SOC476A

Applications of Social Demography

Course Instructor :

Dr. Esha Chatterjee

Submitted By :

Akshat Gupta (200085)
Ravi Patel (200777)
Akash Biswas (200074)
B. Naveen Teja (200266)

Name	**Contribution**
Akshat Gupta	Research, Running Regression
Akash Biswas	Research, Running Regression
Ravi Patel	Research, Creating Project Report
B. Naveen Teja	Research, Creating Project Report

Indian Institute of Technology, Kanpur

November 4, 2023

Assessing the Effects of Infrastructure Development on Health Indicators in Indian States

Akshat Gupta (200085), Ravi Patel (200777), Akash Biswas (200074), Naveen Teja (200266)[a,b]

[a] *Indian Institute of Technology Kanpur,*
[b] *Research Synopsis, submitted to Prof. Esha Chatterjee for the course SOC476A*

Abstract

This paper aims to analyze the impact of infrastructure development on demographic indicators like infant mortality rate, child birth rate and total fertility rate across different Indian states, using a large panel dataset encompassing all the Indian states from the year 2005-2015. We would be using Telephone Lines per 100, Village Electrification, and Road network lengths as the infrastructure indicators to conduct our research. To achieve this, we will run a regression across different Indian states for the time period 2005 - 2015.

Keywords: demography, fertility rates, IMR, infrastructure, PCA

Contents

1. Objective

Analyzing the development of a country is a multifaceted task that requires a comprehensive assessment of various factors. Infrastructure and demographic indicators play a pivotal role in this analysis, as they provide valuable insights into a nation's progress and potential. Our paper deals with three infrastructure indicators from the telecommunications sector (telephone lines per 100), transportation sector (road network length), and power sector (village electrification).

A **well-developed road network** is fundamental for economic growth and social development. Roads facilitate the transportation of goods, people, and services, linking rural and urban areas. They enhance accessibility to markets, healthcare, education, and employment opportunities.
Telephone penetration is a critical indicator of a nation's connectivity and access to information. In the modern age, it goes beyond voice communication and includes internet access through mobile devices. Higher penetration rates correlate with improved communication, business opportunities, and access to education and healthcare services.
Electrification is a key driver of economic development and an essential component of improving living standards. Access to electricity enhances productivity, supports modernization of agriculture, enables better healthcare services, and extends hours for education and economic activities.

These indicators provide a good idea about the infrastructure level of a country and are often-discussed parameters when the development of a country is discussed. Our main objective in this paper to test how these infrastructure indicators affect certain demographic indicators like infant mortality rate, crude birth rate, and total fertility rate.

Infant Mortality Rate (IMR) is a statistical measure that reflects the number of deaths of infants (children under the age of one year) per 1000 live births in a given year or period. IMR is a crucial indicator of the quality of healthcare, access to medical services, and overall living conditions for infants and mothers within a population. It is used to assess the state of maternal and child health and is often used to compare healthcare systems and development levels across different regions or countries.

Crude Birth Rate (CBR) is a demographic measure that represents the total number of live births in a given year within a specific population, typically per 1,000 people, without adjusting for age or gender. CBR provides a basic measure of the population's reproductive behavior and growth. It is used to understand population dynamics, assess the need for family planning programs, and evaluate the impact of fertility-related policies and interventions.

Total Fertility Rate (TFR) is a demographic indicator that estimates the average number of children a woman would give birth to during her lifetime under prevailing age-specific birth rates. It is often expressed as the number of children per woman. TFR is a key measure in demography and population studies. It helps classify countries or regions into demographic transition stages, from high fertility to low fertility. TFR is crucial for understanding future population growth, labor force dynamics, and the potential demographic challenges related to an aging or growing population.

The goal of this paper is understanding how these demographic variables are affected by changes in different infrastructure domains. This insight will be a valuable source of information that can help guide formulation of public policy

2. Literature Review

There has been considerable amount of research done analyzing the impact of growth in infrastructure metrics on the overall development in various facets of society, be it urbanisation, health outcomes or more general trends in demography. Here we provide a brief overview of the prominent work that has been done both at home, and abroad.

