Infrastructure, Institutions, and the Conservation of Biodiversity in India

Raahil Madhok*
University of Minnesota

January 22, 2025

Abstract

Anthropogenic land use change is the leading threat to biodiversity. This paper studies how infrastructure expansion degrades biodiversity and the role of local institutions in mitigating species loss. Combining new data from India on infrastructure-driven deforestation with one million birdwatching diaries, I document a sizeable infrastructure-biodiversity tradeoff. Forest encroachment by transport, irrigation, resettlement camps, and mining projects account for 20% of total species loss. The tradeoff is especially acute in already-fragmented landscapes, and species diversity does not recover in the medium run. Yet the extent of species loss is more than halved when local institutions enable marginalized communities, who are often excluded from project planning, to mobilize around their interests. Informed consent between developers and tribal communities is a key mechanism, underscoring the importance of inclusive institutions for balancing development and conservation.

Keywords: sustainable development, economic development, infrastructure, biodiversity, conservation, institutions, political economy.
JEL Codes: Q01, Q56, Q57, Q20, O13.

---

*University of Minnesota, Department of Applied Economics. Email: rmadhok@umn.edu. I thank the Editor and two anonymous referees for helpful feedback. I also thank Sumeet Gulati, Patrick Baylis, Frederik Noack, Siwan Anderson, Eyal Frank, Ryan Abman, Teevrat Garg, Meredith Fowlie, Sonja Kolstoe, Bianca Cecato, and Tatiana Zarate, as well as seminar audiences at the World Bank, NBER Summer Institute, Vancouver School of Economics, Columbia IPWSD, CREEA, AERE, WEAI, PacDev, and Camp Resources for valuable comments. Veena Vinod provided excellent research assistance. All errors are my own.

1 Introduction

Global infrastructure spending totalled $USD 2.3 trillion in 2015 (Oxford Economics, 2017). Although crucial for economic growth, infrastructure expansion narrows the frontier between human activity and fragile ecosystems. The ecological threat from encroachment is especially acute in the tropics, home to two-thirds of Earth's biodiversity yet where over 60% of global infrastructure spending occurs (FAO and UNEP, 2020). This is exacerbated by the fact that millions of indigenous people—who have supported biodiversity for millennia—are displaced by, disaffected by, or excluded from project planning.

Economists have long sought how to reduce environmental costs of development (Grossman and Krueger, 1995; Dasgupta et al., 2002; Copeland and Taylor, 2004). Biodiversity receives little attention in this literature (Frank and Schlenker, 2016), let alone grassroots solutions for balancing development and conservation. Filling this gap thus requires not only estimates of the ecological threat from infrastructure, but also the role of local institutions for neutralizing it.

My first goal is to provide a deeper understanding of the extent to which infrastructure expansion drives biodiversity loss. I call this the infrastructure-biodiversity tradeoff. The second goal is to investigate the role of decentralized forest governance in mitigating the tradeoff. Better understanding these socio-ecological and institutional processes can assist countries in meeting the dual objectives of development and conservation.

The broad setting is the tropics, where over half of global deforestation occurs (Pacheco et al., 2021). India notably avoided widespread forest loss despite recording rapid economic growth (Forest Survey of India, 2019). It is unclear whether this was due to concerted tree-planting or changing definitions of forest cover. Even if development did leave forests unscathed, important inhabiting species may still become threatened and require policy attention. Elusive measurement of such species has led to biodiversity being overlooked in previous studies (Foster and Rosenzweig, 2003; Burgess et al., 2012).

The first part of this paper estimates the infrastructure-biodiversity tradeoff in India's forests between 2015-2020. This constitutes a valuable setting for three reasons. First, India is among the planet's most biodiverse countries, home to 8% of global biodiversity and 12% of bird diversity (Venkataraman and Sivaperuman, 2018; Jayadevan et al., 2016). Second, India's biodiversity is documented by active "citizen scientists" who upload sightings on species-specific (e.g. eBird) or general (e.g. iNaturalist) platforms. India boasts the highest eBird membership of any developing country, with their geocoded uploads serving as a new, high-resolution biodiversity repository unmatched in the literature. Third, India publicly reports forest encroachments by infrastructure. Deforestation for building roads, mines and other projects now account for 17% of yearly forest loss (authors calculation). The Forest Act (1980) mandates environmental review of such projects before construction. The review process underwent a transparency initiative in 2014, unlocking new administrative data for estimating threats to biodiversity.

To measure infrastructure, I digitize the universe of deforestation permits awarded to firms that passed environmental review. This includes 7,000 scraped from a public portal and 2,000 digitized by hand. Each one describes a forest patch diverted for construction and uniquely bundles infrastructure and deforestation into a single variable. For analysis, permits are aggregated into a cumulative measure of district-monthly forest area diverted for development. This new data improves on satellite measures because the latter overlooks the source of deforestation. Pixel values are also annual aggregates, which masks deforestation throughout the year. In contrast, my data directly measures infrastructure-driven deforestation and features monthly landscape changes as projects roll out. I do, however, use satellite data to verify that *approved* projects trigger *actual* deforestation.

To measure biodiversity, I obtain one million geocoded birdwatching diaries from eBird, the world's largest crowd-sourced platform for wildlife sightings (Sullivan et al., 2009). Birds are a credible indicator for ecosystem health, sensitive to environmental change, and documented with high precision (Morrison, 1986; Fraixedas et al., 2020). Each diary reflects a birdwatching session (i.e., a "trip") and lists the date, GPS coordinates, and a taxonomy of species sightings. I count the number of species in each diary, yielding a biodiversity dataset with unparalleled spatiotemporal resolution, spanning 95% of districts from the Himalayas to the Western Ghats.

The matched panel enables a two-way fixed effects (TWFE) design to estimate the impact of infrastructure on bird species diversity (hereafter, species diversity) in a typical Indian district. I decompose estimates by project category to show which types of infrastructure are the least and most harmful. I also stratify districts by baseline forest cover to reveal whether projects have different effects in pristine or already-fragmented habitats.

Despite the promise of citizen science, its opportunistic nature yields more sample selection than typical administrative data. eBirders tend to visit more biodiverse locations, especially in the Western Ghats. There is also a Siberian bird migration in winter, and a lull in birdwatching activity during monsoons, which induces seasonality. Lastly, users possess varying abilities and learning rates, complicating inference from cross-user comparisons. I employ district fixed effects to address site choice, state-month fixed effects to address seasonality, and user-by-year fixed effects to purge ability and learning biases.

Endogenous sorting of birds, birdwatchers, and projects is the main threats to identification, even with the fixed effects. If construction pushes birds into less-fragmented districts, the control group becomes contaminated. Similarly, if users sort towards biodiverse districts, then estimates are upward biased. I address both issues with spatial lags. Species immigration into a district appears uncorrelated with nearby development. The number of district users also appears unchanged when nearby districts become developed. While these tests help rule out endogenous bird and birdwatcher behaviour, omitted variables bias from strategic project siting remains a key identification concern.

The main analysis yields four key findings. First, infrastructure development triggers substantial species loss. Ten km^2 of infrastructure encroachments reduce species diversity by 4%, as observed by the average eBird user. In contrast, the portion of these projects falling on *non-forest* land has no impact on species diversity, suggesting that habitat loss is a key mechanism. In aggregate, approximately 20% of the observed decline in species diversity over the study period can be attributed to development in India's forests.

Second, nearly all project categories drive the infrastructure-biodiversity tradeoff. The top three most harmful are resettlement, transport, and irrigation projects. Resettlements are akin to camps for relocating displaced communities. The negative impact of mining is surprisingly small, which I show is due to low eBird activity in mining districts. The mining impact doubles when the sample is restricted to higher-activity districts.

Third, some species are more threatened by infrastructure development than others. I manually match bird taxonomies with their IUCN Red List status and physiological traits, and then count the number of times users observe species in each category. Poisson estimates show a sharp decline in the abundance of common and vulnerable species following infrastructure expansion. Forest birds and non-migrant species are also especially sensitive, whereas wetland birds are unaffected.

Lastly, I find that species are more resilient to infrastructure development in intact forests. Heterogeneity by baseline forest cover shows that the infrastructure-biodiversity tradeoff is halved in districts with one standard deviation higher initial forest cover. This evidence therefore supports earmarking degraded landscapes for protection.

Despite a variety of robustness checks, parallel trends, and no evidence of endogenous sorting, causal interpretability of estimates may remain in question since infrastructure is non-random. Yet the infrastructure-biodiversity tradeoff is also apparent under an instrumental variables (IV) design based on close elections. Since winner identity in close elections is essentially a coin toss, I use the fraction of district constituencies where an incumbent just barely won in close elections as an instrument for infrastructure. The second stage, once again, shows that forest encroachment prompts species decline. One concern is that the exclusion restriction assumption—that incumbents influence local ecology only through sanctioning forest diversion for infrastructure—is quite strong. Another is that close-election estimates do not generalize to non-competitive districts. I thus view this approach as a validation of coefficient sign rather than a second set of main estimates.

The second part of the paper studies which institutions minimize biodiversity loss. India is home to 200 million members of forest-dependent tribes, who have stewarded biodiversity on traditional forests for millennia. Today, they are among the country's most economically vulnerable, politically excluded, and face livelihood loss as forests are handed over to commercial interests. I study whether inclusive institutions that emphasize decentralized decision-making can mitigate the infrastructure-biodiversity tradeoff.

Data are from Banerjee and Iyer (2005) and indicate whether district institutions favour elites (extractive) or are more inclusive of the masses. The measure is based on whether historic tax collection was via a middleman. Banerjee and Iyer (2005) find that non-middleman areas feature higher equality today and better ability of the disenfranchised to mobilize. If tribal groups can better protect their livelihoods—which hinges on protecting forests—in inclusive districts, then better conservation outcomes are expected there.

The infrastructure-biodiversity tradeoff estimated in the first part of the paper is significantly smaller in inclusive districts. Implied magnitudes are large; the tradeoff is 78% smaller in these districts, where disaffected groups can better engage in the development process. Results are independent of tribal population share, suggesting that heterogeneity reflects institutional differences, not population differences. These results underscore the importance of inclusive forest governance in achieving sustainable development.

The paper concludes by probing mechanisms through which inclusive institutions mitigate the infrastructure-biodiversity tradeoff. I extract unique data from project permits reporting whether tribes were consulted and whether supplemental cost-benefit analyses were commissioned during project review. I find that projects approved in inclusive districts are associated with significantly higher rates of informed consent and environmental scrutiny. These results indicate that community participation in project planning, along with higher environmental standards, are key features of inclusive institutions that balance development and conservation.

Literature Contributions This paper contributes to three literatures. My main contribution is to provide the first country-wide evidence that infrastructure expansion triggers local species loss. Most economics studies that quantify infrastructure externalities estimate pollution costs (Currie et al., 2015; Hanna and Oliva, 2015). A handful have estimated forest loss: Asher et al. (2020) and Garg and Shenoy (2021) find surprisingly little effect of infrastructure on forest cover in India, and Baehr et al. (2021) also find muted effects in Cambodia. While this suggests that ecosystems are resilient to infrastructure[1], my results indicate otherwise using detailed species-level data.

The most similar paper is Liang et al. (2021), who study GDP and biodiversity in the United States[2]. In contrast to GDP, which subsumes underlying mechanisms, my data captures infrastructure development at the forest frontier. Despite the differences, our results are consistent: development drives species loss.

The second contribution of this paper is to extend the ecology literature by expanding the spatiotemporal scope of data and integrating empirical techniques from economics. In ecology, field workers often count species in small transects with different levels of human activity in one time period. Although citizen science dramatically improves spatial and temporal coverage, much interest from ecologists has been in characterizing its endogeneity rather than using it for causal inference (Callaghan et al., 2019; Kelling et al., 2019). I advance this literature by quasi-experimentally estimating the infrastructure-biodiversity tradeoff across a large developing nation over six years.

