MTH416A REGRESSION ANALYSIS

# ANALYSIS OF ECONOMIC DEVELOPMENT INDICATORS

**Guided By:**

Dr. Sharmishtha Mitra

**Team Members:**

Anupreet Porwal

Ashish Jain

Nalin Dhameliya

Shreesh Laddha

---

*Date : 12-04-2015*

# Outline

- Introduction
- Methodology
  - Model I - MLR model
  - Model II - MLR model with dummy variable
- Conclusion

# Introduction

- To model the dependence of GDP growth rate on a variety of other economic development indicators for 121 countries, using Multiple Linear Regression (MLR) model

- Data gathered from the world Economic development data set ( Source - International Monetary Fund (IMF) - International Financial Statistics, 1996 ) of 121 countries containing twelve economic development indicators

| Indicator and abbreviation | Aspect of Economic development |
|---|---|
| GNP per capita at PPP (GNPPER) | Income level |
| GDP growth rate (GDPGR) | Growth of economy |
| Gross domestic investment as percentage of GDP (DOMINV) | Level of investment |
| GDP deflator (GDPDFL) | Inflation |
| Agriculture value added as percentage of GDP (AGRVLAD) | Structure of output |
| Industry value added as percentage of GDP (INDVLAD) | Structure of output |
| Export of goods and services as percentage of GDP (EXP) | Openness of economy |
| General government consumption as percentage of GDP (GOVCON) | Role of government |
| Resource balance as percentage of GDP (RESBL) | Net borrowing/lending on account of merchandise trade |
| Domestic credit provided by the banking sector as percentage of GDP (DOMCRDT) | Private sector financing |
| Ratio of gross international reserve to imports (GIRIMP) | Strength of foreign exchange reserve |
| Number of months of import cover (IMPCOV) | Strength of foreign exchange reserve |
| Interest spread (INTSPRD) | Efficiency of financial market |

# Methodology - Model I

- Check for Multicollinearity
- Variable Selection
- Outlier Removal
- Residual Analysis

Dependent Variable: gdpgr gdpgr

| | |
|---|---|
| Number of Observations Read | 121 |
| Number of Observations Used | 121 |

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 12 | 69.20889 | 5.76741 | 12.26 | <.0001 |
| Error | 108 | 50.79111 | 0.47029 | | |
| Corrected Total | 120 | 120.00000 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 0.68578 | R-Square | 0.5767 |
| Dependent Mean | -8.2645E-12 | Adj R-Sq | 0.5297 |
| Coeff Var | -8.29788E12 | | |

Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Variance Inflation |
|---|---|---|---|---|---|---|---|
| Intercept | Intercept | 1 | -3.7077E-12 | 0.06234 | -0.00 | 1.0000 | 0 |
| gnpper | gnpper | 1 | -0.20344 | 0.10882 | -1.87 | 0.0643 | 3.02179 |
| dominv | dominv | 1 | 0.45798 | 0.09087 | 5.04 | <.0001 | 2.10697 |
| gdpdfl | gdpdfl | 1 | -0.39684 | 0.07710 | -5.15 | <.0001 | 1.51682 |
| agrvlad | agrvlad | 1 | 0.16056 | 0.10340 | 1.55 | 0.1234 | 2.72805 |
| indvlad | indvlad | 1 | -0.05112 | 0.08486 | -0.60 | 0.5481 | 1.83729 |
| exp | exp | 1 | -0.14053 | 0.08017 | -1.75 | 0.0825 | 1.64012 |
| resbl | resbl | 1 | 0.45364 | 0.09724 | 4.67 | <.0001 | 2.41257 |
| domcrdt | domcrdt | 1 | 0.14186 | 0.09115 | 1.56 | 0.1226 | 2.11998 |
| girimp | girimp | 1 | 0.11704 | 0.11905 | 0.98 | 0.3277 | 3.61619 |
| govcon | govcon | 1 | -0.04302 | 0.07522 | -0.57 | 0.5685 | 1.44361 |
| intsprd | intsprd | 1 | -0.18063 | 0.07127 | -2.53 | 0.0127 | 1.29617 |
| impcov | impcov | 1 | 0.11152 | 0.12308 | 0.91 | 0.3669 | 3.86548 |

The REG Procedure
Model: MODEL1
Dependent Variable: gdpgr gdpgr

| | |
|---|---|
| Number of Observations Read | 115 |
| Number of Observations Used | 115 |

