The Road to Crime

Analyzing the Impact of Rural Road Connectivity on Crime in India

Aarushi Sharma

May 9, 2022

ABSTRACT

In 2001, 28% of India's population lacked access to paved roads (World Bank, 2021). At the same time, the country has been ranked the most dangerous place for women, in terms of public safety (Thomson Reuters Foundation, June 25, 2018). To improve connectivity in India, the government of India launched a $40 billion rural road construction program named the PMGSY, with an aim to connect 115,000 Indian villages. I use administrative data from the Population Census, the PMGSY and official crime records to answer the research question: *Does constructing rural roads impact crime in India, especially crime against women?* Using a Difference-in-Differences identification strategy, I calculate the effect of PMGSY roads on crime in areas that were newly connected under the program. I observe no consistent relationship. I further implement a Multiple Linear Regression to include the effect of non-PMGSY roads on crime. I find that a shift in a district's connectivity from the 25th to the 75th percentile, is associated with an increase of 13.68% in crime rate.

The Road to Crime

Analyzing the Impact of Rural Road Connectivity on Crime in India

By Aarushi Sharma
under the direction of Professor Margaret Lay

A Thesis
presented to the faculty of Mount Holyoke College
in partial fulfillment of the requirements for the degree of
Bachelor of Arts with Honors

Department of Economics
Mount Holyoke College
South Hadley, Massachusetts 01075, USA

May 2022

In loving memory of my Baba, Kedar Nath Sharma, who loved and supported me unconditionally. He would have been so happy to see me graduate college.

ACKNOWLEDGEMENTS

This thesis is one of my most challenging and exciting academic experiences. Looking back to when I was apprehensive of my learning abilities for getting rejected by every engineering university, I can now connect the dots because they led me here — research in development economics! Thank you to the community I built at Mount Holyoke College that appreciated and motivated my curiosity and revived the academic confidence I never thought I will achieve.

My biggest gratitude goes to my advisor, and my mentor Professor Margaret (Gretchen) Lay. I met Gretchen in the semester when everything went online. Irrespective of our circumstances, she recognized and appreciated my interest in thinking about social sciences quantitatively. She supported and trusted my work. I cannot thank her enough for always keeping me on my toes with her constructive criticism and helpful feedback. Beyond academics, Gretchen is my role model, as I see her balance roles being a professor, a scholar, a mother and most importantly a friend. Thank you for being calm and patient when nothing seemed to work.

The Economics Department at Mount Holyoke supported my love for Economics. I cannot thank Jonah Yuen enough, who not only taught me all the essential econometric skills, but also engaged in thinking about challenges of my research, even when he had no obligation to do so. I am grateful to Professor Johannes Norling for helping me with all the GIS for my analysis and also for organizing my thesis process. A huge thanks to Prof Sarah Adelman who in Intro to Econometrics class, taught me that hypothesis testing can be explained beautifully using cartoons. I absolutely enjoy and will miss Prof Lucas Wilson's laughter and his advice during the crazy course registrations. I also thank Prof Ted Gilliland for his valuable feedback on my empirical technique when its problems seemed unsolvable. Thanks Dominique for the countless cups of coffee and green tea, and making the study spaces in Skinner so enjoyable!

Being able to understand and apply mathematics has been a valuable skill. The credit goes to my high school math teacher Mr. Yash Gupta who taught me how to focus on simplicity and the fundamentals when thinking about math. I am extremely grateful for the mathematical guidance provided by Professor Lidia Mrad. From her brilliant pedagogy, I learned the skill of writing mathematical proofs with precision. Taking her class introduced me to so many women in math, who create an exemplary learning community.

My heartfelt thanks to all my friends in the Unites States who never let me miss home. Tanishk, thanks for all the bubble tea, the laughter and the support you give me! Hannah, thanks for being my best buddy and for the endless banter that always makes my day. Max, thank you for your amazing friendship, for believing in me more than I do, and for all the valuable explanations of hard economic concepts. A huge thanks to Deepika Baskar, Sohini Bhatia, Rehat Thussu, Ameya Jain, and Siddhanth Satish for making my senior year so memorable — you all are delightful! Thanks to Savita Diggs and Maddison Erbabian for being wonderfully supportive alumnae. I can never forget the warmth and the support I received in India at Lady Shri Ram College for Women and, back home in Bhopal. Thanks to Anuna, Rashmeet and Pooja for being my closest friends. You made transferring to the US so painful. Thanks to my friends at DPS — I love you all!

Above all, I am grateful to my family — Mumma, Papa and Bhaiya — for making my life so meaningful and for providing an everlasting support system. Thanks Mumma and Papa for trusting my academic endeavors and supporting all my decisions, even if it involved sending me thousands of miles away from you. Thank you Bhaiya, from you I understood the value of learning and working hard. Thanks for letting me make mistakes and follow my dreams!

Contents

List of Tables

List of Figures

1 INTRODUCTION

The World Bank, in its more than 60 years of functioning, has considered infrastructure as one of its top priorities. Between 1970 and 2005, infrastructure-related lending to developing countries varied between one-third to two-thirds of the Bank's total lending, amounting to eight billion US dollars in 2005 (Straub, the World Bank, 2008). At the same time, finding ways to improve public safety, especially safety for women, has been a policy challenge in developing countries. Making women feel safe in the streets and neighborhoods of their country is essential to meet the fundamental Sustainable Development Goal (SDG) of gender equality and women empowerment. Does the massive infrastructural investment have a positive externality of improving public safety?

This paper studies the context of a country that has been ranked the world's most dangerous country for women: India (Thomson Reuters Foundation, Jun 25, 2018). According to the Population Census of India conducted in 2001, 300,000 of the country's 600,000 villages lacked access to paved roads, implying that over one billion people had no access to urban markets. In the wake of the poor state of rural roads, the government of India launched a country-wide, $40 billion, rural road construction project named Pradhan Mantri Gram Sadak Yojana (Prime Minister's Rural Road Construction Project, or PMGSY). This paper aims to empirically analyze the PMGSY program to answer the question: *Does constructing rural roads impact crime in India, specifically crime against women?*

There could be many channels via which rural roads impact crime in developing countries. [1] Providing more roads impacts the opportunity cost of engaging in crime. On one hand, it can increase the opportunity cost. Time spent in otherwise planning and conducting illegal criminal activities, is spent in

[1] The Directed Acyclic Graph in the Appendix section

education or earning money legally after road construction, thus reducing crime. Evidence-based research conducted in the context of India and other developing countries shows causal relationships between rural road connectivity and structural mobility out of low-income agricultural jobs to high-income jobs in the tertiary sector (Asher, Novosad, 2020; Banerjee, Duflo, 2012). New paved rural roads under the PMGSY program increase middle school enrollment, make children stay longer in schools and make them perform better in standardized tests (Adukia, Asher et al., 2017). On the other hand, better connectivity might decrease the opportunity cost of crime like drug trafficking, and transport of other illegal goods because of the reduced cost of travel. It can increase the ease of commute of gangs from one village to the other, thus increasing crime (Zhang, Ingale et al., 2016).

Another important channel via which roads influence crime is their impact on inequality. They can decrease the urban-rural inequality by connecting villages to urban areas. Equality of opportunity may lead to a decrease in crime. At the same time, the benefits of roads unequally distributed within a district (consisting of multiple villages) can increase inequality. In this scenario, roads can promote crime because disadvantaged groups in unequal societies rely on illegal means to distribute resources more equally (Bharadwaj, 2014). Beyond an influence on the occurrence of crime, roads can affect the reporting of criminal charges. Road construction might lead to building of more police stations. These institutions, with the increased number of educated individuals might lead to more reporting, and thus an increase in the rate of crime.

One existing paper directly examines the relationship between road infrastructure and crime in India. Using instrumental variable estimation strategy and data from two waves of the Indian Human Development Survey, Jain and Biswas (2021) find that building roads in rural parts of India reduces crime.

There are no other studies conducted in the area to my knowledge. My paper uses administrative data on crime and roads to answer the same research problem. I scrape the crime data from official annual crime reports, and thus create open source data for other crime-related research. The paper contributes to the understudied literature of the impact of infrastructural investments on improving public safety.

A challenge to estimating the impact of road-placement programs is their endogenous nature. Allocation of roads is affected by social and political factors. I first use a multiple linear regression with appropriate controls for the estimation. This method does not measure the causal impact of connectivity on crime. I implement a Difference-in-Differences (DD) design that removes the endogeneity to find a causal relationship. This becomes possible by exploiting the timing of the PMGSY policy implementation and its coincidence with the available data. I further extend the same strategy to estimate the effect of treatment of roads on the treated districts using data specific to roads constructed under the PMGSY. A major challenge in this estimation technique is the lack of data before the launch of the program. Thus, I cannot check for parallel trends and establish the robustness of the estimates. Finally, to overcome this limitation, I motivate the potential use of an Instrumental Variable to estimate the effect of treatment on the treated. The policy allocation of roads according to a population-based rule [2] provided randomness and thus an opportunity to construct an instrumental variable. All estimation methods are implemented at the district-level.

The linear regression estimates a positive relationship between crime and connectivity provided by both PMGSY and non-PMGSY roads. A district that increased connectivity from the 25th to the 75th percentile of the distribution

[2] The rule instructed that villages with more than 1000 people should get roads before villages with more than 500 people.

would experience an increase in crime of 13.68%. The difference-in-differences measures no clear relationship between crime and connectivity. Results from difference-in-differences that estimate the effect of the PMGSY roads also do not indicate any consistent and statistically significant relationship between crime and roads.

The structure of the paper is as follows. Chapter 2 provides a report of patterns of crime in India. It shows the geographic and temporal trends in criminal offences, and understands them using legal, literary, cultural and data-based context. Chapter 3 explores the history of road construction, and then shifts the focus to the PMGSY program. It also aims to understand temporal and geographic variations in road construction in India until the year 2015. Chapter 3 describes the identification strategy, shedding light on empirical challenges, the specification, and its results. Chapter 4 concludes.

2 CRIME IN INDIA

This chapter provides a report of patterns of crime in India. I explore trends and characteristics of criminal offences in India using legal, literary, cultural and data-based context. My goal is to explore temporal and geographic variations in crime which will be helpful to compare with the construction of roads under the PMGSY program.

2.1 Understanding Crime in India

The Criminal Procedure Code of India divides crime into two categories: Cognizable and Non-cognizable. A 'cognizable offence' is one in which the police may arrest a person without a warrant. They are authorized to start investigation into a cognizable case on their own and do not require orders from the court. A 'non-cognizable offence' is an offence in which police cannot register a First Information Report (FIR) [3], investigate or arrest without the prior permission from the court.

Cognizable Offences fall under either the Indian Penal Code (IPC) or the Special and Local Laws (SLL). SLL includes "elite" crimes such as possessing Arms, Gambling, Indian Passport, and Copyright violations. [4] Stakeholder involved in elites crime are less likely to benefit from a program that improves road connectivity. This thesis includes only IPC cognizable offences from years 2001 to 2013. I exclude SLL crimes because the goal of my research is to analyze how rural road development is linked to the occurrence of crime because of their impact on people's economic situation. Most IPC crimes are non-elite.

[3] An FIR is a written document prepared by the police when they receive information of a cognizable offence at the district level. Police take up investigation only after filing an FIR.

[4] See page 27 of Crimes in India, 2002 for a complete list.

2.2 Data and Issues

The National Crime Records Bureau (NCRB) of India publishes annual reports called 'Crime in India' (CI). The data is based on complaints or FIRs (Cognizable Offences) filed with the police and not convicted cases. These are collected by the State Crime Records Bureaux (SCRBx) from the District Crime Records Bureaux (DCRBx) and then sent to NCRB at the end of every calendar year. The first edition of 'Crime in India' was published in the year 1953. These reports contain statistical information on cognizable offences. Years prior to 2001 have data on reported crime at the state level. Years from 2001 and later have data on crimes under various categories (further elaborated in the section "Types of Crime") at the district level as reported in police stations during the reference year.

I use these data to analyze geographic crime patterns in India because: (i) the data is official (administrative) and was published in the form of PDF tables which are easily scrapable using the Python programming language [5], (ii) the district level granularity of the data is helpful (the more granular the better) since roads were constructed to connect villages.

However, A major limitation of the NCRB data is undercounting [4]. The published data only takes into account the 'principal offence' in every FIR, that is, the charge that attracts the maximum penalty (Rukminin S, the Hindu, September 13, 2013). This was confirmed by R. Rajasekaran, deputy director of the NCRB in 2013. An example of a major mistake is counting the Nirbhaya rape case of December 16, 2012 as murder and not rape because the 'principal offence' is murder that attracts the maximum penalty of death. As a result rape ending in murder is recorded as murder. The introduction of "Rape-and-Muder " and "Attempt to Rape" categories in later years (2014 onward) was

[5] Details of extraction in the Appendix section

[4] Reportedly, the murder data is fairly accurate (Dreze and Khera, 2000)

an attempt to address this problem (The Hindu, 2013).

Table 1 provides summary statistics of the crime rates at the district level of total crime in India. I calculate the crime rates using district populations from the population census of 2001 and 2011 [6]. The extracted dataset contains district-level data on levels of reported crime covering 35 states and union territories [7]. The number of districts increase as data improves over years. So, there are 669 observations in 2001, increasing to 709 in 2008 and then to 763 in 2013.

2.3 Trends in Total Crime

2.3.1 Temporal

Table 1 shows that the average crime rate increased from 150 to 159 crimes committed per 100,000 people as we go from the year 2001 to 2013. This increase though is not uniform. There is a large variance across years. For example, the crime rate was as low as 139 crimes per 100,000 people in 2007. Moreover, there are large variations within each year between districts. As an example, in 2013, the difference in crime rates at the 75th and the 50th percentile is almost 77 crimes per 100,000 people. The standard deviation in the crime rates that year is 110. This shows that crime is not uniformly distributed across districts — some districts have higher incidence of crime than others over years.

2.3.2 Geographic

Figures 1 and 2 are heatmaps of state-wise variation in rate of total crime for years 2001 and 2011 respectively. The color coding is according to the quartiles of crime rate from 2001. In 2001, crime rate is the highest in the central state of Madhya Pradesh, the western state of Rajasthan, north eastern state of

[6] Details of merging the datasets in the Appendix

[7] In the years under consideration (2000-2013), Telangana did not form from Andhra Pradesh and Jammu Kashmir was a state and not a Union Territory.)

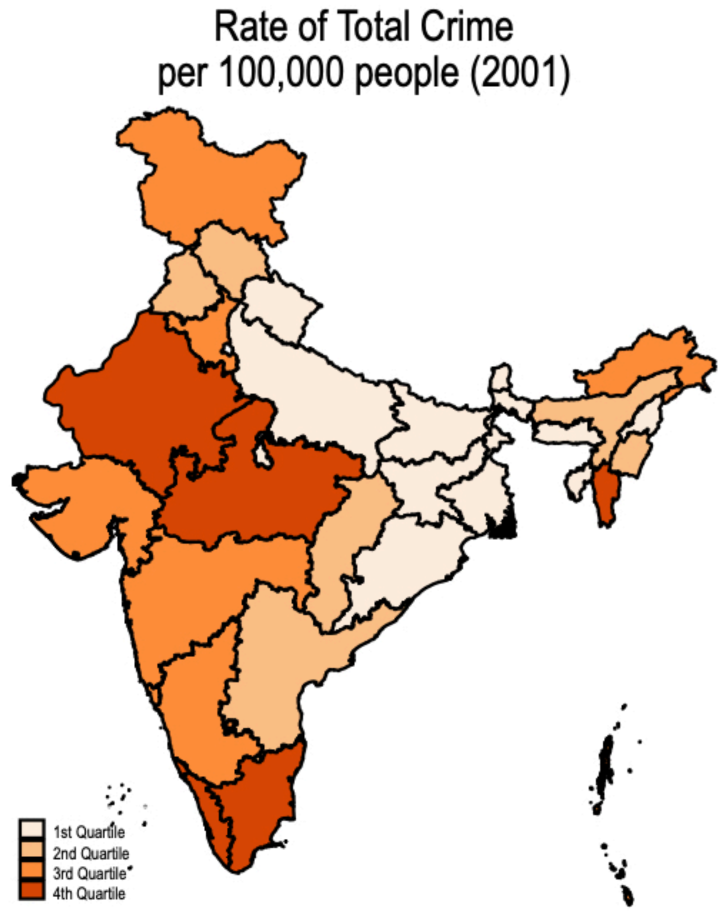

Figure 1: State-level Rate of Total Crime in 2001

Mizoram and southern states of Kerela and Tamil Nadu. In 2011, more states shift to the highest quartile of crime. These states are Goa, and Karnataka in the south, and Haryana in the north. Overall rate of total crime increased everywhere in India from 2001 to 2011 with higher concentration in southern, central and western states.

