



Econometrics Year II Research Project

# **Assessing the Impact of Literacy and Income levels on Crime Rates in India**

Jagriti Agnani (231720)  
Mehak Singh (231731)  
Kashish Thakur (231724)

B.A. (H) Economics  
Gargi College, University of Delhi  
Year II

## INDEX

Page	Table of Contents
2-4	<b>I. Abstract</b> 1.1 Overview 1.2 Time Period and Beta Coefficients
5-7	<b>II. Introduction</b> 2.1 Current Landscape 2.2 Methods
8-10	<b>III. Literature Review</b> 3.1 Theoretical Framework 3.2 Theoretical Grounding
11-16	<b>IV. Research Methodology and Data</b> 4.1 Data Sources 4.2 Variables 4.3 Regression Equation 4.4 Research Design 4.5 Assumptions (Simple Linear) 4.6 Data Tables 4.6.1 Table 1.1 4.6.1 Table 1.2 4.7 Limitations of Methodology
17-24	<b>V. Results and Interpretations</b> 5.1 Outcomes 5.2 Macroeconomic Interpretation 5.3 Microeconomic Interpretation 5.4 Research Gaps
25-26	<b>VI. Conclusion</b> 6.1 Key Takeaways 6.2 Suggestive Policies
27	<b>VII. References and Bibliography</b>

## **I) Abstract**

### **1.1 Overview**

The rising crime rates in India remain a pressing issue of national concern. From petty theft to serious violent offenses, the increasing levels of crime not only disrupt public order but also impede the socio-economic development of the country. While policymakers and researchers alike have investigated various determinants of crime including but not limited to urbanization, policing efficiency, social inequality, and unemployment—**education** continues to stand out as a potentially transformative factor in shaping individual behavior and societal outcomes.

This research paper attempts to empirically examine the relationship between education, income, and crime, focusing on data collected from different states of India for the years 2020 and 2021. Using data collected from various sources, we use the regression model to determine whether higher literacy is associated with lower crime.

The core objective of this study is to evaluate whether improvements in literacy rates and per capita income levels across Indian states have any statistically significant influence on crime rates. The analysis is conducted using a panel data regression model with state-level data across two consecutive years. We hypothesize that higher literacy and increased economic well-being represented by per capita income should ideally lead to lower incidences of crime. This hypothesis stems from the widely held belief that education and financial stability are key instruments of social progress and personal development, both of which are thought to reduce the likelihood of engaging in criminal activity.

### **1.2 Time Period and Beta Coefficients**

In line with this, we collected secondary data from government sources such as the National Crime Records Bureau (NCRB), the Ministry of Education, and the Ministry of Statistics and Programme Implementation (MoSPI) and DataGov. The dataset includes the total reported crime rates per 1,00,000 population, state-wise literacy percentages, and per capita income (in rupees) at constant prices. By comparing this data across the two years of 2020 and 2021, we aim to uncover any consistent patterns or anomalies that could offer insight into how education and economic conditions affect crime trends in India.

Although the preliminary results may not reveal a statistically significant relationship, the study provides a basic understanding of how education might affect social behavior. These findings underscore the need of integrating the role of education in reducing crime across the country.

Interestingly, while our initial expectations pointed towards a strong inverse relationship where higher literacy and income levels would correspond with lower crime rates. The

empirical results did not align with this assumption. The beta coefficients obtained from our regression models revealed a more nuanced and, at times, counterintuitive picture. In some cases, states with higher literacy levels also reported higher crime rates, which could be due to a range of intervening variables such as urban density, migration, or increased reporting and awareness in more educated regions. Likewise, higher per capita income did not uniformly predict lower crime rates, suggesting that the relationship between economic prosperity and social harmony is more complex than traditionally perceived. This also means possible specification biasedness in the model.

One of the most intriguing aspects of the study is the role played by other latent factors that mediate the relationship between education, income, and crime. For instance, employment opportunities, family stability, mental health, and substance abuse are all critical variables that can influence criminal behavior but are harder to quantify across all states consistently. The study discusses these potential mediators at length, emphasizing the need to interpret the findings within a broader sociological and psychological context. It is possible that in some states, educational attainment may lead to higher expectations for employment and income, and if those expectations are unmet, the resulting frustration could contribute to social unrest or criminal activity.

Furthermore, we consider the time period of 2020–2021, which coincided with the COVID-19 pandemic and its resulting economic disruptions, as an important contextual factor. The pandemic not only caused job losses and income insecurity for millions of Indians but also led to a spike in certain types of crimes, such as domestic violence and cybercrime. Therefore, our results must be understood within this exceptional socio-economic backdrop, which may have temporarily distorted or amplified some trends.

Another important contribution of this paper is its emphasis on state-wise heterogeneity. India is a country of immense diversity, with states differing significantly in terms of governance, cultural norms, education systems, economic structures, and law enforcement capabilities. Therefore, a one-size-fits-all explanation for crime cannot be applied across all regions. Our findings reinforce this view, showing that while some states do follow the expected inverse relationship between literacy, income, and crime, others deviate significantly, thereby demanding localized policy responses.

Overall, the Beta Coefficient has proven our expectations to be otherwise, the interpretation of which forces us to rethink about the factors that continuously affect crime in India and how we can improve quality of life as well as safety in the country.

## **II) Introduction**

### **2.1 Current Landscape**

Crime remains an enduring and multifaceted challenge in India, with implications that reach far beyond immediate legal or safety concerns. Its impacts touch upon the nation's development trajectory, economic growth, and the sense of personal security that is vital to a functioning democracy. While overall crime rates in India fluctuate year by year, a persistent concern is the significant variation in crime levels across different states and union territories. This regional heterogeneity has raised important questions about the socio-economic, cultural, and structural determinants of crime in the Indian context.

The contribution of education, particularly literacy, in containing crime rates is usually assigned to enhanced awareness, better employment prospects and enhanced social responsibility. Therefore, after assessing data from the 2011 census, we now shifted to using newer data (2020-21) to explore the direct effects of literacy rates on crime rates in each state. Per Capita Income added as an additional explanatory variable, since higher income residents have better opportunities and are able to afford quality education.