The third contribution is to extend research at the intersection of political economy and conservation. A seminal literature shows how historic institutions shape modern economic development (Nunn, 2009), yet few have considered biodiversity outcomes[3]. In contrast, the conservation literature acknowledges that institutions can moderate economy-environment tradeoffs, yet few have tested the claim (Börner et al., 2020). I advance this literature by credibly estimating of the role of institutions in reducing species loss. In doing so, I am also able to pin down mechanisms. Duflo and Pande (2007) use the same institutional data to show that the negative poverty impact of dams is muted in inclusive districts, arguing that the poor can better access compensation in these districts. Lee (2019) confirm that inclusive districts have better state capacity. My paper ties together this literature by providing evidence of a mechanism with "teeth": inclusive institutions feature higher rates of informed consent between authorities and minorities.

The next section provides background on infrastructure-driven deforestation in India. Section 3 introduces the construction permit and citizen science data. Section 4 documents stylized facts from the data and Section 5 outlines the research design. Section 6 presents estimates of the impact of infrastructure expansion on biodiversity. Section 7 explores the role of institutions for mitigating the tradeoff. Section 8 concludes.

[1]Kaczan (2020) find that road building in India reduces tree cover in remote areas and increases tree cover in peri-urban areas. The net effect may explain the small effects found in previous studies.

[2]Related studies include Liang et al. (2020), which studies effects of pollution on bird abundance Noack et al. (2021), and Noack et al. (2019), which studies the impact of farm size on bird diversity.

[3]Prior work has studied institutions and water conservation (Libecap, 2011; Hagerty, 2021) and forest conservation (Börner et al., 2020; Lal et al., 2021). Tsuda et al. (2023) study place-based policy and resource depletion. Noack et al. (2021) show that land institutions predict farm size and bird diversity.

2 Background

Forest Act (1980) Protects Forests India's Forest (Conservation) Act (1980) protects forests from "conversion to non-forest uses" (MoEFCC, 1980). Infrastructure is the main regulated activity because it fragments important habitats. When non-forest sites are unfeasible, the Act permits infrastructure encroachments pending a rigorous environmental review. It also sets up a forest advisory committee (FAC) of government officials and forestry experts to rule on construction proposals. Projects involving any amount of deforestation, on any land recorded as forest irrespective of ownership, undergo review.

Despite the Act's intent, huge swathes of India's forests have been transferred to public and private firms. Between 1985-2014, about 4,000 km^2 of forest were clearcut for the construction of 23,000 infrastructure projects. Total deforestation during this period was 24,223 km^2 (Meiyappan et al., 2017)[4], implying that infrastructure intrusions accounted for 17% of India's deforestation during the three decades preceding this study.

Informed Consent Since 2006 Infrastructure encroachment is especially harmful to India's Scheduled Tribes. Numbering at 200 million, tribes are forest-dependent custodians of biodiversity, yet have been historically excluded from forest-related development decisions. In 2006, the landmark Forest Rights Act (FRA) granted tribes forest management rights, including the right to informed consent with developers. As shown in Section 7, enforcement in unequal across districts. I exploit this variation to study whether infrastructure is more sustainable under inclusive institutions.

Project Approval Process The journey of a project proposal is known as the Forest Clearance (FC) process (MoEFCC, 2003). There are two stages: stage-I approval is granted after rigorous environmental review (see Appendix S3.1 for details). To receive stage-II approval, firms pay a state-specific amount into a tree-planting fund. Despite potential to offset my findings, the fund is fraught with issues and tree-planting is rarely carried out. A recent audit finds that just 7% of land secured for afforestation between 2006-12 had been planted in 2013 (MoEFCC, 2013). Other studies have found no existence of plantations during field visits (World Rainforest Movement, 2019). Sanctioned afforestation thus poses minimal threat to my research design.

Another stage-II requirement is for the District Forest Office to submit evidence of tribal informed consent to the central Ministry of Environment. After fundraising and

[4]Forest loss $\approx$ 18,000 km^2 from 1985-'05 (Meiyappan et al., 2017), and 6223km^2 from '06-14 (Forest Watch)

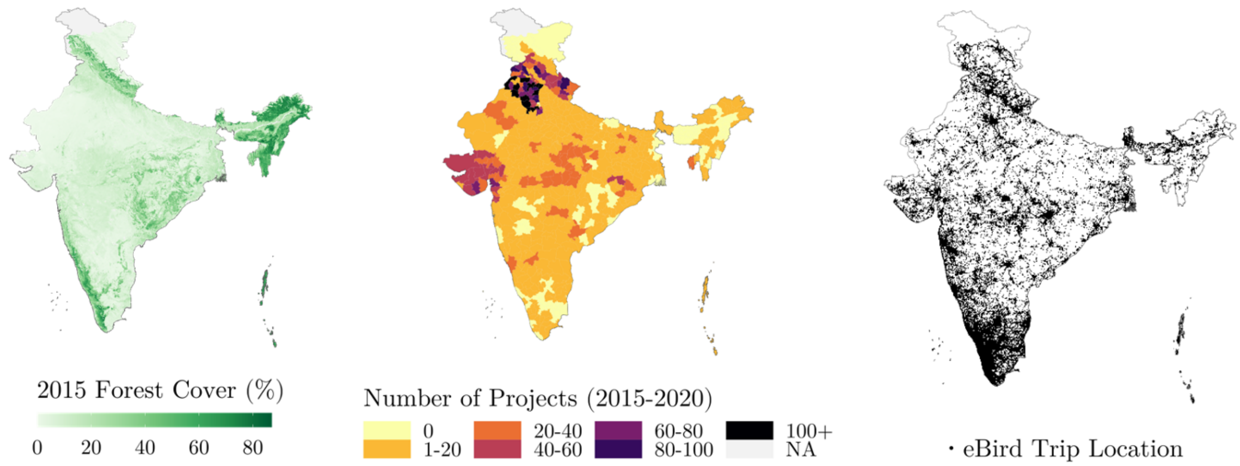

Figure 1: Infrastructure Encroachments and eBird Activity

Note: Panel A is a map of 2015 forest cover (Townshend et al., 2017). Pixels are shaded by percent forest cover. Panel B maps the number of infrastructure projects approved for construction between 2015-2020. Panel C shows GPS coordinates of all birdwatching trips recorded through eBird during the study period.

FRA compliance, stage-II approval is awarded and the firm begins deforestation. In Section 4.1, I verify that stage-II approvals trigger actual deforestation using satellite data.

Project Timelines Projects take many years to complete after permit approval. 60% of Indian projects experience time overruns, with dispute settlement between owners and contractors being the main cause of delay (Salunkhe and Patil, 2014; Prasad et al., 2019). Disputes take around 6.5 years to resolve (Construction Federation of India, 2015). With a six-year study period, and other delays likely occurring, it is unlikely that projects will complete construction during the study period. Non-completion means that my estimates will capture species loss from habitat loss during construction, not broader general equilibrium effects once the project is operational.

3 Data

I estimate the infrastructure-biodiversity tradeoff by drawing on several new datasets. I use newly digitized FC permits to measure development in India's forests. Species diversity is from eBird, a popular e-notebook for birdwatchers. The final panel covers all of India from 2015 to 2020. This section describes the data and provides summary statistics.

Table 1: Summary Statistics of Forest Infrastructure Projects (2015-2020)

	Num. Projects	Mean Size (ha.)	SD (ha.)	Total Area (ha.)
Electricity	882	27.5	228.4	24,274.5
Irrigation	430	57.5	252.5	24,731.1
Mining	229	148.2	253.7	33,927.6
Other	4,448	2.4	34.5	10,486.4
Resettlement	44	71.5	92.9	3,147.2
Transportation	2,296	10.9	32.5	24,986.0
Total	8,329	14.6	110.6	121,552.9

Note: Data are at the project level. Total area is the summed deforestation area of all projects in a category.

3.1 Forest Clearances

Data Collection Administrative data on infrastructure-driven deforestation rarely exist, and previous work mainly relies on satellite data. Yet satellites have difficulty distinguishing anthropogenic intrusions from natural sources (e.g. forest fires). They also report annual data, which mask within-year encroachments and their ecosystem impacts.

I construct a dataset of monthly infrastructure encroachments using newly digitized stage-II FC proposals approved between 2015-2020. Proposals submitted after the review process went digital in 2014, and approved during the study period (N=6,597), were scraped from the online portal (the digital subsample) [5]. Another 1,732 submitted pre-digitization, but approved during the study period, were digitized by hand (the manual subsample). These 8,329 projects comprise the universe of industrial forest encroachments in India during the study period. Figure S1 shows an example approval letter authorizing 185 ha. of deforestation for an irrigation project in Rajasthan. Figure 1B shows the full spatial distribution: projects encroach into both sparse and dense forests (Figure 1B), with dense forests in the North suffering the most encroachment.

Variables and Summary Statistics Both the digital and manual subsamples report deforestation and project category (road, mine, etc.). District-wise deforestation is provided for multi-district projects (e.g. transmission lines). Digital applications additionally report non-forest land diversion, ownership (public, private, neither), and shape (linear, nonlinear)[6]. Digital applications also report whether a cost-benefit analysis was commissioned and whether informed consent was obtained, enabling analysis of how institutions mediate ecosystem impacts (Section 7). Appendix S3.2 provides more data details.

[5]Data are publicly available at www.parivesh.nic.in/

[6]Linear projects are contiguous in terms of land (roads). Nonlinear projects are non-contiguous (mines)

The 8,329 projects collectively triggered 122,000 ha. of deforestation between 2015-2020 (Table 1). The average encroachment is 14 ha., roughly 20 soccer pitches. Mines and resettlements are few in number but massive in size. Mines account for 3% of projects but 30% of deforestation. Resettlements are least common but second to mines in size. Contrastingly, "other" projects (not in the listed categories) are most common, but reflect small patches (see Appendix S3.2)[7]. Transportation is the only category that is both numerous and accounts for a large (20%) share of total deforestation. Appendix S3.3 provides summary statistics by project ownership and shape.

Panel Data Structure Project data are aggregated to the district and year-month level, both overall and by category (e.g. deforestation in January 2018 for electricity projects in Delhi). I do this because the district is the only consistent location identifier. Districts are also one administrative unit below the state and form a natural unit for local policy implementation. The panel is balanced by zero-filling project approvals in districts not in the full sample (Figure 1B). This is reasonable since all projects undergo the FC process, and the full sample contains the universe of approvals.

3.2 eBird

eBird entered India in late-2014 and only requires a smartphone. Each birdwatching session (hereafter, "trip") is GPS-tracked and includes a taxonomy of species sightings called a checklist. Checklists are vetted by ornithologists on each upload (Sullivan et al., 2009). eBird is revolutionary for research because it documents both species observations and the observation process. The latter includes: trip date, duration, protocol (e.g. stationary or travelling), and whether all observed species were recorded, called a complete checklist. These data help identify checklists best reflecting the local species pool.

Sample Selection My sample frame is the eBird Basic Dataset (eBird Basic Dataset, 2019) for 2015-2020, comprising all trips during this period. I follow the eBird manual (Strimas-Mackey et al., 2020) to identify representative checklists which, here, means checklists best reflecting local species diversity at the user's location. To do this, the manual suggests keeping complete, stationary and travelling protocols (99% of sample) as well as lists collected in $<$ 5hrs and with group size ≤ 10. Next, I link trip coordinates to 2011 district borders, which provides a matching key and reveals off-coast boating trips, which are dropped. This leaves 1,049,930 trips by 17,634 users across 628 districts (out of

[7] The most common "other" projects are approach roads (driveways) and fibre optic lines.

640). Figure 1C plots the trips: areas with high forest cover and development activity are popular. Users are also active in South and Central India. Activity is low in the Southeast, despite moderate forest cover, likely due to the remoteness of this region.

Outcomes and Aggregation Despite having 1 million trips, the final panel aggregates to the user-district-month level for consistency with the unit of analysis in the infrastructure data. Aggregation reduces the effective sample to 161,896. The main outcome is mean species richness—the number of unique species observed on a trip—across each user's trips in a district-year-month. Richness proxies the number and stability of ecosystem services and is a widely used biodiversity metric (Fleishman et al., 2006). I also manually code each species with their IUCN Red List status[8] (low concern, vulnerable, etc.), habitat type (forest, wetland, etc.), and other traits[9] (Appendix S4.4). This enables me to study which species are most sensitive to infrastructure development.