Figures 3 and 4 show the rate of crime at the district-level in 2001 and

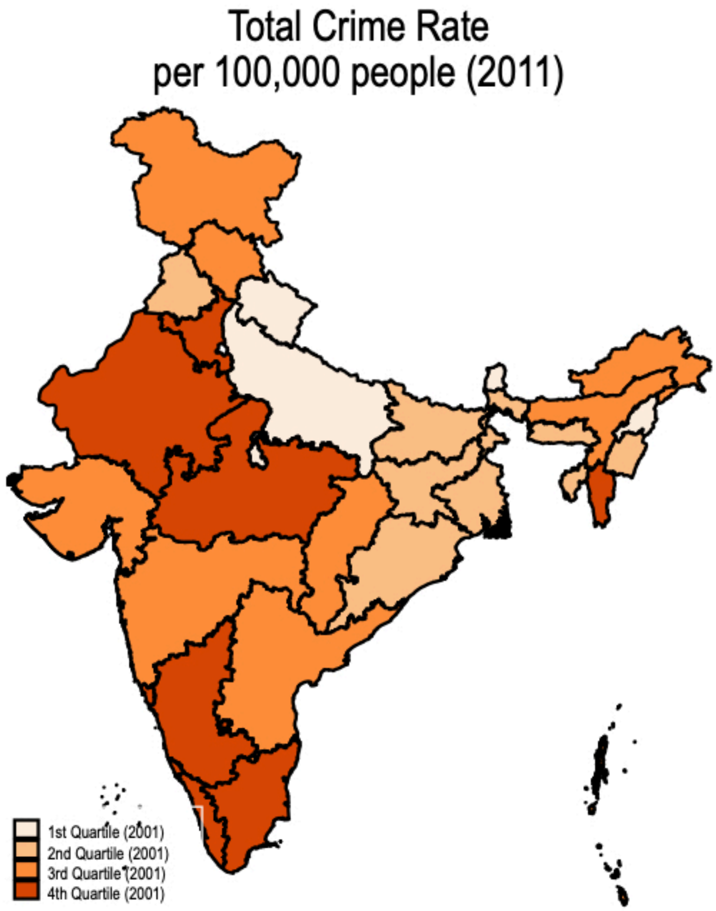

Figure 2: State-level Rate of Total Crime in 2011

2011 respectively. All the districts within states at the center and the south show similar crime rate, however the western state of Rajasthan has significant variation between districts. The westernmost parts of Rajasthan have a lower crime rate as compared to its eastern areas both in 2001 and 2011.

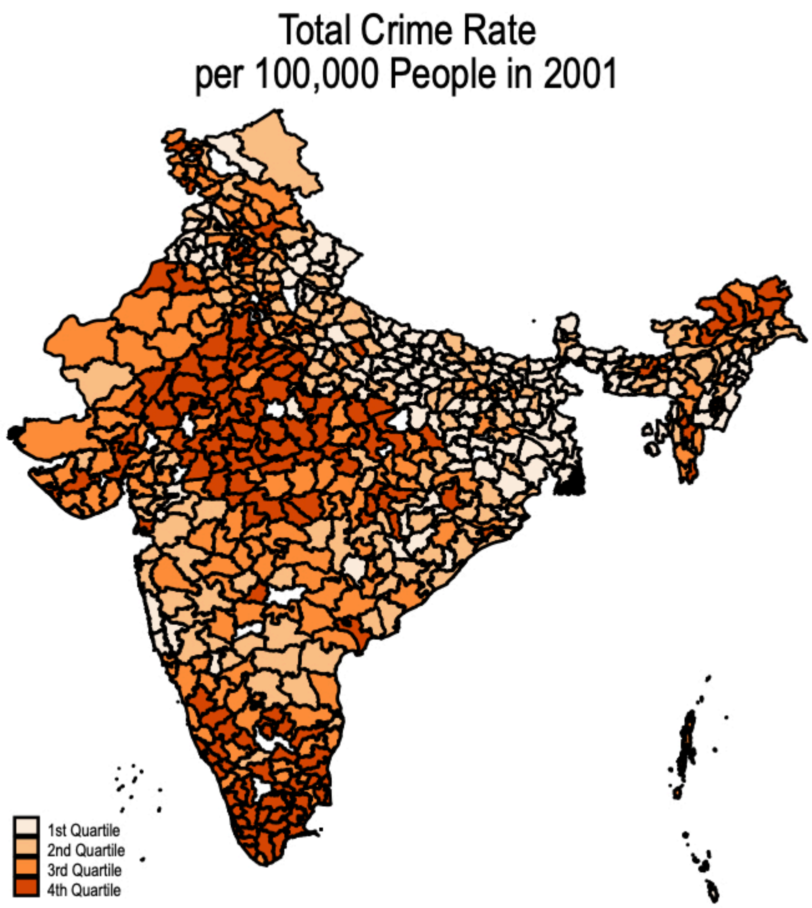

Figure 3: District-level Rate of Total Crime in 2001

2.4 Types of Crime

Crimes reported by the NCRB can be separated into seven main categories: crime against property, causing lethal hurt, crime against women, economic crime, rioting, murder and kidnapping.

2.4.1 Crime Against Property

Offences against property consist of any activities related to theft, extortion, robbery, burglary, and breach of trust among others. These offences and the

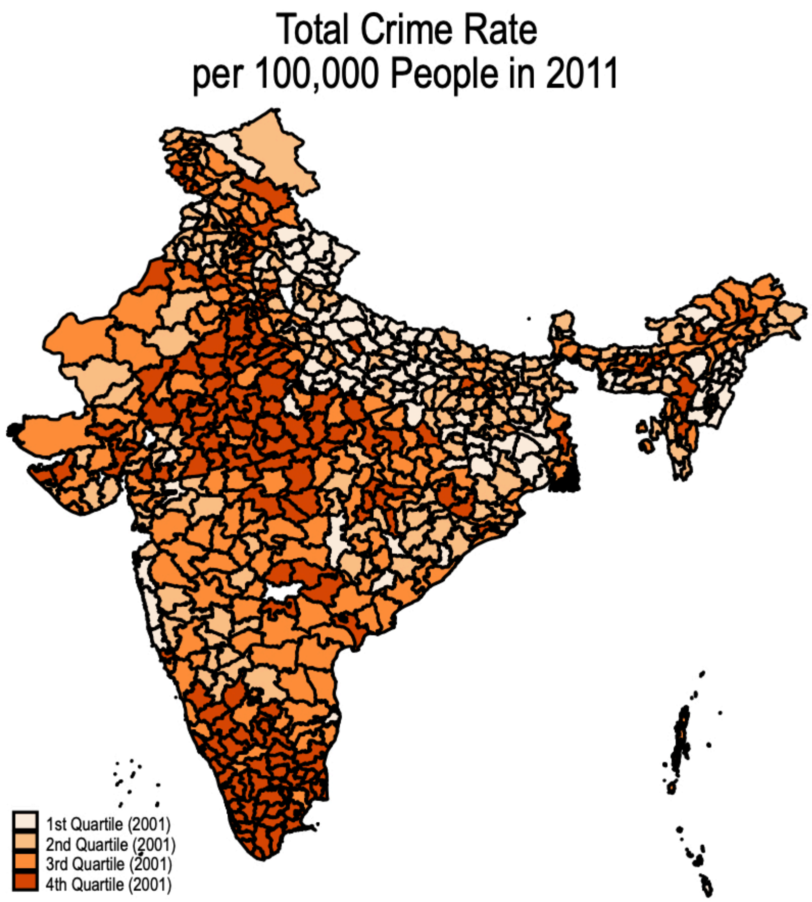

Figure 4: District-level Rate of Total Crime in 2011

punishments relating to them are explained in details in sections 378 to 460 of Chapter XVII of the Indian Penal Code. Robbery [8] is the act of taking or attempting to take anything of value from the care, custody, or control of a person or persons by force or threat of force or violence and/or putting the victim in fear. Burglary is considered as an unlawful entry of a structure to commit a felony or a theft including attempted forcible entry. Criminal Breach of Trust is the dishonest misappropriation or conversion of property someone was

[8] The definition is the same for dacoity

Table 1: Crime Rate (per 100,000 people) of Total Crime at the District-level

	Min	Mean	p25	p50	p75	SD	Max	Count
2001	.34	150.37	82.56	126.73	201.01	102.28	743.59	669
2002	.55	152.66	80.08	128.62	206.23	108.95	771.20	671
2003	.14	146.08	69.23	123.37	201.11	108.00	796.80	681
2004	.93	151.03	80.15	132.76	205.76	103.09	852.62	682
2005	.63	152.00	78.49	131.79	212.14	101.70	654.04	686
2006	.76	155.79	84.11	137.74	218.24	103.38	754.46	693
2007	.49	139.06	74.18	123.96	189.18	90.49	593.19	695
2008	.34	142.95	78.24	133.04	195.27	91.62	620.73	709
2009	.17	142.74	77.26	131.98	193.44	90.92	651.19	715
2010	.34	146.93	78.46	134.31	199.15	99.98	784.03	725
2011	.25	149.36	80.96	133.23	203.78	103.63	1055.27	737
2012	.34	149.48	76.79	133.72	205.28	102.55	728.07	753
2013	0	159.57	80.68	142.68	219.18	110.23	679.00	763
Total								9179

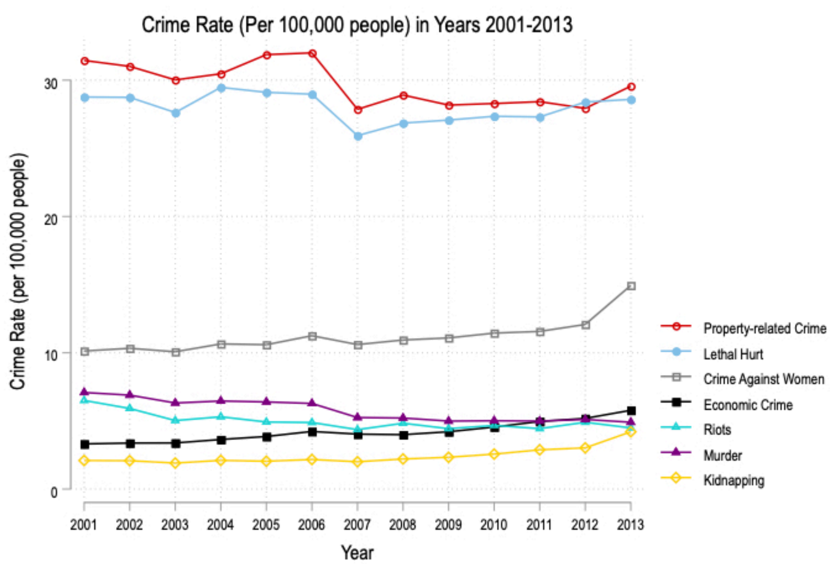

Figure 5: Trend in Crime Rates by Type of Crime (2001 - 2013)

entrusted with, for their own use. Arson is the malicious burning of the dwelling of another person. Finally, Theft is the unlawful taking, carrying, leading, or riding away of property from the possession or constructive possession of another while Automobile Theft, in particular, is the theft or attempted theft of a motor vehicle.

In this thesis, I categorize robbery (or dacoity), preparation of robbery, burglary, theft (includes automobile theft), breach of trust and arson as crime against property. Table 2 shows that in years 2001 to 2013, almost 36% of the total crime was property crime which makes this category take the largest piece in the pie of all types of crime.

Most property related crime in India occurs as a consumption smoothing strategy for low income culprits (Bharawaj, 2014). A bad day of not finding work or a bad year of less income sometimes forces the poor to fulfill basic necessities by engaging in such criminal activities. Out of the over 1 billion population of India, more than 273 million people (21.9%) are poor[9] (World Bank, 2011). The high proportion of property crime relative to total crime may be attributable to the fact that India comprises of a significant population below the poverty line.

The relative-deprivation theory explains that sometimes individuals commit property crimes to communicate to the State that the system they are forced to live in is inherently biased against them and their socioeconomic standing in the society (Chester, 1976; Hughes and Carter, 1981; Stack, 1984). Thus, even individuals with employment and legitimate income earning opportunity will be inclined towards committing crime in light of deprivation of basic needs and general inequality in society.

Figure 5 shows the time trends over years 2001 to 2013. While Crime Against Property remained the most common crime, it decreased overall from 2001 to 2013. This could be due to overall improvement in incomes in India over these years that decreased poverty.

[9] Poverty headcount ratio at US$1.90 a day (2011 PPP). World Bank, Development Research Group

2.4.2 Causing Lethal Hurt

This category is defined as any hurt which endangers life or which causes the sufferer to be during the space of twenty days in severe bodily pain, or unable to follow his ordinary pursuits (Chapter XVI, Section 320, the Indian Penal Code). Illegal hurt includes the offence of giving grievous hurt to someone, and causing death by negligence under lethal hurt. Note that grievous hurt also includes cases of acid attack which mostly takes place to harm women. Since the data did not document acid attacks separately, it couldn't be included in the category of "Offences Against Women" (the following section).

Acid Attack, more formally known as vitriolage, is any act of throwing or using sulfuric, nitric, or hydrochloric acid in any form on a person to cause permanent or partial damage or deformity or disfigurement to any part of the body of the person (Section 3(b), Relief and Rehabilitation of Offences on Women and Children, National Commission for Women). It reflects a deeply entrenched misogyny against women by husbands, disgruntled lovers and other male relatives. Patel (2014) used newspaper reports from January 2009 to December 2013 to find that 65.49% cases of acid violence during this period were found against women aged under 27 years, 15.51% cases were reported from rural areas. Thus, acid attacks are more frequent in urban areas.

Table 2 shows that cases of lethal hurt comprised on an average 30% of all types over years 2001 to 2013. It is the second highest in terms of proportion after crime against property. Figure 5 shows that over years the crime rate of lethal hurt cases decreased from 2001 to 2011. There is an increase after 2011, especially in 2012, the crime rate of lethal hurt is higher than property crime.

2.4.3 Crime Against Women

The Indian Penal Code includes all offences of rape and sexual assault under "Offences Affecting the Human Body" (Chapter XVI, Section 375 to 376 E). In our data, offenses against women include registered cases of rape, dowry deaths, sexual assault not leading to rape, husband cruelty and women trafficking. Table 2 shows that on an average 13% of the total cases that occurred between 2001 and 2013, could be categorized as crime against women. As detailed in Section 2.2 this figure is underestimated because crimes against women that end in murder are categorized as murder and it does not include acid attacks.

In India, the treatment of women as private property, to be protected by men of their family, social, communal and caste groups is implicit in the culture (Kannabiran 1996; Dasgupta 1989; Desai and Krishnaraj 1987 among others). Rape (and generally crime against women) can be associated with this male dominated socio-economic, legal and political order where women are required to be protected because of the values placed upon their virginity and chastity (Atray 1988; Verma 1990; Nagla 1993 among others).

Sexual assault of women is thus correlated with their social status, communal, ethnic and caste identities (Mukherjee et al, 2001). Common occurences of mass rapes of lower caste women by upper caste men (Maydeo 1990) or of rural and tribal women by policemen and army personnel are examples of subduing women (Kelkar 1992). Marital rape is justified by husbands and their families as culturally appropriate ways of maintaining authority and control over a woman's behavior (Dhruvarajan 1989; Mahajan 1990; Mukhopadhyay and Garimella 1998).

An existing empirical study explores the potential solutions to the problem of rape and sexual assault in India. Bhalotra and Prakash (2021) examine the impact of establishing women police stations (WPS) on reporting of gender-based violence. They use administrative crime data and create a natural experiment due to staggered implementation across Indian cities, to find that the opening of WPS is associated with an increase in police reports of crimes against women by 29%, a result driven by domestic violence. They also find that the opening of WPS increases women's labor supply, consistent with women feeling safer once the costs of reporting violence fall.