Research by Glaeser et al. (2001) examined how improvements in infrastructure, particularly transportation and communication, contribute to urbanization globally. They found that better infrastructure can lead to increased migration to cities, impacting urban demographics. In India, a lot of interesting work has been done pertaining to the impact of investment in infrastructure. A 2004 exploratory study by Indrani Gupta and Arup Mitra found that literacy and industrialisation seem to improve both health outcomes and growth, and to reduce poverty. Andrew J. Barenberg, Deepankar Basu & Ceren Soylu (2017) conducted a panel data study on the impact of public health expenditure on the infant mortality rate (IMR), after controlling for other relevant covariates like political competition, per capita income, female literacy, and urbanisation. They found out that a one percent increase in public health expenditure reduced IMR by 9 infant deaths per 1000 live births. They also observed that female literacy and urbanization reduced IMR.

Another notable work in the Indian context in this topic was done by Ghosh and De (2004). They tested how different categories of infrastructure affect the development in different Indian states. One of their key findings was that there are still alarming levels of interstate differences in social, physical and economic infrastructure, and these differences contribute significantly to regional income disparities. We hope to build on this existing literature and derive robust results determining the impact of telecommunications, power and transport infrastructure on demographic indicators like crude birth rate, infant mortality rate and total fertility rate.

3. Theoretical Framework

There has been a long-standing consensus in the field of population studies, health studies and sociology in general, that economic development can be treated as a determinant of health. The economic development and health framework is based on the premise that there is a **strong relationship between a country's economic development and the health of its population**. This framework highlights the idea that improvements in economic conditions can have **positive effects** on various health outcomes. A great deal of research has been conducted using this theoretical framework, like **Subramanian, Belli, Kawachi 2002**, and other papers mentioned in the literature review. The economic development and health framework are often associated with broader human development theories, such as the Human Development Index (HDI). Health is considered one of the key components of human development alongside income and education. We will be testing this framework in the Indian context by analysing how different health outcomes (TFR, IMR and CBR in our case) are affected by differing levels of infrastructure development across Indian states. We will create different proxies of infrastructure variables from variables like Telephone density, Road network length and Village Electrification levels.

4. Methodology & Data Sources

We are using the year-wise and state-wise data compiled by the Reserve Bank of India on various infrastructure and socio-demographic indicators, provided via the **Handbook of Statistics on Indian States**. We have not considered the states Sikkim, Mizoram, Arunachal Pradesh, Nagaland, Manipur, and the Union Territiores due to insufficient data. We would note here that systematic and wide-ranging data collection is needed to conduct more robust research.

We will be using the data provided on **State-wise Telephones per 100 Population**, **State-wise Length of Roads** and **State-wise Village Electrification** as our infrastructure indicators. We will be testing the impact of these infrastructure indicators on the following three demographic indicators: **State-wise Birth Rate**, **State-wise Infant Mortality Rate**, and **State-wise Total Fertility Rate**. We will testing the relationship using data from time period 2005-2015.

We can model our research question using the following equation:

$$\ln \mathrm{Y}_{it} = \alpha_i + \beta_1 \ln\mathrm{SGDP}_{it} + \beta_2 \ln\mathrm{TL}_{it} + \beta_3 \ln\mathrm{RD}_{it} + \beta_4 \ln\mathrm{VE}_{it} + \epsilon_{it}$$

Here Y_{it} represents state-wise and year-wise IMR, CBR and TFR in three seperate regressions that we conduct. $SGDP_{it}$ represents the state-wise and year-wise GDP. TL_{it} represents the state-wise and year-wise Telephones per 100 Population. RD_{it} represents the Road Network per unit State Area for each state across the time period. VE_{it} represents the village electrification levels across state and time. Also, we have used an **individual fixed effects** model with α_i representing individual-specific fixed effect. We use this model to prevent omitted variable bias, by accounting for state-specific effects that also have an impact on the health outcomes.