The question we still stand to ask is: Does literacy/income help decrease crime or not? That is our concern. National Crime Records Bureau (NCRB) and Census data (2011), DataGov and MoSPI are used to research Indian states' data for the year 2020-2021. This research takes as its starting point the premise that education and economic well-being are fundamental to understanding patterns of criminal behavior. More specifically, it investigates the extent to which literacy and per capita income influence crime rates at the state level in India. Drawing on data from 2020 and 2021, this study leverages regression analysis to uncover whether there is any statistically significant correlation between these socio-economic indicators and crime. By focusing on these two years, the study captures the period immediately following the COVID-19 pandemic's peak, when states were grappling with both health crises and economic dislocations, factors that may have intensified or altered crime trends. What sets this research apart is not only its focus on Indian states but also its attempt to examine education and income together, rather than in isolation.

In India, education quality and access vary widely across states. While Kerala boasts a literacy rate above 85%, other states such as Bihar and Jharkhand still face significant challenges in both school enrollment and adult literacy. Similarly, income levels also differ sharply, with states like Goa and Maharashtra reporting high per capita income while states like Uttar Pradesh and Madhya Pradesh lag behind. This variability provides a unique opportunity for comparative analysis. The broader question we pose is whether the presumed

crime-reducing effects of education and income hold true across this heterogeneous national landscape.

## **2.2 Methods**

The methodology adopted for this paper primarily involves the use of Ordinary Least Squares (OLS) regression analysis to determine the extent and direction of the relationship between the dependent variable (crime rate) and independent variables (literacy rate and per capita income). The inclusion of both education and income as explanatory variables allows for a broader understanding of socio-economic determinants of crime. In addition to regression analysis, we also conduct correlation tests to check for multicollinearity between literacy and income, ensuring the robustness of our model. Past research in other countries has shown a negative correlation between schooling and crime, but there is limited literature along these lines in India. In this paper, we try to shed some light and add to the overall debate on education and crime in India. The paper is organized into different sections like a literature review, data and variables employed, methodology, results, and finally a conclusion with policy implications.

Also, the socio-economic heterogeneity of India as a nation means that it offers a compelling case to study. States differ quite significantly not only in literacy levels, but in levels of income, urbanization, police effectiveness, and social organization — all of which have the potential to impact crime in various ways. By studying a range of large states, this work incorporates this diversity and allows a more meaningful analysis of whether literacy is making an equal contribution to crime across states.

Although the regression results may not demonstrate a uniformly strong or statistically significant negative relationship between literacy/income and crime, the broader patterns suggest that education and economic well-being still play vital roles in shaping social behavior. More importantly, these findings open the door for further interdisciplinary research and policy analysis. Future studies might include additional variables such as urbanization rates, unemployment, gender ratios, or caste-based discrimination to build a more comprehensive model. Moreover, qualitative research involving interviews, case studies, and ground-level data can supplement the quantitative findings and provide deeper insights into the lived experiences behind crime statistics.

This study also aims to contribute to the ongoing policy conversation around crime prevention in India. For decades, crime control strategies in the country have been largely reactive and enforcement-based, focusing on increasing policing, surveillance, and incarceration. However, without addressing the root socio-economic causes of crime, such efforts have had limited long-term impact. This research argues for a more holistic

approach—one that views education and equitable economic growth as long-term investments in public safety.

A key contribution of this paper is its call for more nuanced and localized policymaking. Given the vast inter-state differences in India, a universal crime prevention model is unlikely to work. States must be encouraged to use data-driven approaches to understand their unique crime profiles and to tailor education and employment policies accordingly. For instance, in some states, increasing female literacy might have a greater impact on reducing certain forms of gender-based violence, while in others, vocational training might be more effective in addressing youth crime.

Our paper ultimately calls for a more holistic approach to crime prevention—one that recognizes education not just as a tool for employment but as a mechanism for inculcating values, emotional intelligence, and conflict resolution skills. Similarly, economic policies must ensure that income growth translates into equitable opportunities and social safety nets. Only by integrating these dimensions can we begin to understand and address the root causes of crime in India. The findings of this study thus serve as a starting point for reimagining crime reduction strategies that go beyond punitive measures and instead focus on building a more educated, empowered, and equitable society.

Although the emphasis of this research is on literacy as the most important independent variable, we realize that crime is a complex problem with numerous interrelated variables playing important roles. What we aim to examine is not to affirm that education holds all the answers to crime, but to examine whether there is a measurable relationship between the two. What can be learned from this kind of analysis can be useful towards developing better-rounded and more data-informed policymaking systems in education and law and order administration systems.

### **III) Literature Review**

#### **3.1 Theoretical Framework**

Understanding the determinants of crime has long been a central concern in both economics and sociology. The interplay between socio-economic variables and criminal behavior continues to shape research agendas, especially in developing countries where disparities in development and access to resources are stark. Among the variables most frequently studied, education and income stand out as powerful, yet complex, factors associated with crime. Their influence is especially pertinent in a country like India, where wide regional disparities in literacy rates and income levels may significantly affect the social fabric and crime dynamics.

One of the foundational theoretical contributions to this discourse comes from Gary Becker's (1968) economic theory of crime, which applies the rational choice framework to criminal behavior. According to Becker, individuals weigh the potential benefits of committing a crime against the probability of being caught and punished, as well as the opportunity cost of foregoing legitimate income through legal employment. In this context, education plays a dual role: it increases access to legal earning opportunities and raises the cost of engaging in crime. Thus, theoretically, higher levels of education should be associated with lower crime rates.

Building on Becker's work, Lochner and Moretti (2004) provide robust empirical evidence using U.S. data. Their analysis demonstrates that high school graduation significantly reduces the likelihood of incarceration and arrest, particularly for violent and property crimes. This study reinforces the view that educational attainment acts as a deterrent to crime by enhancing both cognitive and economic capacities. The mechanisms they propose include higher legal wages, stronger social ties, and greater risk aversion among the educated.

Turning to the Indian context, Kumar and Das (2016) conducted a comprehensive state-level analysis of the socio-economic determinants of crime. Their findings suggest a clear negative relationship between literacy levels and crime, particularly for crimes against property and



violent offenses. They argue that states with higher educational attainment tend to have more stable social environments, better employment structures, and stronger institutional trust—all of which help in mitigating crime. Similarly, Rajan and Mishra (2018) focus on the role of economic prosperity and find that per capita income is inversely correlated with crime incidence. According to their findings, economic security contributes to crime deterrence by improving access to necessities and reducing incentives for illicit behavior.