One concern is that aggregating over trips masks within-district sorting. I test for this in Section 5.3.3 and find no evidence that users sort within districts after project approval.

Data Representativity Over 1,600 trips by 100 users are recorded in the average district (Table S1). Users are quite active, recording in four districts, two states, and six year-months during the study period. This within-user variation is the cornerstone of my empirical strategy (Section 5). About 23 species are recorded on the average trip, while traveling a wide area. eBirders cover 20% of district area in a typical month and over half of district area over the study period. Wide spatial coverage helps ensure that eBird data reflects the local species pool[10].

Overall, my sample selection procedure yields a replicable biodiversity data product that is a contribution in itself, especially for empirical environmental economists. Data on observer effort enables researchers to separate species observations from the observation process. Trip coordinates enable aggregation of species diversity to any spatial unit. Lastly, several ground-truthing studies show that eBird data strongly correlates with structured bird census data (Horns et al., 2018; Munson et al., 2010; Callaghan et al., 2018), suggesting that eBird is a reasonable guide to local biodiversity. Section 4 presents stylized facts about the analysis sample and how it can be used for causal inference.

[8] Dataset of taxa names of IUCN status obtained from: `https://datazone.birdlife.org/home`.

[9] Physiological details of Indian birds obtained from State of India's Birds database.

[10] It is more important that eBirders collect representative data than themselves be representative of the population. Nevertheless, in the absence of demographic data, I provide a detailed characterization of users in Appendix S5 by matching their approximate home locations to the Demographic and Health (DHS) survey. Perhaps unsurprisingly, users are from more urban and better-off places than the typical Indian.

3.3 Covariates

Environmental Covariates The first set of covariates are environmental and include temperature and rainfall. Accounting for weather is important since it affects species detection. Temperature (°C) is from the ERA5 product on a 0.125° grid (Hoffmann et al., 2019). Rainfall (mm) is from the NASA GPM Level 3 product on a 0.1° grid (Huffman et al., 2019). I extract means over cells within a district, weighted by cell overlap fraction.

Observer Effort The second set of covariates captures effort and includes: trip distance, duration, experience, protocol, group size, and spatial coverage. Duration (minutes) and distance (km) are recorded by eBird. Experience is the cumulative number of trips recorded by each user at the end of every month. Protocol equals one if the user is moving and zero if stationary. Group size is the birdwatching party size. Spatial coverage is the fraction of 10km grid cells in a district traversed by users. This accounts for projects opening inaccessible forest patches (e.g., through supply roads), which may draw users to new sites. It also enables me to characterize representativity of eBird species lists (Fact 4, Section 4). All effort variables are aggregated to their means during aggregation and directly included as control variables in the main specification[11].

Economic Spillovers The third set of covariates captures broader economic activity induced by project-building. This helps disentangle the effect of infrastructure per se from changes to market structure prompted by the projects. Such spillovers should be minimal in any case since projects are unlikely to complete construction during the study period (Section 2). Even otherwise, market spillovers "help" the research design as they reflect alternate channels threatening species diversity, including noise and air pollution.

In the absence of district GDP data, I control for nightlight radiance to capture broader economic activity (Henderson et al., 2012). Data are obtained from the VIIRS satellite (Elvidge et al., 2017). Note that nighlights are a "bad control" if projects themselves affect nightlights, in which case both variables partially subsume the treatment. To check this, Table S2 regresses log nightlights on project permits at the district-year level and reveals near-zero point estimates and no correlation even up to two years later. This suggests that nightlights can be used to control for broader economic changes. I also control for state fixed effects in case projects affect connected industries in the state (Section 5.1).

[11]Spatial coverage and experience are not transformed since they are already at the user-district-yearmonth, and user-yearmonth level, respectively. Since protocol is a trip-level indicator, it transforms into the proportion of stationary trips after data aggregation.

Table 2: Correlation between approved and actual deforestation

	Log Forest Cover	Log Forest Cover	Forest Share
	(1)	(2)	(3)
Log(Infrastructure+1)	-0.024**	-0.024**	
	(0.010)	(0.010)	
Infrastructure Encroachment (share of forest)			-0.365***
			(0.086)
Nightlights	No	Yes	Yes
District FEs	✓	✓	✓
State × Year FEs	✓	✓	✓
Observations	3822	3822	3822
R^2	0.988	0.988	0.993

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Data are aggregated district-yearly. In columns 1 and 2, the outcome is log forest cover (km^2) and the explanatory variable is cumulative approved deforestation (km^2). A value of one is added before log-transforming infrastructure to account for zero values. In column 3, the outcome is forest cover divided by district land area, and the explanatory variable is share of forest encroached by infrastructure. Standard errors clustered by district.

4 Empirical Patterns

I next present four stylized facts that make the data ideal for studying the infrastructure-biodiversity tradeoff. The first verifies that project approvals trigger real deforestation. The second and third illustrate shortcomings and remedies for using citizen science for causal inference. The fourth fact is that users are very mobile, providing spatial variation for identification. These facts motivate the empirical strategy in Section 5.

4.1 Fact 1: Approved deforestation triggers actual deforestation

Throughout the paper, authorized deforestation is assumed to mean actual deforestation. I test this using remotely-sensed validation data with the following equation:

$$ForestCover_{dst} = \alpha + \beta \cdot Infrastructure_{dst} + \Gamma X'_{dst} + \gamma_d + \theta_{st} + \epsilon_{dst} \quad (1)$$

where d indexes districts and t indexes years. $Infrastructure$ is cumulative km^2 of approved deforestation and $ForestCover$ is actual forest cover obtained from the VCF satellite (Townshend et al., 2017). To harmonize the scale, I convert forest cover percentages to km^2[12]. X'_{dst} is a control for nightlight intensity, which disentangles infrastructure from

[12] I convert to km^2 by multiplying cell values (% forest cover) by pixel area and summing over districts.

other drivers of forest loss. γ_d and θ_{st} are district and state-year fixed effects. $\beta < 0$ tests whether approved deforestation translates into actual deforestation.

Forest cover declines as districts approve more projects (Table 2). Column 1 specifies a log scale since $Infrastructure_{dst}$ aggregates over *differently sized* projects in a district[13]. The point estimate implies that a 1% increase in approved deforestation leads to a 0.02% decline in observed forest cover, and is robust to controlling for nightlights (column 2). Column 3 specifies a linear relationship using forest share of land area as the outcome and share of forest encroached by infrastructure as the explanatory variable. The coefficient remains negative and statistically significant. The point estimate implies that 1pp. more approved deforestation leads to a 0.37pp. decline in forest cover. All three point estimates are reasonable since infrastructure is only one source of total deforestation.

4.2 Fact 2: eBird usage is higher in winter and in more biodiverse places

Although citizen science data is revolutionizing biodiversity monitoring, loose restrictions on when, where, and by whom data are collected yields more endogeneity than most administrative datasets. First, there is stark seasonality arising from the ability to record trips at any time. Figure 2A shows sharp peaks in collective species richness (left axis) in winter when Siberian birds migrate to India, and a trough during lulls in activity (right axis) during monsoons. I address seasonality with state-month fixed effects so that time-invariant differences across months, such as seasonal species fluctuations, are eliminated. This also assumes migratory patterns vary regionally. In a robustness check, I test whether seasonality materializes at the sub-state level.

A second source of endogeneity in citizen science data is site selection. The ability to record species from anywhere can trigger heightened activity in certain districts. Figure 2B shows that eBird users record more trips in districts with higher "true" species diversity, measured by intersecting historic bird range maps (BirdLife International, 2018). Species checklists will be longer in these districts since the species pool is larger. Previous studies account for this with a variety of fixed landscape covariates (Kelling et al., 2015), whereas I use district fixed effects to rule out site selection more confidently.

[13]Small encroachments are unlikely to be captured by the satellite, introducing noise at low levels of $Infrastructure$. Also, if small projects are sited in highly forested areas (and overlooked by the satellite), then β is upward biased. A log scale helps avoid these problems since values represent *exponential* changes in approved deforestation, which are more likely to be captured by satellite.

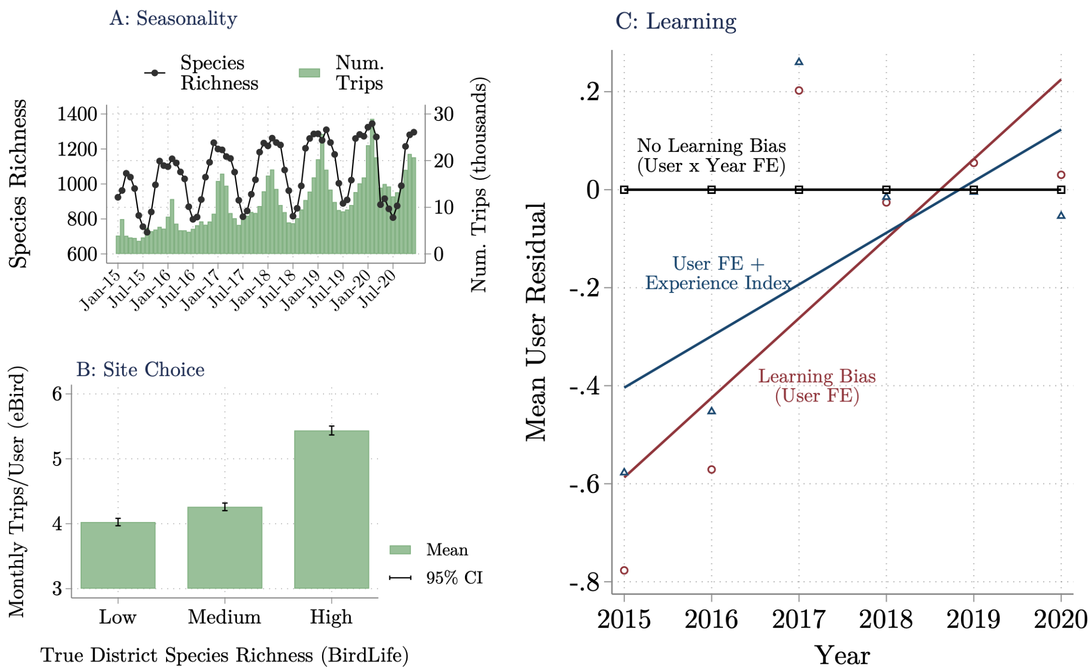

Figure 2: Biases in Citizen Science

Note: The left y-axis of Panel A shows total species richness across all users. The right y-axis shows total number of trips. Panel B shows mean number of trips per user-month in three quantiles of *true* species richness, obtained from historic range maps. In Panel C, red circles plot mean residuals per user from regressing species richness on user, district, state-month, and year fixed effects. Blue triangles control for experience. Black squares partial out user-year, district, and state-month fixed effects.

4.3 Fact 3: Learning is a crucial source of bias in citizen science

Besides seasonality and site selection, another bias arises from pooling users with different abilities (Fitzpatrick et al., 2009). To address this, previous studies construct a fixed ability score for eBird users based on random effects (Kelling et al., 2015). Instead, I compare species richness *within the same user*, making an ability score superfluous.

Circles in Figure 2C show species richness residualized on user, state-month, and district fixed effects. An upward trend remains, evidence that users' ability may improve over time. Triangles add a control for experience, which increments with each trip. The "learning curve" flattens, but is not fully absorbed. This suggests that learning is also driven by longer-term unobservables. For example, a novice may detect the same common species month-to-month, gradually listing rarer species after learning their songs.

My solution hinges on restricting variation to within-user-by-year. This removes user-specific annual trends, including accumulated trips, months per year of activity, and other long-term learning indicators. It also allows for differential learning curves between users. Partialling out user-by-year fixed effects mechanically flattens the trend (black line), implying that residual variation is stripped of the learning bias. This is the variation that I use to estimate infrastructure-biodiversity tradeoff in Section 5.

4.4 Fact 4: Users are highly mobile across space and time

One concern with high-dimensional fixed effects is that they absorb a lot of identifying variation. District and state-month fixed effects leave monthly within-state deviations from district means, e.g. the amount by which a district in Kerala is more species diverse than normal in a month compared to its neighbour. User fixed effects remove additional variation by restricting district comparisons to those traversed by individual users. Therefore, identification hinges on users being sufficiently mobile. Table S1 showed that users visit multiple districts and states over the study period. Figure S2 plots spatial variation within the year—the same variation used in the main analysis. About 30% of users visit multiple states and districts, and over 40% are are active in multiple months of the year.