Since one might find a high degree of association between different infrastructure variables, we will be conducting **Principal Component Analysis** to derive modified infrastructure indices to get more robust results. We will be regressing our dependent variables, i.e., CBR, IMR and TFR across different states for three different time periods, once using these principal components and once without to compare the results.

5. Empirical Results & Discussion

Testing the impacts of our infrastructure variables on **Infant Mortality Rate (without PCA)**, we find the following:

Fixed-effects (within) regression		Number of obs	= 208
Group variable: id		Number of groups	= 22
R-sq:		Obs per group:	
within	= 0.7806	min	= 6
between	= 0.2641	avg	= 9.5
overall	= 0.3411	max	= 10
		F(4,182)	= 161.86
corr(u_i, Xb)	= -0.1215	Prob > F	= 0.0000

logIMR	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logSGDP	-.0129604	.0162816	-0.80	0.427	-.0450855	.0191646
logVL	.0987842	.0837954	1.18	0.240	-.0665512	.2641196
logTL	-.1328628	.0119866	-11.08	0.000	-.1565135	-.1092122
logRD	-.190692	.038352	-4.97	0.000	-.2663637	-.1150202
_cons	4.330037	.222603	19.45	0.000	3.890822	4.769251
sigma_u	.32442004					
sigma_e	.07530665					
rho	.94887199	(fraction of variance due to u_i)				

F test that all u_i=0: F(21, 182) = 166.90 Prob > F = 0.0000

Figure 1: Fixed Effects Panel Regression results for IMR (without PCA)

- We can see that IMR decreases with an increase in State-GDP, Telephone Network Density and Road Network Length. These results are in line with the theoretical framework.
- A unit increase Road Density leads to a **0.2 unit** decrease in Infant Mortality Rate, whereas a unit increase in Telephone lines per 100 leads to around **0.13 unit** decrease in IMR.
- However we observe that IMR actually increases with Village Electrification Levels, but this result is not statistically significant (p-value = 0.240).

Testing the impacts of our infrastructure variables on **Crude Birth Rate(without PCA)**, we find the following:

- We can see that CBR decreases with an increase in State-GDP, Telephone Network Density and Road Network Length. These results are in line with the theoretical framework.
- A unit increase Road Density leads to a **0.06 unit** decrease in Crude Birth Rate, whereas a unit increase in Telephone lines per 100 leads to around **0.03 unit** decrease in CBR.
- However we observe that CBR actually increases with Village Electrification Levels, but this result is not statistically significant (p-value = 0.650).

Fixed-effects (within) regression		Number of obs	= 208
Group variable: id		Number of groups	= 22
R-sq:		Obs per group:	
within	= 0.8513	min	= 6
between	= 0.2602	avg	= 9.5
overall	= 0.2560	max	= 10
		F(4,182)	= 260.45
corr(u_i, Xb)	= 0.1110	Prob > F	= 0.0000

logCBR	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logSGDP	-.0054314	.003773	-1.44	0.152	-.0128759	.002013
logVL	.0086699	.0194181	0.45	0.656	-.0296437	.0469835
logTL	-.0356503	.0027777	-12.83	0.000	-.0411309	-.0301697
logRD	-.0627286	.0088874	-7.06	0.000	-.0802642	-.045193
_cons	3.237254	.0515844	62.76	0.000	3.135473	3.339034
sigma_u	.17770933					
sigma_e	.017451					
rho	.99044892	(fraction of variance due to u_i)				

F test that all u_i=0: F(21, 182) = 877.82 Prob > F = 0.0000

Figure 2: Fixed Effects Panel Regression results for CBR (without PCA)

Testing the impacts of our infrastructure variables on **Total Fertility Rate(without PCA)**, we find the following:

