Project Report - Assignment 1

Estimation Of Growth Rate Using India's Socio Economic Data

ECON F342 - Applied Econometrics

A report by **Group 1**

Name	ID
Himanshi Badola Jain (Group Leader)	**2020B3AA2159H**
Yashraj Mehta	2020B3A72226H
Akash Goel	2020B3A31858H
Nishant Mishra	2020B3AA0529H
Ashish Kusumanchi	2020B3A71445H
Shubham Agrawal	2020B3A70505H
Kanha Toshniwal	2020B3A30759H
Soham Barui	2020B3AA1170H

Instructor In-Charge: **Dr. Rishi Kumar**

March 2023

TABLE OF CONTENTS

Section I: Introduction

Economic Growth rate is one of the indicators of overall development. For developing countries like India, growth rate becomes more important due to the following reasons:

- **Poverty reduction:** India is the world's 5^{th} largest economy. But when you look at its per capita income it says a different story because of its dense population. India has a large population living below the poverty line, and sustained economic growth is the key to its reduction. Economic growth leads to better and more job opportunities which raise income and thus increases their standards of living.
- **Infrastructure Development:** Economic growth will lead to an increase in tax collection which will help the government develop roads, airports, coast-line ports and waterways. This will make the transport sector of the country much more efficient leading to higher productivity.
- **Increases Foreign Investment:** A high economic growth rate tends to attract foreign investors. This can bring in much-needed capital, technology and expertise to help spur economic growth.
- **Better Public Services:** With better revenue, governments can focus more on public services like basic healthcare and educational services. This can help poor and at-risk citizens improve their quality of life.
- **Competitiveness:** Liberalisation and free trade have levelled the playing field for foreign players to compete with domestic ones. For domestic competitors to compete, they need to be able to innovate. Economic growth can help them invest in R&D and efficient scaling solutions.

It is easy and intuitive to discuss about improving economic growth at a larger, country level. But we need to keep in mind that while cities like Mumbai thrive, cities like Shravasti also exist, where 74.38% of the population is deemed to be multi-dimensionally poor.

This is why we need to narrow down our analysis to district-wise economic growth rates. By doing this we can identify regional disparities and allocate resources judiciously and target districts that are lagging and address their challenges specifically. This will not only help in uplifting poorer districts but also help in achieving uniform growth. If sustained, this can also reduce inequality to some extent. Jobs could be created, quality of life and public infrastructure could be improved as well.

Design: This research includes the formulation of a system explaining the relationship between district-wise growth rates and variables like sex ratio, female literacy rate, urban population etc which will be discussed in detail.

Confirm: The analysis seeks to see which factors affect district-wise economic growth the most.

The organization of this paper is such that Section II is on related work, Section III discusses the methodology, and Section IV presents the results and in-depth discussion. Section V concludes the paper.

Section II: Literature Review

While there are a lot of studies conducted by various researchers to determine the effect of various factors on district-wise growth, many of them have been state-specific and not broader on a country level. This also means that their suggestions and conclusions have been separate for every district, but only in that state.

Dr. I Sivakumar (2020) concluded his paper by suggesting that India needed to focus on deregulating sectors that poor depend upon. He also believed that resource allocation needs to be improved in underprivileged regions of Central and Eastern India.

Rampul Ohlan (2012) assessed the level of development with a different approach. He assessed the level of development separately for agriculture, industrial and infrastructural sectors, with districts further classified into four categories. He concluded that while there is vast amounts of economic disparity in every district, southern states have lesser disparity compared to their northern counterparts. He also suggests that the level of agricultural development influences overall socioeconomic development, but the level of industrial development does not significantly influence overall socioeconomic development and agriculture.

Amod Sharma (2012) took a different approach in his paper, he focused only on the northeastern states. While he did study overall socio-economic development, he also studied agricultural, infrastructure and industrial development separately for each state. His paper also claimed that Infrastructure development is related to agricultural development.

Prem Narain (2009) had an approach like that of Amod's. But his study was focused solely on the state of Andhra Pradesh, where he ranked the district of West Godavari first in terms of overall socio-economic development. His paper also had separate studies for industrial and agricultural development.

Prem Narain (2000) was a paper similar to his 2009's study on Andhra Pradesh, only this time it was on Tamil Nadu. The level of development was obtained with the help of a composite index devised using a combination of forty-two socio-economic indicators. He found the district of Chengalpattu MGR to be ranked first. He also found that the northern districts of Tamil Nadu were better developed. He also concluded that the level of public infrastructure like healthcare and education did not have a significant impact on agricultural development.

Section III - Data and Methodology

A. **Description of the dataset**:

The data for this study has been sourced from Kaggle which is derived from the Indian Census Data and it consists of data from 618 districts ranging for the year 2015-2016. to analyze the Indian socio-economic growth in terms of the following parameters:

- **Growth rate:** It measures the growth rate of a particular district.
- **Urban Population percentage** - It is the percentage of the urban population in a district out of its total population.
- **Sex ratio-** The number of females per 1000 males is the sex ratio.
- **Female literacy-** The percentage of the female population which is educated out of the total female population.