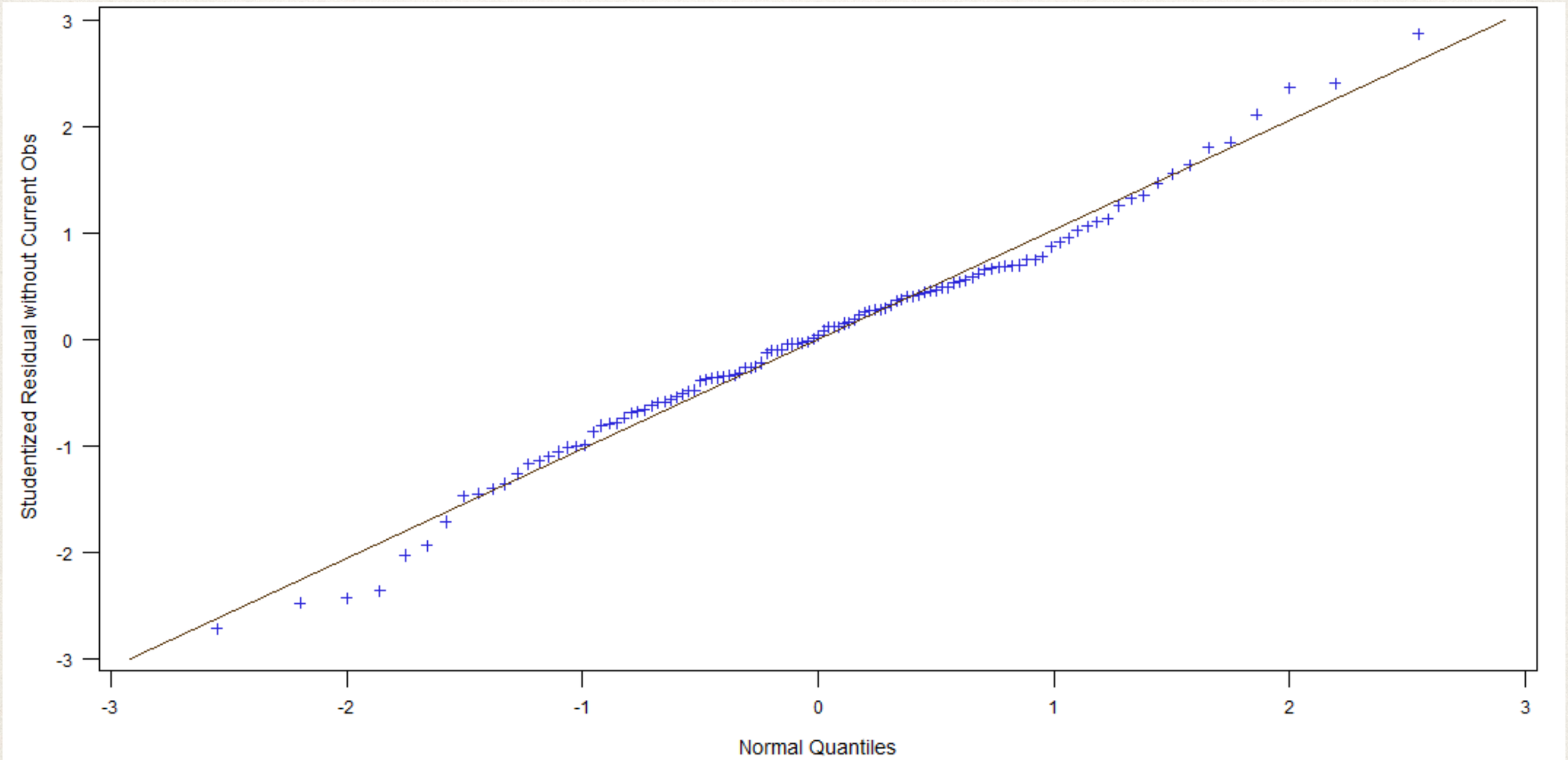
Analysis of Variance

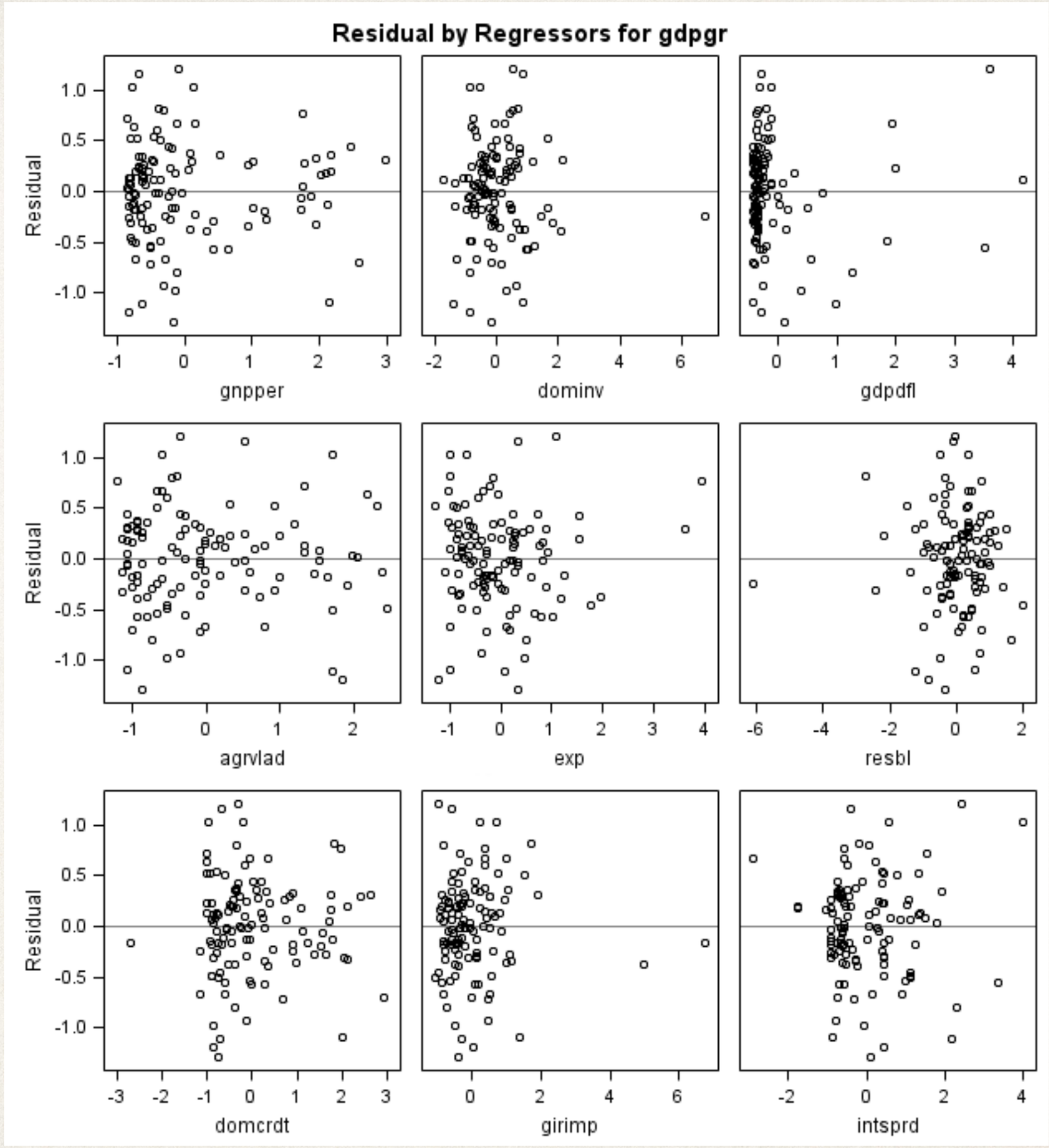
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 9 | 63.12584 | 7.01398 | 27.64 | <.0001 |
| Error | 105 | 26.64298 | 0.25374 | | |
| Corrected Total | 114 | 89.76883 | | | |

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

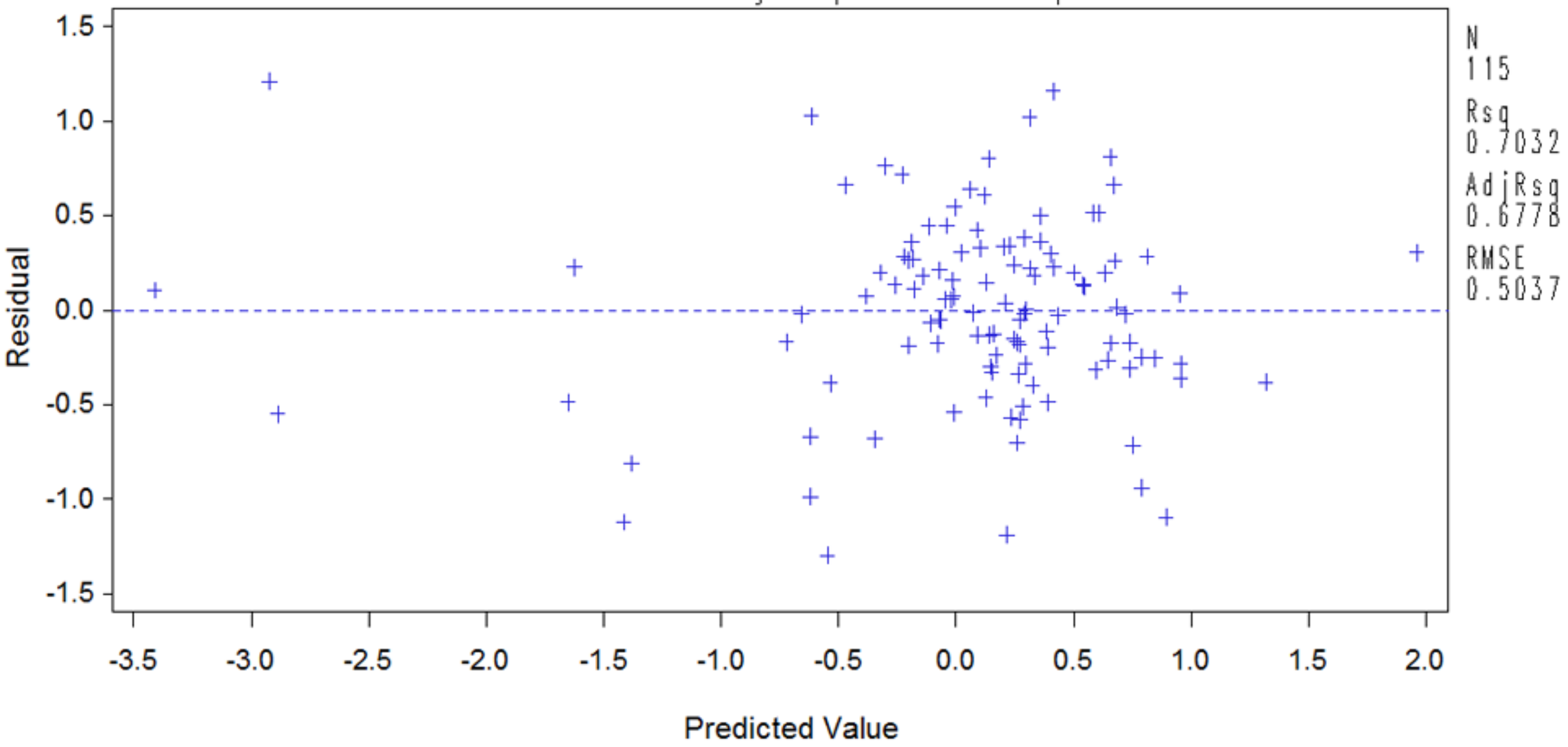
Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Variance Inflation |
|---|---|---|---|---|---|---|---|
| Intercept | Intercept | 1 | 0.01649 | 0.04740 | 0.35 | 0.7287 | 0 |
| gnpper | gnpper | 1 | -0.25869 | 0.08052 | -3.21 | 0.0017 | 2.82889 |
| dominv | dominv | 1 | 0.28397 | 0.06702 | 4.24 | <.0001 | 1.94955 |
| gdpdfl | gdpdfl | 1 | -0.62187 | 0.07027 | -8.85 | <.0001 | 1.32240 |
| agrvlad | agrvlad | 1 | 0.13117 | 0.07404 | 1.77 | 0.0794 | 2.37461 |
| exp | exp | 1 | -0.19852 | 0.06810 | -2.92 | 0.0043 | 1.52191 |
| resbl | resbl | 1 | 0.21967 | 0.07087 | 3.10 | 0.0025 | 2.07742 |
| domcrdt | domcrdt | 1 | 0.19937 | 0.06729 | 2.96 | 0.0038 | 2.07385 |
| girimp | girimp | 1 | 0.16068 | 0.04979 | 3.23 | 0.0017 | 1.15101 |
| intsprd | intsprd | 1 | -0.15570 | 0.05407 | -2.88 | 0.0048 | 1.35555 |