Fixed-effects (within) regression		Number of obs	= 208
Group variable: id		Number of groups	= 22
R-sq:		Obs per group:	
within	= 0.7968	min	= 6
between	= 0.3103	avg	= 9.5
overall	= 0.3193	max	= 10
		F(4,182)	= 178.39
corr(u_i, Xb)	= 0.0031	Prob > F	= 0.0000

logTFR	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logSGDP	-.0086436	.0077371	-1.12	0.265	-.0239096	.0066223
logVL	-.091593	.0398199	-2.30	0.023	-.1701611	-.013025
logTL	-.0544903	.0056961	-9.57	0.000	-.0657292	-.0432514
logRD	-.1061529	.018225	-5.82	0.000	-.1421125	-.0701934
_cons	1.268463	.1057819	11.99	0.000	1.059746	1.477179
sigma_u	.20967219					
sigma_e	.03578603					
rho	.97169422	(fraction of variance due to u_i)				

F test that all u_i=0: F(21, 182) = 287.78 Prob > F = 0.0000

Figure 3: Fixed Effects Panel Regression results for TFR (without PCA)

- We can see that TFR decreases with an increase in State-GDP, Telephone Network Density, Village Electrification and Road Network Length. These results are in line with the theoretical framework.
- A unit increase Road Density leads to a **0.1 unit** decrease in Total Fertility Rate, whereas a unit increase in Telephone lines per 100 leads to around **0.05 unit** decrease in TFR. A unit increase in Village Electrification also leads to a **0.09 unit** decrease in TFR.
- Differing from the previous two observations, impact of Village Electrification Levels on TFR is in line with theory and statistically significant.

Now, we conduct these regressions again by using **Principal Componenet Analysis (PCA)**. We divide our infrastructure variables into the following three components:

Variable	Comp1	Comp2	Comp3
logTL	**0.6388**	**-0.1132**	**-0.7610**
logRD	**0.5110**	**0.8019**	**0.3096**
logVL	**0.5752**	**-0.5867**	**0.5701**

Figure 4: Principal Component Analysis

Here **Comp1** denotes a variable which is moderately and almost equally correlated with all our three infrastructure variables. **Comp2** denotes a variable is which is largely correlated with Road Network per State Area variable and moderately negatively-correlated with Village Electrification Level. **Comp3** denotes a variable which is largely negatively-correlated with Telephone Density and moderately correlated with the remaining two infrastructure variables.

Testing the impacts of our infrastructure variables on **Infant Mortality Rate (with PCA)**, we find the following:

```
Fixed-effects (within) regression               Number of obs     =        208
Group variable: id                              Number of groups  =         22

R-sq:                                           Obs per group:
     within  = 0.7806                                         min =          6
     between = 0.2641                                         avg =        9.5
     overall = 0.3411                                         max =         10

                                                F(4,182)          =     161.86
corr(u_i, Xb)  = -0.1215                        Prob > F          =     0.0000

------------------------------------------------------------------------------
      logIMR |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logSGDP |  -.0129604   .0162816    -0.80   0.427    -.0450855    .0191646
         pc1 |  -.1753668   .0172371   -10.17   0.000    -.2093769   -.1413566
         pc2 |  -.1369692   .0309726    -4.42   0.000    -.1980807   -.0758576
         pc3 |   .0613633   .0194502     3.15   0.002     .0229863    .0997402
       _cons |   3.918342   .2383448    16.44   0.000     3.448068    4.388616
-------------+----------------------------------------------------------------
     sigma_u |  .32442004
     sigma_e |  .07530665
         rho |  .94887199   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(21, 182) = 166.90                   Prob > F = 0.0000
```

Figure 5: Fixed Effects Panel Regression results for IMR (with PCA)

- We can see that IMR decreases with an increase in State-GDP, pc1 and pc2. However IMR increases with an increase in pc3.
- We can see that these results are in line with previous results, with components that are more correlated with Road Networks and Telephone Densities (pc1 and pc2) have negative effects on IMR.