Table S3 presents the identifying variation more formally. It summarizes regressions of species richness on different fixed effects and reports residual variation (column 1) and the standard deviation of residual variation (column 2). One-fifth of the variation in species richness is explained by seasonality and site choice (second row). About half is explained when user heterogeneity and learning are also accounted for (third and fourth row). Overall, substantial identifying variation remains—driven by users traveling across space and time—after removing important biases in citizen science data. The residual standard deviation is 12-13 species in the most saturated specifications, providing a wide margin for identification. These findings underscore the richness of crowd-sourced data.

5 Empirical Strategy

The main analysis leverages panel fixed effects. Development projects fragment district forests, and eBird users venture to these districts to record birds. I compare species diversity *within* users' trips as they travel around. This strategy exploits rich cross-sectional and time-series variation, giving rise to plausible control groups within users: their recorded species diversity in districts and time periods with different levels of encroachment.

5.1 Main Specification

I estimate the following equation to reveal the infrastructure-biodiversity tradeoff:

$$SR_{idsym} = \alpha + \beta_1 \cdot Infrastructure_{dsym} + \Gamma X'_{idsym} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idsym} \quad (2)$$

where SR_{idsym} is mean species richness observed by user i across their trips in district d of state s during year y and month m. As such, SR_{idsym} is a continuous measure representing an average of underlying trip-level species counts. I also estimate specifications where the outcome is the *frequency* that user i observes common, vulnerable, and endangered species[14]. $Infrastructure_{dsym}$ is cumulative area of infrastructure encroachments. X_{idsym} is a vector of weather and effort covariates described in Section 3. It includes spatial coverage, which accounts for projects opening up inaccessible parts of the forest. It also includes nightlights, which controls for market spillovers in the district. User-year fixed effects, ϕ_{iy}, absorb cross-user differences in ability and account for individual-specific learning. District fixed effects, γ_d, absorb mean checklist length and other other fixed district attributes. State-by-month fixed effects, θ_{sm}, control for state-specific seasonality.

Identifying Variation ϕ_{iy} require that user i visit at least one district in two months of the year to qualify for the estimation sample. In this case, only time variation is exploited. Users active for just one month are also included as long as they visit > 2 districts, in which case identification is from cross-sectional comparisons. In general, users are more active than these limiting cases (Figure S2) and both temporal and spatial variation across user i's trips to different districts over months of the year are used for identification.

Conditional on covariates and fixed effects, β_1 identifies the impact of infrastructure on species diversity. It captures the impact of infrastructure-driven forest loss, not general equilibrium effects of infrastructure, because nightlights are a covariate and because projects are unlikely to complete construction during the study period. If species relocate within the district, then they may be spotted by user i on another trip or by other users, leaving species diversity unchanged[15]. $\beta_1 < 0$ is thus even more striking as it implies the species and its ecosystem services are displaced from the district altogether.

OLS Estimator β_1 is estimated in a TWFE setup via OLS. Callaway et al. (2024) outline three key assumptions for OLS to recover causal treatment effects in this setting. First,

[14] Poisson regression is used for estimation in specifications where the outcome is a count variable.

[15] The fact that 10 users are active in the typical district-month, together covering 20% of district area (Table S1), helps ensure that the local species pool is reported, even if one user misses a species.

treatment must be continuous, which fits my setting since $Infrastructure_{dsym}$ is measured in km^2. Second, units must not anticipate treatment, an assumption supported below with an event study (Figure 3). The third assumption is strong parallel trends: the marginal change in species richness among users visiting a district with a given infrastructure level is the change that all districts would experience if they had that amount of infrastructure. Otherwise, the OLS estimate contains selection bias. While this is fundamentally untestable, I lean on the fact that user fixed effects exploit within-user comparisons across districts with different treatment intensities. In Section 5.3.3, I demonstrate that users' location choices are uncorrelated with infrastructure, which minimizes selection bias since the treatment is plausibly as-if randomly distributed across potential outcomes.

Clustering Standard errors are clustered by biome in the main analysis. From an ecological view, this is the most appropriate cluster because biomes delineate biological communities with shared eco-climatic characteristics that are unobserved in my model. These characteristics may generate arbitrary correlation of ϵ_{idsym} within a biome. Maps of India's 12 biomes are obtained from the Nature Conservancy[16] (Figure S3). For districts spanned by many biomes, I select the one with the largest overlapping area as the cluster.

From an econometric view, clustering by district is more appropriate since deforestation varies at the district level. Although unobserved ecological components of biodiversity are unlikely to adhere to political boundaries, I report estimates clustered by district in the robustness checks. I also cluster by state and report Conley standard errors as a compromise between biome and district clustering.

5.2 Additional Specifications

I decompose $Infrastructure_{dysm}$ in Equation 2 into six separate categories: electricity, transportation, mining, resettlement, irrigation, and other. This not only reveals particularly harmful categories, but also those with negligible, or even positive, impacts on species diversity. I estimate the following specification:

$$SR_{idsym} = \alpha + \sum_{k=1}^{6} \beta_{1k} \cdot Infrastructure_{kdsym} + \Gamma X'_{idsym} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idsym} \tag{3}$$

where the term under summation is cumulative forest area diverted for projects of category k. Remaining terms and subscripts are defined as in Equation 2. β_{1k} measures

[16] I use the "Terrestrial Ecoregion" files accessed from https://worldmap.maps.arcgis.com

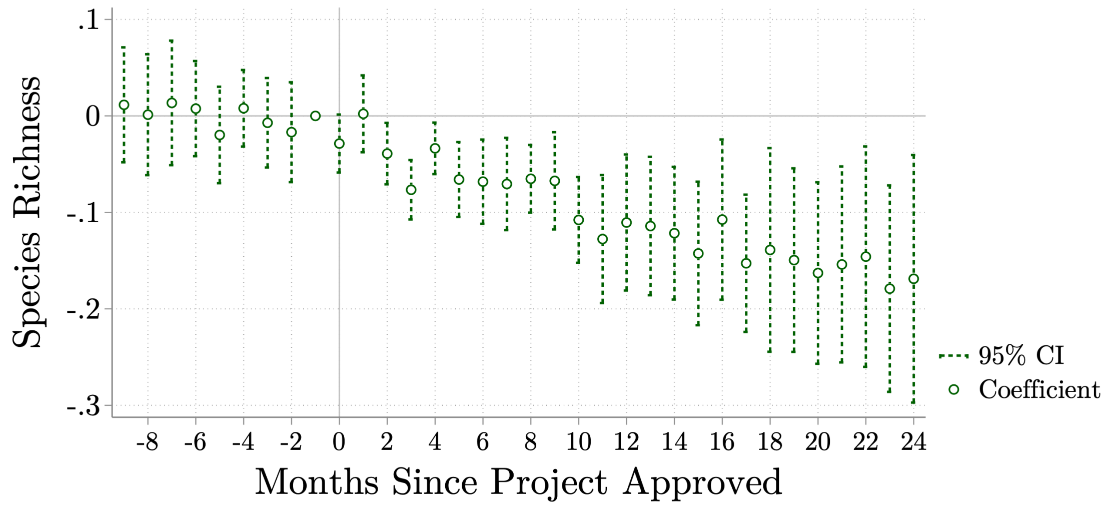

Figure 3: Event Study Results

Note: Dark green circles are event study coefficients (Equation 5). The outcome is log of mean species richness observed by users in a district-month. x-axis is number of months since project approval. The regression includes project, state-month, and year fixed effects, linear district time trends, as well as district-level controls for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log group size, and log spatial coverage. Standard errors clustered by biome.

the impact of infrastructure category k on species richness. I use the same approach to estimate impacts by project ownership (public, private) and shape (linear/nonlinear).

Additionally, I estimate heterogeneity by initial forest cover as follows:

$$\begin{aligned} SR_{idsym} &= \alpha + \beta_1 \cdot Infrastructure_{dsym} + \beta_2 \cdot (Infrastructure_{dsym} \cdot EQ_d) \\ &\quad + \sum_{k=1}^{6} \beta_{3k}(Infrastructure_{dsym} \cdot Share_{kd}) + \Gamma X'_{idsym} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idsym} \end{aligned} \tag{4}$$

where EQ_d is a fixed measure of initial ecosystem quality in district d. It is measured with 2014 forest cover and, in a robustness check, with bird diversity from range maps (Section 4.2). Since certain project categories may dominate particular landscapes (e.g., mines in intact forests), I control for the interaction of infrastructure with the baseline share of projects in category k to disentangle area effects from category effects. β_2 reveals whether the infrastructure-biodiversity tradeoff is accentuated or muted in more pristine landscapes, independent of project type.

5.3 Identifying Assumptions

5.3.1 Assumption I: Parallel Trends

The parallel trends assumption states that, in the absence of treatment, species diversity in districts that received an additional km^2 of projects must be on the same outcome path as districts that never received the increment. Since treatment is continuous and comprises multiple project approvals (events) in a district, I unravel the data into a stacked project-district-month panel and estimate the following event study:

$$Log(SR_{dsym}) = \sum_{\tau=-9}^{24} \beta_{1\tau} \cdot \mathbb{1}[t - e_{pdsym} = \tau] + \Gamma X'_{dsym} + y \cdot \mu_d + \alpha_p + \theta_{sm} + \eta_y + \epsilon_{pdsym} \quad (5)$$

where p indexes projects, d indexes districts, s indexes states, and ym indexes year-months. The outcome is mean species richness observed by users in a district-year-month. e_{pdsym} is the date that project p was approved in district d. $\tau = -1$ is the reference period. X'_{dsym} are the same covariates from Equation 2 at the district level, and $y \cdot \mu_d$ are linear district time trends. Each $\beta_{1\tau}$ captures mean species richness τ months relative to the project approval date. Lack of pre-trends are indicated by $\beta_{1\tau} = 0 \ \forall \, \tau < 0$. Standard errors are clustered by biome for the same reason as the main specification (Section 5.1).

Figure 3 displays estimates of Equation 5. Coefficients fluctuate tightly around zero during the pre-period. Yet two months after approval, coefficients turn sharply negative and continue downward. The lack of pre-trends support the parallel trends assumption and suggest that projects are not selectively sited. It also supports the no-anticipation assumption, which requires that users do not alter their behaviour prior to the treatment.

While parallel trends help rule out endogenous project placement, omitted variables bias remains a threat to validity since the research design relies on fixed effects for identification. Appendix S4.1 provides additional evidence to minimize this concern; neither bureaucratic nor geographic characteristics consistently predict the likelihood or timing of project approval, building confidence that the timing of permit allocation is plausibly random (Table S16). Appendix S4.2 shows that infrastructure placement does not appear to be influenced by lagged bird diversity (Table S17), helping rule out reverse causality.

5.3.2 Assumption II: No Spatial Spillovers

β_1 in Equation 2 is unbiased assuming no interference between units, known as the stable-unit treatment value assumption (SUTVA) (Imbens and Rubin, 2015). This requires that species richness in district d depend on infrastructure in district d only. However, SUTVA is violated in my context since habitat loss triggers species dispersal to other districts.

To account for spatial spillovers, I add a control for the "spatial lag of X" (Elhorst and Vega, 2015), $SLX^B_{dsym} = \sum_{j \in \Omega_B} \frac{Infrastructure_{jsym}}{distance_{dj}}$, where $distance_{dj}$ is the distance between district d and j. This term measures spatially-weighted development in other districts $j \neq d$, and has two important features. First, inverse distance weights assume that birds relocate more toward nearby districts. Second, birds are assumed to disperse within the biome, B, denoted by Ω_B, the set of districts j in the same biome as d. I also test robustness to spillovers within arbitrary, potentially biome-spanning, circles of radii up to 500km.

After including SLX^B_{dsym} in Equation 2, its coefficient captures changes to species diversity in district d when other districts in the biome become relatively more fragmented. Conditional on this, β_1 is purged of spillover bias.