Studentized Residual without Current Obs
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Normal Quantiles


Residual by Regressors for gdpgr
Residual
gnpper
dominv
gdpdfl
agrvlad
exp
resbl
domcrdt
girimp
intsprd

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



# Methodology - Model II

- Check for Multicollinearity
- Variable Selection
- Outlier Removal
- Residual Analysis

## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 15 | 71.77679 | 4.78512 | 10.42 | <.0001 |
| Error | 105 | 48.22321 | 0.45927 | | |
| Corrected Total | 120 | 120.00000 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 0.67769 | R-Square | 0.5981 |
| Dependent Mean | 3.0944E-16 | Adj R-Sq | 0.5407 |
| Coeff Var | 2.190063E17 | | |

## Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Variance Inflation |
|---|---|---|---|---|---|---|---|
| Intercept | Intercept | 1 | 0.28021 | 0.17577 | 1.59 | 0.1139 | 0 |
| gnpper | gnpper | 1 | -0.02375 | 0.15052 | -0.16 | 0.8749 | 5.91977 |
| dominv | dominv | 1 | 0.47262 | 0.09181 | 5.15 | <.0001 | 2.20224 |
| gdpdfl | gdpdfl | 1 | -0.37057 | 0.08207 | -4.52 | <.0001 | 1.75997 |
| agrvlad | agrvlad | 1 | 0.07474 | 0.11997 | 0.62 | 0.5346 | 3.76055 |
| indvlad | indvlad | 1 | -0.05499 | 0.08443 | -0.65 | 0.5163 | 1.86253 |
| govcon | govcon | 1 | -0.04359 | 0.07845 | -0.56 | 0.5796 | 1.60804 |
| exp | exp | 1 | -0.15294 | 0.07965 | -1.92 | 0.0576 | 1.65773 |
| resbl | resbl | 1 | 0.45185 | 0.09795 | 4.61 | <.0001 | 2.50704 |
| domcrdt | domcrdt | 1 | 0.13993 | 0.09316 | 1.50 | 0.1361 | 2.26743 |
| impcov | impcov | 1 | 0.14192 | 0.12567 | 1.13 | 0.2614 | 4.12678 |
| girimp | girimp | 1 | 0.08544 | 0.12291 | 0.70 | 0.4885 | 3.94738 |
| intsprd | intsprd | 1 | -0.18026 | 0.07158 | -2.52 | 0.0133 | 1.33884 |
| hdev | hdev | 1 | -0.65924 | 0.33897 | -1.94 | 0.0545 | 6.22337 |
| mdev | mdev | 1 | -0.35207 | 0.23234 | -1.52 | 0.1327 | 2.65196 |
| ldev | ldev | 1 | -0.01173 | 0.21599 | -0.05 | 0.9568 | 1.89218 |

The REG Procedure
Model: MODEL1
Dependent Variable: gdpgr gdpgr

| | |
|---|---|
| Number of Observations Read | 114 |
| Number of Observations Used | 114 |