Another major category of this type of crime is Dowry Death. According to IPC, any crime "Where the death of a woman is caused by any burns or bodily injury or occurs otherwise than under normal circumstances within seven years of her marriage and it is shown that soon before her death she was subjected to cruelty or harassment by her husband or any relative of her husband for, or in connection with, any demand for dowry, such death [shall be called] "dowry death", and such husband or relative shall be deemed to have caused her death." (Chapter XVI, Section 304 B, Indian Penal Code).

Traditionally, marriage in India is endogamous and patrilocal (Bloch and Rao, 2002). That is no individual can marry outside of their community and the bride leaves her father's home to live with the husband's family. The bride's family must arrange for the dowry (gifts given to the groom's from the bride's family) and dowry costs have been increasing over years (Bloch and Rao, 2002).

Once a bride moves in with the groom's family, she becomes a potential hostage. Since divorce is impossible (marriage is considered final), she cannot move back to her parents' home. In some situations parents may prefer violence against their daughters over the dishonor that divorce brings (Musa, 2012). This makes husbands resort to violence to extract further transfers beyond the original dowry, failure to pay might the lead to the wife's death.

There are patterns of domestic violence leading to dowry death. Sekhri and Storeygard (2014) examine the response of dowry deaths to weather variability. A one standard deviation decline in annual rainfall from the local mean increases dowry deaths by 7.8%. Thus deaths tend to increase during economic hardships. Bloch and Rao (2002) use probit estimates to show that husbands are more likely to beat their wives, resulting in death, when the wife's family is rich because there are more resources to potentially extract. They further find that the probability of violence decreases by husband's greater satisfaction with the marriage. This is indicated by higher numbers of male children, another opportunity to seek dowry from their future wives. These patterns show that dowry deaths are strongly linked with economic incentives. Bhattacharya et al, (2021) use spatial panel regression results on dowry death to show that neighbourhood influences variations of dowry death. They further find that female work force participation and presence of police stations significantly reduces the incidence of dowry deaths.

Figure 5 shows that not only is crime against women, the third most common crime in terms of proportion, but it also has consistently increased over the years. Interestingly, there is a large jump from 2012 to 2013. This may be because after the 2012 Delhi gang rape and murder incident [10], NCRB decided to expand crime against women to include rape leading to murder.

Figures 6 and 7 show the rate of crime against women across different states in India, in years 2001 and 2011. In 2001, Madhya Pradesh in the center and Mizoram in northeast have the highest rate of crime against women. In 2011, more states shifted to higher quartiles of crime rate. Mostly, crime against women increased in western states of Rajasthan and Haryana, southern states of Kerela and Andhra Pradesh and eastern states of Orissa, West Bengal and the north eastern states. This may be due to better reporting and improvement

[10] On 16 December 2012, a 22-year old physiotherapy intern (Nirbhaya, a named used to protect the victim's identity) was beaten, gang-raped and tortured in a private bus where seven men raped the victim. The incident generated widespread national and international coverage and was widely condemned. (Source: Wikipedia)

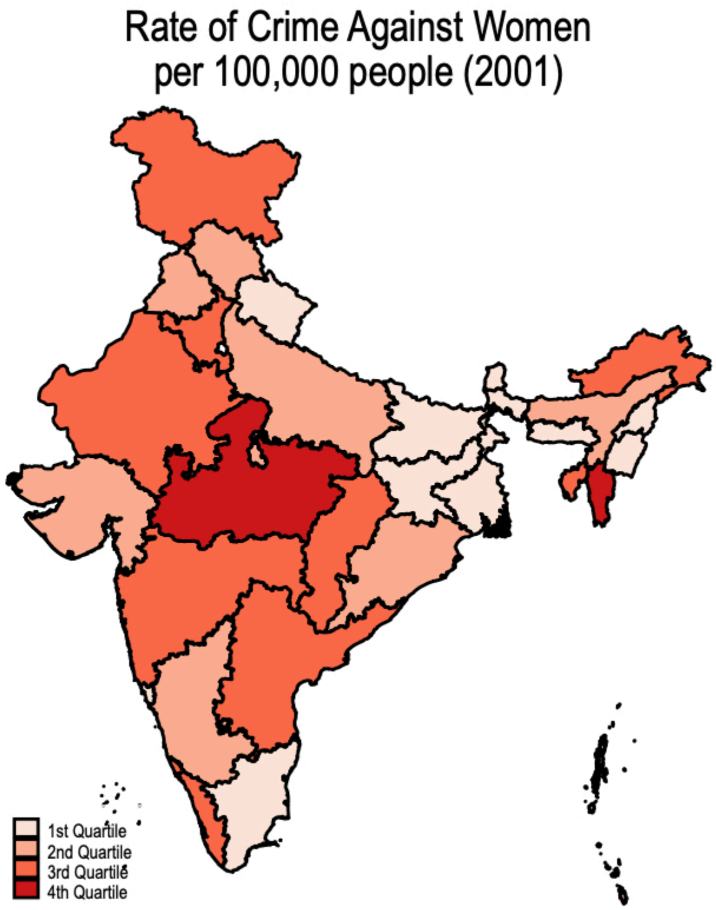

Figure 6: State-level Rate of Crime Against Women in 2001

in data collection over years.

Figures 8 and 9 show rate of crime against women at the district level. Overall in both 2001 and 2011, we get similar results of higher crime rate at the center and towards the west. It is interesting to see some trends within states, for example, the state map shows that the crime rate increased in the state of Tamil Nadu from 2001 to 2011, but the district-level maps suggest that

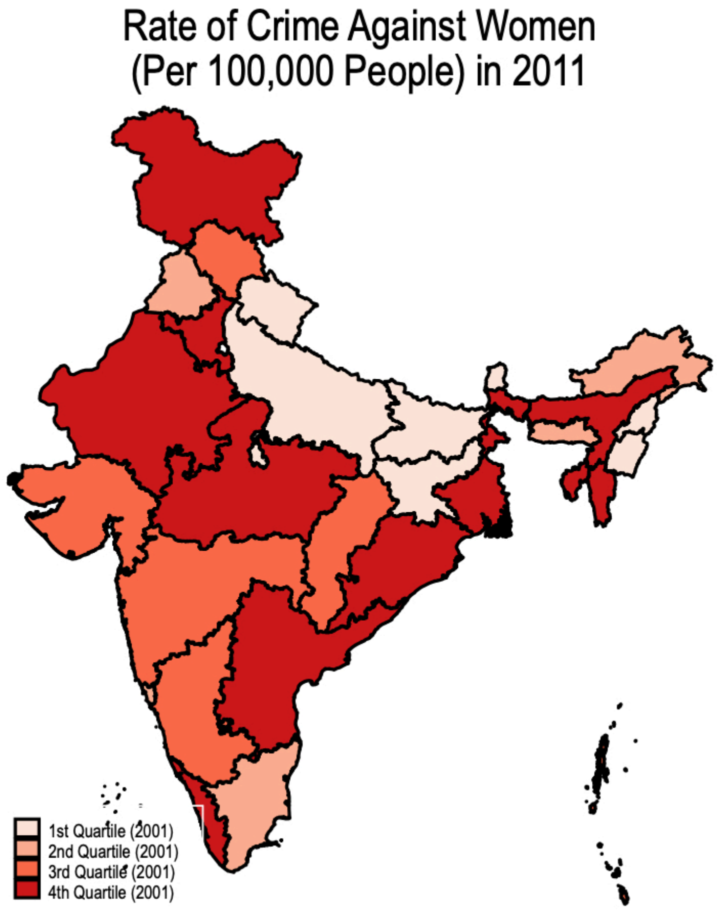

Figure 7: State-level Rate of Crime Against Women in 2011

the increase in rates was disproportionate between districts. Similar to total crime, crime against women in Rajasthan has a higher concentration in eastern districts.

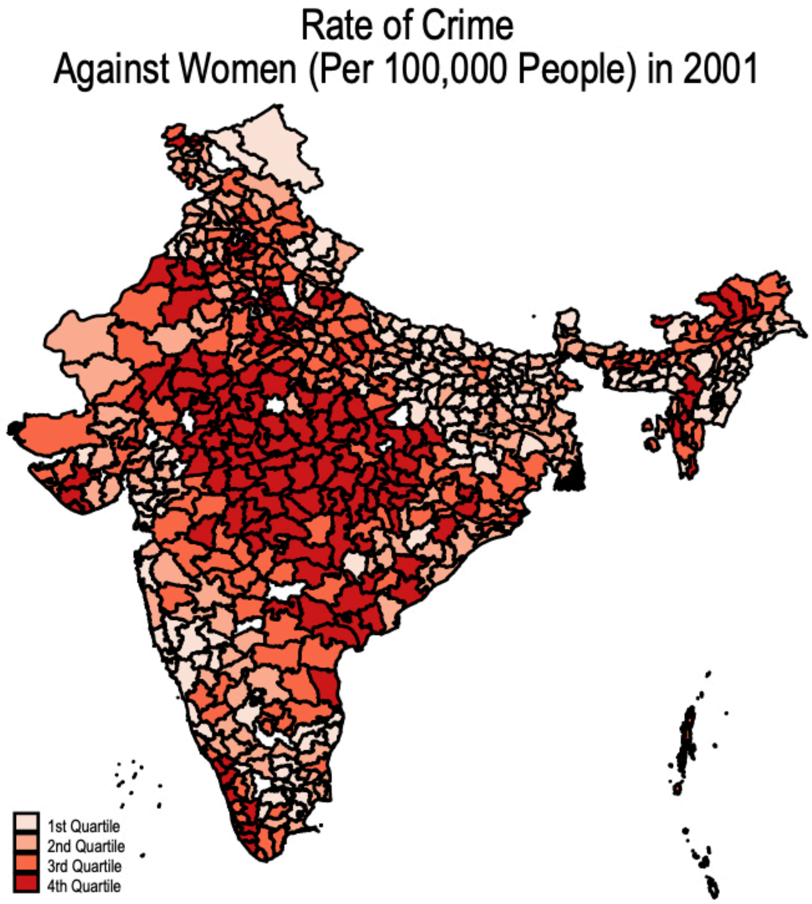

Figure 8: District-level Rate of Crime Against Women in 2001

2.4.4 Economic Crime

I classify Cheating and Counterfeiting under the category of Economic Crime. The Indian Penal Code defines counterfeiting and Cheating as "Offences Related to Documents and to Property Marks". (Chapter XVIII, Sections 463 through 477A, Indian Penal Code). I call these economic offenses, because most of them are related to illegally forging economic documents. This is a white collar crime, an exception in cognizable IPC offences where generally the culprits belong to the low-income population of India.

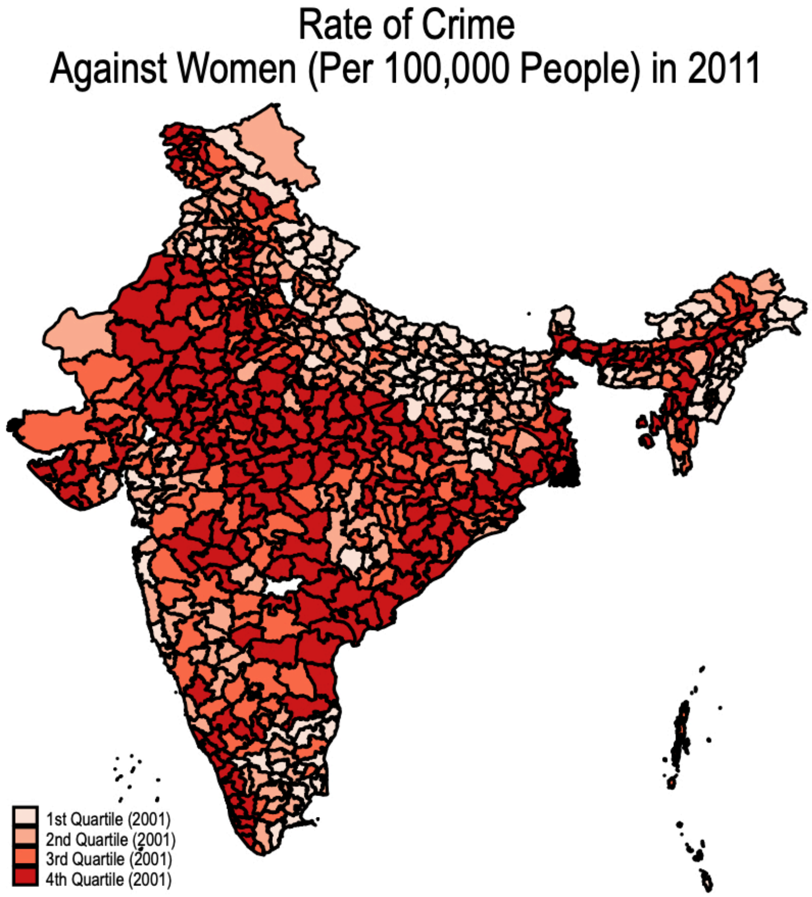

Figure 9: District-level Rate of Crime Against Women in 2011

2.4.5 Rioting

Rioting is considered an offence against public tranquility. It's definition —"Whenever force or violence is used by an unlawful assembly, or by any member thereof, in prosecution of the common object of such assembly, every member of such assembly is guilty of the offence of rioting." (Chapter VIII, Section 146, the Indian Penal Code). Riots can be a result of religion based or caste based differences.

India is largely composed of Hindus (79.8%) who are in majority, followed by Muslims (14.2%), Christians (2.3%), Sikhs (1.7%), Budhhists (0.7%) and

Jains (0.4%) (Census of India: Data on Religion, C-01 Population by Religious Community, India, 2011. Values calculated by Mehar Singh, 2019). Most religion-based riots in India are because of Hindu-Muslim and Hindu-Christian differences [11]. These differences have existed historically, but the seed of hatred was sown in the 1980s when the issue of Hindu to Islam conversions popularly known as the Meenakshipuram conversions was highly politicised. (Engineer, 1992). This policitization also led to two major movements, that is the Shah Bano case, [12] and the Babri Masjid movement [13] in an agressive manner which made the average Hindu hostile towards Muslims. (Upadhyay and Robinson, 2012).

Since then there have been multiple instances of religious killings in riots of Hindu-Muslim differences. Major riot incidents in the time period of our data set include The Gujarat Riots of 2002 which led to more than 2000 deaths that was started by the burning of a train causing the death of Hindu pilgrims abroad. In the Kandhamal violence of 2008 over 500 churches were burnt down and over 54,000 people were left homeless due to their homes getting ransacked. The Muzaffarnagar riots of 2013 is another Hindu-Muslim riot example that ended up in 62 deaths and 93 causalities. (All numbers and timeline sourced from Wikipedia: List of Riots in India)

Hate crime might also lead to riots between two parties. This term refers to "unlawful, violent, destructive, or threatening conduct in which the perpetrator is motivated by prejudice toward the victim's putative social group" (Green et al., 2001). In India, the vernacular caste system (division in the community on

[11] Jains, Sikhs and Budhhists many overlapping religious beliefs with Hindus, which reduces possibility of violence between the groups

[12] Controversy of a divorced muslim woman named Shah Bano's right to alimony from her husband. When the supreme court ruled that she will received alimony, the ruling evoked criticism among the Muslim community who cited the Qu'ran to show conflict with Islamic law

[13] Demolition of a mosque by Hindu nationalists who believed the mosque was built on the birthplace of Rama, a hindu god.

the basis of occupation) created a social hierarchy where historically, the upper castes have had an advantage over the lower castes in terms of opportunities and economic stability (Deshpande, 2010).