5.3.3 Assumption III: No Sorting of eBird Users Across or Within Districts

Another threat to identification in Equations 2-4 is endogenous user sorting. If project development incentivizes users to birdwatch in more biodiverse areas, then β_1 is biased upwards. I test for *cross-district* sorting with the following specification:

$$log(Users_{dsym}) = \alpha + \beta_1 Infrastructure_{dsym} + \beta_2 SLX^N_{dsym} + \Gamma X'_{dsym} + \gamma_d + \theta_{sm} + \mu_y + \epsilon_{dsym} \tag{6}$$

where $Users_{dsym}$ is the number of users active in district d during year-month ym. These are the same sample of users that identified species loss in Equation 2. The third term is the previously defined spatial lag, except Ω spans the nation, N. This enables users to sort anywhere in India following project construction in district d, but with lower probability toward further away destinations. I also test specifications that restrict sorting to 100km, 200km, and 500km of district d. Remaining terms are as in Equation 2. $\beta_2 > 0$ implies that users sort into district d when other districts become relatively more fragmented.

I test for *within-district* sorting by estimating Equation 6 with spatial coverage as the outcome and omitting the spatial lag. Spatial coverage is the percent of district grid cells "birdwatched" in a given time period. If infrastructure pushes users into new parts of the district, then spatial coverage will increase and $\beta_1 > 0$. As discussed next, I find no evidence of cross- or within-district sorting, improving confidence in the TWFE design.

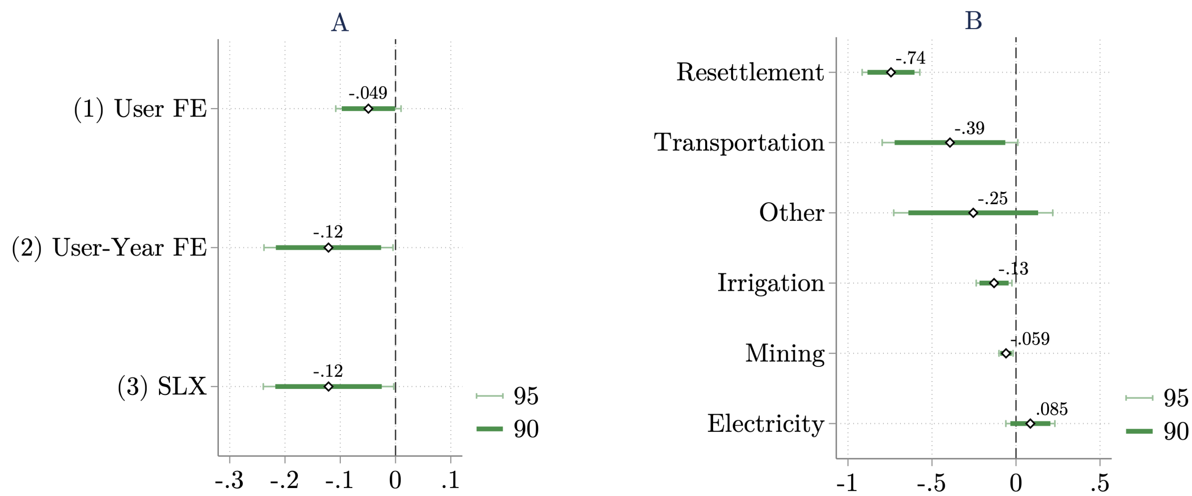

Figure 4: Estimates of the Infrastructure-Biodiversity Tradeoff in India

Note: The outcome is mean species richness across users' trips in a district-month. Panel A shows coefficients on cumulative area of infrastructure encroachments in a district-month. Specification (1) includes user, district, state-month, and year fixed effects. Specification (2) includes user-by-year, district, and state-month fixed effects. Specification (3) controls for spatial spillovers within the biome (Sec. 5.3.2). Panel B is a single regression with deforestation decomposed into project categories. Shaded bars denote confidence intervals. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log experience, log group size, and log spatial coverage. Standard errors clustered by biome.

6 Main Results

This section presents evidence on the impact of infrastructure on biodiversity. Species diversity is significantly threatened by infrastructure, driven by lower abundance of common and vulnerable species. Resettlement, transport, irrigation, and mining projects are particularly harmful. Lastly, species diversity does not rebound in the medium run.

6.1 Estimates of the Infrastructure-Biodiversity Tradeoff

Main Estimates Figure 4A illustrates the infrastructure-biodiversity tradeoff. Full tabular estimates are in Table S4. Specifications (1) and (2) of the figure estimate Equation 2 with and without the learning curve, respectively. The main coefficient, β_1, is negative in both specifications, indicating that infrastructure intrusions reduce local species diversity. The upward learning curve counteracts species declines in specification (1), yielding an attenuated coefficient. Removing this counterbalancing pressure with user-by-year fixed effects in specification (2) yields a steep decline in species richness.

An additional km^2 of infrastructure-driven deforestation in a district causes users to observe 0.12 fewer species, equivalent to 0.5% of the average bird list. To put this in perspective, eBird users observed 0.8 fewer species at the end of the study period compared to the start. The average district had 1.14 km^2 of forest *newly* displaced by infrastructure during this time, implying a loss of $1.14 \times -0.12 = 0.14$ species. Thus, infrastructure accounted for $0.14/0.80 \approx 17.5\%$ of species loss across India between 2015-2020.

Sensitivity: Spatial Spillovers Biases from spatial spillovers are minimal. Specification (3) of Figure 4A shows that species loss is unchanged when accounting for species displacement within the biome. Conditional on the direct effect, the spillover coefficient is positive but insignificant (Table S4, column 3). Table S5 tests robustness to allowing spillovers to materialize over different distances. In all cases, estimates of species loss remain stable, significant, and virtually equivalent to the main estimate. Spillovers similarly remain positive but noisy. This increases confidence that the lack of spillovers is pervasive, not an artifact of the within-biome assumption. These results do not mean species do not relocate following habitat loss. It means they do so in a way that is uncorrelated with local infrastructure development.

Sensitivity: Controlling for Observables Columns 4-6 of Table S4 probe sensitivity by successively adding controls. When observer behaviour and nightlights are removed (column 4), the coefficient remains negative but loses precision. When behaviour is added back, the tradeoff reappears (column 5), suggesting that behaviour is a key source of bias. The coefficient is equivalent to the main specification, which also controls for nightlights (column 3). Equal point estimates with or without nightlights implies that species loss is driven by habitat loss, not economic spillovers. This is unsurprising given that projects remain mostly incomplete during the study period (Section 2).

Column 6 adds diversion of *non-forest* land for infrastructure as a covariate. It has no impact on species diversity, underscoring habitat loss as the key mechanism driving species loss as opposed to other infrastructure-driven disturbances such as pollution. Although the impact of non-forest land diversion is statistically insignificant, we cannot reject the null hypothesis that its magnitude equals the deforestation effect.

More robustness tests are in Section 6.6. Among many others, these include: controlling for alternative forms of seasonality, accounting for a changing user base, and alternative species diversity metrics. I also investigate spatial correlation more systematically.

Ruling out Sorting Across and Within Districts Estimates do not appear to be driven by cross-district or within-district sorting. Table S6 tests for cross-district sorting by estimating Equation 6. Infrastructure measures are standardized for comparability between direct and spillover effects. Users do not sort into district d when other districts j within 100km become relatively more fragmented (column 1, second row). Neither does development in d push users elsewhere (first row). Lack of spillovers are visible under distance cutoffs up to 500km (columns 2-3), suggesting that users are highly mobile (Fact 4, Section 4.4), but not because of infrastructure development.

Column 4 tests for within-district sorting by dropping the spatial lag term and using the percent of "birdwatched" grid cells in a district as the outcome in Equation 6. The coefficient is negative and insignificant[17] , suggesting that eBird users continue visiting the same birdwatching locations as districts undergo development. This implies that they rarely birdwatch near construction sites, although I am unable to verify this since exact project coordinates are unavailable. Taken together, the lack of sorting across and within districts supports causal interpretation of the main estimates.

6.2 Estimates by Project Category

Figure 4B presents estimates of Equation 3. Coefficients describe the impact of a marginal encroachment by projects of that category, conditional on that by other categories. Five out of six categories trigger species diversity loss. Four of them—resettlement, transport, irrigation, and mining—do so with statistical precision.

Resettlements threaten species the most. An example is the diversion of 2.85 km^2 of forest in Madhya Pradesh for relocating a village previously located in a nearby Tiger Reserve[18]. The coefficient is largest likely because resettlements comprises a package of projects, including access roads and shelters, such that the magnitude reflects a sum of coefficients on other categories. Another reason may be that it is the only category directly linked to human activity. If one km^2 of habitat loss for building resettlements is associated with spillover economic activity *not captured by nightlights*, then it will result in more species loss than one km^2 of other projects. Without project GPS coordinates or details on what is inside each resettlement, I am unable to formally test these hypotheses.

The negative impact of "other" projects is imprecise. These are the smallest projects on average, but feature a standard deviation 17 times greater than the mean, the largest ratio of any category (Table 1). When aggregated to the district, a marginal encroachment thus comprises many underlying patch sizes. Coefficient magnitude is likely driven by large projects, where marginal encroachments comprise a single patch, and the noise by smaller projects, each too small to affect species diversity with precision. The same logic

[17] The coefficient is also insignificant when the outcome is the convex hull area around user's trips.

[18] The project was approved in April 2017 and includes housing, playgrounds, and roads. Site report: http://forestsclearance.nic.in/writereaddata/SIR/06022017561SBScan_02-06-2017_1501.pdf.

may explain the noisy impact of electricity projects, which have the second highest noise-to-signal ratio. The positive coefficient may be explained by the large number of hydro-electric dams, which create reservoirs that may attract previously unseen waterbirds.

Although mining appears to threaten species minimally, the coefficient is likely attenuated since mines are often sited in remote areas where few eBird users travel. Half of sample mines are in Odisha, Madhya Pradesh, and Chhattisgarh, with 27% in Odisha alone. The median number of users and trips in Odishi mining districts is under half of the national median. The few users who travel there may be a selected sample that undercount the species pool (Section 3.2).

Table S7 probes sensitivity of the estimates and further investigates mining effects. Similar to main estimates, category-wise estimates materialize when observer behaviour is accounted for (column 2) and remain stable when controlling for economic activity (column 3). To test the conjecture about the small mining effect, I restrict the sample to districts with high eBird activity, measured as above-median numbers of users recording above-median trips per user. If the bias is mining-specific, only the mining coefficient should be accentuated. Indeed, mining projects are twice as harmful in the high-activity sample and other coefficients remain virtually unchanged (column 4). This implies that non-mining projects are sited in places with sufficient eBird activity to begin with.

Appendix S4.3 present additional results by project ownership and shape (Table S18).

6.3 Estimates by Species Characteristics

Having established that infrastructure drives species loss, I next investigate which species are under threat. Table 3 shows estimates of Equation 2 with the outcome measured as counts of common, vulnerable, and endangered species[19]. Since outcomes are counts, coefficients should be interpreted in terms of abundance and not diversity.

Low-concern and vulnerable species are under most threat from infrastructure expansion. The coefficient in columns 1 and 3, where outcomes are in levels, is negative and significant. Magnitudes relative to the mean imply that an additional unit of infrastructure causes users to observe 1% fewer low-concern and vulnerable species. Since the outcome is a count, columns 2, 4 and 6 report Poisson estimates. I use the pseudo-maximum likelihood estimator to adjust standard errors (Wooldridge, 1999). Again, lower species abundance is detected for low-concern and vulnerable species. Since observing endangered species is rare, there is insufficient variation to detect effects in columns 5 and 6.

[19] IUCN lists species as: least concern, near threatened, vulnerable, endangered, and critically endangered. I combine least concern and near threatened as well as endangered and critically endangered.

