Indian Statistical Institute, Kolkata

An Analysis of Unbalanced Economic Growth in India and Its Causes

Submitted by:

Rahul Konar
Roll Number: BS2338

Supervisor:

Prof. Ayanendranath Basu

November 8, 2024

Contents

Acknowledgements

The author extends heartfelt gratitude to Prof. Ayanendranath Basu for providing the opportunity to undertake this project, and for his constant guidance and support. Sincere thanks are also due to Prof. Samarjit Das, Professor,Economic Research Unit at the Indian Statistical Institute, for his invaluable support, guidance, and assistance in obtaining the dataset. His encouragement and insights have been instrumental throughout this project.

I would also like to thank my colleagues and friends for their encouragement and constructive feedback, which has motivated me throughout this journey. Finally, I extend my appreciation to the academic community at the Indian Statistical Institute, Kolkata, for providing a conducive learning environment.

1 Introduction

India, despite being one of the fastest growing economies in the world, grapples with a widening income gap that reflects disparities across various sections of society. Evidence of India's unbalanced growth is apparent from the numerous studies that find richer states are growing faster, so that state average incomes are diverging. This pattern of divergence might be regarded as unusual given that there are no political barriers to migration, approximately free trade, and a common set of federal institutions.

In this project, we seek to analyze the nature of unbalanced growth by using data on India's 575 districts. In particular we wish to see whether the pattern of divergence across states is similar within states, and, if so, how geographical factors, infrastructure, and other possible factors affect these district-level differences.

The nature of economic growth is measured through a concept of **convergence**, as defined below.

1.1 Convergence

Convergence refers to the process by which poorer regions grow faster economically than their rich counterparts. It is of mainly two types:

1.1.1 β Convergence

According to the Solow-Swann growth model, absolute β convergence refers to the phenomenon where all economies, in the long run, converge to the same common steady state in terms of growth rate and GDP per capita.

Conditional β convergence is a weaker condition of convergence, where all economies converge to their unique steady state,depending on various structural characteristics like savings rates, population growth, technology, education, and institutions, in the long run. It takes into account the fact that growth rate of an economy depends not only on its initial income level but also on country-specific factors.

1.1.2 σ Convergence

According to Barro and Sala-i-Martin(1990,1991), σ convergence is defined as a decline in the variance of district-level per-capita log incomes across time.

2 Data Visualisation

The absolute β convergence is calculated as the regression coefficient β in

$$\log(y_{i,t}) - \log(y_{i,0}) = \alpha + \beta \log(y_{i,0}) + \epsilon_i$$

where $y_{i,t}$ denotes the per capita income of region i at time t, and $y_{i,0}$ denotes the per capita income of region i initially.

A **negative** value of β denotes **absolute convergence**,that is, it implies a tendency for poorer regions to catch up with richer regions in terms of economic growth, whereas a positive value of β denotes divergence.

2.1 Histograms

Keeping the number of bins equal to 10,(we choose this by Sturges' Rule) we try to fit a normal distribution to the distribution of natural log of Per Capita GDP in 2001 and 2008.The histograms suggest that the normal distribution is a good fit.