Testing the impacts of our infrastructure variables on **Crude Birth Rate(with PCA)**, we find the following:

- We can see that CBR decreases with an increase in State-GDP, pc1 and pc2. However IMR increases with an increase in pc3.

```
Fixed-effects (within) regression               Number of obs     =        208
Group variable: id                              Number of groups  =         22

R-sq:                                           Obs per group:
     within  = 0.8513                                         min =          6
     between = 0.2602                                         avg =        9.5
     overall = 0.2560                                         max =         10

                                                F(4,182)          =     260.45
corr(u_i, Xb)  = 0.1110                         Prob > F          =     0.0000

------------------------------------------------------------------------------
      logCBR |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logSGDP |  -.0054314    .003773    -1.44   0.152    -.0128759     .002013
         pc1 |  -.0541487   .0039944   -13.56   0.000    -.0620299   -.0462674
         pc2 |  -.0439913   .0071774    -6.13   0.000    -.0581528   -.0298297
         pc3 |   .0115876   .0045073     2.57   0.011     .0026944    .0204808
       _cons |   3.107465   .0552322    56.26   0.000     2.998487    3.216443
-------------+----------------------------------------------------------------
     sigma_u |  .17770933
     sigma_e |    .017451
         rho |  .99044892   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(21, 182) = 877.82                   Prob > F = 0.0000
```

Figure 6: Fixed Effects Panel Regression results for CBR (with PCA)

- We can see that these results are in line with previous results, with components that are more correlated with Road Networks and Telephone Densities (pc1 and pc2) have negative effects on CBR.

Testing the impacts of our infrastructure variables on **Total Fertility Rate(with PCA)**, we find the following:

```
Fixed-effects (within) regression               Number of obs     =        208
Group variable: id                              Number of groups  =         22

R-sq:                                           Obs per group:
     within  = 0.7968                                         min =          6
     between = 0.3103                                         avg =        9.5
     overall = 0.3193                                         max =         10

                                                F(4,182)          =     178.39
corr(u_i, Xb)  = 0.0031                         Prob > F          =     0.0000

------------------------------------------------------------------------------
      logTFR |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logSGDP |  -.0086436   .0077371    -1.12   0.265    -.0239096    .0066223
         pc1 |  -.0965688   .0081911   -11.79   0.000    -.1127306    -.080407
         pc2 |  -.0660217   .0147183    -4.49   0.000    -.0950622   -.0369812
         pc3 |    .005942   .0092428     0.64   0.521    -.0122948    .0241789
       _cons |   .9711428   .1132624     8.57   0.000     .7476665    1.194619
-------------+----------------------------------------------------------------
     sigma_u |  .20967219
     sigma_e |  .03578603
         rho |  .97169422   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(21, 182) = 287.78                   Prob > F = 0.0000
```

Figure 7: Fixed Effects Panel Regression results for TFR (with PCA)

- We can see that TFR decreases with an increase in State-GDP, pc1 and pc2. However IMR increases with an increase in pc3.
- Although the results for pc3 are not statistically significant.
- We can see that these results are in line with previous results, with components that are more correlated with Road Networks and Telephone Densities (pc1 and pc2) have negative effects on TFR.