Table 3: Estimates by IUCN Threat Status

	Least Concern		Vulnerable		Endangered	
	(1) Level	(2) Poisson	(3) Level	(4) Poisson	(5) Level	(6) Poisson
Infrastructure (km^2)	-0.961***	-0.012***	-0.007*	-0.009***	0.002	0.006
	(0.280)	(0.003)	(0.004)	(0.003)	(0.002)	(0.008)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Outcome Mean	90.541	90.545	0.589	0.741	0.167	0.378
User x Year FEs	✓	✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓	✓
State × Month FEs	✓	✓	✓	✓	✓	✓
Observations	161896	161889	161896	128511	161896	71284
R^2	0.517		0.395		0.377	

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome is a frequency count for the number of times a user observed a species of each type during a district-year-month. Forest infrastructure is cumulative area of infrastructure encroachments in a district-month. Poisson regressions are estimated with a pseudo-maximum likelihood estimator. All specifications include user-year, district, and state-month fixed effects as well as controls for temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log experience, log group size, and log spatial coverage. Standard errors clustered by biome.

Appendix S4.4 presents estimates by habitat specialization and migratory status. Perhaps unsurprisingly, forest birds are the most sensitive to infrastructure encroachment into forests, whereas there is no impact on wetland birds (Table S19). Non-migrant species are also especially threatened by development projects, potentially reflecting their inability to adapt through relocation. These results suggest that the overall decline in species *diversity* in Figure 4A is driven by lower *abundance* of common and vulnerable species, forest habitat specialists, and resident species.

6.4 Heterogeneity: Species are More Resilient in Intact Forests

I next investigate impacts by baseline ecosystem quality, which has implications for whether conservation should target intact or fragmented landscapes. Table 4 presents estimates of Equation 4. Columns describe treatment heterogeneity using two measures of ecosystem quality. Both are standardized so that a one-unit change can be compared.

Species are more resilient to infrastructure development in pristine areas. The adverse impact of infrastructure on species diversity is halved in districts with one standard deviation higher initial forest cover[20] (column 1). To account for potential selection of

[20] Forest cover (% of a pixel) is from the VCF satellite product on a 250m grid (Townshend et al., 2017)

Table 4: Treatment Effects by Baseline Forest Intactness

	(1)	(2)	(3)	(4)
Infrastructure (km^2)	-0.130***	-0.107***	-0.133***	-0.108***
	(0.019)	(0.028)	(0.024)	(0.032)
Infrastructure (km^2) $\times$ Baseline Forest Cover	0.065*	0.056		
	(0.030)	(0.032)		
Infrastructure (km^2) $\times$ Baseline Species Richness			0.052**	0.046**
			(0.020)	(0.019)
Infrastructure $\times$ Category Shares	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes
User x Year FEs	✓	✓	✓	✓
District FEs	✓	✓	✓	✓
State $\times$ Month FEs	✓	✓	✓	✓
Observations	161896	161896	161896	161896
R^2	0.690	0.690	0.690	0.690

Note: * $p < .1$, ** $p < .05$, *** $p <.01$. The outcome is mean species richness across users' trips in a district-month. $Infrastructure$ is cumulative area of infrastructure in a district-month. Baseline district forest cover is for in 2014. Baseline species richness is measured at the district level by overlapping species range maps. All specifications include user-year, district, and state-month fixed effects as well as controls for temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log experience, log group size, and log spatial coverage. Columns 2 and 4 additionally include six interactions terms of infrastructure with the baseline district share of projects in each category. Standard errors clustered by biome.

projects into certain habitats (e.g., mines disproportionately sited in pristine forests), column 2 controls for the interaction of infrastructure with the baseline project category distribution. Results are very similar, although precision of the interaction slightly declines (p=0.108). Remaining columns test sensitivity to measuring baseline ecosystem quality with species range maps. The tradeoff reduces by a similar amount (column 3) and is robust to controlling for project category (column 4).

In Appendix S4.5, I test heterogeneity by districts' importance for conserving bird habitat, a measured developed by BirdLife International. Species are more resilient in districts' designated as Important Bird Areas (Table S20), in line with the previous results. The degree to which species loss is muted in these districts is also very similar.

Overall, since species loss is largest when baseline ecosystem quality is low, these results support stronger protections for degraded landscapes. The findings also corroborate existing theory from ecology (Hanski, 1998).

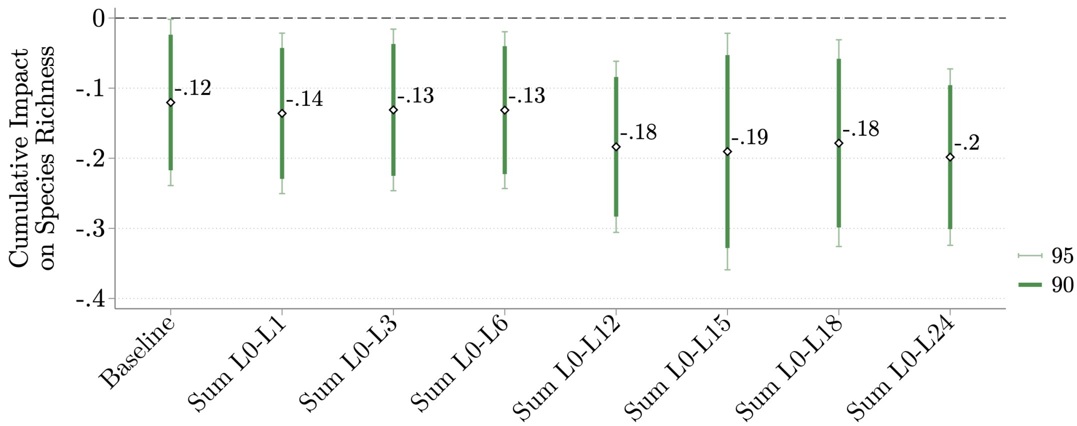

Figure 5: Cumulative Dynamic Lag Results

Note: "Baseline" repeats the main result with user-year, district, and state-month fixed effects. "Sum L0-L1" adds the first lag of forest diversion to the main specification and reports the sum of coefficients on the first lag and baseline effect. "Sum L0-L3" sums up to the third lag, and so on. Shaded bars are confidence intervals. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log experience, log group size, and log spatial coverage. Standard errors clustered by biome.

6.5 Dynamics: Species Diversity Loss is Persistent in the Medium Run

My estimates thus far reflect how species respond to habitat loss within the month. This overlooks lagged effects, either reflecting a delay for species diversity to equilibrate (Odum, 1969), or, a delay between project approval and forest clearing. To separate these channels, I investigate dynamics up to two years. Since Table 2 showed that logging is observed within the year of project approval, lagged effects within the year may be driven by driven by either channel, whereas lags beyond one year likely reflect delayed species responses since deforestation is likely to have already occurred by then.

Figure 5 presents estimates of Equation 2 with lags up to two years. White diamonds are the sum of baseline and lagged coefficients, which measures *net impacts* of infrastructure several periods later. Species declines are triggered in the month of project approval and persist thereafter. The cumulative impact three months later ("Sum L0-L3") is nearly equal to the baseline effect, with stable point estimates up to six months. A slight lagged effect is also observed after one year and persists through the second year[21]. However, we cannot interpret this as evidence of delayed species responses since confidence intervals overlap point estimates across all periods. We therefore cannot reject the null that second-year coefficients are the same as the first year.

The main takeaway is that species loss does not recover in the medium run. This has

[21] Persistence even extends through the 3rd year (not shown), although data loss lowers precision.

two implications. First, it provides further evidence that my estimates capture responses to permanent habitat loss. If species were responding to temporary disturbances, such as noise or air pollution, then Figure 5 would have featured a U-shape as species return after the disturbance dies down. Second, the dynamic results also highlight the ineffectiveness of compensatory afforestation requirements, would have also been indicated by a U-shape. However, as described in Section 2, tree-planting rarely takes place, posing minimal threat to the research design.

6.6 Additional Robustness Checks

This section shows robustness to different estimation samples, diversity metrics, and removing outliers. The next section shows robustness to an alternative research design.

Sample Restrictions Table S8 columns 1-4 show estimates from alternative samples. Whereas the main specification accounts for within-user behaviour changes, it does not account for a changing user base. Column 1 thus fixes the sample to only those users who signed up in 2015 (N=2,938 users). The estimate is very similar, suggesting that users who joined later are similar to veteran users.

Column 2 drops users' home districts (see Section S5 for computation), which tests whether they display different recording practices at home and away. The point estimate is very similar, suggesting that they do not. Column 3 drops districts with low eBird activity, measured as districts with below-median number of users who record below-median number of trips per user. Again, the coefficient is remarkably similar, suggesting that my estimates are not biased by peculiarities in places with sparse eBird usage[22].

Lastly, column 4 drops 2020, the year COVID-19 swept the globe. During India's lockdown, "balcony birdwatching" was popularized and eBird sign-ups quadrupled (Madhok and Gulati, 2022). Estimates remain stable, which is unsurprising since year fixed effects absorb time shocks and the protocol covariate captures the shift indoors.

Removing Outliers I transform the sample in two ways to remove outliers. In column 5, I drop India's top three "mega-projects": 1) the world's largest lift irrigation[23] project, located in Telangana and requiring 3,168 ha. of deforestation, 2) a 4,000 MW coal plant, also in Telangana, that requires 4,334 ha. of deforestation, and 3) the world's largest concrete dam, located in Arunachal Pradesh, that requires 5,056 ha. of deforestation. Coefficient

[22] Similarly, point estimates are robust to splitting the sample by North/South, although precision is lower in the North since data are sparser (Fig. 1C).

[23] Lift irrigation is where water is transported by pumps rather than by exploiting natural flow.

size doubles but falls within the lower bound of the baseline estimate. The larger magnitude may be from dropping the two irrigation mega-projects, which may have created water habitat that attract new species. Dropping these releases this offsetting pressure on the coefficient, leading to greater species loss.

In column 6, instead of dropping mega-projects, I apply the inverse hyperbolic sine transformation[24] to $Infrastructure_{dysm}$ (Bellemare and Wichman, 2020). The coefficient implies that a 1% increase in infrastructure leads to a loss of 0.006 species. For comparison, 1% of $Infrastructure_{dysm}$ evaluated at the mean is 0.0114 km^2. Applying this to the baseline coefficient yields $0.0114 \times -0.12 \approx -0.0014$ species. The discrepancy likely arises from different functional assumptions: IHS assumes diminishing marginal effects of habitat loss whereas the baseline does not.

Alternative Diversity Measures Species richness has been criticized for its simplicity. Somewhere with one pigeon and 99 crows, and another with fifty of each, both have a richness of two despite the latter being more even. I compute two metrics that account for evenness:

$$SH_j = -\sum_{s=1}^{S} p_{sj} ln(p_{sj}) \qquad\qquad SI_j = 1/\sum_{s=1}^{S} p_{sj}^2$$

where p_{sj} is the proportion of all observations on eBird checklist j belonging to species s. The Shannon Index (left) increases in diversity. The Simpson Index (right) reflects the probability that two randomly drawn birds belong to the same species (Magurran, 2013). I use $1 - SI_j$ so that it also increases in diversity. One issue is that eBird counts are imprecise given difficulties recording quickly moving flocks. About 90% of counts in my sample are approximated to the nearest tenth, and 10% of checklists have missing counts.

Columns 7 and 8 show an infrastructure-biodiversity tradeoff using these alternative measures, but coefficients are imprecise as expected. Infrastructure impacts on Shannon and Simpson diversity are 1.7% and 4.0% of their means, respectively.

Regression Weights Column 9 adds regression weights equal to the number of trips underlying each observation[25]. This ensures that observations influence the coefficient in proportion to their measurement precision. I implement this test because species richness is a mean over users' trips in a district-month and part of the error variance in Equation 2 may thus be explained by differences in the number of underlying trips. Figure S6 shows

[24] This uses the function $arcsinh(x) = ln(x + (x^2 + 1)^{1/2})$.

[25] I truncate at the 99th percentile first. As an example outlier, the maximum number of trips in a district-yearmonth is 3779 by one user i.e., 126 trips *per day*.

the cumulative distribution: 10% of observations are a mean over more than 10 trips (precisely measured) and 90% are over fewer than 10 trips (imprecisely measured). The coefficient is virtually unchanged and remains significant at the 10-percent level.

Other Robustness Tests Appendix S4.6 demonstrates robustness to a variety of alternative fixed effects. Appendix S4.7 investigates spatial error correlation (Conley, 1999) and finds that precision remains similar even when allowing for spatial correlation up to 1000km. Lastly, Appendix S4.8 reports the same robustness checks above with infrastructure decomposed by project category. Coefficients on most categories remain negative.