However, the relationship between education and crime is not always straightforward. Several scholars caution against drawing overly simplistic or linear conclusions. For example, Jaitman and Machin (2013) contend that while education may generally reduce crime, its impact is conditional upon other structural variables. These include urbanization, inequality, youth unemployment, and quality of education—all of which can dilute or even reverse the expected effects. In highly urbanized or economically unequal settings, educated individuals may still turn to crime due to unmet expectations, lack of quality jobs, or frustration with systemic inefficiencies.

In addition, Gaviria and Raphael (2001) in their cross-country analysis argue that income inequality, rather than absolute income, may have a stronger association with crime. Their findings imply that in societies where disparities are pronounced, even educated individuals might experience social exclusion or relative deprivation, increasing the risk of deviant behavior. This adds another layer of complexity to the education-crime-income nexus, especially in a socio-economically stratified country like India.

Another important contribution comes from the work of Chakraborty and Mukherjee (2020), who examine crime trends in post-liberalization India. They suggest that the rapid economic changes since the 1990s have led to an uneven distribution of opportunities. While urban centers have seen growth in education and employment, rural and semi-urban areas continue to face infrastructural and institutional deficits. This duality may partly explain why crime patterns often defy expectations based on simple literacy or income metrics.

The current study builds upon these foundations by focusing specifically on Indian states over the two-year period of 2020–2021, a period marked by post-pandemic recovery, educational disruptions, and economic uncertainty. Unlike much of the existing literature, which tends to rely on older or pre-2010 data, this study uses recent government datasets to provide a more updated picture of the education-crime relationship in modern India. In doing so, it attempts to capture how newer socio-economic realities interact with traditional predictors of crime.

### **3.2 Theoretical Grounding**

The study draws upon a range of sociological and economic theories to frame its investigation. First, Human Capital Theory, as articulated by Becker (1968), forms the

economic backbone of the analysis. It posits that investment in education increases individual productivity and potential earnings, thereby making crime a less attractive option.

Second, Social Control Theory (Hirschi, 1969) suggests that education fosters conformity by instilling societal norms and strengthening attachments to institutions such as schools, families, and workplaces. Individuals who are better integrated into these social structures are less likely to engage in deviant behavior, as they have more to lose from non-conformity.

Third, Structural Strain Theory (Merton, 1938) offers an alternative explanation particularly relevant for unequal societies like India. This theory posits that crime arises when individuals experience a gap between societal expectations (e.g., success, prosperity) and the legitimate means to achieve them (e.g., quality education, fair employment). In such cases, crime can be seen as an alternative path to achieving culturally valued goals. This framework is especially relevant when considering regions where educational attainment does not necessarily translate into meaningful economic opportunities, thus generating frustration and potential deviance.

Overall, the literature suggests a nuanced and often context-dependent relationship between education, income, and crime. While theoretical models and empirical evidence from developed countries suggest a clear negative relationship, the Indian context—marked by stark regional, economic, and institutional diversity demands a more differentiated analysis.

This study seeks to contribute to the growing body of literature by offering a recent, India-specific investigation of the topic using empirical data and robust econometric techniques. By examining whether the relationships suggested in international studies hold true in India's federal and highly varied socio-economic framework, this research not only fills a gap in the existing literature but also offers insights that may be useful for policy development at both state and national levels.

#### **IV) Methodology and Data**

This study attempts to explore the correlation between the literacy and crime rates and per capita income of all Indian states and union territories for the years 2020 and 2021. The assumption is that higher literacy and per capita will come with a reduced rate of crime and hence a negative or inverse relationship. To account for the economic variation between states, per capita income is taken as the second explanatory variable.

The data set includes details for Indian states. For every state, we have information for total registered crime, literacy, and per capita income for both the selected years. We have used crime figures from the National Crime Records Bureau (NCRB), literacy from state government reports and statistical abstracts, and income from the Ministry of Statistics and Programme Implementation (MoSPI). Given the complexity of social behavior and regional heterogeneity in India, a quantitative empirical approach using panel data enables us to test for both temporal and cross-sectional variations in crime and its socio-economic determinants.

We utilized Ordinary Least Squares (OLS) regression in STATA to quantify the relationship between variables. The dependent variable is the number of offenses of a given year, and the most significant independent variable is the literacy rate. We utilized per capita income as the control variable to improve the accuracy of the model and to reduce omitted variable bias.

##### **4.1 Data Sources:**

The study utilizes secondary data from credible and publicly available government sources to ensure reliability and consistency across all states:

- **Crime Rate Data:** Sourced from the *National Crime Records Bureau (NCRB)*, which publishes annual state-wise crime statistics in India. We use the total IPC recorded crimes per 1,00,000 population as the dependent variable, including both violent and non-violent crime categories.
- **Literacy Rate:** While the last complete Census was conducted in 2011, projected literacy data based on interpolated estimates from government surveys and the *DataGov India portal* were used to approximate state-wise literacy rates for 2020 and 2021. Literacy is measured as the percentage of literate individuals in the population aged 7 and above.
- **Per Capita Income:** Data on Net State Domestic Product (NSDP) at constant prices with 2011 as the base year, was obtained from the *Ministry of Statistics and Programme Implementation (MoSPI)*. Per capita income was calculated by dividing NSDP(Net State Domestic Product) for states and Total Income for UTs by the estimated state population for both years.

This combined dataset covers 28 Indian states and 8 Union Territories over a 2-year period, forming a balanced panel.

#### **4.2 Variables:**

- **Dependent Variable:**  
Crime Rate  $((\text{Total IPC Crimes} / \text{Total Population}) \times 1,00,000)$ .
- **Independent Variable:**  
Literacy rate (percentage of literate individuals aged 7 and above).
- **Control/Explanatory Variable:**  
Per Capita Income (Total Income/Total Population).  
Unemployment\*

\* Unemployment has not been considered as an independent variable in this study due to lack of reliable sources for data., although it is a necessary explanatory variable.