- **Area in sq. km -** Area of the district in square kilometres.

- **School h_sec -** The number of schools with minimum higher secondary education is captured by this parameter.

- **Teacher_qualified -** The number of qualified teachers present in the district.

B. Econometric Model

- **Equation**

$$\begin{gathered} growth_rate = \beta_0 + \beta_1 sex\_ratio + \beta_2 female\_literacy + \beta_3 area\_dummy + \beta_4 school\_dummy + \beta_5 log\_urban_pop + \beta_6 sqrt\_teacher\_qualified + \\ \delta_{jk} + \delta_{hp} + \delta_{pun} + \delta_{ch} + \delta_{uk} + \delta_{hr} + \delta_{dl} + \delta_{raj} + \\ \delta_{up} + \delta_{br} + \delta_{sik} + \delta_{ap} + \delta_{nag} + \delta_{mani} + \delta_{miz} + \\ \delta_{tri} + \delta_{meg} + \delta_{ass} + \delta_{wb} + \delta_{jk} + \delta_{odi} + \delta_{ch} + \\ \delta_{mp} + \delta_{guj} + \delta_{dnd} + \delta_{dad} + \delta_{ap} + \delta_{kar} + \delta_{goa} + \\ \delta_{lk} + \delta_{ker} + \delta_{tn} + \delta_{dad} + \delta_{pondi} + \delta_{ani} \end{gathered}$$

- **VARIABLE** - growth_rate
  Description - Dependent variable growth rate of district.
  Reason - To study the effect of growth rate on different variables like sex ratio, and area of the district

- **VARIABLE** - sex_ratio
  Description - The sex ratio is defined as the number of females per 1000 males in a given population.
  Reason - To find out that whether districts with better sex ratios have better growth rates.

- **VARIABLE** - female_literacy
  Description - The female literacy rate is the percentage of people ages 15 and above who can read and write with an understanding of a short, simple statement about their everyday life.
  Reason - It is expected that with an increase in female literacy rates, the growth rate should also increase.

- **VARIABLE** - area_dummy
  Description - Districts with an area of more than 5000 sq km take 1 as the value, and less than 5000 sq km will take 0 as the value.
  Reason - To study whether or not the district's area affects the growth rate.

- **VARIABLE** - school_dummy
  Description - Districts with a number of higher secondary schools more than 100 will take 1 as the value, and less than 100 will take 0 as the value.
  Reason - The economic model suggests that the growth rate should increase with an increasing number of schools. We will verify that with regression.

- **VARIABLE** - log_urban_pop
  Description - We have taken the log transformation of the percentage of the urban population of that district to remove the skewness.
  Reason - It is expected that with an increase in the percentage of the urban population, the growth rate should increase.

- **VARIABLE** - sqrt_teacher_qualified
  Description - We have taken the square root of the total number of teachers available in that district to obtain a normal distribution.
  Reason - To find out whether districts with a better number of highly qualified teachers have better growth rates.

C. Model Specifications

- We performed theBreusch-Pagan Test totest for Homoscedasticity, and the result indicated towards Heteroscedasticity, so we ran a linear regression with robust standard errors.

- We checked for multicollinearity with the help of the Variance Inflation Factor Test.

- We checked for normality by performing the Shapiro-Wilk Test.

- We checked for the Omitted variable by Reset test

D. Graph matrix of the regressors:

growth_rate	-0.04	-0.07	0.06	-0.06	-0.29	-0.3
-0.04	area_dummy	0.39	0.1	0.07	0.09	-0.21
-0.07	0.39	sqrt_teacher_qualified	0.31	0.18	0.08	-0.05
0.06	0.1	0.31	school_dummy	0.25	-0.1	0.07
-0.06	0.07	0.18	0.25	log_urban_pop	-0.05	0.48
-0.29	0.09	0.08	-0.1	-0.05	sex_ratio	0.18
-0.3	-0.21	-0.05	0.07	0.48	0.18	female_literacy

The correlation between the explanatory variables is shown above in the graph matrix. None of the variables is strongly correlated (except for female literacy and log_urban_pop). A few notable observations from this graph matrix are

- There is a positive correlation between the school dummy variable and the growth rate. Therefore higher the number of schools, the higher is the growth rate.
- There is a positive correlation between sqrt_teacher_qualified and log_urban_pop, which is evident from the fact that if more qualified teachers are there in some districts, then students will come to that district to obtain an education.
- There is a positive correlation between sqrt_teacher_qualified and sex_ratio, which should have been the case.
- There is a negative correlation between log_urban_pop and sex ratio.
- There is a positive correlation between sex_ratio and female_literacy which should have been the case.