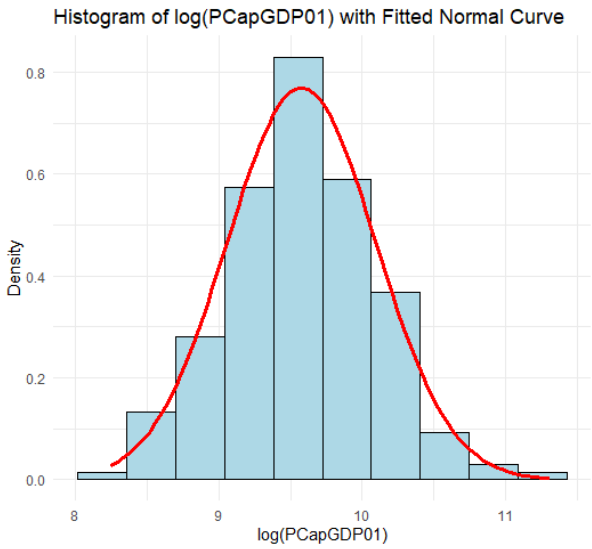

Figure 1: Fitting Normal Distribution to log of Per Capita GDP in 2001

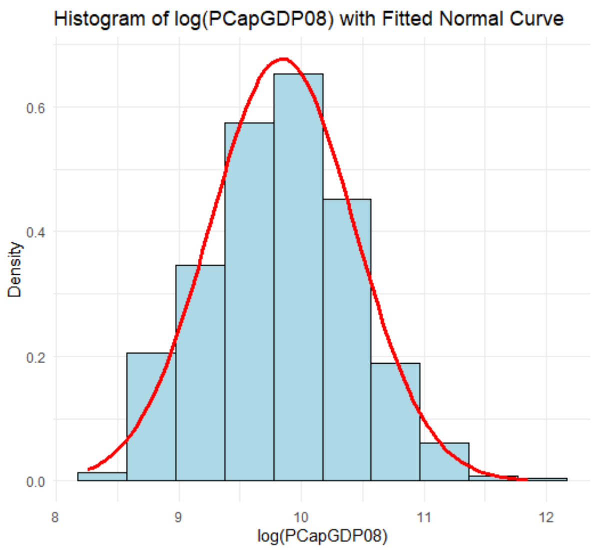

Figure 2: Fitting Normal Distribution to log of Per Capita GDP in 2008

We carry out Pearson's goodness of fit test to test our hypothesis, but find no evidence in the data to reject our null hypothesis, which is that the log transformed data is normally distributed.

2.2 Q-Q Plots and Residual Plot

We obtain the following Q-Q plots for log transformed Per Capita GDP in 2008 and 2001.

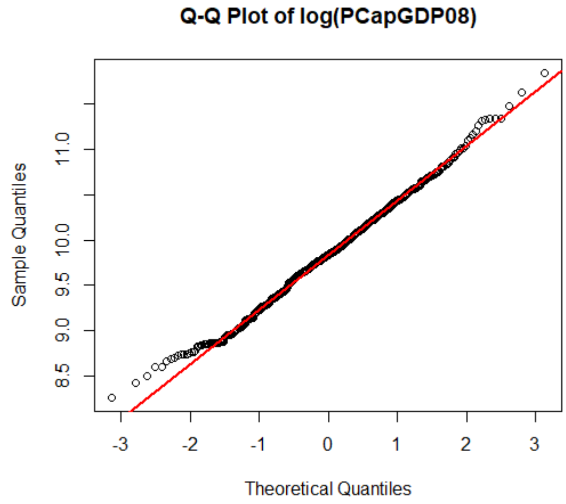

Figure 3: Q-Q plot of log transformed per capita GDP,2008

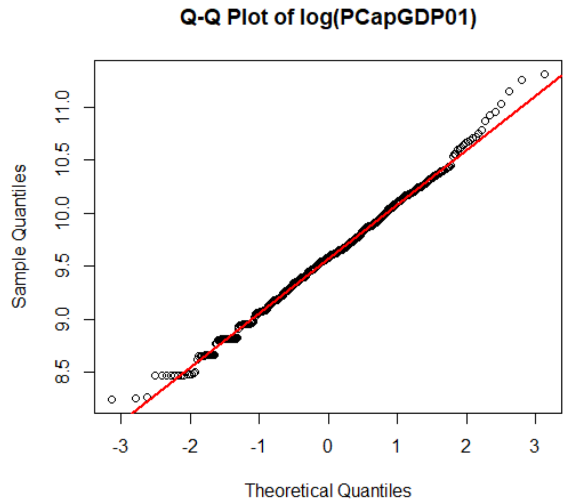

Figure 4: Q-Q Plot of log tansformed per capita GDP,2001

We also obtain the following residual plot after carrying out the regression

$$\log(y_{i,t}) - \log(y_{i,0}) = \alpha + \beta \log(y_{i,0}) + \epsilon_i$$

We observe from the residual plot that the residuals are evenly distributed about the X-axis. The residual plot suggests that the log transformed per capita GDP in both years, 2001 and 2008, is homoscedastic.The Q-Q plots along with the histograms suggest that the log transformed data follows a normal distribution. For further analysis, we make the following assumptions:

- The log transformed per capita GDP in 2008 and 2001 are both normally distributed.
- The log transformed data is homoscedastic.
- ϵ_i are iid $N(0, \sigma^2)$

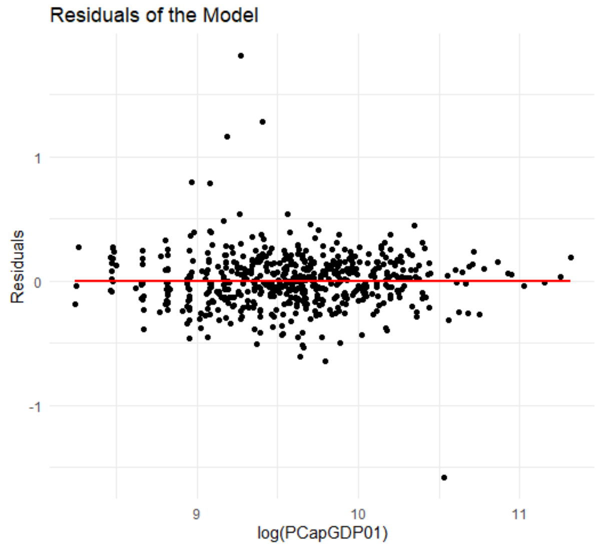

Figure 5: Plot of residuals after regression

3 Analysis of Absolute Convergence

We carry out simple linear regression, with the model

$$\log(y_{i,t}) - \log(y_{i,0}) = \alpha + \beta \log(y_{i,0}) + \epsilon_i$$

to estimate β, our measure of absolute convergence. The β coefficients for different states are as shown in Table 1 (See Appendix).

3.1 Results

There is evidence of divergence all over India, with $\beta = 0.0484$, with a P-value of 0.00872, which is statistically significant at the 1% level. In certain states like Assam, Chattisgarh, Kerala and Rajasthan, there is significant absolute β convergence of per capita GDP. In other states like Haryana, Odisha and Uttar Pradesh, there is significant within state divergence. In most of the states, the estimated β convergence coefficient is insignificantly different from zero. Thus, there is no strong evidence of absolute convergence, within individual states or across the country as a whole.

4 Analysis of σ convergence

Table 2 shows the variance of district log per capita incomes in the two periods, 2001 and 2008. It can be seen that there was a 30.7% increase in the variance of log per-capita incomes across districts—from 0.27 to 0.35. Thus there has also been σ-divergence. We can decompose the India-wide variance of per-capita incomes across all districts, v^T , is equal to the sum of the within-state variance(v^W) and between state variance(v^B), $v^T = v^W + v^B$. We derive this decomposition in the following way: Let y_{ij} be the underlying variable (say, per capita logged income) of j^{th} district in i^{th} state, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, K$.
We define,

$$N = \sum_{i=1}^{K} n_i$$

$$\bar{\bar{y}} = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n_i} y_{ij}$$

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

$$\text{Total Sum of Squares}(TSS) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$$

$$\text{Within Sum of Squares}(WSS) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

$$\text{Between Sum of Squares}(BSS) = \sum_{i=1}^{K} n_i (\bar{y}_i - \bar{\bar{y}})^2$$

Clearly, $TSS = WSS + BSS$. $v^B = \frac{BSS}{N}, v^T = \frac{TSS}{N}, v^W = \frac{WSS}{N}$. Thus,

$$v^T = v^B + v^W$$

This variance decomposition shows that there has been a similar increase in σ-divergence both within states and between states.

5 Causes of Divergence in Growth

We find no strong evidence of absolute β convergence or σ convergence, across the country as a whole or within states. This pattern of divergence might be regarded as unusual given that there are no political barriers to migration, approximately free trade, and a common set of federal institutions.