6.7 Robustness: Instrumental Variable Estimates

Next, I show that results are robust to a widely-implemented IV design based on close races between incumbent and runner-ups in State elections (Appendix S6 for more details). I instrument project approvals with the fraction of close-election constituencies in a district with incumbent winners. Since these elections are essentially won by coin toss, places where the incumbent just barely won and lost should be statistically similar in terms of economic prospects and other confounders that were previously a concern.

The exclusion restriction assumes that, conditional on controls and fixed effects, district-level incumbent strength affects local biodiversity only by sanctioning forest diversion for infrastructure. I acknowledge that this is a strong assumption. A second concern is that estimates do not generalize to non-competitive districts. For these reasons, I view this design as a robustness check on coefficient sign rather than another set of main estimates.

The 2SLS strategy compares eBird observations within users travelling to districts where the incumbent just barely won and lost:

$$\text{First Stage:} \quad Infrastructure_{dsym} = \mu_1 IC_{dsy} + \mu_2 C_{dsy} + \mu_3 (f(M_{dsy}) \cdot I_{dsy}) + \mu_4 f(M_{dsy}) + \mu_5 I_{dsy} + \mu_6 E_{sy} + \Gamma X'_{idsym} + \phi_i + \gamma_d + \theta_{sm} + \nu_y + \epsilon_{idsym} \tag{7}$$

$$\text{Second Stage:} \quad SR_{idsym} = \beta_1 \widehat{Infrastructure}_{dsym} + \beta_2 C_{dsy} + \beta_3 (f(M_{dsy}) \cdot I_{dsy}) + \beta_4 f(M_{dsy}) + \beta_5 I_{dsy} + \beta_6 E_{sy} + \Gamma X'_{idsym} + \phi_i + \gamma_d + \theta_{sm} + \nu_y + \epsilon_{idsym} \tag{8}$$

where $Infrastructure_{dsym}$, SR_{idsym}, subscripts, and fixed effects are the same as Equation 2. Infrastructure is instrumented with IC_{dsy}, the share of constituencies in district d where the incumbent party won in a close race during the last election. Elections are close if the win margin is within 2 pp. C_{dsy} is the share of close-election constituencies where the incumbent ran, which helps disentangle the plausibly random election outcome from the potentially non-random factors that led to a close election in the first place. In the spirit of a fuzzy regression disctontinuity design, I include the mean win margin (close or not), M_{dsy}, which enters linearly, and with second- and third-order polynomials, $f(M_{dsy})$, as robustness checks. I also control for the interaction of $f(M_{dsy})$ and I_{dsy}, an indicator for whether incumbents ran in the district. Lastly, I control for election year, E_{sy}, and the same covariates, X'_{idsym}, as Equation 2. User-by-year fixed effects are too strict since eBirders are unlikely to traverse many closely contested districts in a year. I thus include user and year fixed effects separately, and test their interaction in a robustness check.

The first stage shows a sharp discontinuous decrease in forest encroachments when incumbents win, and the reduced form shows an accompanying increase in species richness (Figure S4)[26]. Second stage estimates (Table S9) are negative, significant, and robust to various polynomials and bandwidths, with F-statistics near rule-of-thumb levels. Reassuringly, estimates are also robust to user-by-year fixed effects (column 6), the same specification that addresses eBird biases in the main research design. Despite limitations of the close-election approach, uncovering an infrastructure-biodiversity tradeoff with this alternative design lends additional credibility to the main findings.

7 The Political Economy of Conservation

Having established that infrastructure expansion degrades biodiversity, this section explores which institutions minimize the tradeoff. I estimate the tradeoff from the previous section as a function of whether districts have inclusive or extractive institutions. I find that the infrastructure-biodiversity tradeoff is smaller under inclusive institutions. I then explore mechanisms by documenting how project authorities interact with tribal groups under both institutional types. Informed consent between developers and tribes, as well as more stringent environmental review, is more common in inclusive districts.

7.1 Data: Measuring Institutional Quality

I begin by categorizing districts as having inclusive or extractive institutions, broadly defined. Data on institutional quality is obtained from Banerjee and Iyer (2005) for 163 districts. They distinguish between two colonial institutions: in *zamindari* districts (N=64), landlords set land taxes, could dispossess peasants for nonpayment, and kept residuals

[26]Figure S5 shows no evidence of manipulation around the cutoff. I fail to reject the null hypothesis of no difference in density at the boundary (Cattaneo et al., 2020).

after paying the British. In *raiyatwari* districts (N=99), cultivators paid taxes without a middleman. Perhaps unsurprisingly, *zamindari* districts perform worse today on several equality and development measures. Persistence of class-based inequality and lower ability of the disenfranchised to mobilize around their interests in *zamindari* districts are key mechanisms explaining the lack of convergence[27].

Building on their paper, I re-conceptualize *raiyatwari* and *zamindari* districts as inclusive and extractive, respectively. To further justify this re-conceptualization, in Section 7.3, I show that tribal communities are more involved in development planning in inclusive districts. I also replicate Banerjee and Iyer (2005) in Appendix S4.9 using additional outcomes. There are fewer protests involving minorities in inclusive (*zamindari*) districts. There are also fewer criminal politicians in these districts, though precision is low.

If disaffected groups are better able to engage in the development process and protect their livelihoods in inclusive districts, then the adverse ecological impacts of infrastructure should be smaller in these districts. I formally examine this hypothesis next.

7.2 Results: Inclusive Institutions Minimize Species Loss

Estimation To investigate the role of institutions in mediating the infrastructure-biodiversity tradeoff, I estimate heterogenous treatment effects with the following equation:

$$\begin{aligned} SR_{idsym} = \alpha + \beta_1 \cdot Infrastructure_{dsym} + \beta_2 (Infrastructure_{dsym} \cdot Inclusive_d) \\ + \Omega(Infrastructure_{dsym} \cdot X'_d) + \Gamma X'_{idsym} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idsym} \end{aligned} \tag{9}$$

where $Inclusive_d$ is a dummy for whether district d had a history of inclusive institutions. X'_d are a set of district-level covariates that enter interacted with $Infrastructure$. Ω thus accounts for heterogeneous effects of infrastructure along dimensions potentially correlated with institutions. All other terms are as in Equation 2. Data are aggregated to 1991 census boundaries to match Banerjee and Iyer (2005). The coefficient of interest is β_1 and β_2, which capture the main infrastructure-biodiversity tradeoff, and any moderation of the tradeoff depending on institutional type. I focus on the hypothesis that $\beta_2 > 0$, i.e., biodiversity is conserved in districts with better institutions.

Threats to Identification The main identification concern is endogenous institutions (Aghion et al., 2004). This is less of an issue in my context because *zamindar* status was based on British politics and not local characteristics (Banerjee and Iyer, 2005). Moreover,

[27] In a follow up paper, Lee (2019) provide additional evidence that state capacity is indeed the most plausible mechanism driving the results in Banerjee and Iyer (2005).

Table 5: The Impact of Infrastructure on Biodiversity by Institutional Type

	(1)	(2)	(3)	(4)	(5)
β_1: Infrastructure (km^2)	-0.551***	-0.447***	-0.434***	-0.394***	-0.551**
	(0.067)	(0.051)	(0.073)	(0.082)	(0.214)
β_2: Infrastructure (km^2) × Inclusive (=1)	0.434**	0.340*	0.315*	0.421**	0.434***
	(0.116)	(0.144)	(0.134)	(0.129)	(0.142)
Infrastructure × Tribal Share	Yes	Yes	Yes	Yes	Yes
Infrastructure × Baseline Forest	Yes	Yes	Yes	Yes	Yes
Infrastructure × High-Activity	No	No	No	Yes	No
User × Year FEs	✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓
State x Month FEs	✓	✓	✓	✓	✓
Spillovers		✓			
Weighted			✓		
Clustering	Biome	Biome	Biome	Biome	District
Observations	58760	58760	58760	58760	58760
R^2	0.704	0.704	0.784	0.704	0.704

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome is mean species richness across users' trips in a district-month. *Inclusive* means the district has historically inclusive institutions. Sample is restricted to 163 districts in Banerjee and Iyer (2005) and aggregated to 1991 boundaries. Tribal share is the fraction of district population belonging to a tribal group as measured in 2011. High-Activity equals one if the district has above-median number of users recording above-median number of trips per user. All specifications include user-year, district, and state-month fixed effects as well as controls for temperature, rain, traveling trips, log nightlights, log duration, log distance, log experience, log group size, and log spatial coverage.

time-invariant differences in the ecology of inclusive and extractive districts are absorbed by district fixed effects. The remaining concern is that infrastructure may exhibit heterogeneous effects along dimensions correlated with institutional type, in which case β_2 is biased. The interaction coefficient Ω controls for this source of endogeneity, and I test sensitivity to several definitions of X'_d as a safeguard.

Results and Robustness Estimates of Equation 9 are in Table 5. All columns control for interactions of infrastructure with baseline tribal population share as well as forest cover. The former separates heterogeneity through population effects from that through institutions. The latter accounts for potentially higher forest cover in inclusive districts, in which case species resilience in these districts may upward bias β_2 (Section 6.4, Table 4).

78% of species loss is erased in inclusive districts. The mitigating effect of inclusive institutions are very similar when controlling for spatial spillovers within the biome (column 2), weighting by number of eBird trips underlying SR_{idsym} (column 3), and adding an interaction between infrastructure and a district dummy for high eBird activity (column 4). The latter accounts for β_2 potentially confounding differences in eBird usage across institution types. Lastly, the mitigating effect remains statistically significant under district clustering (column 5). Since the moderating role of institutions is independent of tribal population, we can conclude that institutions empowering disaffected people, not their population per se, determine the extent of sustainable development.

Table S10 conducts additional robustness tests performed in Section 6.6. Columns 1-4 show that estimates are generally robust to alternative fixed effects that account for sub-state seasonality and flexible learning rates. The interaction is noisy in column 1, likely due to demanding user-by-month fixed effects. Column 5 shows that treatment heterogeneity is very similar on the sample of users who signed up for eBird in 2015, suggesting that estimates are not driven by a changing user base. The main effect becomes noisy, perhaps due to the stringent sample restriction. Column 6 shows stable estimates when accounting for COVID. To deal with outliers, columns 7 and 8 show that estimates are stable when dropping mega-projects and when using IHS on the infrastructure variable, respectively. Lastly, estimates are robust to clustering by state (column 9).

The results emphasize the role of inclusive institutions in mitigating anthropogenic pressures on ecosystems. However, it is difficult to glean specific policy lessons since the muted tradeoff may operate through many channels. I investigate mechanisms next.

7.3 Policy Mechanisms: Tribal Rights and Informed Consent

Informed Consent between Developers and Tribes Why are development projects built more sustainably in districts with historically inclusive institutions? I explore two important mechanisms: developers are more likely to incorporate the voices of tribal people, and more likely to undergo stringent environmental review, in inclusive districts.

Banerjee and Iyer (2005) argue that the absence of a landed gentry in inclusive districts left a legacy enabling "elites and the masses to act together in the collective interest". Lee (2019) shows that more state contact with cultivators in inclusive districts left a legacy of better state capacity compared to extractive districts where the state was absent. This suggests that tribes can better mobilize around their interests in inclusive districts.