Section IV: Results And Discussions

A. F-Test

It is a test done to check the joint significance of the variables used in the model.
H_o :All the coefficients are jointly equal to 0

```
F-statistic: 9.143 on 40 and 576 DF,  p-value: < 2.2e-16
```

As p-value is lesser tha 0.01, we can reject the null hypothesis at 99% confidence interval and conclude that all the variables used are jointly significant.

B. Omitted Variable Bias

It is to check if our model has included all the important variables.
To check for this, RESET test was done
The null hypothesis is
H_o : Model has omitted variable bias

```
        RESET test

data:  final_reg
RESET = 0.72562, df1 = 2, df2 = 574, p-value = 0.4845
```

As seen from the above test, the p-value is greater than 0.1. So the null hypothesis can be rejected even at a 90% confidence interval. Hence there does not exist the omitted variable bias in our regression.

C. Assumptions Of OLS

1) Linear in parameters

The model should be linear in parameters meaning that all the β_i should be linear. Here since we have only a single β_i term for every variable, we don't find any non-linearity in the coefficients. Hence, we can say that our model is linear in parameters.

2) Random sampling
We have collected the data from Kaggle, which was sourced from the Indian census data from the government websites. And it takes a wide range of data and forms a sample which can represent the whole population. Hence it follows the assumption of random sampling.

3) Multicollinearity
Multicollinearity is the condition when several of the independent variables of a dataset are correlated to each other. This would generally mean that a random variable can be predicted using another variable using a linear polynomial equation. We use the VIF(Variation Inflationary Factor) test to check for multicollinearity in our data.

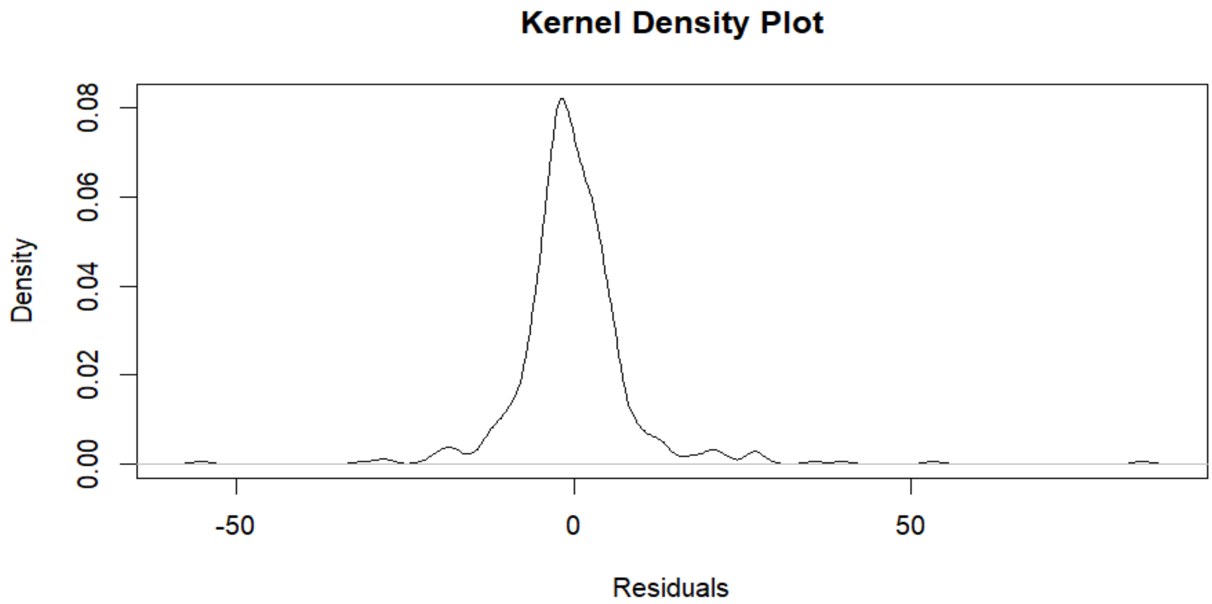
	GVIF	Df	GVIF^(1/(2*Df))
sex_ratio	2.247006	1	1.499002
female_literacy	3.095426	1	1.759382
area_dummy	1.931764	1	1.389879
school_dummy	2.182432	1	1.477306
log_urban_pop	2.130952	1	1.459778
sqrt_teacher_qualified	3.494452	1	1.869345
factor(state_name)	58.362784	34	1.061629

The VIF values are shown in the third column of the above table. Popular literature suggests that a Mean VIF of less than 10 satisfies our purposes. Our dataset presents all the VIF values of the variables less than 10. This suggests that we can safely assume that there is ***no multicollinearity*** present in our model.