The systematic discrimination created by the caste systems are the roots for hate crime and rioting between groups. For example, in Bihar, which is a state in eastern India, upper-caste landowners formed a private army in 1994 called the Ranvir Sena to "protect" themselves from the lower castes which led to 20 massacres by early 1999. (Deshpande, 2010). Sharma (2013) finds that among the largely violent caste-based hate crimes, it is the property related crimes that seek to deprive the victim of their property symbolic of their material progress that are affected by the relative standards of living. A 1% increase in the expenditure gap between the upper and lower caste people (1% more for upper caste), thus decreases crime rates by 0.34%. The upper caste feel less threat when there is high inequality in the society making them commit less hate crime. (Sharma, 2013)

2.4.6 Murder

The Indian Penal Code defines culpable homicide and murder as "Whoever cause[ing] death by doing an act with the intention of causing death, or with the intention of causing such bodily injury as is likely to cause death, or with the knowledge that he is likely by such act to cause death, commits the offence of culpable homicide." (Chapter XVI, Offences of Affecting Life, Section 299 and Section 300). For the purpose of this paper, I combine the data on reported murder, attempt to murder, homicide, and culpable homicide not amounting to murder under one broader category of murder. The homicide trend is the only crime category following the global crime trend (Ansari et al, 2015). Note that I will use Murder and Homicide as synonyms.

The nature of homicide in India is not very different from the rest of the world, except of a kind known as "Honor Killing". Honor Killings happen when patterns of human behavior, particularly among women are considered as bringing dishonor to their families and communities. The lost honor is repaid by killing them. (Dublish and Khan, 2021). This criminal offence comes under the category of "culpable homicide". Honor-based violence and killing in India is a centuries-old occurrence and in fact has a well established socio-cultural acceptance (Hossain & Welchman, 2006, p.3).

Since honor killing is socially accepted in many communities, unlike other forms of crime, community members, elders, and especially councils (eg: Khap Panchayats in Haryana) patronise it and defend the killers. The role of the state is rejected and legal action is seen as intrusion in socio-cultural norms and familial patterns (Dublish and Khan, 2021).

Dublish and Khan (2021) consider the sample of honor killings in the state of Haryana [14] where they found that inter-caste romantic affairs are the leading cause of honor killing (47% of the cases), while relationship between girls and boys of the same lineage [15] make 17% of the killings. They also find that these killings are strongly biased against girls: 52% killings were only girls, 10% only boys and 38% was killing of both the girl and the boy.

There are some patterns in the reported cases of murder. Most of the convicts of homicide are young men. There are rare cases of murder by women, the cases usually registered in response to harassment or infidelity. Infanticide, especially of girl child by the mother is another category of murder crime committed by women, but the incidence is highly unrecorded. (Dreze and Khera, 2000)

The main motive behind murder committed by men is to acquire property and fight over women that account for a substantial portion of reported crime (Dreze and Khera, 2000). Driver (1961) performed a case study of 144 convic-

[14] They use 100 honor killings documented by the newspaper 'The Tribune' in the state of Haryana from 2005 to 2013 and then survey these particular cases.

[15] Hinduism strictly forbids marrying or any relationship in the same lineage/clan

tions for murder in central India and found that victim and offender belonged to the same caste in 84 percent of the cases. The main motives for murder were disputes over property, living arrangements, sexual matters, and transgressions of social norms.

Most empirical work using homicide data to my knowledge has been done by Dreze and Khera in 2000. Three significant patterns emerged from their robust regression analysis study. First, murder rates in India bear no significant relation to urbanization or poverty. Second, education appears to moderate influence on criminal violence. Third, the strongest correlate of the murder rate is the female-male ratio: districts with higher female-male ratios have lower murder rates.

Table 2: Year-wise Country-wide Proportion (in %) of Crime

Year	Property	Lethal Hurt	Crime Against Women	Economic Crime	Riots	Murder	Kidnapping	Total Number
2001	37.12	31.76	10.87	4.35	7.14	6.66	2.11	1067946
2002	36.63	32.46	11.10	4.55	6.57	6.60	2.08	1050033
2003	37.08	32.64	11.62	4.88	5.64	6.17	1.97	1015708
2004	37.20	32.53	11.74	4.90	5.50	6.00	2.14	1090984
2005	37.24	32.48	11.82	5.18	5.20	5.95	2.11	1080625
2006	36.96	32.02	12.45	5.48	5.16	5.76	2.18	1098596
2007	36.08	31.79	13.31	5.82	5.16	5.46	2.37	1160467
2008	36.82	31.25	13.24	5.63	5.34	5.28	2.45	1236416
2009	36.82	30.78	13.47	6.02	5.01	5.20	2.70	1255804
2010	35.94	31.05	13.47	6.27	5.19	5.12	2.96	1300759
2011	35.50	30.95	13.50	6.62	5.04	5.10	3.29	1359543
2012	33.99	31.79	13.67	6.80	5.25	5.15	3.35	1420944
2013	34.48	29.29	15.69	7.05	4.64	4.63	4.21	1555097

The proportions were calculated by aggregating the levels of the crime category for the whole country for each year and then determining its proportion as a percentage of the sum of all crime categories. The sum does not include the category "Other Crime" in the NCRB Data

3 RURAL ROADS IN INDIA

This chapter is an overview of the rural surface road network in India. I explore the history of road construction, and then shift the focus to the PMGSY program. My aim through this chapter is to understand temporal and geographic variations in road construction in India until the year 2015.

3.1 Road Development in India Before 2001

In the 18th century and the beginning of the 19th century, most Indian roads were built by Mughal emperors. These roads were metalled[16] and were constructed to improve the postal system, and trade of commodities within the empire (Farooque, 1973).

During the late 19th century, as the British colonized India, roads were constructed for inland travel and to transport goods (Donaldson, 2018). In the early days of the British Raj in India, the management of public works was under the control of the Military Board of Imperial Government (Public Works Department, Government of India). Most of the inland travel was facilitated by bullocks. However, high-quality roads were sparse and the roads were impassable during the monsoon season (Donaldson, 2018). To improve the management and state of affairs in public works, in 1855, Lord Dalhousie formed the Central Public Works Department to facilitate road construction and maintenance (Public Works Department, Government of India [17]). However, by the 1866 the focus of this department shifted from road construction to railways in order to exploit the abundant gains from the trade of resources (like raw cotton) in India (Donaldson, 2018).

The advent of railways led to the decline of the road system. Any road

[16] A metalled road in the 1800s had a level surface made of small pieces of stone. Today, metalled roads are made up of cement, concrete or coal tar.

[17] History of Public Works

improvements were confined only to feeder roads that supported railways (Das, 2018). The British government identified the underdeveloped road transport as an infrastructural barrier. In 1929 the Jayakar committee, appointed by the British, was the first organized effort of a national level road construction scheme (Thomas, 1984). The committee recommended formation of the Indian Roads Congress in 1934 to organize a sequence of plans: the Nagpur Plan (1943-62), the Bombay Plan (1961-81) and the Lucknow Plan (1981-2001). Each subsequent plan was more advanced than the previous one.

The goal of all three plans was to facilitate organized road construction. The Nagpur Plan (1943-62) classified the name of various types of roads as national highways, provincial highways, district roads and rural roads and prescribed standards, norms and targets for road development under these categories (Thomas, 1984). The Bombay Plan (1961-81) targeted that no village be more than 2.5 kilometers from a road in developed agricultural areas, 5 kilometers from a road in semi-developed areas and 8 kilometers from a road in underdeveloped and uncultivated areas (Das, 2018). The Lucknow Plan (1981-2001) guided states to prepare their own plans based on differences in geography, demographics and social structure (Das, 2018). One of the advantages of the half-century long organized road construction was the strict classification of different kinds of roads, thus making it easier to identify rural roads.

3.2 What is a Rural Road?

According to the Nagpur Road Development Plan of India (1943-1961) and Lucknow Road Development Plan (1981-2001), rural roads are the tertiary road system comprising Other District Roads and Village Roads (Sikdar 2000). According to this definition, a rural road connects villages with their nearest market centers or another village that is already connected to a market center, thus linking it to the main network. These roads are characterized by low traffic volume, which primarily consist of two wheeled vehicles (cycles, motorbikes), animal drawn vehicles, and pedestrians (Khanna and Justo, 2009; Sikdar 2000).

Rural Roads can be categorized based on their usage in different seasons of the year (all weather [18] and fair weather roads), pavement status (paved [19] and unpaved road) and pavement surfacing (surfaced and unsurfaced roads). Until the year 2000, rural roads in India were unpaved, earthen or gravel (Rao, 2000). The unpaved roads become inaccessible during rainy season and people struggle to commute to other places (Fukubayashi and Makoto, 2014). Thus, paved roads (also known as pucca roads in India) are essential for uninterrupted commute via roads.

According to the World Bank (2007) rural road connectivity can be measure using two indices: Road Density and Road Availability. Medhi (2017) formulated a method of calculating these indices using the following formulae:
(i) Rural Road Density = Rural Road Length per 100 sq. km. of Rural Area
(ii) Rural Road Availability = Rural Road Length per Lakh[20] Rural Population

3.3 Status of Rural Roads before 2001

Rural roads were included for the first time in the Fifth Five-Year Plan (1974-79) under the Minimum Need Programme. Prior to 2000, rural development programs executed by village-level municipality called village panchayats were primarily responsible for building rural roads (Medhi, 2017). Sometimes, market committees and sugar cane societies also built rural roads for their own benefit.

Although the length of rural roads increased under these programs, the lack of a systematic construction prevented all villages from being connected by

[18] All Weather Roads are those which are negotiable during all weather, except at major river crossings where the interruption of traffic is permissible up to a certain extent.

[19] Paved Roads are roads with hard pavement, and should fall under the classification Water Bound Macadam (WBM) roads

[20] 1 lakh = 100,000

a road. The quality of roads was also not adequate for heavy-duty vehicles which lead to fast deterioration of the roads. At the end of the 20th century, about 330,000 of India's 825,000 villages, and habitations (hamlets or subvillages) lacked an all-weather road access (Government of India Report, 2006). Most of the poorly connected rural communities were in 10 states: Assam, Bihar, Chhattisgarh, Himachal Pradesh, Jharkhand, Madhya Pradesh, Orissa, Rajasthan, Uttar Pradesh, and West Bengal. About 28% of India's population had limited access to roads and thus to education, hospitals and other economic and social opportunities (World Bank, 2021).

3.4 The PMGSY

As a solution to the lack of connectivity, in October 1999, the president of India announced a new program to build all-weather roads to connect villages and habitations. The National Rural Road Development Committee was constituted in January 2000 to move this plan forward and launched the Pradhan Mantri Gram Sadak Yojana (PMGSY) or the Prime Minister's Village Road Program in December 2000. It was estimated that about 178,000 habitations would be provided with connectivity under the program (World Bank, 2021). Thus, the PMGSY is the first ever nationwide road construction program in India.

3.4.1 Objective

According to the PMGSY guidelines (2015), the primary objective of the PMGSY is to provide connectivity, by the way of an all-weather road [21](with necessary culverts and cross-drainage structures, operable throughout the year), to the eligible unconnected habitations [22] in rural areas with a population of 500 persons

[21]All-weather roads and paved roads are used as synonyms in the context of this thesis.

[22]a cluster of people living in the area whose location does not change over time. According to the PMGSY administrative database, a village mostly has more than one habitation.

and above (Census 2001). Thus, the goal was to connect the greatest number of Indian villages by roads.

3.4.2 Selection and Sequencing of Location

The road construction that began toward the end of 2001 had a target of providing road access to every habitation with a population of more than 1000 by 2003 and to every habitation with a population of more than 500 by 2007 [23]. The priorities can be summarized as:

- **Priority I:** Habitations with Population size of 1000+
- **Priority II:** Habitations with Population size of 500-999
- **Priority III:** Habitations with Population size of 250-499

The government focused on routes terminating in habitations rather than routes passing through habitations onto larger roads. Many projects under the scheme upgraded pre-existing roads which involved building the base and the surface courses to a desired technical specification and/or improving the geometry of the road (upgrading) as required by traffic conditions (PMGSY Program Guidelines, 2015).

3.4.3 Data

Two data sets provide information about rural road connectivity — the Population Census and the PMGSY data. The Population Census consists of data on the presence of paved roads in years 2001 and 2011. It also provides data on village characteristics such as the village area, literate population, number of schools in the village, and presence of power supply. The PMGSY data has

[23] This threshold is 250 in hill states (the Northeastern states, Sikkim, Himachal Pradesh, Jammu and Kashmir, and Uttarakhand)

information on year-wise PMGSY road allocation, both for newly constructed roads and upgraded roads.

These data are procured from the Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG) database, Version 1.0. SHRUG is an open access repository currently comprising dozens of data sets covering India's more than 500,000 villages and 8000 towns using a set of common geographic identifiers that span 25 years.

3.4.4 Implementation

Since the provision of all-weather roads is conceived to contribute towards poverty alleviation, the Ministry of Rural Development (MoRD) administers and manages the PMGSY. The program is run by state/union territory governments because the Constitution of India considers rural roads as a state responsibility. The execution happens on a state-by-state basis, meaning if one state connects all villages at one population threshold, they go ahead and connect at the next threshold level.

The road construction was fully funded by the central government until 2014-15. After that period, the construction costs have been shared by the central and state governments, with the central government funding 60% in plain areas and 90% in special category states. States are responsible for maintaining the constructed roads.

An important feature of the PMGSY is its strong national focus on rural road development through the National Rural Infrastructure Development Agency (NRIDA) under the Ministry of Rural Development (MoRD). NRIDA oversees and coordinates all technical aspects and facilitates systematic monitoring of program implementation in the states and union territories.

3.4.5 Efficiency of Implementation

The World Bank (2021) reports that the PMGSY connected 90% of its initial targets (originally planned to be connected by 2007) by 2017. Lack of resources and corruption can be some reasons of the decade long delay in results. Lehne, Shapiro, and Eynde (2018) use regression discontinuity to study the causal relationship between winners of close elections and named contractors of PMGSY road building contracts. They find that contracts with a name matching the winning politicians increase from 4% to 7% in the term after a close election compared to the previous term, and that this raises the cost of road construction and the likelihood that roads go missing both of which are implicit in corruption. On similar lines, Asher, Nagpal, Novosad (2017) find that villages with higher distance from administrators are 1.2% less likely to have paved roads, and in general administrative remoteness reduces the provision of public goods. This is a consequence of political preference in resource allocation.

Another possible reason of inefficient implementation can be inequality within the country across different states. The program was too standardized and many states did not have the capacity (in terms of infrastructure and implementation efficacy) to achieve the highly ambitious goals in a timely manner. Economists Sam Asher and Paul Novosad worked closely with the National Rural Roads Development Agency to identify a subset of states that followed the population threshold prioritization rule according to the national PMGSY guidelines. These states are Chhattisgarh, Gujarat, Madhya Pradesh, Maharashtra, Odisha and Rajasthan.

In early 2000s, Madhya Pradesh, Rajasthan, Uttar Pradesh, Bihar and Odisha were the least connected states of the country (The World Bank, 2020). Though with a delay of 10 years, by December 2017, connectivity in most of the above states increased considerably (The World Bank, 2021). Table 3 shows the percentage of connected habitations of the most connected states in 2017 (figures from the World Bank Report, 2021).

Table 3: PMGSY Progress in Select States as of 2017

State	Eligible Habitations	Connected Habitations	% Connected Habitations
Bihar	34,586	27,590	80%
Chhattisgarh	10,191	9,368	92%
Madhya Pradesh	18,404	17,826	97%
Rajasthan	16,694	16,165	97%
West Bengal	18,641	12,557	67%

These data are drawn from a report by the World Bank published in 2021

According to the extracted data from the census of 2001, 22.5% (130,235 out of 579,546 villages) of villages had a population of more than 500 and no roads, while 10.53% (61,043 out of 579,546 villages) had a population of more than 1000 and no roads. The goal of the first phase was to provide roads to the 61,043 villages by 2003 and to the 130,235 villages by 2007.

Table 4 uses the PMGSY data to show the trend in percentage of villages connected by the PMGSY between years 2001 to 2015 out of the eligible villages in the country, under each category. The implementation was not perfect. By 2003 only 3.86 % of the planned roads were built; by 2007 only 13.05% of the eligible villages were connected by roads.

The overall implementation of the program improved over time, with only 15,500 kilometers of completed rural roads in the first year (2001) to 52,400 kilometers in 2017 at a cost of about rupees 1.88 trillion ($27 billion) (World Bank, 2021). This makes the PMGSY one of the most successful infrastructural development programs in India since independence.