There may be several reasons for unbalanced growth. One possibility is that, it reflects policy failures such as poor governance, different levels of public infrastructure across states, or corruption. Alternately, it may also result due to natural causes, as emphasised in the new economic geography (NEG) literature. The new economic geography (NEG) literature states that, unbalanced growth may also be a natural outcome in a growing economy. Differences in incomes can arise due to trade and migration costs, and economies of scale associated with agglomerations.

We try to explain this pattern of divergence through association of a district's growth rate with spatial factors like distance from nearest urban agglomeration, and other socio-economic factors.

5.1 Role of Urban Agglomerations

A key insight from the NEG literature is that the combination of increasing returns and factor mobility results in a spatial concentration of economic activity, or agglomerations (Fujita et al., 2001; Helpman & Krugman, 1985; Krugman, 1991; Krugman & Venables,

1995). Agglomerations reflect the existence of increasing returns combined with migration and factor movements. Consequently much of the growth process, such as technology adoption and capital accumulation, occurs in cities. The likely role of urbanization, and urban agglomerations specifically, in explaining the growth of districts in India is therefore twofold. First, more urbanized districts will be able to benefit themselves from increasing returns and hence may be wealthier or experience faster growth. Second, UAs will be able to offer higher prices for exports and higher wages for migrants due to increasing returns. Hence UAs may also benefit neighboring districts through these forward linkages.
To analyse the role of agglomerations statistically, we consider the variable Minimum Distance, D_i, defined as the distance between district i and the closest UA. We take the seven largest cities in India, namely Delhi, Mumbai, Kolkata, Chennai, Bangalore, Hyderabad, and Ahmedabad to be our urban agglomerations (UA's).

5.2 Market Access

Under Dixit-Stiglitz preferences, Market Access(MA) is defined as the GDP weighted average distance to all external markets.

$$MA = \sum_{j=1}^{N} w_j d_{ij}$$

where $w_j = \frac{Y_j}{Y}$, $Y = \sum_{j=1}^{N} Y_j$, Y_j is the GDP per capita income in region j, N is the total number of districts, d_{ij} is the minimum distance between district i and j.

5.3 Model

The empirical model:

$$\ln y_i(t) - \ln y_i(0) = \alpha_0 + \alpha_1 \ln y_i(0) + \alpha_2 \ln D_i + \eta \mathbf{X}_i + \epsilon_i$$

where, $y_i(t)$ denotes the per capita GDP at time t in region i, D_i denotes minimum distance from nearest urban agglomeration, $\mathbf{X}_i$ is a vector of other characteristics of region i including market access, and ϵ_i is a district specific random shock.(for example, climate, institutions, endowments) The coefficient α_1 is interpreted as the conditional convergence coefficient, and $\frac{\alpha_2}{\alpha_1}$ denotes the elasticity of long-run income with respect to minimum distance. The empirical model can be derived from the partial adjustment model, while taking the gravity parameter into consideration.

5.4 Statistical Analysis

- We first ignore $\mathbf{X}_i$ and run a multiple regression with only logarithm of minimum distance. We run a Breusch-Pagan test to check for heteroscedasticity in the model. We obtain a p-value of 0.9886, thus the data does not provide sufficient evidence to suggest that the data is heteroscedastic. For further analysis, we assume the data to be homoscedastic.

- We observe, that the coefficient of conditional convergence is positive, with a p-value of 0.0306, statistically significant at the 5% level, indicating conditional divergence all over India.

- Next, we consider $\mathbf{X}_i$ to be the logarithm of the percentage of urban households in a given district. It is a measure of initial urban population. We find no significant results.
- We consider the logarithm of the net irrigated area in districts, and drop minimum distance. We find that Irrigated land is a significant explanatory variable and has a negative sign. This suggests that more irrigated land per capita is associated with slower transitional growth. The results are significant at the 5% level.
- We consider the logarithm of the percentage of households with electricity along with logarithm of net irrigated area. We observe an α_1 insignificantly different from 0, but negative coefficient for logged net irrigated area and positive coefficient for logged percentage of households with electricity access. The results are significant at the 5% level. This suggests that districts with better public infrastructure grow faster.