4) Zero Conditional Mean
This condition is satisfied when the explanatory variables used in regression does not contain any information about the residual or the unobserved factors.
Also:

$$E(ui|xi) = 0$$

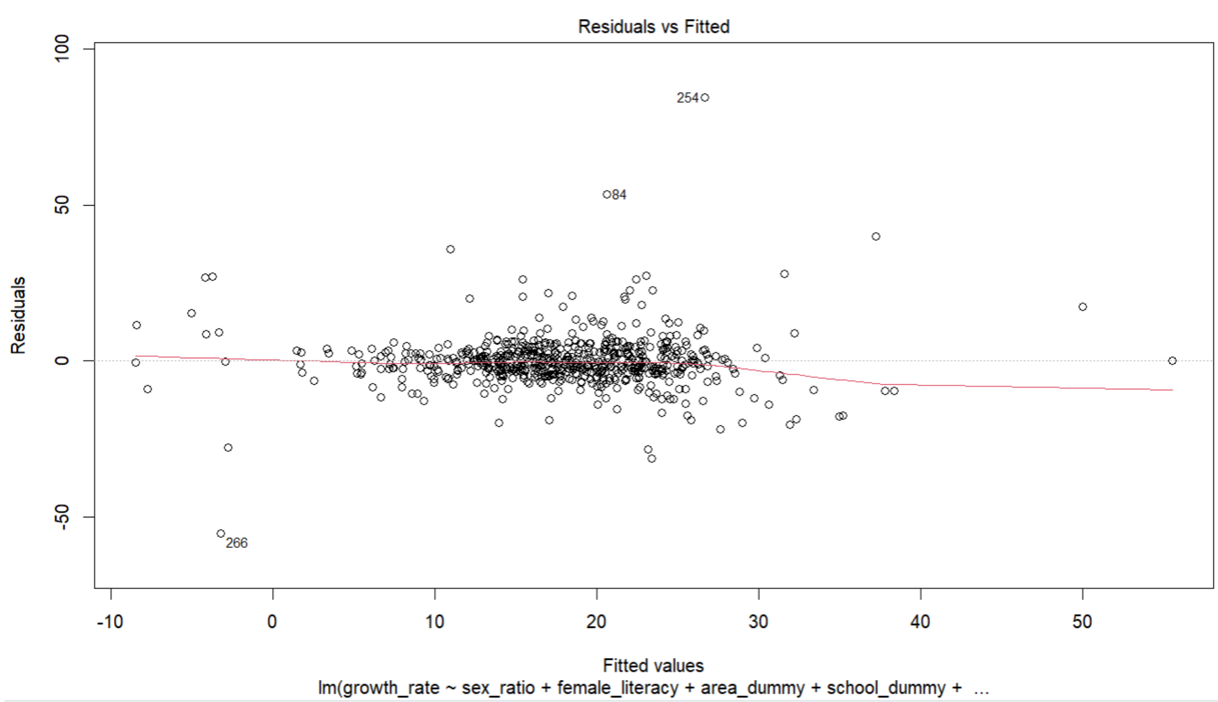
To show this, a plot of residuals density was made. As it is centered at mean = 0, we can conclude that it satisfies zero conditional mean.



5) Homoscedasticity of error term

A sequence of random variables is homoscedastic if all independent variables have the same variance, which is also finite. This can be checked through:

(a)Scatter plot of residuals vs the Fitted values of the dependent variable



From the above graph we can see that there is a deviation of the red line from the straight horizontal line so heteroscedasticity is present in our model.

(b)Breusch-Pagan Test

We check if the data is heteroscedastic by using the Breusch-Pagan test, in which the null hypothesis states that there is a constant variance in the error terms, and the error terms are normally distributed.

```
        studentized Breusch-Pagan test

data:  final_reg
BP = 109.31, df = 40, p-value = 2.364e-08
```

Since the p-value < 0.01, We reject the null hypothesis of constant variance even at 0.01(1%) level, we will say that there is ***heteroscedasticity*** in the dataset that we have used.