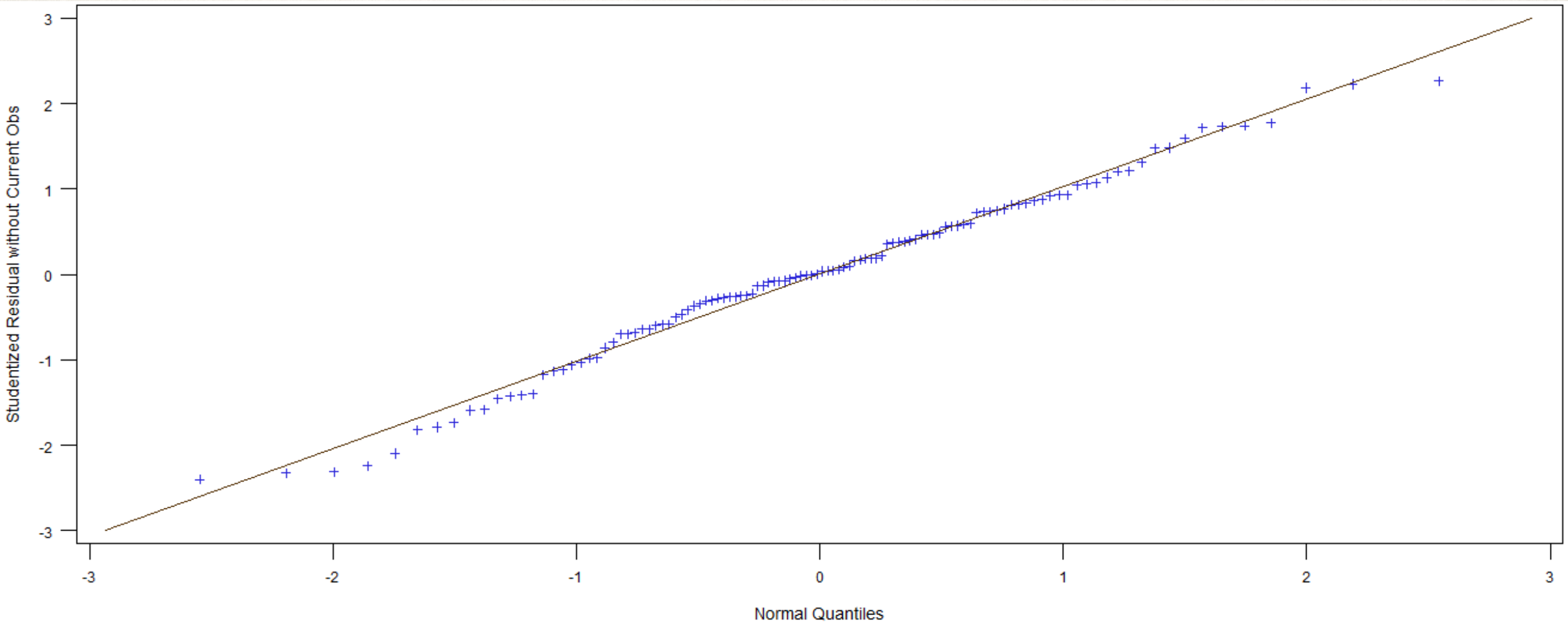
Analysis of Variance

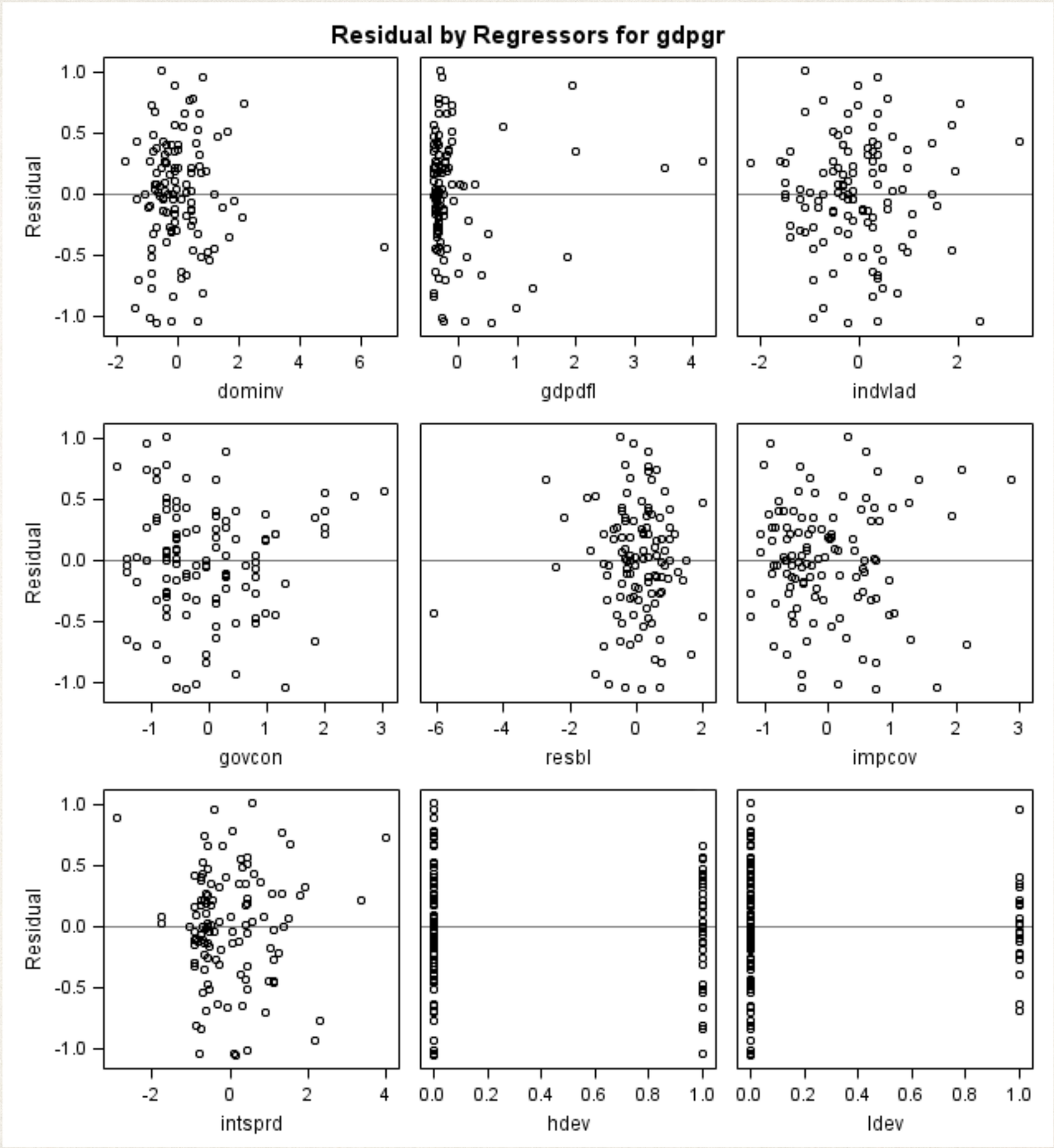
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 9 | 64.99875 | 7.22208 | 32.04 | <.0001 |
| Error | 104 | 23.44416 | 0.22542 | | |
| Corrected Total | 113 | 88.44291 | | | |

