of the chosen regressors in our model. More regressors like infrastructure level, labour market, health and education level of society and other relevant economic indicators may help explain the movement in GDP growth rate with a better precision.

References

[1] Chatterjee, S., Hadi, A., Price, B. (2000), *Regression analysis by example. New York: Wiley.*

[2] Resources for SAS - http://www.ats.ucla.edu/stat/sas/

[3] HDI Data source - http://hdr.undp.org/en/content/table-2-human-development-index-trends-1980-2013

[4] Wikipedia, Investopedia