Remedial Measures

Now to rectify this we rerun the regression using robust standard errors.

```
                                        Estimate Std. Error t value  Pr(>|t|)    
(Intercept)                            59.102390  14.551839  4.0615 5.551e-05 ***
sex_ratio                              -0.034237   0.015717 -2.1784 0.0297809 *  
female_literacy                        -0.288878   0.065278 -4.4254 1.152e-05 ***
area_dummy                             -1.765019   1.007703 -1.7515 0.0803873 .  
school_dummy                            3.041831   1.188241  2.5599 0.0107229 *  
log_urban_pop                           1.604221   0.901256  1.7800 0.0756056 .  
sqrt_teacher_qualified                  0.058985   0.024316  2.4257 0.0155836 *  
factor(state_name)ANDHRA PRADESH       -9.015112   3.333053 -2.7048 0.0070375 ** 
factor(state_name)ARUNACHAL PRADESH    12.085044   6.717156  1.7991 0.0725213 .  
factor(state_name)ASSAM                 2.379836   1.860715  1.2790 0.2014155    
factor(state_name)BIHAR                 4.979360   2.236503  2.2264 0.0263731 *  
factor(state_name)CHANDIGARH           -1.358936   0.729210 -1.8636 0.0628903 .  
factor(state_name)CHHATTISGARH          4.263894   3.344627  1.2748 0.2028768    
factor(state_name)DADRA AND NAGAR HAVELI 34.166584  1.612726 21.1856 < 2.2e-16 ***
factor(state_name)DAMAN AND DIU        25.524194  12.863437  1.9842 0.0477028 *  
factor(state_name)DELHI                 3.595828   2.246057  1.6010 0.1099359    
factor(state_name)GOA                  -3.028090   1.708170 -1.7727 0.0768052 .  
factor(state_name)GUJARAT               0.446584   2.177023  0.2051 0.8375389    
factor(state_name)HARYANA              -0.285232   3.371523 -0.0846 0.9326085    
factor(state_name)HIMACHAL PRADESH      2.147257   1.934019  1.1103 0.2673519    
factor(state_name)JAMMU AND KASHMIR     4.545402   2.354839  1.9302 0.0540678 .  
factor(state_name)JHARKHAND             6.081584   2.587081  2.3508 0.0190718 *  
factor(state_name)KARNATAKA            -2.946964   3.069268 -0.9602 0.3373815    
factor(state_name)KERALA               -2.161771   2.982490 -0.7248 0.4688562    
factor(state_name)LAKSHADWEEP          -3.575626   1.421315 -2.5157 0.0121499 *  
factor(state_name)MADHYA PRADESH        1.253426   2.265335  0.5533 0.5802679    
factor(state_name)MAHARASHTRA          -0.520442   2.333914 -0.2230 0.8236213    
factor(state_name)MANIPUR               8.924652   2.530715  3.5265 0.0004546 ***
factor(state_name)MEGHALAYA            20.540562   2.812429  7.3035 9.379e-13 ***
factor(state_name)MIZORAM              14.760950   6.183193  2.3873 0.0172956 *  
factor(state_name)NAGALAND            -16.106001   7.281767 -2.2118 0.0273704 *  
factor(state_name)ODISHA                0.061178   2.376631  0.0257 0.9794724    
factor(state_name)PONDICHERRY          27.602621  12.587698  2.1928 0.0287191 *  
factor(state_name)PUNJAB               -5.489499   1.738940 -3.1568 0.0016785 ** 
factor(state_name)RAJASTHAN            -3.472387   3.190916 -1.0882 0.2769574    
factor(state_name)SIKKIM               -3.068284   2.849287 -1.0769 0.2819934    
factor(state_name)TAMIL NADU           -1.124434   2.577129 -0.4363 0.6627734    
factor(state_name)TRIPURA               6.864783   2.716614  2.5270 0.0117719 *  
factor(state_name)UTTAR PRADESH         0.062678   2.092309  0.0300 0.9761121    
factor(state_name)UTTARAKHAND           0.953944   3.522578  0.2708 0.7866354    
factor(state_name)WEST BENGAL          -3.338743   2.564576 -1.3019 0.1934815    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

After using robust standard errors, the data is homoscedastic and the variables are still significant at 95% confidence interval.

6) Normality of Residuals

It means that the errors will be independent of the variables used and will also be normally distributed about the population regression function
To check for then, 2 methods were used Shapiro-Wilk Test and a graphical method of using the QQ plots.

7) Shapiro- Wilk test can be used to determine the normality of the data
H_0: Data is part of a normal distribution.

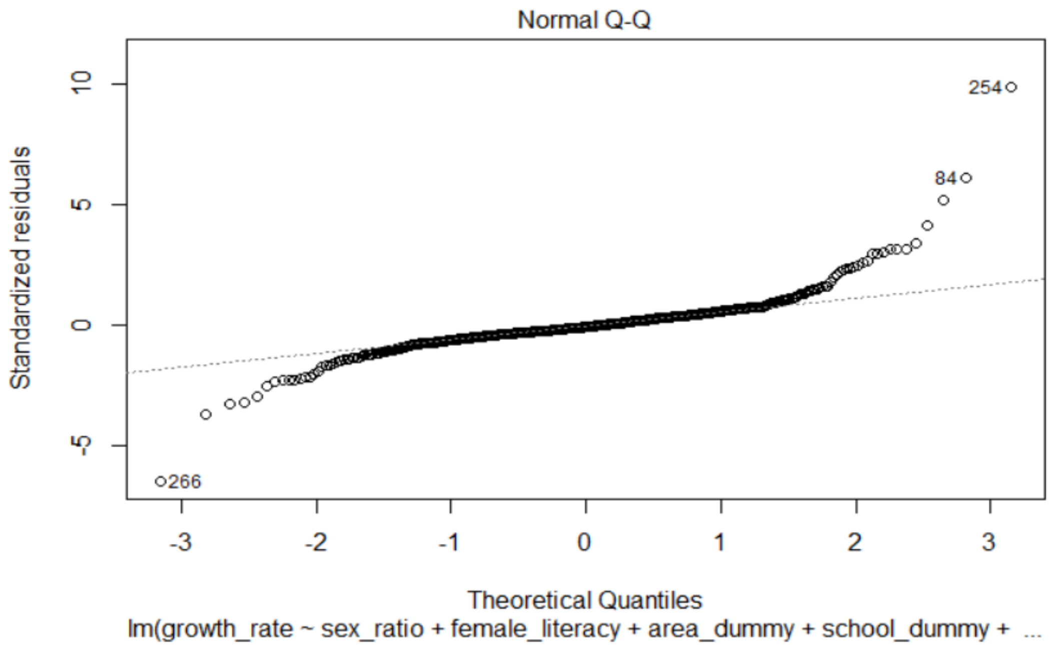
```
        Shapiro-Wilk normality test