6. Visualizations

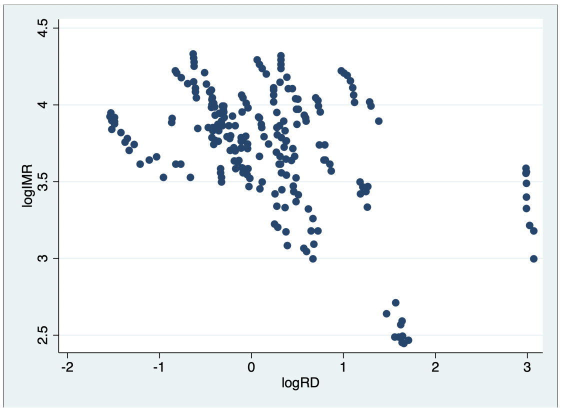

Figure 8: logIMR v logRD

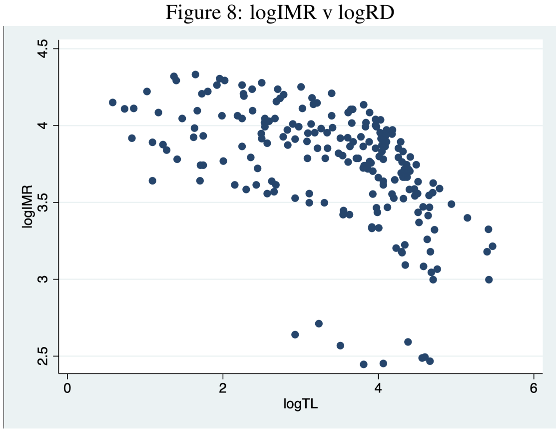

Figure 9: logIMR v logTL

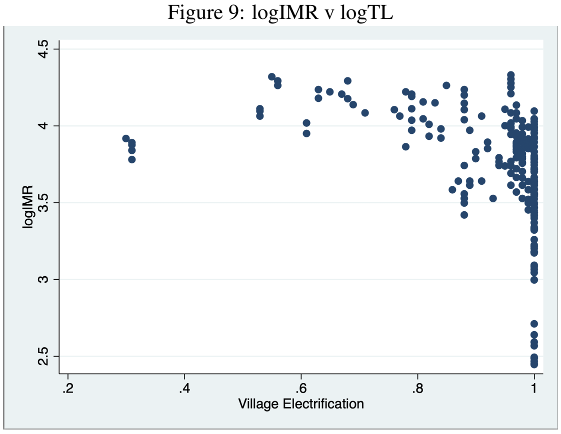

Figure 10: logIMR v logVL

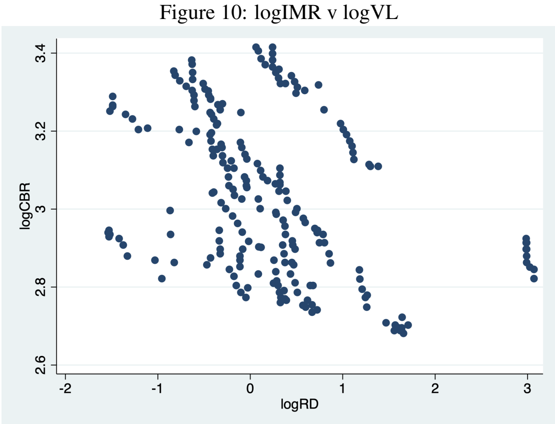

Figure 11: logCBR v logRD

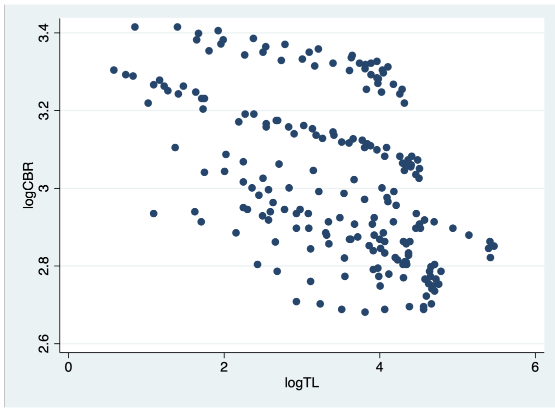

Figure 12: logCBR v logTL

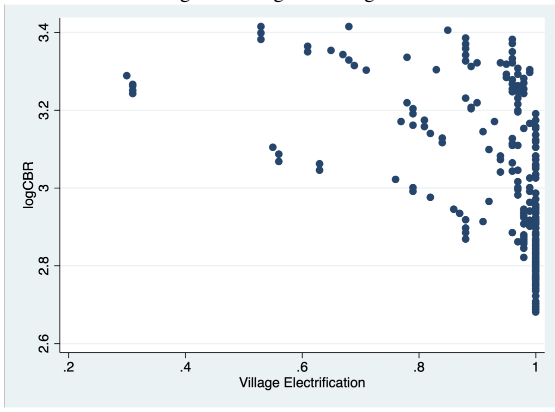

Figure 13: logCBR v logVL

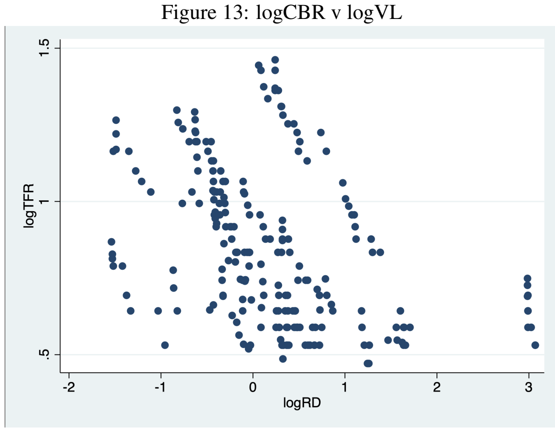

Figure 14: logTFR v logRD

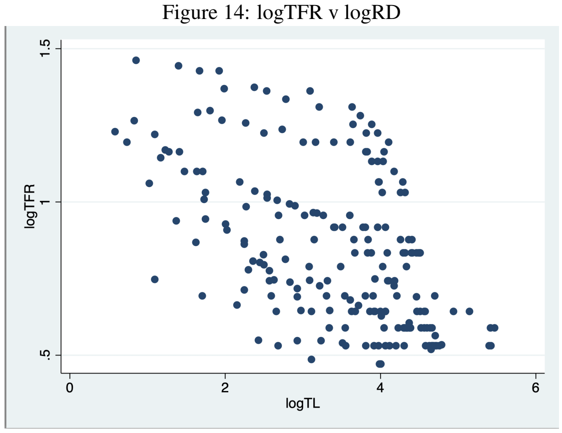

Figure 15: logTFR v logTL

7. Conclusion

We conducted regression for the data employing panel regression and PCA. Using both the methods, we can draw the following conclusions

- Decrease in IMR, CBR & TFR can be attributed to Road network density followed by Telephone Density.
- This makes sense as an improved road network provides better accessibility to healthcare facilities that can understandably lead to lower levels of Infant Mortality Rate, Crude Birth Rate and Total Fertility Rate.
- Improved Telephone Network can signify better outreach among the population, and also indicate higher income levels. This should be understandably correlated with lower levels of IMR, CBR, and TFR.
- One possible explanation for no significant correlation between Village Electrification and other health indicators might be due to a case of diminishing returns to scale. Much of the progress in improving village electrification was already achieved before 2005, the year from which our dataset was taken.