3.4.6 Social and Economic Impact

Existing literature on the PMGSY, has explored the impact of connectivity on local economic development, structural transformation of the economy, educa-

Table 4: Yearwise Village Level Connectivity Under the PMGSY

Year	Pop > 500	Pop > 1000
2001	0.11	0.18
2002	0.77	1.29
2003	2.41	3.86
2004	4.46	6.95
2005	6.73	9.78
2006	8.92	12.12
2007	13.05	15.05
2008	16.34	18.57
2009	20.17	22.97
2010	23.49	26.79
2011	26.01	29.32
2012	28.33	31.80
2013	30.06	33.51
2014	32.31	35.18
2015	32.46	35.39
Observations	130235	61043

The table shows percentage of eligible villages under each category connected by the PMGSY

tion, and forest cover. These studies exploit the population rule of the PMGSY policy to structure their empirical strategy.

Asher, Novosad et al. (2020) use a fuzzy regression discontinuity design to show that a new rural roads under the PMGSY moves on an average, 10% workers out of agriculture, with no major change in agricultural outcomes, income, assets or employment in village firms. Thus, rural roads lead to migration from rural to urban areas for jobs.

Adukia, Asher et al, (2017) implement a panel fixed-effects regression that exploits the timing of the PMGSY road construction, within the set of all villages that received new roads by 2015. They found that a new paved road leads to a 7% increase in middle school enrollment in a village (95% confidence interval: 4.1 – 9.9 percent) and results in children staying longer and performing better in standardized exams. This provides evidence for rural roads improving attendance of educational institutions and literacy of a village.

Aggarwal (2018) studies the impact of rural roads on a variety of economic outcomes at the district level using the exogenous variation in the timing and placement of the PMGSY paved roads. The study found that roads lead to better integration between rural and urban markets thus reducing prices of goods imported from urban areas and improving availability. Teenagers drop out of schools to join labor force as urban markets become accessible. This means that the increase in school enrollment associated with rural roads is not observed across all age groups.

While road connectivity has many positive externalities including but not limited to economic growth and opportunity, the construction can have negative effects on the environment and thus society because it entails deforestation. Garg, Asher, and Novosad (2020) estimate the ecological impact of transportation infrastructure in India. They exploit the PMGSY implementation to use regression discontinuity, and the exact timing of road construction to implement difference-in-differences. Furthermore, they implement a straight-line instrument using data from construction of major highways in India. They estimate a 0.5% reduction in forest cover during rural-road construction which reverses soon after roads are complete, while there was a 20% decline in forest cover in a 100 kilometer band around a new highway creating an effect that persists eight years. It thus seems that even though rural road construction has short-term effects on deforestation, it does not negatively impact the forest cover in the long-term. Perhaps, the localized effect of rural road construction does not average over to higher levels, unlike highways which are constructed to increase connectivity nationally.

One study estimates the impact of rural road connectivity on crime. Jain and Biswas (2021) examine the relationship between road infrastructure and crime rate in rural India using a nationally representative survey. They use an instrumental variable and observe that building roads in rural India reduces crime via the potential channels of improved street lighting, better public bus services and higher employment among others. This study does not use data from the PMGSY because of the limitation of inconsistent village identifier data. None of the studies conducted so far in my knowledge have yet considered the impact of PMGSY on crime at the district level using administrative data.

3.5 Trends in Rural Road Connectivity

I use data from the Population Census of 2001 and 2011 to analyze the trends in overall connectivity in the country, including but not limited to road connectivity provided particularly under the PMGSY program. Rural road connectivity of a geographic level (state or district) is measured as the fraction of people at that level with access to a rural road.

3.6 Temporal

Overall connectivity increased from 2001 to 2011. Figures 10 and 11 present the probability density function of the percent of the population with roads. Comparing the two shows an increase in the percentage of population with road access from 2001 in 2011. Figures 12 and 13 present the same probability distribution function as the district-level. They show a similar albeit, smoother upward shift. Thus, connectivity increased between 2001 and 2011.

3.7 Geographic

The temporal increase in state-level connectivity is not distributed equally across all states. The variation in state wise connectivity can be seen in Figures 14 and 15. In 2001, the central and Eastern states of Madhya Pradesh, Jharkhand and Bihar were amongst the lowest quartile of connectivity. Table 5 shows that

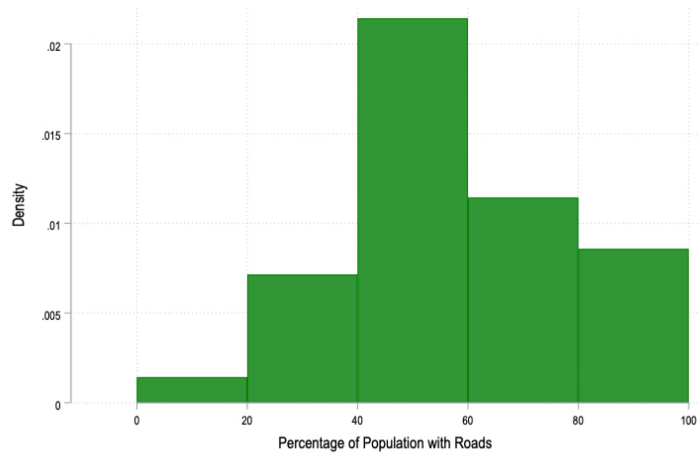

Figure 10: State-Level Connectivity in 2001

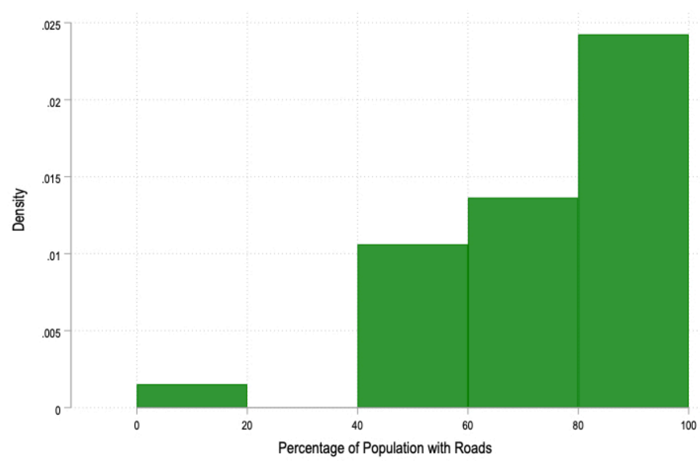

Figure 11: State-Level Connectivity in 2011

on an average only 30% of villages in these states had roads in 2001. Overall connectivity was highest in the west, north and south. In 2011, connectivity increased widely in all the states. Especially, the central and eastern states went from the lowest quartile to the highest quartile. [24] According to table 5,

[24] quartiles calculated as per the values of connectivity in 2001

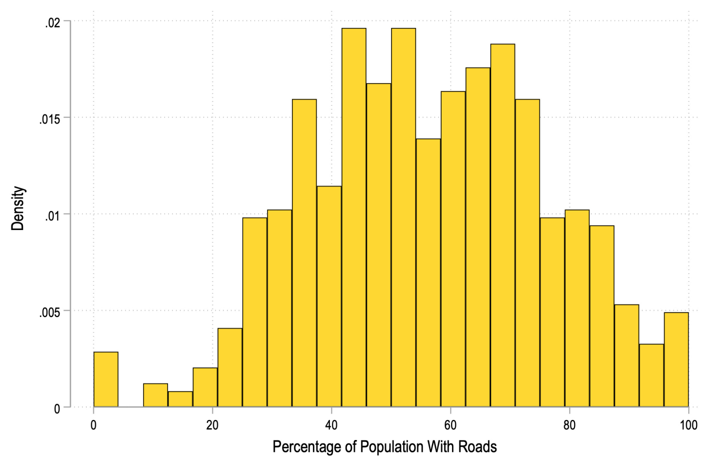

Figure 12: District-Level Connectivity in 2001

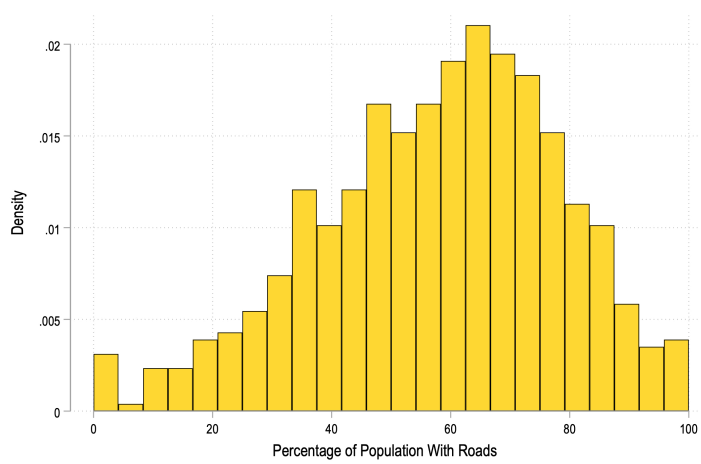

Figure 13: District-Level Connectivity in 2011

in 2011 on an average 60% of villages in the states of Madhya Pradesh, Bihar and Jharkhand were connected — an improvement of 30% percentage points. These states coincide with the states where the PMGSY implementation was comparatively better (Novosad, Asher, 2020).

Figures 16 and 17 show the district-level variation in connectivity. Even

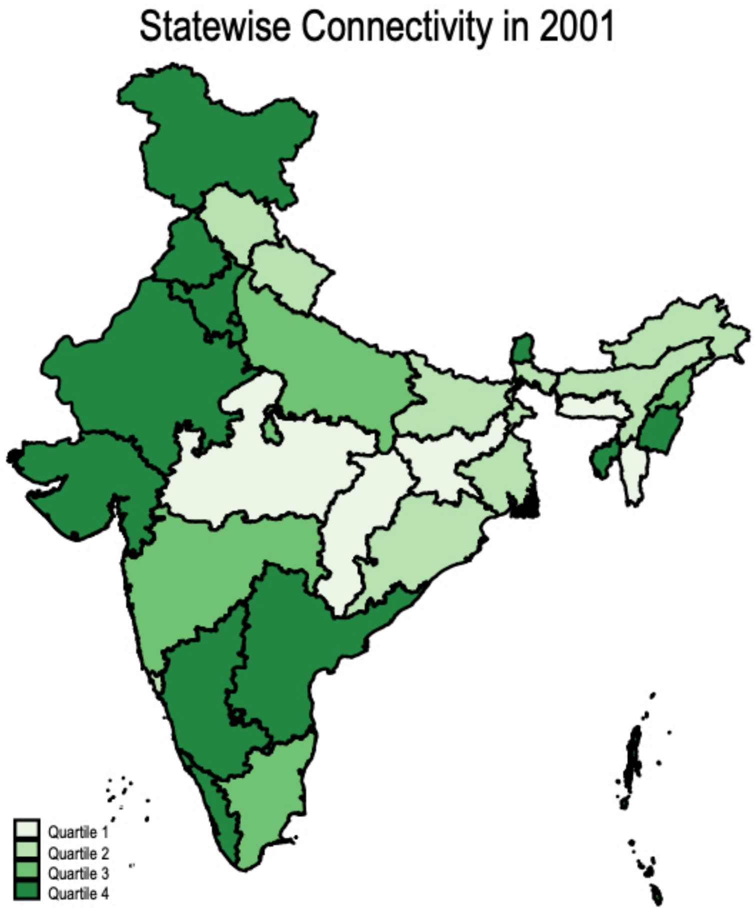

Figure 14: State-level Road Connectivity in 2001

though as previously seen, from 2001 to 2011, connectivity increased all over the country at the state level, there are large variations within the states, at the district level. In the central and eastern states, the increased connectivity was concentrated only to some districts. In fact, in some areas like the western part of Rajasthan, connectivity decreased. This is an indication of the potential unequal distribution of roads under the policy.

Table 5: State-wise Percentage of Connected Villages in 2001 and 2011

State/UT	2001 Percentage	2001 Total Villages	2011 Percentage	2011 Total Villages
Northern				
Haryana	98.56	6611	97.12	6604
Himachal Pradesh	45.17	13968	55.62	17490
Jammu & Kashmir	67.17	6287	56.04	6306
Punjab	99.62	11632	92.70	12151
Rajasthan	52.49	39405	24.32	39673
North Eastern				
Arunachal Pradesh	35.22	3273	25.36	3352
Assam	48.36	23048	18.93	24466
Manipur	41.17	2157	33.67	2290
Meghalaya	36.40	5420	42.57	5490
Mizoram	30.83	665	36.05	674
Nagaland	55.85	1205	27.42	1262
Sikkim	70.21	423	81	421
Tripura	84.71	824	80.89	832
Central				
Chhattisgarh	32.62	19567	62.09	19546
Madhya Pradesh	32.2	51524	52.91	51877
Uttar Pradesh	59.66	96794	65.96	97256
Uttarakhand	25.64	15520	32.31	15696
Eastern				
Bihar	38.20	37411	63.88	39050
Jharkhand	21.13	29096	95.16	29371
Odisha	39.97	45772	61.91	47421
West Bengal	46.84	37567	36.26	37278
Western				
Dadra Nagar Haveli	98.57	70	96.92	65
Daman & Diu	100	23	100	17
Goa	96.22	344	96.56	320
Gujarat	82.31	16128	95.43	16073
Maharashtra	82.71	40712	86.37	40776
Southern				
Andaman Nicobar Islands	59.94	377	57.55	384
Andhra Pradesh	71.21	244051	75.97	24196
Karnataka	70.65	26476	89.19	26679
Kerala	100	1320	100	991
Puducherry	100	92	100	90
Tamil Nadu	94.67	14906	96.15	14940
Total Villages		572678		583045

Table constructed using the dummy variable on presence of a paved road at the village level, in census data. Percentages are of the total number of villages in each state The division into regions (north, south ...) is based on zonal councils.