data:  final_reg$residuals
W = 0.80427, p-value < 2.2e-16
```

Since the p-value <0.01 we can reject the null hypothesis and say our data is not normally distributed.

8) QQ plots

It is a normal probability plot of residuals. It helps to visually test the normality of residuals. If the plot follows a nearly straight line, then it can be said to be normally distributed. But if it follows an S-shaped curve (as in our case) then that means errors are large in both directions and the residuals are not normally distributed.



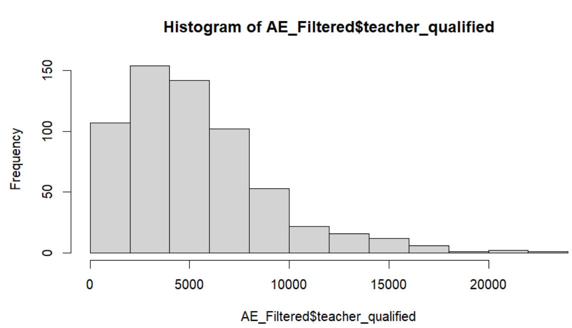
There are 3 outliers present in our model:

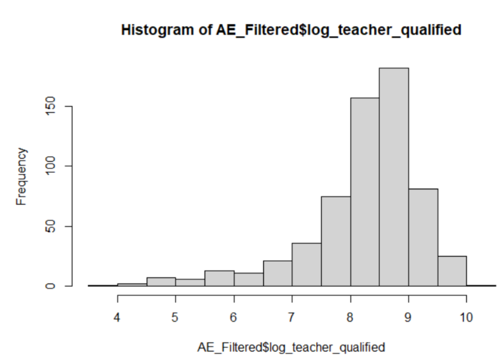
1. District Kurung Kumey from the state of Arunachal Pradesh (254)
2. District Kiphire from the state Nagaland (266)
3. District Gurgaon from the state Haryana (84)

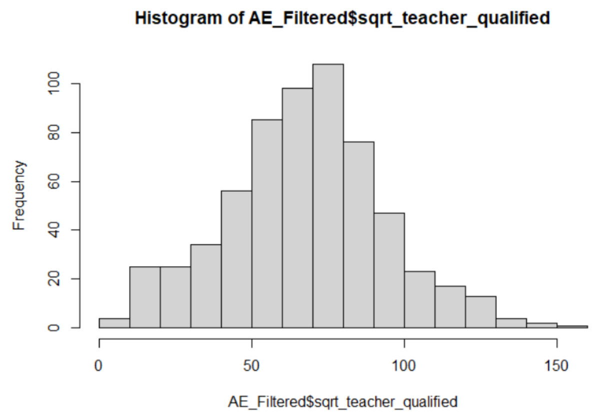
Remedial Measures

As the residuals were not normally distributed, we tried some transformations

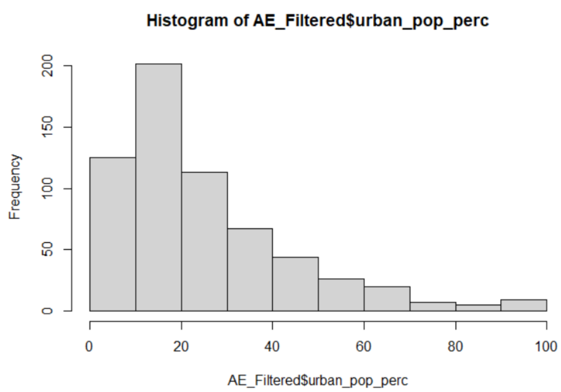
i) **teachers_qualified** was changed to **log_teacher_qualified**. But as it still showed some skewness, it was again changed to **sqrt_teacher_qualified.**

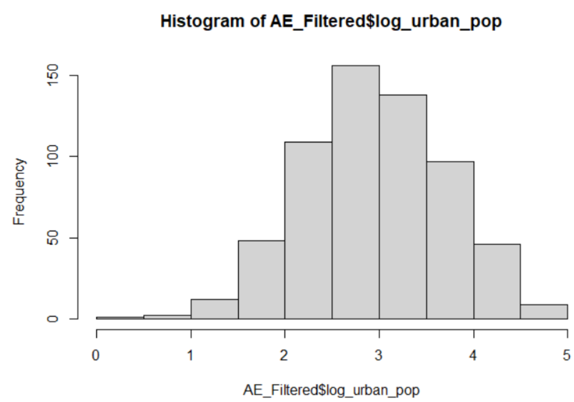






ii) **urban_pop_perc** was changed to **log_urban_pop**





Even after these transformations, the normality was not entirely removed. But as seen from the Kernal plot, the residuals are nearly normally distributed, so there is no further need to account for normality in residual terms**.**

D. Summary Statistics

	growth_rate	sex_ratio	female_literacy	area_dummy	school_dummy	log_urban_pop	sqrt_teacher_qualified
Min	-58.39	533.0	30.97	0	0	0.2469	5.916
1st Quartile	12.25	904.2	56.09	0	0	2.4423	51.884
Median	17.21	946.0	63.74	0	0	2.9714	68.760
Mean	17.93	942.7	64.62	0.3328	0.1734	2.9614	67.882
3rd Quartile	22.78	979.8	73.57	1	0	3.4968	83.791
Max	111.01	1176	98.28	1	1	4.6052	151.03

Section V: Estimation And Interpretation

- **Adjusted R^2**