#### **4.3 Regression Equation**

The regression model is specified as :-

$$\text{Crime}_{it} = \beta_0 + \beta_1(\text{Literacy Rate}_{it}) + \beta_2(\text{Per Capita Income}_{it}) + \epsilon_{it}$$

i :- denotes the state

t :- represents the year (2020 or 2021)

LiteracyRate<sub>i</sub>/PerCapitaIncome<sub>i</sub> :- represents the vector of control variables

e<sub>i</sub> :- is the population disturbance/error term

β<sub>1</sub> and β<sub>2</sub>:- represent the coefficients

B<sub>0</sub> represents the intercept

This model is estimated for each year separately, providing four distinct regression outputs.

The sign and magnitude of the coefficients β<sub>1</sub> and β<sub>2</sub> will indicate whether literacy and income have a positive or negative association with crime rates across Indian states.

We ran two model specifications to test the robustness of results, including:

1. **Bivariate regressions** (crime on literacy, crime on income individually),
2. **Multivariate regressions** (including both literacy and income),

#### 4.4 Research Design

Type of Research: Quantitative, Explanatory

Model Used: OLS Regression

Time Period: 2 separate years: 2020 and 2021

Geographic Scope: Indian states chosen based on availability of data and socio-economic diversity

Each year is analyzed as an independent cross-section, and results are compared across time to observe any consistent patterns or shifts in the relationship.

#### 4.5 Assumptions of OLS Model

To ensure the reliability of OLS estimates, we tested for the standard classical linear regression assumptions:

- **Linearity:** Scatter plots and Linear Fit Plots were used to inspect potential non-linear patterns. We also *log transformed* our variables to improve goodness of fit.
- **No multicollinearity:** Variance Inflation Factor (VIF) values for the independent variables were calculated. As expected, some correlation exists between literacy and income, but it remained within acceptable levels ( $VIF < 5$ ).

#### Multicollinearity Test

**. estat vif**

Variable	VIF	1/VIF
logLR2020	<b>1.25</b>	<b>0.800050</b>
logPCI2020	<b>1.25</b>	<b>0.800050</b>
Mean VIF	<b>1.25</b>	

(2020)

**. estat vif**

Variable	VIF	1/VIF
logLR2021	<b>1.24</b>	<b>0.803439</b>
logPCI2021	<b>1.24</b>	<b>0.803439</b>
Mean VIF	<b>1.24</b>	

(2021)

- **Homoscedasticity for constant variance in residuals**

Homoscedasticity Test: The Breusch-Pagan Test was used in STATA

**. estat hettest**

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity  
 Ho: Constant variance  
 Variables: fitted values of logCR2020

chi2(1) = **3.24**  
 Prob > chi2 = **0.0720**

(2020)

```
. estat hettest
```

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of logCR2021

```
chi2(1)          =      2.19
```

```
Prob > chi2     =    0.1392
```

(2021)

- **Normally distributed errors**

- **Independence:** Since each data point corresponds to a distinct state and year, independence of observations was assumed.

OLS regression is appropriate here because the aim is to estimate the average effect of literacy on crime under these assumptions. Results are interpreted as correlations, not causation.

## 4.6 Data Tables

	A	B	C	D	E	F
1	S. No.	State/Union Territory	Literacy Rate % (2020)	Literacy Rate % (2021)	Crime Rate 2020	Crime Rate 2021
2	1	Andaman and Nicobar Islands	86.27	86.6	637.1	332.2
3	2	Andhra Pradesh	66.4	70.1	452.7	420.4
4	3	Arunachal Pradesh	66.95	69.2	164.5	197.7
5	4	Assam	85.9	72.2	349.5	379
6	5	Bihar	70.9	71.8	211.3	228
7	6	Chhattisgarh	77.3	70.3	352.9	373.7
8	7	Chandigarh	86.43	86	271.6	247.1
9	8	Dadra & Nagar Haveli and Diu	82.36	81.6	51.3	52.3
10	9	Delhi	88.7	86.2	1309.6	1479.9
11	10	Goa	87.4	88.7	281.1	191.6
12	11	Gujarat	82.4	78	1011.4	1044.2
13	12	Haryana	80.4	75.6	658.6	697.3
14	13	Himachal Pradesh	86.6	82.8	280.2	254.3
15	14	Jammu and Kashmir	77.3	79.4	216.7	235.7
16	15	Jharkhand	74.3	71.3	166.8	157.3
17	16	Karnataka	77.2	75.4	225.7	244.4
18	17	Kerala	96.2	94	1568.4	1477.2
19	18	Ladakh	80.2	85.9	136.1	187.6
20	19	Lakshadweep	92.28	91.8	216.2	188.2
21	20	Madhya Pradesh	73.7	70.5	511.1	560.8
22	21	Maharashtra	84.8	82.3	435.8	433.5
23	22	Manipur	79.85	79.2	95	101
24	23	Meghalaya	75.48	74.4	114.7	104
25	24	Mizoram	91.58	91.3	189.6	262.2
26	25	Nagaland	80.11	79.6	69.4	67.2
27	26	Odisha	77.3	72.9	295.2	339.4
28	27	Puducherry	86.55	85.8	512.6	290.1
29	28	Punjab	83.7	75.8	274.6	242
30	29	Rajasthan	69.7	71.3	331.2	357.6
31	30	Sikkim	82.2	81.4	100.4	107.4
32	31	Tamil Nadu	82.9	80.1	1808.8	989.5
33	32	Telangana	72.8	78.7	393	420.5
34	33	Tripura	87.75	87.2	115.1	117.3
35	34	Uttarakhand	87.6	78.8	506.8	262.4
36	35	Uttar Pradesh	73	77.2	287.4	304.9
37	36	West Bengal	80.5	76.3	186.6	185
38						

(Table 1.1, raw data)