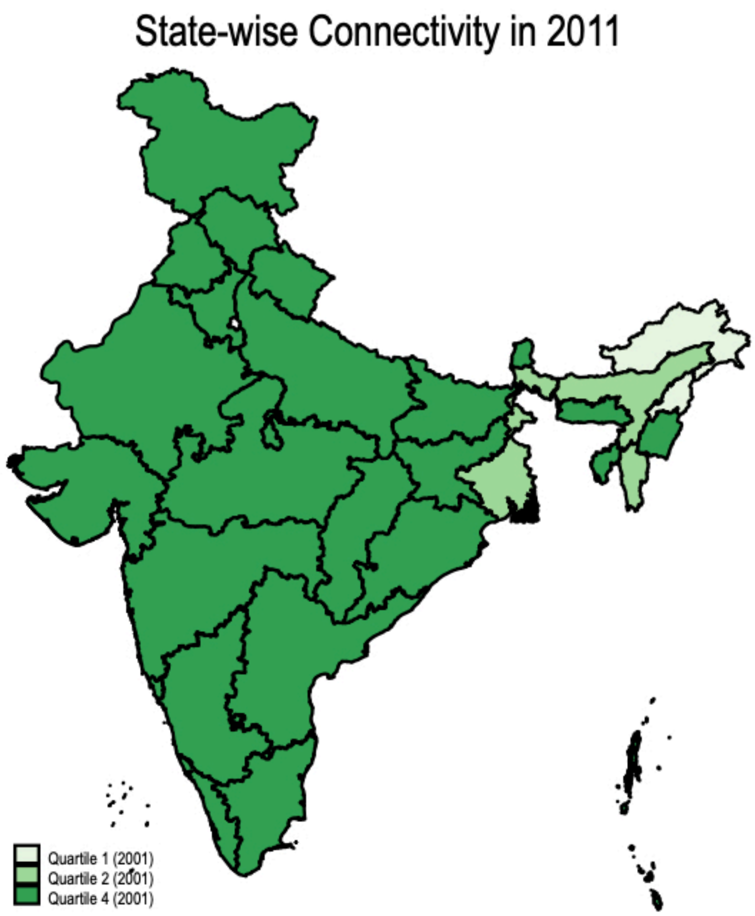

Figure 15: State-level Road Connectivity in 2011

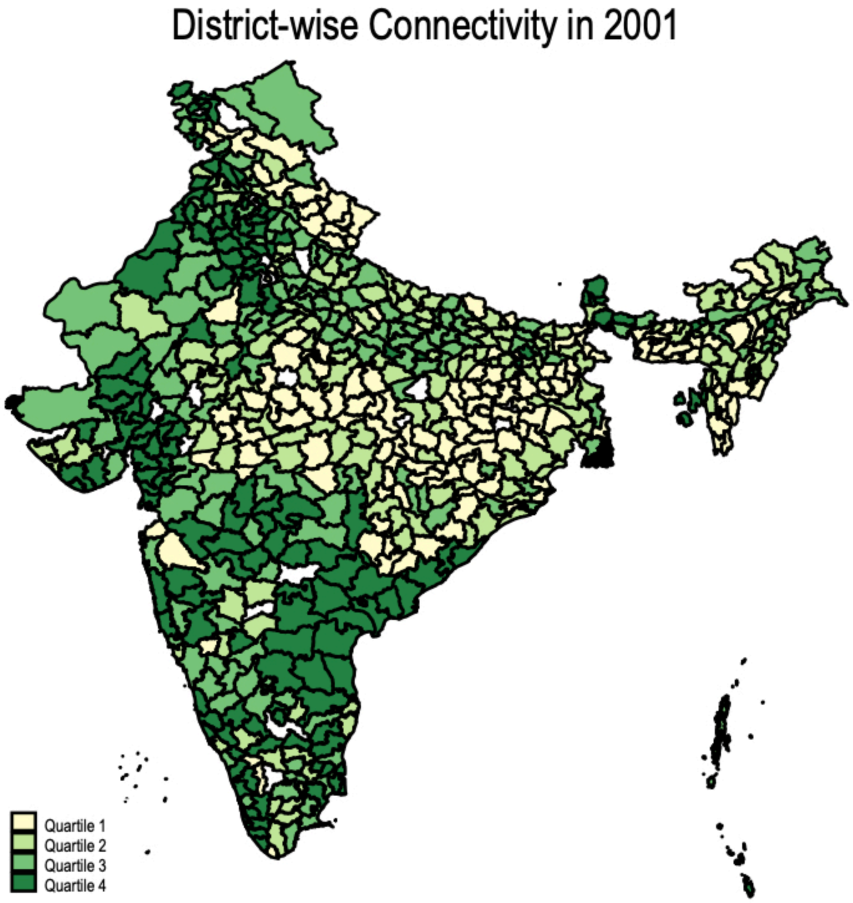

Figure 16: District-level Road Connectivity in 2001

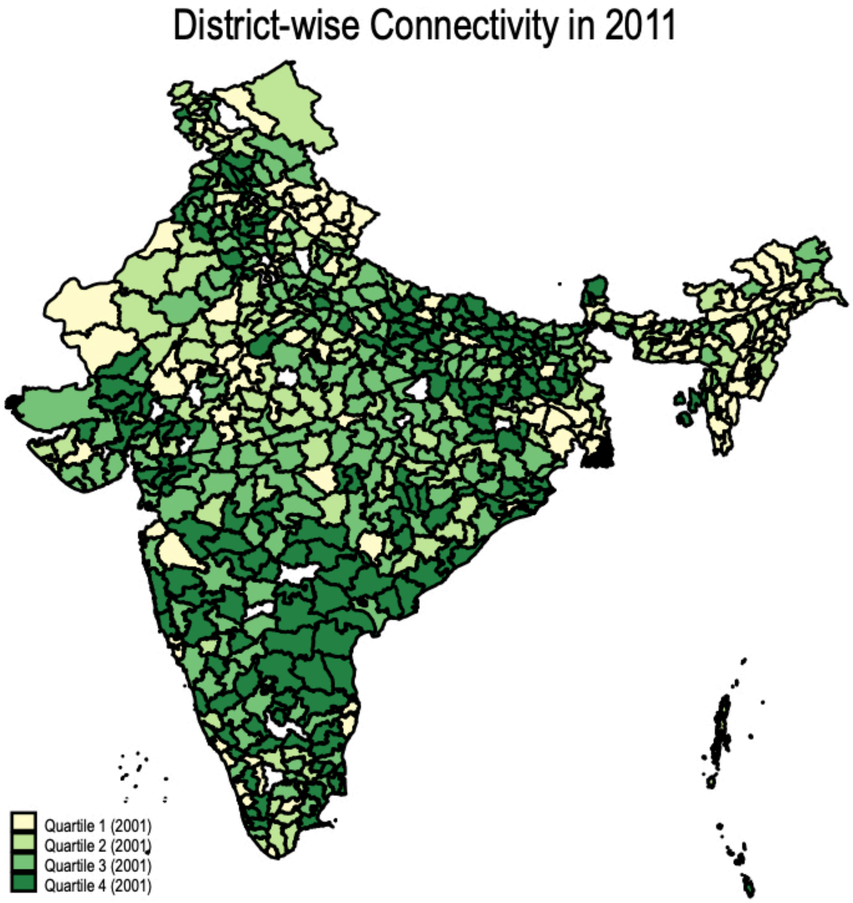

Figure 17: District-level Road Connectivity in 2011

4 IDENTIFICATION STRATEGY

So far, I have discussed the trends in crime and roads in India. Both crime and road connectivity increase over time. However, this increase is not uniform geographically all over the country. Are the two related? Simple visual or descriptive analysis cannot answer the research question of interest. I understand the relationship between roads and crime using econometric empirical strategies.

The National Crime Records Bureau (NCRB) collects data on criminal charges reported under different crime categories at the district level. [25] Since the smallest identifiable unit is the level of crime in each district of India, my empirical analysis can only be carried out at the district level. I thus aggregate the village level road connectivity variables to the district level to employ any identification techniques. Since the question of interest is to determine the impact of rural roads on crime, I only consider rural districts (districts that have 100% population living in villages) for my analysis.

4.1 Multiple Linear Regression

I first explore how roads, irrespective of whether they were allocated under the PMGSY, are related to crime. Thus, the independent variable of interest is connectivity at the district level, calculated as the fraction of the district's population living in a village that has access to a paved road, using the formula:

$$Connectivity_{(district,year)} = \frac{\Sigma(Population\ with\ Road)_{(village,district,year)}}{Population_{(district,year)}} \quad (1)$$

Since the Population Census of India is decadal, data on the presence of roads in villages (locations with 100% rural population) from the census of 2001 and 2011, aggregated to the district-level using equation (1), measures connectivity

[25] refer to the chapter on "Crime in India" for more details

in 2001 and 2011.

A simple linear regression estimates the relationship between connectivity and crime rate in district d, year t.

$$log(Crime\ Rate)_{dt} = \beta_0 + \beta_1 Connectivity_{dt} + \epsilon_{dt} \tag{2}$$

Figure 18 and 19 plot the scatter of rate of crime against women and total crime on connectivity respectively. The fitted values show a clear positive correlation between the two variables. This implies that increase in connectivity is correlated with higher incidence of crime.

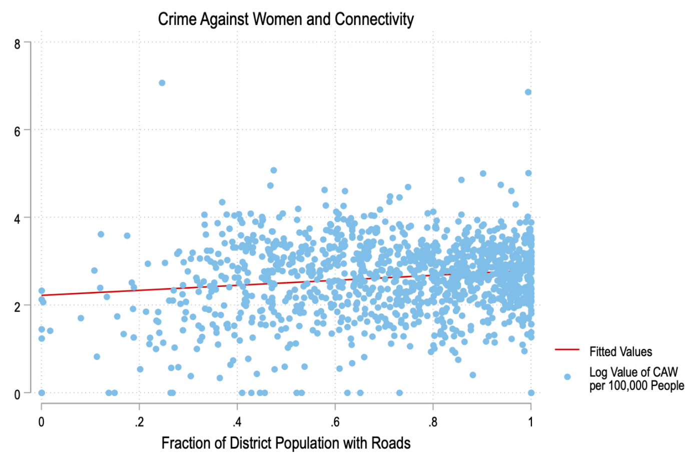

Figure 18: Log of Rate of Crime Against Women Versus Connectivity

4.1.1 Specification

The simple linear regression estimate is not precise. Influences on crime rate other than connectivity are captured by the error term. The estimate of β_1 in equation (2) is unbiased as long as the error term is uncorrelated with connectivity. However, in this case, it cannot be argued that the error term, ϵ_{dt} in equation

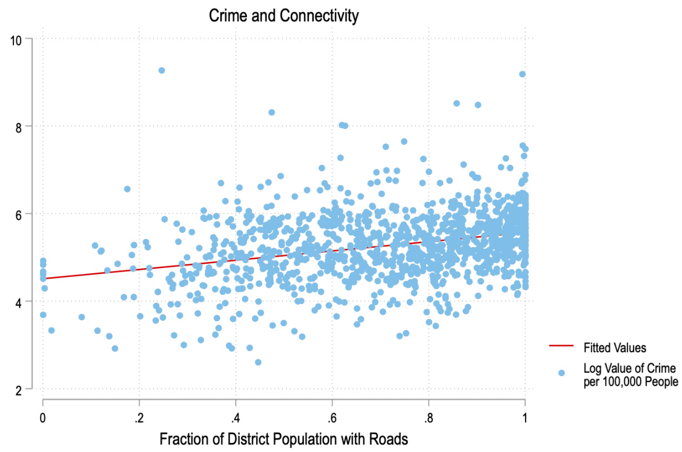

Figure 19: Log of Rate of Total Crime Versus Connectivity

(2) is not correlated with the independent variable $Connectivity_{dt}$. There can be missing variables that determine the rate of crime and are correlated with connectivity. In other words, there is a possibility of Omitted Variable Bias. Findings from studies conducted so far hint toward some such missing variables. Higher population density (population per square km) might increase connectivity and crime rate by increasing social interactions (Glaeser, Sacerdote, et al, 1996) [26]. Better connectivity improves school attendance and thus literacy of rural population (Adukia, Asher et al., 2017) which in turn impacts crime rate (Lochener and Moretti, 2004). Higher fraction of a village belonging to historically and culturally disadvantaged groups like the Scheduled Caste or Scheduled Tribe (SC/ST) increases the probability of getting roads through poverty alleviation programs (Government of India, 2012) and also affects crime (see section on 'Hate Crime' in the chapter on "Crime in India"). Power supply can increase the ease of constructing roads (the reverse might also be true, roads can increase power supply), while less crime occurs with improved supply

[26] In fact, higher connectivity can cluster more people, thus increasing population density

of electricity (women feel safer in areas with street lights). All of these factors generate omitted variable bias and need to be controlled.

I thus use a multiple linear regression to estimate how changes in road connectivity (not necessarily under the PMGSY road construction program) are related to crime by running the following regression equation at the district-level for years 2001 and 2011.

$$log(Crime\ Rate)_{dt} = \beta_0 + \beta_1 Connectivity_{dt} + \theta_t + \nu X_{dt} + \mu_s + \epsilon_{dt} \tag{3}$$

where the dependent variable is the log value of crime reports (crime against women or overall crime) per 100,000 people in district d, year t. Connectivity measures the fraction of the district d's rural population living in a village accessible via paved road in year t (Calculated using equation (1)). The coefficient of interest, β_1, estimates the impact of greater connectivity on crime. $X_{d,t}$ is a district level vector of district controls measured in the year t. θ_t and μ_s are year and state fixed effects. The district-level controls include the log of population density, the log of literacy rate (per 100,000 people), the log of share of inhabitants who belong to a scheduled caste or scheduled tribe, and the fraction district d's population with supply of electricity.

4.1.2 Results

Table 6 show how connectivity is related to crime against women. By controlling for time invariant factors within a state, the effect of connectivity decreases (see columns (1) and (2)). This indicates that more controls are required. Column (3) is a more precise specification since it includes all the controls with state fixed effects. The relationship now becomes insignificant.

Table 7 shows estimates of equation (2) with the log value of rate of total crime as the dependent variable. Coefficient on the variable 'Connectivity' shown in column (3) estimates that an increase in connectivity of a district from the 25th (55%) to the 75th percentile (91%) in the country, increases total crime by 13.68%. Notice that in both the tables 6 and 7 as controls are introduced, the coefficient estimate decreases in magnitude or shows a change in sign. This supports the argument of omitted variable bias in the specification without controls.

4.1.3 Limitations

Though a multiple linear regression indicates the existence of a potential relationship between crime and connectivity, it is not perfect. There still can be omitted variable bias in equation (3) because there is not enough data to include every possible control variable. Moreover, it does not necessarily imply that the relationship is causal. It could merely be a correlation. This method does not estimate the impact of the PMGSY because connectivity can be influenced by any other form of road construction (private companies constructing roads for transporting materials to and fro their factories as an example).

4.2 Difference-in-differences

Estimation of the causal impact of the PMGSY is challenging. The high cost of road construction under the program means that policymakers cannot randomly allocate only some regions to get the treatment. Thus, road placement is endogenous and is correlated with economic and political characteristics of locations (Blimpo, Harding, and Wantchekon 2013; Lehne, Shapiro, and Vanden Eynde 2018). I overcome this challenge by taking advantage of the timing of road placement under the policy to implement a Difference-in-Differences (DD) strategy.

Table 6: Multiple Linear Regression on Log of Rate of Crime Against Women

	(1)	(2)	(3)
Connectivity	0.549***	0.501***	-0.0329
	(0.109)	(0.113)	(0.123)
Year = 2011	0.124*	0.121***	0.0254
	(0.0490)	(0.0359)	(0.0415)
log(Population Density)			0.0199
			(0.0362)
log(Literacy)			0.459***
			(0.114)
log(SC/ST)			-0.132***
			(0.0367)
Electricity			0.419***
			(0.0985)
Constant	2.172***	1.591***	-2.013
	(0.0831)	(0.322)	(1.349)
State FE	No	Yes	Yes
Observations	1188	1188	1118
Adjusted R^2	0.027	0.480	0.457

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table follows specification (3).

$log(rate\ of\ crime\ against\ women)$ is the dependent variable.

Connectivity (fraction of population with a road) is the independent variable.

Population Density is population per sq km

Literacy is the number of literate population per 100,000 people.

SC/ST is the Scheduled Caste and Tribe population per 100,000 people.

Electricity is the fraction of district population with power supply.

Table 7: Multiple Linear Regression on Log of Rate of Total Crime

	(1)	(2)	(3)
Connectivity	1.057***	0.842***	0.380***
	(0.0950)	(0.103)	(0.112)
Year = 2011	0.00511	0.00242	-0.138***
	(0.0429)	(0.0326)	(0.0377)
log(Population Density)			0.0561
			(0.0328)
log(Literacy)			0.711***
			(0.103)
log(SC/ST)			-0.0758*
			(0.0334)
Electricity			0.377***
			(0.0894)
Constant	4.509***	4.246***	-2.767*
	(0.0727)	(0.292)	(1.225)
State FE	No	Yes	Yes
Observations	1188	1188	1118
Adjusted R^2	0.094	0.481	0.498

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table follows specification (3) with $log(crime\ rate)$ as dependent variable.

Population Density is population per sq km

Literacy is the number of literate population per 100,000 people.

SC/ST is the Scheduled Caste and Tribe population per 100,000 people.

Electricity is the fraction of district population with power supply.

4.2.1 Specification

I use the rounds of data (2001 and 2011) provided by the Population Census. Since road construction under the PMGSY started after 2001, 2001 is a pre-period. The first phase of road construction should have been finished by 2007, so 2011 is a post period. This pre and post period set up of the data can be used to estimate the causal relationship. I implement the following DD specification:

$$log(Crime\ Rate)_{d,t} = \beta_0 + \beta_1 Connectivity_{d,t} + \delta(Connectivity_{d,t} \times POST) + \theta POST + \nu X_{d,t} + \mu_s \tag{4}$$

where POST is an indicator variable equal to 1 in the year 2011 and the rest of the variables have the same definitions as equation 3. The dependent variable is the log value of crime (crime against women or total crime) per 100,000 people in district d, year t. The coefficient of interest is δ which estimates the change in crime rate in the post period as compared to the pre-period with a change in connectivity.