```
Residual standard error: 8.939 on 576 degrees of freedom
Multiple R-squared:  0.3884,    Adjusted R-squared:  0.3459
```

The Linear Regression gives an adjusted R-squared of 0.3459 which is interpreted as the goodness of fit measure. This means that our model explains 34.59% of the actual variance in the population.

- **Intercept:** This is the hypothetical value of the gdp growth rate when all the dummies are zero, full male population and hence zero female literacy rate. This is unusually high at 54.940453 growth and that is because of the negative coefficients we have for other variables.

- **Sex_ratio:** From the regression, we see that if the sex ratio increases by one unit(if there is one more female per 1 Thousand males), the growth rate will decrease by -0.034250 units keeping everything constant. Sadly, this is significant at a 95% interval and possibly could be attributed to the gender wage gap where women are paid less than men.

- **female_literacy:** From the analysis, if the female_literacy increases by one unit, the growth rate will decrease by 0.290555 units, keeping all the other variables constant. This is highly significant at a 99.9% level, counterintuitive probably because of the dataset chosen or the time at which the data is collected, as this should not be happening.

- **area_dummy:** As the name states this is a dummy variable which is 1 for large districts(>5000 sqkm) and 0 for small districts. We can see that larger districts decrease the growth rate by 1.772433 units. This is counterintuitive as large districts should have more land available for doing economic activities that raise growth rate but nevertheless this is insignificant at a 95% confidence interval. This could again be because of how district boundaries are designed to keep arable land constant or maybe signifying resource course, but this is not significant so all these reasons are just possible speculations.

- **School_dummy:** We have categorized districts into two categories 1(if it has more than 100 higher secondary schools) and zero otherwise. We can see districts with a higher number of schools have a growth rate that is higher by 3.015275 units than districts with fewer schools.
  This is again as per general expectations that the more number of schools, the higher the % of educated people hence a better economy. This is significant at 95% confidence interval.

- **Log_urban_pop:** if the urban population increases by 1 percent the growth rate would increase by 0.016 units. This is again expected as urban economic activities generate more value than rural activities but this is not significant based on the regression analysis.

- **Sqrt_teacher_qualified:** We cannot arithmetically interpret sq root transformations but if the root of the number of teachers in a district increases by 1 unit the growth rate increases by 0.059192 units. This confirms that more teachers imply better educational opportunities and hence a better economic growth rate. This is significant at 95% confidence level.

- **factor(state_name):** This is a qualitative variable that changes based on the state(we assume 36 states in india : 28 states+8 union territories) in which the district is located with the base state aa Andaman and Nicobar Islands. Based on this we can interpret other states for example if we consider Arunachal Pradesh has a growth rate that is 16.182626 units more than Andaman and Nicobar if somehow both states have everything else similar. The states and their coefficients if significant are mentioned in the following table.

State	Estimate	Std. Error	t value	Pr(>|t|)	Significance Star
factor(state_name) ARUNACHAL PRADESH	16.18263	7.000085	2.3118	0.021141	*
factor(state_name) ASSAM	6.491325	3.4317	1.8916	0.059048	.
factor(state_name) BIHAR	9.076026	3.582935	2.5331	0.011569	*
factor(state_name) CHHATTISGARH	8.332553	4.586673	1.8167	0.069784	.
factor(state_name) DADRA AND NAGAR HAVELI	38.17412	3.43678	11.1075	< 2.2e-16	***
factor(state_name) DAMAN AND DIU	29.53576	13.24319	2.2303	0.026114	*
factor(state_name) DELHI	7.67544	3.590335	2.1378	0.032952	*
factor(state_name) HIMACHAL PRADESH	6.288555	3.418304	1.8397	0.06633	.