- We also found that Village Electrification only had a significant relationship with Total Fertility Rate, which can be understood as improved electrification leading to better mass media access and programs pertaining to family planning.
- The state's GDP also accounts for the decrease in the IMR, CBR & TFR but, the weight is much less. This shows that higher economic development is on average a decent indicator of improved health.
- We find that even in the Indian context, states with higher levels of economic development, tend to have on average better health outcomes, with lower birth rates, lower infant mortality rates and lower fertility rates.

8. References

- Buddhadeb Ghosh, and Prabir De. "How Do Different Categories of Infrastructure Affect Development? Evidence from Indian States." Economic and Political Weekly, vol. 39, no. 42, 2004, pp. 4645–57. JSTOR, http://www.jstor.org/stable/4415682.
- Gupta, I. and Mitra, A. (2004), Economic Growth, Health and Poverty: An Exploratory Study for India. Development Policy Review, 22: 193-206. https://doi.org/10.1111/j.1467-7679.2004.00245.x
- Andrew J. Barenberg, Deepankar Basu & Ceren Soylu (2017) The Effect of Public Health Expenditure on Infant Mortality: Evidence from a Panel of Indian States, 1983–1984 to 2011–2012, The Journal of Development Studies, 53:10, 1765-1784, DOI: 10.1080/00220388.2016.1241384