Since I also have access to the PMGSY road data, as another approach, I run a similar model outlined in equation (4) to estimate the differential results seen in areas with connectivity specifically provided by roads constructed under the PMGSY. This method essentially estimates the treatment effect on the treated. The specification using the PMGSY data becomes:

$$log(Crime\ Rate)_{dt} = \beta_0 + \beta_1 (Connectivity'_{dt}) + \delta_t (Connectivity'_{dt} \times Year_t) + \theta_t Year_t + \nu X_{dt} + \mu_s \tag{5}$$

where, $Connectivity'_{dt}$ is the fraction of population in district d connected by a road constructed under the PMGSY between years 2001 and t. The dependent variable is the log value of crime reports (crime against women or overall crime) per 100,000 people in district d, year t. The coefficients of interest are δ_t which estimate the change in crime rate in the year t as compared to the omitted year of 2001. All control variables are the same as in equation 4 except because of the multiple years, data for control variables from the population census of 2001 are used for years 2001 to 2006, and from the population census of 2011 for years 2007 to 2013.

It should be noted that since the data are aggregated to the district level, the DD design is not based on comparing treated villages with untreated villages, but on comparing districts with varying intensity of treatment under the PMGSY.

4.2.2 Assumptions

A DD design relies on the assumption that if no treatment had occurred, the difference between the treated and the untreated groups would have remained the same in the post-treatment period as it was in the pre-treatment period. This is the assumption of **Parallel Trends**. The pre-treatment period in the specifications under equation (4) and (5) is just one year i.e. 2001. Administrative data before 2001 are not openly available which makes checking for parallel trends difficult.

Another important assumption is that there should not be a coincident change, other than more roads, that affects crime in areas with higher connectivity. The specification includes necessary controls of population density, literacy, electricity, and share of marginalized communities to make this assumption hold true.

4.2.3 Results

Tables 8 and 9 show the estimates of the specification in equation 4 for rate of crime against women and rate of total crime as dependent variables. The estimate of δ has a negative values for crime against women (column (3), table 8), but a positive value for total crime (column (4), table 9). However these estimates are statistically insignificant, and hence provide no evidence of a consistent relationship.

Tables 10 and 11 show the results of the estimation based on equation 5. The estimates of the coefficients on the interaction terms for both crime against women and total crime are negative. Notice that the magnitudes are smaller by a tenth for crime against women as compared to total crime. Maybe, the effect of PMGSY roads was less powerful in influencing crime against women. These estimates are also statistically insignificant, thus indicating no clear relationship.

4.2.4 Limitations

The difference-in-difference design used in this paper has limitations because of the data available. The fundamental concern with a DD setup is that the trends before the treatment might not be parallel, thus invalidating the results. Since the program started in 2001, data on crime and connectivity before 2001 are required to estimate parallel trends. I do not have access to such data yet. Thus, the results might be biased based on the unknown trends before 2001. Moreover, the assumption that no other factor will affect crime in areas with high connectivity is not perfect. There might be variables which could not controlled for because of limited data that affect crime in areas with higher connectivity.

4.3 Exploring Instrumental Variable

To overcome the limitation of parallel trends in Difference-in-Differences, I explore the econometric technique of instrumental variables as a potential tool to estimate the effect of treatment on the treated using the PMGSY road data.

Table 8: Difference-in-Differences on Rate of Crime Against Women

	(1)	(2)	(3)
POST × Connectivity	-0.248	-0.226	-0.148
	(0.217)	(0.174)	(0.180)
POST	0.301	0.282*	0.135
	(0.163)	(0.129)	(0.140)
Connectivity	0.668***	0.613***	0.0406
	(0.151)	(0.142)	(0.152)
log(Population Density)			0.0202
			(0.0362)
log(Literacy)			0.453***
			(0.114)
log(SC/ST)			-0.132***
			(0.0367)
Electricity			0.412***
			(0.0988)
Constant	2.089***	1.513***	-1.993
	(0.110)	(0.328)	(1.349)
State FE	No	Yes	Yes
Observations	1188	1188	1118
Adjusted R^2	0.027	0.480	0.457

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table follows specification (4).

log(rate of crime against women) is the dependent variable.

POST is an indicator for years 2011, 2012, 2013

POST × Connectivity is the independent variable of interest.

Population Density is population per sq km

Literacy is the number of literate population per 100,000 people.

SC/ST is the SC/ST population per 100,000 people.

Electricity is the fraction of district population with power supply.

Table 9: Difference-in-Differences on Rate of Total Crime

	(1)	(2)	(3)
POST × Connectivity	-0.0960	0.0656	0.146
	(0.190)	(0.158)	(0.163)
POST	0.0736	-0.0444	-0.247
	(0.142)	(0.117)	(0.127)
Connectivity	1.103***	0.810***	0.307*
	(0.132)	(0.129)	(0.138)
log(Population Density)			0.0558
			(0.0329)
log(Literacy)			0.717***
			(0.103)
log(SC/ST)			-0.0757*
			(0.0334)
Electricity			0.384***
			(0.0898)
Constant	4.477***	4.268***	-2.787*
	(0.0965)	(0.297)	(1.225)
State FE	No	Yes	Yes
Observations	1188	1188	1118
Adjusted R^2	0.093	0.481	0.498

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Table follows specification (4).
$log(rate\ of\ total\ crime)$ is the dependent variable.
POST is an indicator for years 2011, 2012, 2013
POST × Connectivity is the independent variable of interest.
Population Density is population per sq km
Literacy is the number of literate population per 100,000 people.
SC/ST is the SC/ST population per 100,000 people.
Electricity is the fraction of district population with power supply.

Table 10: Difference-in-Differences on Rate of Crime Against Women (PMGSY)

	(1)	(2)	(3)
Year=2002 × Connectivity	-0.00166	-0.00235	-0.00185
	(0.00280)	(0.00273)	(0.00272)
Year=2003 × Connectivity	-0.00166	-0.00237	-0.00174
	(0.00275)	(0.00267)	(0.00266)
Year=2004 × Connectivity	-0.00154	-0.00240	-0.00177
	(0.00274)	(0.00266)	(0.00265)
Year=2005 × Connectivity	-0.00156	-0.00236	-0.00173
	(0.00273)	(0.00266)	(0.00264)
Year=2006 × Connectivity	-0.00149	-0.00232	-0.00167
	(0.00272)	(0.00265)	(0.00264)
Year=2007 × Connectivity	-0.00160	-0.00242	-0.00181
	(0.00272)	(0.00264)	(0.00263)
Year=2008 × Connectivity	-0.00154	-0.00230	-0.00169
	(0.00271)	(0.00264)	(0.00263)
Year=2009 × Connectivity	-0.00151	-0.00226	-0.00164
	(0.00271)	(0.00264)	(0.00262)
Year=2010 × Connectivity	-0.00138	-0.00212	-0.00149
	(0.00271)	(0.00264)	(0.00262)
Year=2011 × Connectivity	-0.00143	-0.00217	-0.00154
	(0.00271)	(0.00264)	(0.00262)
Year=2012 × Connectivity	-0.00137	-0.00214	-0.00152
	(0.00271)	(0.00264)	(0.00262)
Year=2013 × Connectivity	-0.00136	-0.00215	-0.00151
	(0.00271)	(0.00264)	(0.00262)
Connectivity	0.00155	0.00220	0.00150
	(0.00270)	(0.00263)	(0.00262)
log(Population Density)			0.0000803***
			(0.0000128)
log(Literacy)			-0.000134***
			(0.0000403)
log(SC/ST)			-0.0000846***
			(0.0000131)
Electricity			-0.0000403
			(0.0000344)
Constant	0.000181***	0.0000979	0.00220***
	(0.0000239)	(0.000109)	(0.000479)
State FE	No	Yes	Yes
Observations	7287	7287	7287

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Difference-in-Differences on Rate of Total Crime (PMGSY)

	(1)	(2)	(3)
Year=2002 × Connectivity	-0.0134	-0.0293	-0.0243
	(0.0263)	(0.0252)	(0.0250)
Year=2003 × Connectivity	-0.0145	-0.0325	-0.0260
	(0.0258)	(0.0247)	(0.0245)
Year=2004 × Connectivity	-0.0135	-0.0324	-0.0257
	(0.0257)	(0.0246)	(0.0244)
Year=2005 × Connectivity	-0.0140	-0.0317	-0.0247
	(0.0256)	(0.0245)	(0.0244)
Year=2006 × Connectivity	-0.0153	-0.0325	-0.0252
	(0.0256)	(0.0244)	(0.0243)
Year=2007 × Connectivity	-0.0163	-0.0328	-0.0248
	(0.0255)	(0.0244)	(0.0243)
Year=2008 × Connectivity	-0.0164	-0.0323	-0.0241
	(0.0255)	(0.0243)	(0.0242)
Year=2009 × Connectivity	-0.0165	-0.0322	-0.0240
	(0.0255)	(0.0243)	(0.0242)
Year=2010 × Connectivity	-0.0154	-0.0315	-0.0231
	(0.0254)	(0.0243)	(0.0242)
Year=2011 × Connectivity	-0.0160	-0.0321	-0.0236
	(0.0254)	(0.0243)	(0.0242)
Year=2012 × Connectivity	-0.0158	-0.0320	-0.0236
	(0.0254)	(0.0243)	(0.0242)
Year=2013 × Connectivity	-0.0159	-0.0321	-0.0236
	(0.0254)	(0.0243)	(0.0242)
Connectivity	0.0173	0.0318	0.0228
	(0.0254)	(0.0243)	(0.0241)
log(Population Density)			0.000839***
			(0.000118)
log(Literacy)			0.000686
			(0.000371)
log(SC/ST)			-0.000688***
			(0.000121)
Electricity			-0.000120
			(0.000317)
Constant	0.00274***	0.00203*	-0.000816
	(0.000224)	(0.00100)	(0.00442)
State FE	No	Yes	Yes
Observations	7287	7287	7287

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The road allocation criteria [27] provides quasi random variation which is used to construct an instrument at the district-level.

4.3.1 The Instrument and its Strength

To construct the instrument, I first make a variable (referred to as $Pop > 500$ in the rest of the paper) that measures the fraction of a district's population that lives in a village with a population of more than 500. The next step is to determine the years to consider as the treatment period. According to policy recommendations, all villages with more than 500 people should have been received a road by the end of the year 2007. Table 12 runs the following regression equation:

$$Connectivity_{d,t} = \beta_0 + \beta_1(Year_t) + \beta_2(Pop > 500)_d + \beta_3(Year_t \times (Pop > 500)_d) + \nu X_{d,t} + \mu_s \tag{6}$$

where 'Connectivity' is measured according to equation (1), except in this case, the roads are constructed specifically under the PMGSY. The controls X_{dt} are consistent with the previous specifications. β_3 shows the differential impact of the $Pop > 500$ variable on connectivity in years 2002 to 2013 as compared to the omitted year 2001. Notice that the differential impact starts to grow in magnitude and becomes significant in the years after 2007. I thus consider years from 2008 to 2013 as the post period.

The instrumental variable is the interaction between $Pop > 500$ and the POST period. To be a good instrument, the interaction term should fulfill some conditions. It should have a (i) causal effect on connectivity, (ii) should be relevant to connectivity, and should follow (iii) exclusion restriction. Since the policy guidelines specifically enforced road construction in villages with more

[27] villages with populations greater than a threshold (1000, 500, 250) get roads. More information in the chapter on roads

than 500 people, the instrument **causes** increase in roads and is thus **relevant**. Moreover, the instrument should influence crime **only through the channel** of more roads. This is the condition of **Exclusion Restriction**. Villages with a population of more than 500 can have more or less crime via channels other than more roads (for example, villages with higher population density have fewer resources per person, which encourages crime). I argue that such an influence will happen similarly in both the pre and the post years. Thus, the interaction term will not be affected by variables that change crime similarly in both the pre and the post periods [28].

4.3.2 Specification

To see whether it will be productive to move forward with this approach, I use the first stage and the reduced form estimates:
First Stage:

$$Connectivity_{d,t} = \gamma_0 + \gamma_1 POST_t + \gamma_2 (Pop > 500)_d + \gamma_3 POST_t \times (Pop > 500)_d + \nu X_{d,t} + \mu_s \tag{7}$$

Reduced Form:

$$log(Crime\ Rate_{d,t}) = \beta_0 + \beta_1 POST_t + \beta_2 (Pop > 500)_d + \beta_3 POST_t \times (Pop > 500)_d + \nu X_{d,t} + \mu_s \tag{8}$$

where Connectivity in district d, year t, is measured according to equation (1) with roads constructed under the PMGSY. POST comprises of years from 2008 to 2013 and the base (omitted) pre-period are years from 2001 to 2007. The controls $X_{d,t}$ are consistent with other specifications, except here there are

[28] Unless someone (condoned by our favorite ”Chowkidar”) decides to say start a riot in villages with more than 500 people, and that too only in the post period. In this case, I will be worried for the villages, and my instrumental variable.

multiple years. Data from the population census of 2001 provide control values from 2001 to 2006 and data from the population census of 2011 for years 2007 to 2013. μ_s are state fixed effects

The first stage estimates how the instrument influences connectivity and the reduced form estimates how the it influences the crime rate. If the instrument works, then the second stage measures the impact of roads.

4.3.3 Results

Table 13 shows the estimates for the first stage according to the specification in equation (7). According to column (3) of the table, a district that has 25% more population living in villages with 500+ people, is likely to have an increased connectivity of 3% in the post period as compared to the pre-period. This result has a strong statistical significance. The estimates from this table confirm the strength of the instrument.

Tables 14 and 15 show the reduced form estimates of equation (8) on logarithmic values of rate of crime against women and rate of total crime respectively. The estimate from column 3 of table 14 shows a very slight positive effect of $Pop > 500$ on crime against women, however this is not statistically significant. Column (3) of table 15 shows a greater magnitude in the negative direction of the coefficient on $Pop > 500$ variable. However, this is also not statistically significant.

4.3.4 Can we use Instrumental Variables?

Since the estimates from the reduced form hold no statistical significance, the instrument is not strong enough to estimate the impact. However, the strong first stage results hint toward the possibility of another approach, for example using multiple instruments.