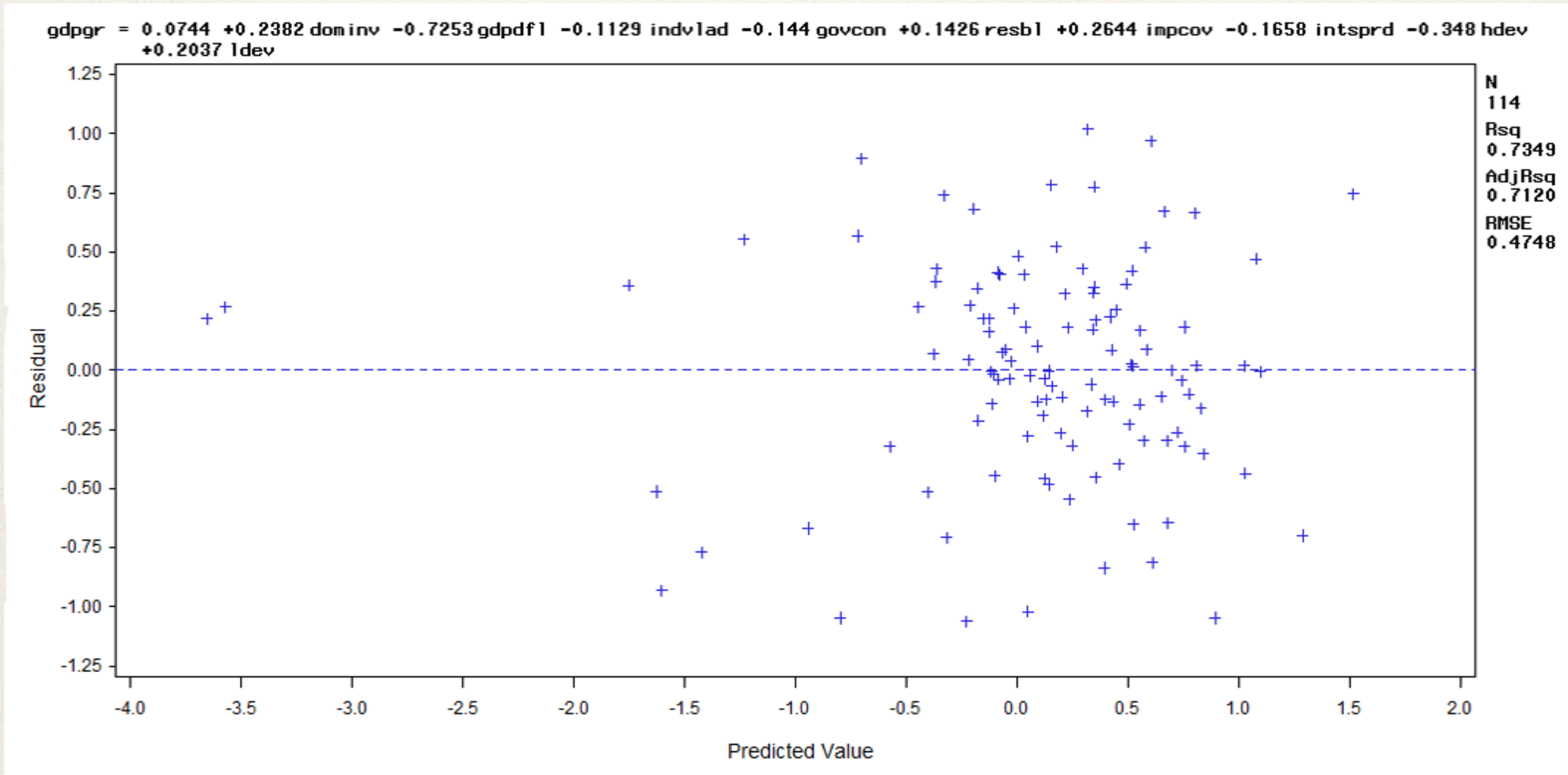
| | | | |
|---|---|---|---|
| Root MSE | 0.47479 | R-Square | 0.7349 |
| Dependent Mean | 0.09883 | Adj R-Sq | 0.7120 |
| Coeff Var | 480.41427 | | |