	A	B	G	H
1	S. No.	State/Union Territory	Per Capita Income (2020, in rupees)	Per Capita Income (2021, in rupees)
2	1	Andaman and Nicobar Is	2,05,568	2,29,570
3	2	Andhra Pradesh	1,68,063	1,97,214
4	3	Arunachal Pradesh	1,81,537	1,90,851
5	4	Assam	86,947	1,03,371
6	5	Bihar	42,083	53,478
7	6	Chhattisgarh	1,06,117	1,22,934
8	7	Chandigarh	2,90,454	3,37,406
9	8	Dadra & Nagar Haveli an	1,66,341	1,32,451
10	9	Delhi	3,22,311	3,76,217
11	10	Goa	4,23,047	4,59,094
12	11	Gujarat	2,07,324	2,41,584
13	12	Haryana	2,24,587	2,64,729
14	13	Himachal Pradesh	1,73,152	1,95,795
15	14	Jammu and Kashmir	1,01,645	1,12,898
16	15	Jharkhand	69,963	88,550
17	16	Karnataka	2,21,781	2,66,866
18	17	Kerala	1,94,432	2,30,280
19	18	Ladakh	99,871	1,04,321
20	19	Lakshadweep	1,87,651	1,62,341
21	20	Madhya Pradesh	1,02,007	1,16,689
22	21	Maharashtra	1,89,889	1,83,704
23	22	Manipur	75,784	98,826
24	23	Meghalaya	90,751	1,07,971
25	24	Mizoram	1,73,521	1,90,965
26	25	Nagaland	1,19,781	1,27,225
27	26	Odisha	1,03,203	1,34,091
28	27	Puducherry	2,08,862	2,31,557
29	28	Punjab	1,50,620	1,70,276
30	29	Rajasthan	1,14,850	1,34,060
31	30	Sikkim	4,15,045	4,66,518
32	31	Tamil Nadu	2,09,628	2,42,339
33	32	Telangana	2,25,734	2,69,000
34	33	Tripura	1,18,401	1,37,032
35	34	Uttarakhand	1,90,558	1,74,526
36	35	Uttar Pradesh	71,307	81,398
37	36	West Bengal	1,41,373	1,21,904
38				

(Table 1.2, raw data)

#### 4.7 Limitations of Methodology

Though OLS provides a clear and interpretable estimation framework, this study acknowledges some limitations:

1. **Endogeneity:** There may be reverse causality, regions with lower crime may attract more investment in education and generate higher income. Instrumental variables were not employed due to data constraints, but this limitation is noted.
2. **Omitted Variable Bias:** Variables like unemployment, urbanization, quality of policing, and population density were not included due to either unavailability or inconsistencies in measurement across states.
3. **Measurement Error:** Crime data in India may suffer from underreporting, especially in less literate or rural areas. This introduces potential bias in interpreting the results.

Despite these constraints, the OLS model still offers meaningful insights into the correlation patterns between education, income, and crime across states in India. The regression results, discussed in the following section, help identify whether these relationships conform to



theoretical expectations or if Indian states demonstrate unique dynamics that challenge global assumptions.

## V) Results and Interpretations

### 5.1 Outcome

#### A) Regression 1: Crime Rate 2020 and Literacy Rate 2020

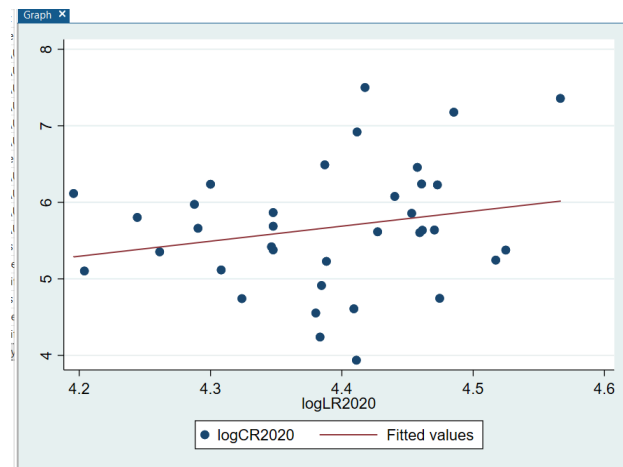
<b>. reg logCR2020 logLR2020</b>						
Source	SS	df	MS	Number of obs	=	<b>36</b>
Model	<b>1.08817368</b>	<b>1</b>	<b>1.08817368</b>	F(1, 34)	=	<b>1.62</b>
Residual	<b>22.791157</b>	<b>34</b>	<b>.670328146</b>	Prob > F	=	<b>0.2113</b>
Total	<b>23.8793306</b>	<b>35</b>	<b>.68226659</b>	R-squared	=	<b>0.0456</b>
				Adj R-squared	=	<b>0.0175</b>
				Root MSE	=	<b>.81874</b>
logCR2020	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
logLR2020	<b>1.965666</b>	<b>1.542781</b>	<b>1.27</b>	<b>0.211</b>	<b>-1.169643</b>	<b>5.100975</b>
_cons	<b>-2.960219</b>	<b>6.774554</b>	<b>-0.44</b>	<b>0.665</b>	<b>-16.72777</b>	<b>10.80733</b>

(Table 1.3)

- Coefficient of 1.965666: A 1% increase in literacy rate is associated with approximately a 1.97% increase in the crime rate, holding other factors constant.
- p-value (0.211) is high as ( $p=0.211 > 0.05$ ), meaning no statistically significant relationship between literacy and crime in 2020.
- R-squared (4.56%): The R-squared value is 0.0456, suggesting that only about 4.56% of the variation in the crime rate is explained by literacy rates. The Adjusted R-squared is even lower at 1.75%, highlighting that the model has very limited explanatory power.

Therefore, Literacy is weakly positively associated with the crime rate in 2020, but the relationship is not strongly statistically significant.

*a) Regression 1: Linear Fit Plot and Scatter Plot*



(Table 1.4)

Interpretation of graph: Scatterplot shows no clear trend. Fitted line has a very mild positive slope. Therefore, a weak positive link between literacy and crime in 2020 which is not strongly conclusive.

*B) Regression 2: Crime Rate 2021 and Literacy Rate 2021*

<b>. reg logCR2021 logLR2021</b>						
Source	SS	df	MS	Number of obs	=	<b>36</b>
Model	<b>.017031487</b>	<b>1</b>	<b>.017031487</b>	F(1, 34)	=	<b>0.03</b>
Residual	<b>20.7690801</b>	<b>34</b>	<b>.610855298</b>	Prob > F	=	<b>0.8684</b>
				R-squared	=	<b>0.0008</b>
				Adj R-squared	=	<b>-0.0286</b>
Total	<b>20.7861116</b>	<b>35</b>	<b>.593888904</b>	Root MSE	=	<b>.78157</b>