factor(state_name) JAMMU AND KASHMIR	8.618025	3.713229	2.3209	0.02064	*

State	Estimate	Std. Error	t value	Pr(>|t|)	Significance Star
factor(state_name) JHARKHAND	10.15547	3.97246	2.5565	0.010829	*
factor(state_name) MANIPUR	13.01684	3.989742	3.2626	0.001169	**
factor(state_name) MEGHALAYA	24.6555	4.133835	5.9643	4.29E-09	***
factor(state_name) MIZORAM	18.81951	6.963364	2.7026	0.007082	**
factor(state_name) PONDICHERRY	31.61018	12.98785	2.4338	0.015243	*
factor(state_name) TRIPURA	10.96583	4.035169	2.7176	0.006774	**

Section VI: Conclusion

In this research, an econometric study was conducted on India's socio-economic data to check the dependence of growth rate on various factors. The Cross-sectional data used in this research was taken from Kaggle.com (derived from District-wise Census data of India: 2015-2016).

The data itself had to be statistically treated for relevant conclusions to be inferred. Robust standard errors were used in hypothesis testing because the data showed heteroscedasticity (found using Breusch-Pagan Test). Independent variables of the dataset were not correlated to each other as VIF showed no multicollinearity. The assumptions for the use of OLS were also verified, and appropriate action was taken to ensure the results carry statistical meaning and significance.

The factors we used to determine the growth rate are - Urban Population percentage, Sex ratio, Female literacy, Area in sq. km, Teacher_qualified (no. of

qualified teachers present in the district), and School h_sec (no. of schools with minimum higher secondary education). The effect of these variables was found to be in line with what is observed in the literature.
The factors that had a negative correlation with the growth rate were found to be - Sex_ratio, female_literacy and area_dummy.
The factors that had a positive correlation with the growth rate were found to be - School_dummy, Log_urban_pop and Sqrt_teacher_qualified.

From the regressive analysis, it can be concluded that states with a higher number of schools and more teachers per district had higher growth rates. Similarly, an increase in urban population also leads to increasing growth rate as the economic value of urban activities is more than rural activities, but at the same time having small effects on gender-related factors like female literacy, sex ratio, etc.

Section VII: References

1. **Access, W. (2018, May 25). *India Socio Economic Data*. Kaggle. Retrieved March 8, 2023, from https://www.kaggle.com/datasets/webaccess/all-census-data**

2. **Sharma, A. (2012). Inter-state disparities in socio-economic development in North East Region of India. *Journal of Agricultural Science*, *4*(9), 236.**

3. **Kumar, K. R., Sivakumar, I., Saravanakumar, N., & Sathishkumar, R. (2020). Regional Disparities and Indian States: A Macro Level Study. *Journal of Critical Reviews*, *7*(13), 87-92.**

4. **Narain, P., Sharma, S. D., Rai, S. C., & Bhatia, V. K. (2000). Regional disparities in socio-economic development in Tamil Nadu. *J. Ind. Soc. Agril. Statist*, *53*, 35-46.**

5. **Narain, P., Sharma, S. D., Rai, S. C., & Bhatia, V. K. (2009). Inter-district variation of socio-economic development in Andhra Pradesh. *J. Ind. Soc. Agril. Statist*, *63*(1), 35-42.**

6. **Gupta, M.D. (2022) *Can't learn, can't work, can't even Cross River: Life in 3 of India's 4 poorest districts, all in up*, *ThePrint*. Available at: https://theprint.in/india/cant-learn-cant-work-cant-even-cross-river-life-in-3-of-indias-4-poorest-districts-all-in-up/841307/ (Accessed: March 8, 2023).**

7. **Ohlan, R. (2013). Pattern of regional disparities in socio-economic development in India: District level analysis. *Social indicators research*, *114*, 841-873.**

8. **Analysts, D.H.S.C. (2021) *Intermediate R - R for survey analysis*, *Chapter 11 Testing regression assumptions*. Available at: https://bookdown.org/jimr1603/Intermediate_R_-_R_for_Survey_Analysis/testing-regression-assumptions.html (Accessed: March 8, 2023).**

9. **Goldblatt, R., You, W., Hanson, G., & Khandelwal, A. K. (2016). Detecting the boundaries of urban areas in india: A dataset for pixel-based image classification in google earth engine. *Remote Sensing*, *8*(8), 634.**