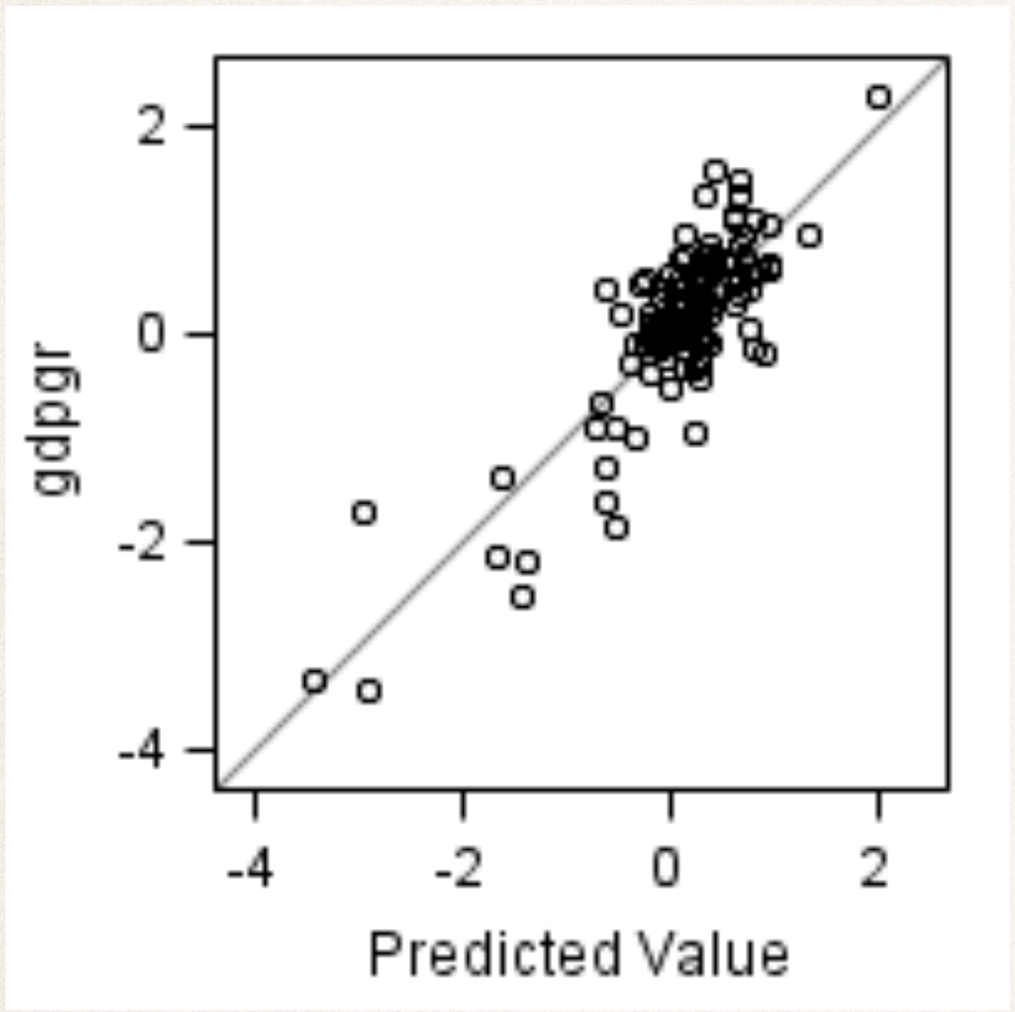
Parameter Estimates

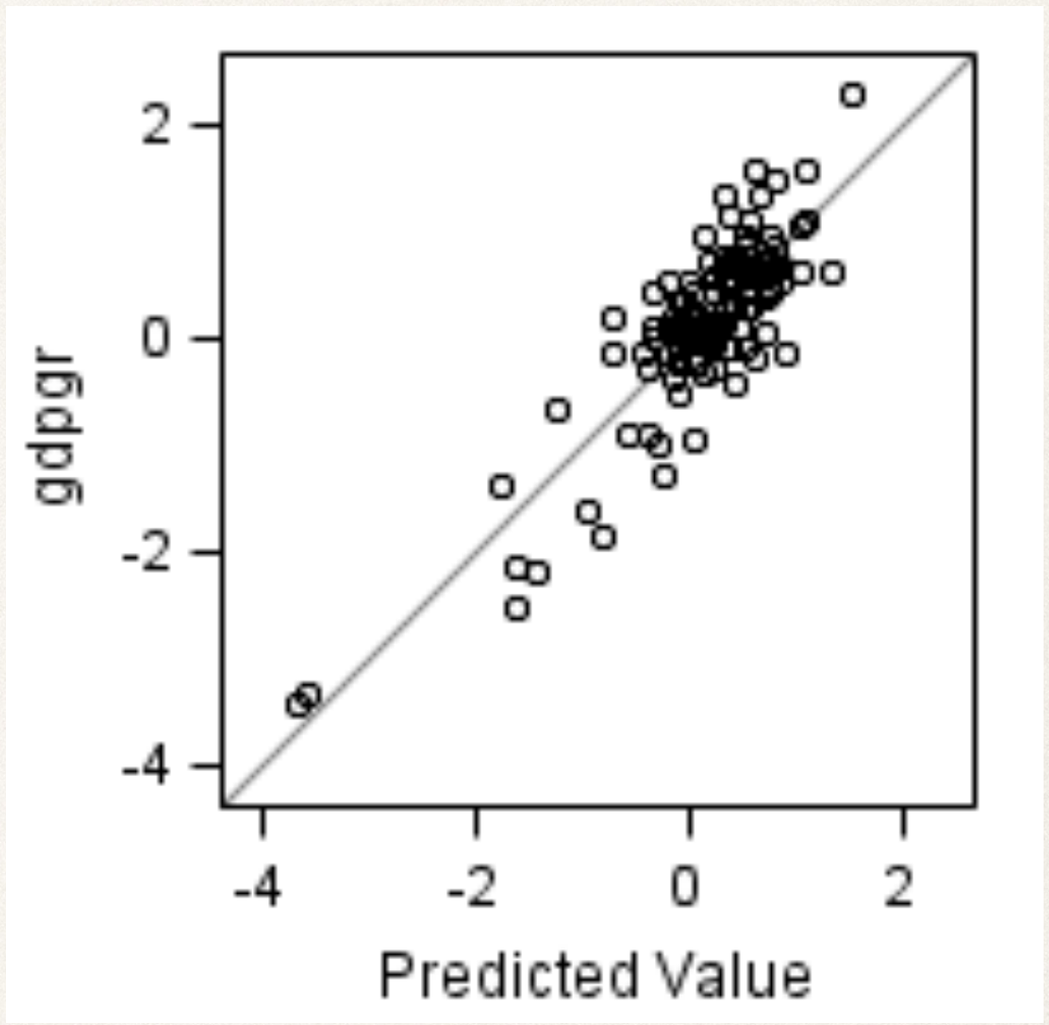
| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Variance Inflation |
|---|---|---|---|---|---|---|---|
| Intercept | Intercept | 1 | 0.07441 | 0.06762 | 1.10 | 0.2737 | 0 |
| dominv | dominv | 1 | 0.23824 | 0.05657 | 4.21 | <.0001 | 1.58294 |
| gdpdfl | gdpdfl | 1 | -0.72528 | 0.07226 | -10.04 | <.0001 | 1.25885 |
| indvlad | indvlad | 1 | -0.11291 | 0.05906 | -1.91 | 0.0587 | 1.45935 |
| govcon | govcon | 1 | -0.14404 | 0.06001 | -2.40 | 0.0182 | 1.54266 |
| resbl | resbl | 1 | 0.14260 | 0.06076 | 2.35 | 0.0208 | 1.78249 |
| impcov | impcov | 1 | 0.26435 | 0.06778 | 3.90 | 0.0002 | 1.37945 |
| intsprd | intsprd | 1 | -0.16582 | 0.05155 | -3.22 | 0.0017 | 1.31761 |
| hdev | hdev | 1 | -0.34797 | 0.12944 | -2.69 | 0.0084 | 1.80281 |
| ldev | ldev | 1 | 0.20366 | 0.12515 | 1.63 | 0.1067 | 1.19033 |


Studentized Residual without Current Obs
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Normal Quantiles


Residual by Regressors for gdpgr
Residual
dominv
gdpdfl
indvlad
govcon
resbl
impcov
intsprd
hdev
ldev


gdpgr = 0.0744 +0.2382 dominv -0.7253 gdpdfl -0.1129 indvlad -0.144 govcon +0.1426 resbl +0.2644 impcov -0.1658 intsprd -0.348 hdev
+0.2037 ldev
N
114
Rsq
0.7349
AdjRsq
0.7120
RMSE
0.4748
Residual
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
-4.0
-3.5
-3.0
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Predicted Value


gdpgr
2
0
-2
-4
-4
-2
0
2
Predicted Value


gdpgr
2
0
-2
-4
-4
-2
0
2
Predicted Value

# Conclusion

- After applying a set of regression specific procedures nine of the 12 regressors were chosen to represent the model I
- For model II, nine out of 15 regressors were chosen to represent it, two of which were dummy variables.
- The value of adj$R^2$ from model I is 0.67 and from Model II is 0.71 which suggests that the inclusion of categorical variables in the regression model helped to improve the model.
- As a remark for further research, the value of adj$R^2$ 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.

# Thank You!