The results are given in Table 3 in the Appendix.

6 Conclusion

We have found that per-capita incomes across India's districts are diverging rather than converging, which is the opposite of most within-country experiences. Despite accounting for distance from nearest urban agglomeration, the pattern of divergence remains. Districts with better public infrastructure grow faster than districts with poor public infrastructure. Thus, in order to achieve economic parity, it is essential for policymakers to improve public infrastructure.

A Appendix

A.1 Tables

Table 1: Regression Results by Region

Region	**β Coefficient**	**p-value**
All India	0.0484	0.00872
Andaman & Nicobar Islands	-0.0829	na
Andhra Pradesh	-0.022033675	0.759529232
Arunachal Pradesh	-0.093114933	0.680055605
Assam	-0.232510535	0.00280455
Bihar	-0.047806577	0.65052894
Chhattisgarh	-1.105645834	0.01250104
Delhi	na	na
Goa	na	na
Gujarat	0.008600784	0.905386801
Haryana	0.232659545	0.059605007
Himachal Pradesh	0.057126183	0.763477663
Jammu & Kashmir	0.032620949	0.718144087
Jharkhand	0.213041675	0.103075957
Karnataka	0.071325401	0.394132481
Kerala	-0.273788007	0.065756436
Madhya Pradesh	-0.003198963	0.966540647
Maharashtra	0.083779032	0.170919705
Manipur	-0.005620872	0.983841575
Meghalaya	0.070819269	0.605890002
Mizoram	0.122782266	0.488297966
Nagaland	-0.110084289	0.588609015
Orissa	0.344271945	0.000176368
Pondicherry	-0.081203801	0.720067002
Punjab	-0.03778236	0.860639969
Rajasthan	-0.23628122	0.005904264
Sikkim	0.062790826	0.837333567
Tamil Nadu	0.062353208	0.484531856
Uttar Pradesh	0.09316612	0.021397064
Uttaranchal	0.05554483	0.617567271
West Bengal	0.022995009	0.768534481

Note:

- na denotes not enough observations to compute value
- p-values are found using t-test.

Table 2: Decomposition of σ divergence

	Total variance	**Within State Variance**	**Between State Variance**
2001	0.27	0.12	0.15
2008	0.35	0.15	0.20
Change	0.08	0.03	0.05

Table 3: Exploration of Conditional Convergence

Independent Variable	α_1	α_2	η_1	η_2
Initial income + Minimum Distance	0.04316	-0.01636	NA	NA
Initial income+Net Irrigated Area	0.04637	NA	-0.01557	NA
Initial income + Net Irrigated Area + Electricity	-0.0009858	NA	-0.0189021	0.0422429

Note:

- Initial Income in this context is the logarithm of per capita income in 2001.
- Net Irrigated Area is the natural logarithm of net irrigated area.
- Electricity denotes the natural logarithm of the percentage of households with electricity.
- NA represents that the variables were not part of the model.

References

1. Barro, R. J., & Sala-i Martin, X. (1991). Convergence across states and regions. Brookings Papers on Economic Activity, 1(1), 107–182.

2. Baumol, W. J. (1986). Productivity growth, convergence, and welfare: What the long run data show. American Economic Review, 76(5), 1027–1085.

3. Sala-i Martin, X. (1996). Regional cohesion: Evidence and theories of regional growth and convergence. European Economic Review, 40(6), 1325–1352.

4. Sala-i Martin, X. (1997). I just ran two million regressions. American Economic Review, 87(2), 178–183.

5. Durlauf, S. N., Johnson, P. A., Temple, J. (2005). Growth econometrics. In P. Aghion, S. N. Durlauf (Eds.). Handbook of economic growth (Vol. 1A, pp. 555–677). Amsterdam: North-Holland.

6. Durlauf, S. N., Quah, D. T. (1999). The new empirics of economic growth. Handbook of macroeconomics, 1, 235–308.