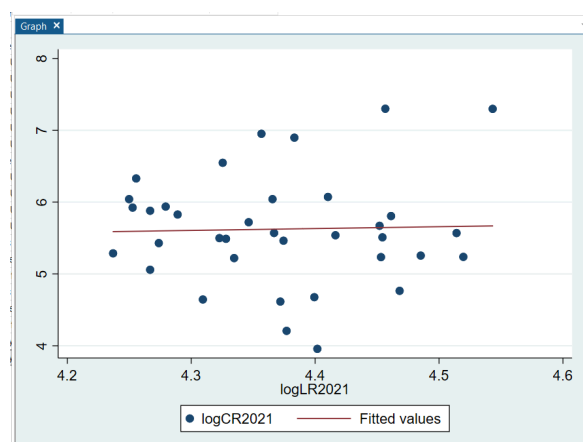
logCR2021	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
logLR2021	<b>.2587446</b>	<b>1.549581</b>	<b>0.17</b>	<b>0.868</b>	<b>-2.890382</b>	<b>3.407871</b>
_cons	<b>4.492962</b>	<b>6.775093</b>	<b>0.66</b>	<b>0.512</b>	<b>-9.275683</b>	<b>18.26161</b>

(Table 1.5)

- Coefficient of .2587446: A 1% increase in literacy rate is associated with approximately a 0.25% increase in the crime rate, holding other factors constant.
- p-value (0.868) is high as ( $p=0.868 > 0.05$ ), meaning no statistically significant relationship between literacy and crime in 2021 as well.
- R-squared (0.08%): The R-squared value is 0.0008, suggesting that only about 0.08% of the variation in the crime rate is explained by literacy rates. The Adjusted R-squared is negative which indicates insignificance.

Here, the correlation is much lower than the significance found in 2020. Possibly due to COVID-19 restrictions.

*a) Regression 2: Linear Fit Plot and Scatter Plot*



(Table 1.6)

Interpretation of graph: Scatterplot has more variability with no trend. Fitted line is almost flat. Matches regression of low  $R^2$  and no significant relationship. Therefore, no significant relationship between literacy and crime rate in 2021.

*C) Regression 3: Crime Rate 2020 and Per Capita Income 2020*

<b>. reg logCR2020 logPCI2020</b>						
Source	SS	df	MS	Number of obs	=	<b>36</b>
Model	<b>2.46763188</b>	<b>1</b>	<b>2.46763188</b>	F(1, 34)	=	<b>3.92</b>
Residual	<b>21.4116988</b>	<b>34</b>	<b>.629755846</b>	Prob > F	=	<b>0.0559</b>
				R-squared	=	<b>0.1033</b>
				Adj R-squared	=	<b>0.0770</b>
Total	<b>23.8793306</b>	<b>35</b>	<b>.68226659</b>	Root MSE	=	<b>.79357</b>

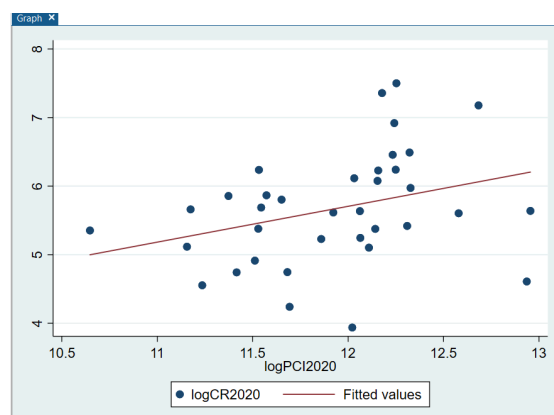
logCR2020	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
logPCI2020	<b>.5224514</b>	<b>.2639319</b>	<b>1.98</b>	<b>0.056</b>	<b>-.0139227</b>	<b>1.058826</b>
_cons	<b>-.5637955</b>	<b>3.151724</b>	<b>-0.18</b>	<b>0.859</b>	<b>-6.96887</b>	<b>5.841279</b>

(Table 1.7)

- Coefficient of .5224514: A 1% increase in the per capita income is associated with approximately a 0.5% increase in the crime rate, holding other factors constant.
- p-value (0.056) is just above the significance level at 5% LOS, hence the data is marginally significant, although pure significance will come at 10% LOS.
- R-squared (10.33%): The R-squared value is 0.1033, suggesting that roughly 10% of the variation in the crime rate is explained by per capita income.

Here, the correlation is much stronger than the significance found in literacy rates in 2020 and 2021. Therefore, per capita income may be a stronger explanatory variable for our model.

*a) Regression 3: Linear Fit Plot and Scatter Plot*



(Table 1.8)

Interpretation of graph: The scatterplot shows a slight upward slope. The fitted line is slightly positive. Matches regression coefficient and p value both. Therefore, slight positive relation between income and crime in 2020 and marginally the most significant in our model till now.

*D) Regression 4: Crime Rate 2021 and Per Capita Income 2021*

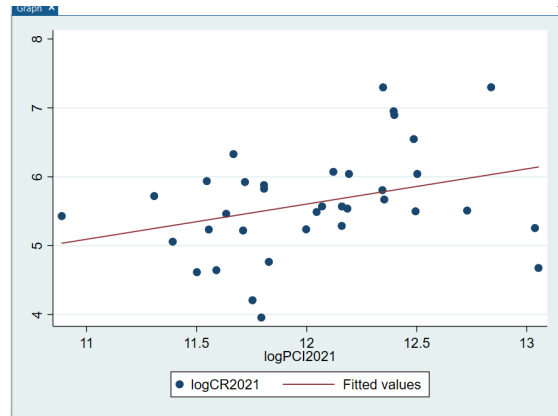
<b>. reg logCR2021 logPCI2021</b>						
Source	SS	df	MS	Number of obs	=	<b>36</b>
Model	<b>2.20508486</b>	<b>1</b>	<b>2.20508486</b>	F(1, 34)	=	<b>4.03</b>
Residual	<b>18.5810268</b>	<b>34</b>	<b>.546500787</b>	Prob > F	=	<b>0.0526</b>
Total	<b>20.7861116</b>	<b>35</b>	<b>.593888904</b>	R-squared	=	<b>0.1061</b>
				Adj R-squared	=	<b>0.0798</b>
				Root MSE	=	<b>.73926</b>
logCR2021	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
logPCI2021	<b>.5110519</b>	<b>.2544179</b>	<b>2.01</b>	<b>0.053</b>	<b>-.0059876</b>	<b>1.028091</b>
_cons	<b>-.5285144</b>	<b>3.065414</b>	<b>-0.17</b>	<b>0.864</b>	<b>-6.758186</b>	<b>5.701157</b>

(Table 1.9)

- Coefficient of .5110519: A 1% increase in the per capita income is associated with approximately a 0.5% increase in the crime rate, holding other factors constant.
- p-value (0.053) is just above the significance level at  $p > 0.05$ , hence the data is marginally significant, although pure significance will come at 10% LOS.
- R-squared (10.61%): The R-squared value is 0.1061, suggesting that roughly 10% of the variation in the crime rate is explained by per capita income.