Table 12: First Stage for $Pop > 500 \times Year$ as the Instrument

	(1)	(2)	(3)
Year=2002 × $Pop > 500$	0.0381	0.0383	0.0383
	(0.0401)	(0.0313)	(0.0313)
Year=2003 × $Pop > 500$	0.0683	0.0705*	0.0700*
	(0.0398)	(0.0311)	(0.0310)
Year=2004 × $Pop > 500$	0.0635	0.0654*	0.0650*
	(0.0398)	(0.0311)	(0.0310)
Year=2005 × $Pop > 500$	0.0730	0.0758*	0.0753*
	(0.0398)	(0.0311)	(0.0310)
Year=2006 × $Pop > 500$	0.0812*	0.0862**	0.0857**
	(0.0397)	(0.0310)	(0.0310)
Year=2007 × $Pop > 500$	0.0916*	0.0976**	0.0953**
	(0.0397)	(0.0310)	(0.0310)
Year=2008 × $Pop > 500$	0.120**	0.125***	0.123***
	(0.0397)	(0.0310)	(0.0310)
Year=2009 × $Pop > 500$	0.145***	0.151***	0.148***
	(0.0392)	(0.0307)	(0.0306)
Year=2010 × $Pop > 500$	0.191***	0.197***	0.194***
	(0.0392)	(0.0307)	(0.0306)
Year=2011 × $Pop > 500$	0.188***	0.194***	0.192***
	(0.0392)	(0.0307)	(0.0306)
Year=2012 × $Pop > 500$	0.196***	0.202***	0.199***
	(0.0392)	(0.0307)	(0.0306)
Year=2013 × $Pop > 500$	0.190***	0.195***	0.193***
	(0.0392)	(0.0307)	(0.0306)
$Pop > 500$	-0.00698	0.0202	0.0107
	(0.0283)	(0.0236)	(0.0239)
log(Literacy)			0.00462
			(0.00532)
log(SC/ST)			-0.00481**
			(0.00180)
log(Population Density)			0.00510***
			(0.00136)
Electricity			-0.0108*
			(0.00485)
Constant	0.00822	-0.117***	-0.122
	(0.0255)	(0.0258)	(0.0672)
Observations	7740	7740	7740
Adjusted R^2	0.401	0.635	0.636

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: First Stage for $Pop > 500 \times Year$ as the Instrument

	(1)	(2)	(3)
POST $\times$ $Pop > 500$	0.112***	0.115***	0.117***
	(0.0163)	(0.0132)	(0.0132)
$Pop > 500$	0.0527***	0.0823***	0.0690***
	(0.0112)	(0.0128)	(0.0134)
POST	0.0389**	0.0366**	0.0232
	(0.0147)	(0.0119)	(0.0121)
POST $\times$ $Pop > 500$	0.112***	0.115***	0.117***
	(0.0163)	(0.0132)	(0.0132)
log(Literacy)			0.0390***
			(0.00516)
log(SC/ST)			-0.00296
			(0.00197)
log(Population Density)			0.00476**
			(0.00149)
Electricity			-0.00831
			(0.00530)
Constant	-0.000763	-0.130***	-0.527***
	(0.0101)	(0.0198)	(0.0639)
State FE	No	Yes	Yes
Observations	7740	7740	7740
Adjusted R^2	0.325	0.557	0.561

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Reduced Form on $log(Rate\ Crime\ Against\ Women)$

	(1)	(2)	(3)
POST $\times$ $Pop > 500$	0.0000792	0.0000807	0.0000737
	(0.0000845)	(0.0000818)	(0.0000820)
$Pop > 500$	0.000149*	0.000198*	0.0000468
	(0.0000582)	(0.0000791)	(0.0000834)
POST	-0.0000411	-0.0000460	-0.0000140
	(0.0000760)	(0.0000737)	(0.0000749)
log(Literacy)			-0.0000807*
			(0.0000320)
log(SC/ST)			-0.0000721***
			(0.0000122)
log(Population Density)			0.0000407***
			(0.00000923)
Electricity			-0.0000270
			(0.0000329)
Constant	0.0000586	-0.0000435	0.00156***
	(0.0000525)	(0.000122)	(0.000396)
State FE	No	Yes	Yes
Observations	7740	7740	7740
Adjusted R^2	0.003	0.065	0.073

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Reduced Form on $log(Total\ Crime\ Rate)$

	(1)	(2)	(3)
POST × $Pop > 500$	0.00125	0.00123	0.00138
	(0.000796)	(0.000757)	(0.000758)
$Pop > 500$	0.00181***	0.00127	-0.000620
	(0.000549)	(0.000732)	(0.000771)
POST	-0.00104	-0.00108	-0.00134
	(0.000716)	(0.000682)	(0.000692)
log(Literacy)			0.000401
			(0.000296)
log(SC/ST)			-0.000710***
			(0.000113)
log(Population Density)			0.000443***
			(0.0000853)
Electricity			0.0000429
			(0.000304)
Constant	0.00112*	0.00114	0.00387
	(0.000494)	(0.00113)	(0.00366)
State FE	No	Yes	Yes
Observations	7740	7740	7740
Adjusted R^2	0.005	0.099	0.108

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.4 Summary

After implementing different strategies, I find that increasing road connectivity in general, including but not limited to PMGSY is correlated with an increase in the occurrence of total crime. The correlation does not exist for crime against women.

Roads constructed under the PMGSY have no statistically significant causal impact on crime according to estimates from administrative data. Though there is evidence of a potential relationship between other variables. Some important results are that higher literacy rates and higher fraction of historically and culturally marginalized population decreases both crime against women and total crime.

5 CONCLUSION AND DISCUSSION

Can infrastructural development pull developing countries out of poverty? What is the economic impact of improving infrastructure? Development Economists and policymakers have been asking whether large investments in constructing roads, railways, and other public transport systems is worth taxpayers' money. Along with this dilemma, developing countries face the challenge of high rates of crime — culturally sanctioned violence and crime against women being one of the most disturbing issues. Do women feel safer in places with better transport infrastructure? This paper attempts to shed light on one such important policy question: Can building more roads have an impact on cases of reported crime, especially crime against women?

The paper uses administrative data to find that reports of cases of crime increased as connectivity improved in India. Crime against women experienced no such effect. An increase in number of police stations and improved provision of education brought about by better connectivity might have led to better reporting of crime. On the contrary, the allocation of roads was not equally distributed across the country. Some districts received more roads than others within the same state. The increased inequality could have led to higher incidence of crime. There is no significant relationship observed between roads provided by the PMGSY and crime. Since the level of analysis is a district in India, the impact of road allocation to villages under the program might not have aggregated to the district-level.

The analysis contributes to the literature of crime and road patterns in India, by using data and data visualization. The identification approach showcases how different econometric techniques can be used to answer the same research question. Importantly, the study makes official crime records openly accessible for any statistical purposes.

There are limitations of the results. The administrative data used for the analysis has notable discrepancies. Previous studies that found a significant relationship between crime and road connectivity in India used data available from surveys conducted by private bodies. [29]. The results I find are not consistent with such studies.

Thus, the paper raises questions on the validity of official data published by the government. Assessing the quality of administrative data in developing countries, especially in India is an understudied component of state capacity. While research has been done on the misreporting of crime in official records provided by the National Crime Records Bureau of India, similar studies on the Population Census — data popularly used by development economists and policy makers — do not exist.

Future research work in the field entails analyzing the authenticity of important official data like the Population Census. Furthermore, the results from this research could be made robust by getting access to more data, especially for years prior to the launch of the PMGSY (years before 2001). Such robust research can have important policy implications as development economists and policymakers seek to find ways to increase both infrastructural integrity and public safety in developing countries.

[29] Jain and Biswas (2021) used data from the Indian Human Development Survey (IHDS) These data are collected by scholars from the University of Maryland, the National Council of Applied Economic Research (NCAER), Indiana University, and the University of Michigan.

6 APPENDIX

6.1 Directed Acyclic Graph

Figure 20 shows the different channels that can influence crime and roads under the PMGSY. This is useful to understand the identification challenges.

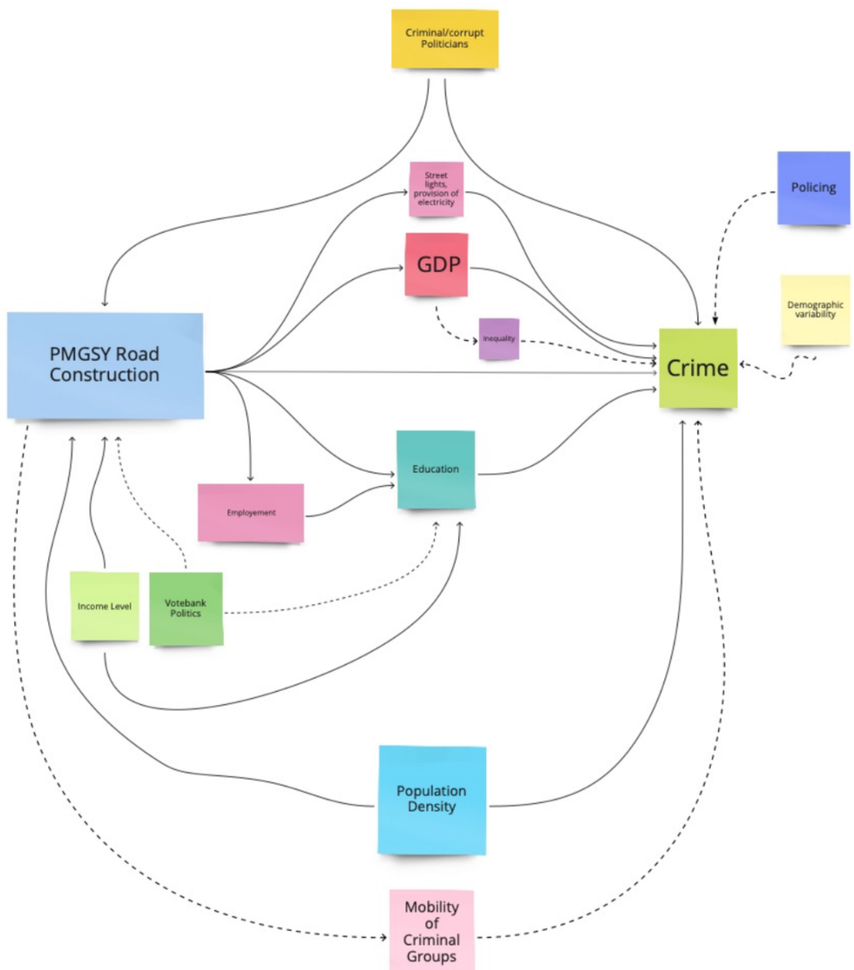

Figure 20: Directed Acyclic Graph Between PMGSY and Crime

6.2 Extracting Crime Data

The National Crime Records Bureau (NCRB) of India publishes annual reports called Crime in India (CI) that contain year wise crime records from previous year publications. The PDF files of annual reported Indian Penal Code (IPC) crimes are available on the official NCRB website. The number of reported criminal cases are available for 825 districts spread over 35 States and Union Territories of India from years 2001 to 2013.

I extracted the district-wise data on incidence of cognizable crimes in police custody from 2001 to 2013 by scraping the NCRB PDF files using Python. I used the `tabula` and `os` packages of Python to enable the conversion from PDF to CSV. To do this, I made a directory of all the PDF files and then looped over them to convert each to CSV using iteration (as shown in Listing 1).

Once I got the 13 .csv versions (years 2001 to 2013) of all the PDF files, I imported them to Stata, and appended them together after converting them to .dta files. The final appended file contains 9385 observations for years 2001 to 2013.

```python
import tabula
import os

os.chdir("./District-wise IPC")
for file in os.listdir("./"):
    print(file)
    if file.endswith(".pdf"):
        convertToFile = file[:-3]+"csv"
        tabula.convert_into(file, convertToFile,
                            output_format="csv", pages="all")
```

Listing 1: PDF to CSV Python Scraping Code

6.3 Merging Crime With Census Data

The crime data have levels of crime at the district level. These figures will be more meaningful if expressed in rates instead of levels. I use population statistics from census to calculate crime rate per 100,000 people. The Indian Population Census happens in every 10 years. For our purposes, the census is available for years 2001 and 2011. Since the crime data and census data are collected by independent bodies, merging them was a challenge.

The names of districts in both crime data and population census data are English transliteration of names of districts in Hindi or the state's regional language. Different English spellings can be pronounced to the same name in a regional language, there were mismatches in the spellings of same districts in the two data sets. Since no one in my knowledge has previously attempted to merge crime with census data at a large scale, there weren't any unique IDs available as identifiers for districts.

I manually constructed a data set with 850 district names matching to a common district name. This key will be available for public use. With a common set of districts to match to, I merged the two data sets keeping the common district names as the unifying column.

WORKS CITED

Adukia, A. (2020). Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction. *American Economic Association*, 348-76.

Ansari, S., Verma, A., & Dadkhah, K. (2015). Crime Rates in India: A Trend Analysis. *International Criminal Justics Review*.

Asher, G., & Garg, T. N. (2020). The Ecological Impact of Transportation Infrastructure. *The Economic Journal*, 1173-1199.

Asher, S., & Novosad, P. (2017). Rural Roads and Structural Transformation.

Asher, S., & Novosad, P. (2020). Rural Roads and Local Economic Development. *American Economic Association*, 797-823.

Asher, S., Lunt, T., Matsuura, R., & Novosad, P. (2021). Development Research at High Geographic Resolution: An Analysis of Night-Lights, Firms, and Poverty in India Using the SHRUG Open Data Platform. *The World Bank Economic Review*, 845-871.

Asher, S., Nadpal, K., & Novosad, P. (2017). The Cost of Distance: Geography and Governance in Rural India.

Bharadwaj, A. (2014). Is Poverty the Mother of Crime? Empirical evidence of the impact of socioeconomic factors on crime in Inida. *Atlantic Review of Economics*, 2174-3835.

Bhattacharyya, A., Haldar, S., & Banerjee, S. (2021). Determinants of Crime Against Women in India: A Spatial Panel Data Regression Analysis. *Millennial Asia*.

Blimpo, M. P., Harding, R., & Wantchekon, L. (2013). Public Investment in Rural Infrastructure: Some Political Economy Considerations. *Journal of African Economies*, 57-83.

Bloch, F., & Rao, V. (2002). Terror as a Bargaining Instrument: A Case Study of Dowry Violence in Rural India. *American Economic Review*, 1029-1043.

Das, S., & Mocan, N. (2016). Analyzing the Impact of the World's Largest Public Works Project on Crime. *NBER Working Paper*, 22499.

Das, U. (2018). Trends in Growth of Roads Development Programmes in India Since 1943. *International Journal of Social Science and Economic Research*.

Dasgupta, A., & Kapur, D. (2017). (When) Do Antipoverty Programs Reduce Violence? India's Rural Employment Guarantee and Maoist Conflict. *Cambridge University Press*, 606-632.

Debnath, A., & Roy, N. (2013). Linkage Between Internal Migration and Crime: Evidence from India. *International Journal of Law, Crime and Justice*, 203-212.

Deshpande, M. D. (2010). History of the Indian Caste System and Its Impact on India Today.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 899-934.

Dreze, J., & Khera, R. (2000). Crime, Gender, and Society in India: Insights from Homicide Data. *Population and Development Review*, 335-352.
Driver, E. D. (1961). Interaction and Criminal Homicide in India. *Social Forces*, 153-158.
Dublish, P., & Khan, Y. (2022). Impact of Honour Killings in Haryana, India. *SSRN*.
Duflo, E., Banerjee, A., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*.
Farooque, A. K. (1973). Roads and Communications in Mughal India.
Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and Social Interactions. *The Quarterly Journal of Economics*, 507–548.
Green, D. P., McFalls, L. H., & Smith, J. K. (2001). Hate Crime: An Emergent Research Agenda. *Annual Review of Sociology*, 479-504.
Jain, R., & Biswas, S. (2021). The road to safety- Examining the nexus between road infrastructure and crime in rural India.
Khanna, G., & Zimmermann, L. (2016). Guns and Butter? Fighting Violence with the Promise of Development. *SSRN*.
Lenhe, J., Shapiro, J., & Eynde, O. (2017). Building connections: Political corruption and road construction in India. *Journal of Development Economics*, 62-78.
Lochner, L. (2010). Education Policy and Crime. *NBER Working Paper*, 15894.
Lochner, L., & Moretti, E. (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review*, 155-189.
Medhi, A. (2017). Rural Roads in India: An Overview of Connectivity Status. *South Asian Journal of Multidisciplinary Studies*.
Mukherjee, C., Rustagi, P., & Krishnaji, N. (2001). Crimes against Women in India: Analysis of Official Statistics. *Economic and Political Weekly*, 4070-4080.
Musa, Sainabou. (2012). Dowry-Murders in India: The Law & Its Role in the Continuance of the Wife Burning Phenomenon. *Mitchell Willimas Law Firm*.
Palasinski, M., Zhang, L., Ingale, S., & Hanlon, C. (2017). Gangs in Asia: China and India. *Asian Social Science*.
Patel, M. (2014). A Desire to Disfigure: Acid Attack in India. *International Journal of Criminology and Sociological Theory*, 1-11.
S.K., K., & Justo. (2009). *Highway Engineering.* Roorkee, India: Nem Chand and Brothers .
Sekhri, S., & Storeygard, A. (2014). Dowry deaths: Response to weather variability in India. *Journal of Development Economics*, 212-223.
Sharma, S. (2013). Hate Crimes in India: An Economic Analysis of Violence and Atrocities against Scheduled Castes and Scheduled Tribes. *American Economic Review*, J15, K42.
Sikdar, P. (2000). Rural Road Network Planning using GIS.
Singh, M. (2019). Religion in India: Religious Composition of Population and Religious Regions. *Research Journal ARTS*, 63-80.

Straub, S., Vellutini, C., & Warlters, M. (2008). Infrastructure and Economic Growth in East Asia. *Policy Research Working Paper, World Bank*, No. 4589.

Thomas, P. K. (1984). *Rural Roads and Poverty Alleviation in India.* Routledge.

Upadhyay, S. P., & Robinson, R. (2012). Revisiting Communalism and Fundamentalism in India. *Economic and Political Weekly*, 35-37.

Vincente, G., Goicoa, P., Fernandez-Rasines, M., & D., U. (2019). Crime against women in India: unveiling spatial patterns and temporal trends of dowry deaths in the districts of Uttar Pradesh. *Royal Statistical Society*.

Y., F., & K., M. (2014). Improvement of rural access roads in developing countries with initiative for self-reliance of communities,. *Soils and Foundations*.