Here, the correlation is mildly stronger than the significance found in literacy rates in 2020 and 2021 and PCI 2020. Therefore, per capita income may be a stronger explanatory variable in the long run.

*a) Regression 4: Linear Fit Plot and Scatter Plot*



(Table 2.1)

Interpretation of graph: The scatterplot shows dispersion. The fitted line is slightly positive. Matches regression: regression coefficient was positive mild statistical significance at p value close to 5% at 95% confidence.

#### E) Multivariate Regression 2020

```
. regress logCR2020 logLR2020 logPCI2020
```

Source	SS	df	MS	Number of obs	=	36
Model	<b>2.61274279</b>	<b>2</b>	<b>1.3063714</b>	F(2, 33)	=	<b>2.03</b>
Residual	<b>21.2665878</b>	<b>33</b>	<b>.644442056</b>	Prob > F	=	<b>0.1478</b>
				R-squared	=	<b>0.1094</b>
				Adj R-squared	=	<b>0.0554</b>
Total	<b>23.8793306</b>	<b>35</b>	<b>.68226659</b>	Root MSE	=	<b>.80277</b>

logCR2020	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
logLR2020	<b>.8025133</b>	<b>1.691197</b>	<b>0.47</b>	<b>0.638</b>	<b>-2.638252</b>	<b>4.243279</b>
logPCI2020	<b>.4591143</b>	<b>.2984965</b>	<b>1.54</b>	<b>0.134</b>	<b>-.1481813</b>	<b>1.06641</b>
_cons	<b>-3.331351</b>	<b>6.646841</b>	<b>-0.50</b>	<b>0.620</b>	<b>-16.85445</b>	<b>10.19175</b>

(Table 2.2)

#### F) Multivariate Regression 2021

```
. regress logCR2021 logLR2021 logPCI2021
```

Source	SS	df	MS	Number of obs	=	36
Model	<b>2.55187979</b>	<b>2</b>	<b>1.27593989</b>	F(2, 33)	=	<b>2.31</b>
Residual	<b>18.2342318</b>	<b>33</b>	<b>.55255248</b>	Prob > F	=	<b>0.1152</b>
				R-squared	=	<b>0.1228</b>
				Adj R-squared	=	<b>0.0696</b>
Total	<b>20.7861116</b>	<b>35</b>	<b>.593888904</b>	Root MSE	=	<b>.74334</b>

logCR2021	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
logLR2021	<b>-1.302582</b>	<b>1.644203</b>	<b>-0.79</b>	<b>0.434</b>	<b>-4.647738</b>	<b>2.042574</b>
logPCI2021	<b>.6112965</b>	<b>.2854058</b>	<b>2.14</b>	<b>0.040</b>	<b>.0306342</b>	<b>1.191959</b>
_cons	<b>3.958749</b>	<b>6.448488</b>	<b>0.61</b>	<b>0.543</b>	<b>-9.160799</b>	<b>17.0783</b>

(Table 2.3)

## 5.2 Macroeconomic Interpretation

In a macroeconomic framework, the relationship between literacy rates, per capita income, and crime rates reflects broader structural dynamics across Indian states. Classical economic theory suggests that higher literacy and income levels should correlate with reduced crime due to better employment prospects, higher opportunity costs of criminal activity, and improved social capital. However, the findings of this study indicate a more nuanced reality.

The regression analyses for 2020 and 2021 reveal that neither literacy nor per capita income significantly reduces crime rates. In fact, both literacy and income show weak positive associations with crime, although the results are largely statistically insignificant. This paradoxical outcome suggests that rising income and education levels may, at least in the short term, generate greater reporting of crimes, heightened awareness, and possibly new types of economic crimes, rather than uniformly reducing crime across states.

From a macroeconomic lens, this phenomenon can be understood through the lens of structural transition. As Indian states urbanize and modernize, the traditional informal controls over behavior weaken, while the legal-institutional framework struggles to keep pace. Higher literacy and income enable individuals to recognize and report crimes more readily, inflating crime statistics during periods of socio-economic transition. Furthermore, income inequality, rather than absolute income, might be a more decisive factor, that is, a better explanatory variable in influencing crime.

Another macro-level explanation could involve regional disparities. Richer or more literate states may also attract greater internal migration, leading to social friction, housing pressures, and strains on law enforcement, thereby raising crime rates independently of local education or wealth improvements.

The relatively low  $R^2$  values in the regressions indicate that much of the variance in crime rates remains unexplained by literacy and per capita income alone, reinforcing the idea that macroeconomic complexities, including governance quality, migration trends, and

urbanization levels, are pivotal factors. Thus, a purely literacy- or income-focused policy will likely be insufficient without addressing these broader systemic challenges.

In conclusion, policymakers must complement economic growth strategies with investments in policing, legal reforms, urban planning, and social cohesion programs to effectively curb crime during India's ongoing economic transition.

### **5.3 Microeconomic Interpretation**

At the microeconomic level, the relationship between literacy, income, and crime operates through individual incentives and constraints. Rational choice theory posits that individuals weigh the costs and benefits of criminal activity; thus, higher education and income should theoretically increase the opportunity cost of engaging in crime. However, the empirical evidence from this study complicates this expectation.

The positive coefficients of literacy and per capita income on crime rates suggest that individual decisions are influenced by factors beyond simple economic calculation. A higher literacy rate may enable individuals to participate in white-collar crimes, cybercrimes, or elaborate fraudulent activities that were less prevalent in low-literacy environments. Similarly, a higher income may enhance access to resources, but if accompanied by rising consumer aspirations and visible wealth disparities, it may also increase the motivation for acquisitive crimes among those left behind.

On a micro scale, social capital and community engagement play critical roles in crime deterrence. Merely raising education levels without corresponding investments in moral education, community networks, and civic engagement can result in educated but socially disconnected individuals, who may be more prone to criminal or antisocial behaviors.

Another dimension is the labor market mismatch. Higher literacy rates do not automatically translate into suitable employment opportunities. Underemployment or mismatch between education and job availability can lead to frustration and deviant behavior. This could partially explain why increased literacy does not significantly suppress crime rates.

Additionally, microeconomic behavior around crime reporting must be considered. As individuals become more literate and economically secure, they are more likely to recognize and report crimes, inflating reported crime rates relative to previous underreporting. In states with traditionally low crime reporting due to illiteracy, stigma, or distrust of law enforcement, improvements in literacy may temporarily result in statistical increases in crime, without necessarily indicating a real increase in criminal behavior.

Finally, income mobility dynamics are critical. If income gains are volatile or precarious, individuals may experience insecurity, further weakening the deterrent effect that stable income would typically have against crime.

Thus, from a microeconomic standpoint, individual incentives, employment structures, societal norms, and the psychological impacts of relative deprivation all interact in complex



ways to mediate the impact of literacy and income on criminal behavior. Simple linear assumptions are insufficient to capture the real-world complexities revealed by this study.

#### **5.4 Research Gaps**

Although more academic attention is focused on the socio-economic determinants of Indian crime, not much has been quantitatively investigated about the direct role of literacy through statistical methods such as Ordinary Least Squares (OLS) previously. The majority of previous research combines education with other economic variables or uses only descriptive analysis. This study bridges this gap by employing an OLS model to investigate the direct link between literacy and crime rates in a three-year panel of Indian states and provides new evidence-based results.

## VI) Conclusion

The hypothesis was that higher literacy and income levels would correlate with lower crime rates. However, the findings revealed a far more complex and nuanced reality. The regression analyses indicated that both literacy and per capita income had weak, and sometimes positive, associations with crime rates. Neither factor exhibited strong statistical significance in predicting crime reductions. Graphical analyses further confirmed that the relationships were dispersed, with no strong linear trends. These results suggest that the classical assumptions linking human capital and economic prosperity to lower crime do not straightforwardly apply across the heterogeneous and dynamic landscape of Indian states and that there may be multiple other factors involved which may only be revealed in the longer run. Policy interventions, given these findings, should not emphasize the improvement of literacy but take a broader approach. Enhancing education access continues to be relevant, especially through primary, secondary, and adult literacy initiatives. But parallel efforts also need to strengthen employment opportunities, economic conditions, law enforcement mechanisms, and social awareness. Vocational training and rehabilitation programs for high-risk groups can also be an important step. A multi-sector approach involving education, economy, and social reforms is thus the need of the hour to check crime and build a safer and more equitable society.

### 6.1 Key Takeaways

- **Complexity:** The relationship between socio-economic indicators and crime is far from linear. States with higher literacy and income may experience higher crime rates, possibly due to better crime reporting mechanisms, migration pressures, urbanization, and inequality.
- **Crime Reporting Sensitivity:** Improved literacy and income can lead to more active crime reporting, thus increasing reported crime rates without necessarily reflecting an actual rise in criminal behavior.
- **Relative Deprivation:** Rising incomes can exacerbate feelings of inequality and social dissatisfaction among those who do not experience similar gains, potentially increasing certain types of crimes.
- **Structural Transformation:** Economic and social modernization processes must be accompanied by robust institutional frameworks to manage the accompanying frictions and prevent crime escalation.
- **Low Predictive Power:** Literacy and income explain only a small fraction of the variance in crime rates, implying that other variables play crucial roles.

## 6.2 Suggestive Policies

- State-Level Crime Data Digitization Missions: Mandate each Indian state to digitize all First Information Reports and create an open, real-time crime dashboard accessible to public and law enforcement. Objective of mandate: To improve transparency and detect real crime patterns rather than relying on underreporting.
- Mandatory Civic Education in School: All education boards must add a "Civic Responsibility and Legal Literacy" subject from Classes 8-10. Objective of mandate: Creating law-conscious citizens from teenage years to reduce future crimes.
- Fast-Track Economic Offence Courts: Set up a network of special courts in each state exclusively to handle white-collar crimes, financial fraud, and cybercrime. Objective of mandate: Speedy justice needed to deter sophisticated crimes.
- Skill-Linked Employment Guarantee Program: Launch an Urban Skill Guarantee Program (USGP) modeled after MGNREGA for unemployed graduates in urban India. Objective of mandate: Educated but jobless youth are a crime risk, hence building headstarts for everyone.

## VII) References and Bibliography

- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2), 169–217.
- Hirschi, T. (1969). *Causes of delinquency*. University of California Press.
- Jha, R., & Kumar, A. (2018). Socioeconomic determinants of crime in Indian states: An empirical analysis. *Indian Journal of Economics and Development*, 14(2), 325–334.
- Lochner, L., & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review*, 94(1), 155–189.
- Machin, S., Marie, O., & Vujic, S. (2011). The crime reducing effect of education. *Economic Journal*, 121(552), 463–484.
- Merton, R. K. (1938). Social structure and anomie. *American Sociological Review*, 3(5), 672–682.
- Rao, D. (2019). Education, unemployment, and crime in India: An exploratory analysis. *Journal of Development Policy and Practice*, 4(1), 23–45.
- Sengupta, S., & Das, S. (2015). Inequality and crime in India: A panel data approach. *Asian Journal of Research in Social Sciences and Humanities*, 5(8), 221–236.
- Sharma, K., & Singh, R. (2020). Urbanization, literacy, and crime: An inter-state study in India. *International Journal of Social Science Research*, 8(1), 101–117.
- Dornbusch, Rudiger; Fischer, Stanley; Startz, Richard (2004). *Macroeconomics*. New York: McGraw-Hill Irwin. pp. 22–23.

-<https://www.ncrb.gov.in/uploads/nationalcrimerecordsbureau/custom/1701607577CrimeinIndia2022Book1.pdf>

-<https://iced.niti.gov.in/>

-<https://mospi.gov.in/literacy-rate-india-state-wise-rgi-nss>

-<https://www.data.gov.in/keywords/Literacy>

-<https://www.ijfmr.com/papers/2023/1/1409.pdf>

-<https://www.indiabudget.gov.in/budget2022-23/economicsurvey/doc/stat/tab85.pdf>