Permit Data The simplest test is whether projects in inclusive districts are more likely to follow the FRA, which requires inclusion of tribes during the permitting process (Section 2). Yet if the policy is binding, then there would be no variation. Recent reports indicate that FRA enforcement is weak, often bypassing consent altogether (Dubey et al.,

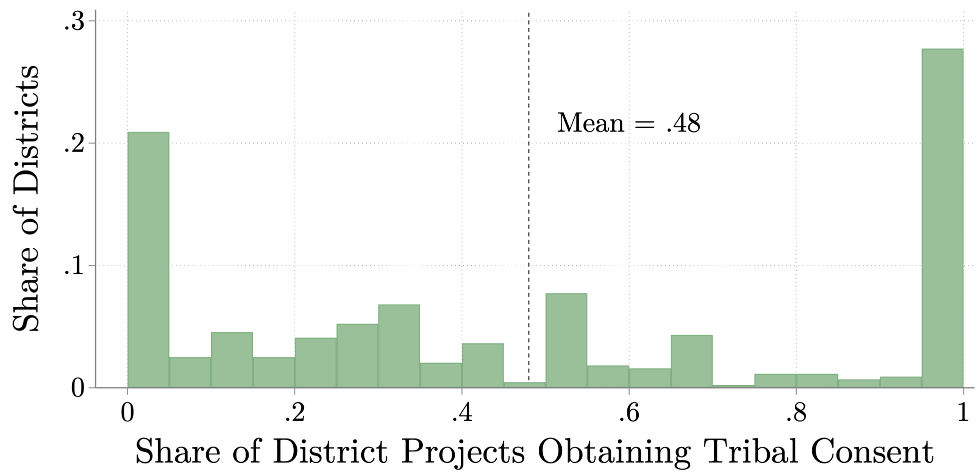

Figure 6: Enforcement of Forest Rights Act (2006)
Note: Data are the share of district projects approved with informed consent by the Gram Sabha during the study period. Sample comprises the 80% of projects that reported informed consent (the digital subsample).

2017). Since the project sample reports whether consent was obtained (Figure S7), I plot the distribution of projects obtaining Gram Sabha consent in Figure 6. The lack of right-tail bunching is evidence of imperfect compliance, implying that there are districts where inclusive development is always, sometimes, and never observed. I exploit this fact to study if inclusive institutions are actually more inclusive.

Two other permit variables highlight mechanisms. The first is whether a supplemental cost-benefit report was commissioned, beyond the standard site monitoring reports. This reflects the rigour of environmental review since commissioning is based on value judgement[28]. The second is whether the project is sited in a protected area buffer.

Estimation I match project permits with the inclusive-extractive dummies and use pooled OLS to compare project characteristics under each district institution. Since institutional type is fixed, I make cross-district comparisons within the same state and time-period:

$$Y_{pdsym} = \alpha + \beta_1 \cdot Inclusive_d + \Gamma X'_{pdsym} + \theta_{sm} + \epsilon_{pdsym} \tag{10}$$

where Y_{pdsym} is a dummy for whether project p approved in district d of state s in year y and month m received informed consent, completed a supplemental cost-benefit report, or was sited near a protected area. $Inclusive_d$ is the institutional dummy from Equation 9.

[28]Value judgment is used for projects > 20 ha., which is more than 90% of my projects. Official guidelines here: http://forestsclearance.nic.in/writereaddata/Addinfo/0_0_7111512571261CostBenefitAnalysisGuidelinesforforestlanddiversion-2017.pdf

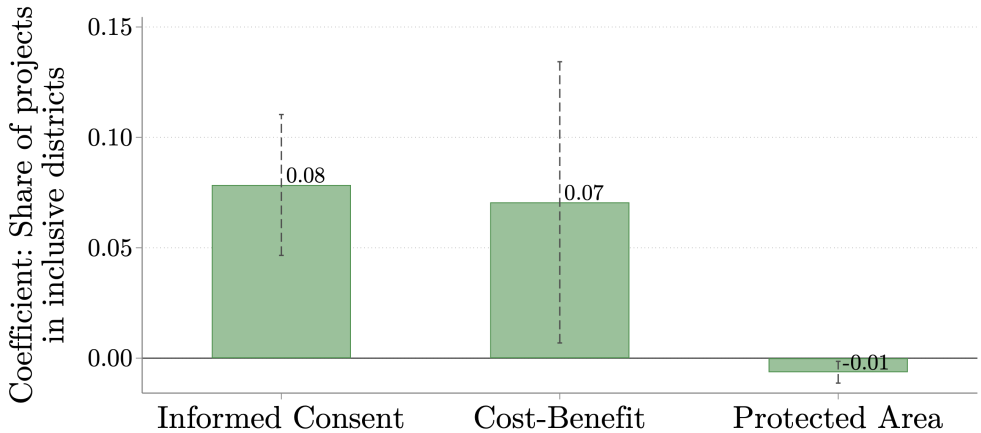

Figure 7: Mechanisms: Informed Consent, Cost-Benefit Analyses, and Project Placement

Note: Data are at the project level. Bars are coefficients from Equation (10). Sample is restricted to 163 districts in Banerjee and Iyer (2005). Informed consent indicates whether the FRA was followed. Cost-Benefit Analyses indicates whether one was done during project review. Protected Area equals one if the project is sited in or near one. All specifications control for: project size, tribal population share, baseline forest cover, latitude, altitude, coastal dummy, and district area. Grey bars are 95% confidence intervals.

X'_{pdsym} is a set of covariates including project size, tribal population share, baseline forest cover, and district size. θ_{sm} are state-month fixed effects. β_1 reveals the proportion of projects with each feature in inclusive districts compared to extractive ones.

Results Projects in inclusive districts are 8 pp. more likely to obtain informed consent from tribal groups compared to projects sited in extractive districts in the same state (Figure 7; Table S11 for table). Forest officers in inclusive districts are also 7 pp. more likely to commission extra cost-benefit reports during project review. Lastly, projects in inclusive districts are 1 pp. less likely to be sited near a protected area. These are three important mechanisms driving the smaller infrastructure-biodiversity tradeoff in Table 5.

Beyond these three, the sorting of conservation-friendly projects toward inclusive districts is another potential mechanism leading to less species loss in these districts. To test whether inclusive and extractive districts receive different types of projects, Table S12 estimates the project category distribution by institutional type using the same data as Equation 10. Values are coefficients from regressing project category indicators on $Inclusive_d$. Overall, project categories are well balanced across institutional types, partially ruling out sorting as a mechanism. There are two exceptions: first, electricity projects are more common in inclusive districts under state fixed effects. Recall that these projects have a positive (albeit insignificant) effect on species diversity (Figure 4B). Second, mining projects are (weakly) less common in inclusive districts, which may partially explain smaller biodiversity loss in these districts (Table 5). Otherwise, there are statistically similar proportions of projects between the two district types.

The minimal role of sorting implies that inclusive institutions themselves have "teeth", as observed through higher rates of informed consent, greater environmental scrutiny of proposed projects, and an overall smaller ecological footprint of projects in these districts (Table 5). Figure 7 also corroborates Banerjee and Iyer (2005) and other studies. Duflo and Pande (2007) use the same institutions classification to claim that populations affected by dams are more effective at demanding compensation in inclusive districts. Lal et al. (2021) show that inclusive governance in India increased tree cover. My results thus point to the mechanisms through which institutions drive conservation. They suggest that engaging forest-dependent communities, along with more stringent checks-and-balances during project approval, are vital for protecting biodiversity.

8 Conclusion

This paper provides rigorous evidence on the impact of infrastructure on biodiversity in a developing nation. It also quantifies the role of institutions in mitigating the tradeoff. Between 2015-2020, development in India's forests accounted for nearly 20% of the decline in bird diversity. Species loss does not rebound in the medium-run, and is accentuated in already-fragmented areas. The tradeoff is more than halved when local institutions amplify the voices of indigenous groups in the development planning process.

My results are especially relevant as emerging economies prioritize the types of infrastructure studied here. Surprisingly, studies from emerging regions find limited ecological costs of infrastructure projects (Asher et al., 2020; Garg and Shenoy, 2021; Baehr et al., 2021). In the absence of biodiversity data, these studies use tree cover to measure ecosystem health, whereas I leverage several million verified species sightings. After accounting for observer biases and spatial spillovers, this novel data yields robust evidence of anthropogenic species decline, and can be used to help inform infrastructure planning.

Results of this paper are policy relevant at both a broad and grassroots level. In places where institutions favour the economically advantaged, infrastructure development is associated with more biodiversity loss. This highlights the need for people-centred conservation policy. India has made strides with the FRA (2006), which promises forest rights to indigenous people and their inclusion in development decisions. Yet nearly two decades later, half of forest rights claims remain legally unrecognized and face other forms of weak enforcement (Ministry of Tribal Affairs, 2022). I find that upholding the FRA helps neutralize the infrastructure-biodiversity tradeoff. In sum, inclusive institutions and procedural justice are critical for meeting the dual objectives of development and conservation.

This paper is not without limitations. First, species richness abstracts from notions of functional diversity, genetic diversity, and other dissimilarity indices (Weitzman, 1992, 1993). Second, with a six year study period, I am unable to study whether species diversity rebounds in the long-run. Lastly, without reliable species values, I am unable to benchmark the cost of infrastructure-driven species loss. Despite these limitations, this study provides useful insights into the dynamics of biodiversity in human-modified landscapes and is relevant for decision-makers tasked with conserving local biodiversity.

References

Aghion, Philippe, Alberto Alesina, and Francesco Trebbi. 2004. Endogenous political institutions. *The Quarterly Journal of Economics* 119 (2): 565–611.

Anukriti, S, Bilge Erten, and Priya Mukherjee. 2022. Female leaders and intrahousehold dynamics: Evidence from state elections in india. *Working Paper* .

Asher, Sam, Teevrat Garg, and Paul Novosad. 2020. The ecological impact of transportation infrastructure. *The Economic Journal* 130 (629): 1173–1199.

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad. 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* 35 (4): 845–871.

Baehr, Christian, Ariel BenYishay, and Bradley Parks. 2021. Linking local infrastructure development and deforestation: Evidence from satellite and administrative data. *Journal of the Association of Environmental and Resource Economists* 8 (2): 375–409.

Banerjee, Abhijit, and Lakshmi Iyer. 2005. History, institutions, and economic performance: The legacy of colonial land tenure systems in india. *American economic review* 95 (4): 1190–1213.

Bellemare, Marc F, and Casey J Wichman. 2020. Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics* 82 (1): 50–61.

BirdLife International. 2018. Bird species distribution maps of the world. Version 2018.1.

Blanchard, Paul, Douglas Gollin, and Martina Kirchberger. 2023. Perpetual motion: Human mobility and spatial frictions in three african countries. *CEPR Discussion Paper No. DP16661* .

Börner, Jan, Dario Schulz, Sven Wunder, Alexander Pfaff, et al. 2020. The effectiveness of forest conservation policies and programs. *Annu. Rev. Resour. Econ* 12 (1): 45–64.

Burgess, Robin, Matthew Hansen, Benjamin A Olken, Peter Potapov, and Stefanie Sieber. 2012. The political economy of deforestation in the tropics. *The Quarterly journal of economics* 127 (4): 1707–1754.

Callaghan, Corey T, John M Martin, Richard E Major, and Richard T Kingsford. 2018. Avian monitoring–comparing structured and unstructured citizen science. *Wildlife Research* 45 (2): 176–184.

Callaghan, Corey T, Jodi JL Rowley, William K Cornwell, Alistair GB Poore, and Richard E Major. 2019. Improving big citizen science data: Moving beyond haphazard sampling. *PLoS Biology* 17 (6): e3000357.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2024. Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637* .

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma. 2020. Simple local polynomial density estimators. *Journal of the American Statistical Association* 115 (531): 1449–1455.

Clots-Figueras, Irma. 2012. Are female leaders good for education? evidence from india. *American economic journal: applied economics* 4 (1): 212–244.

Cole, Shawn. 2009. Fixing market failures or fixing elections? agricultural credit in india. *American economic journal: applied economics* 1 (1): 219–250.

Conley, Timothy G. 1999. Gmm estimation with cross sectional dependence. *Journal of econometrics* 92 (1): 1–45.

Construction Federation of India. 2015. India construction sector ? moving forward.

Copeland, Brian R, and M Scott Taylor. 2004. Trade, growth, and the environment. *Journal of Economic literature* 42 (1): 7–71.

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. Environmental health risks and housing values: Evidence from 1,600 toxic plant openings and closings. *American Economic Review* 105 (2): 678–709.

Dasgupta, Susmita, Benoit Laplante, Hua Wang, and David Wheeler. 2002. Confronting the environmental kuznets curve. *Journal of economic perspectives* 16 (1): 147–168.

Dixit, Avinash, and John Londregan. 1996. The determinants of success of special interests in redistributive politics. *the Journal of Politics* 58 (4): 1132–1155.

Donald, Paul F, Lincoln DC Fishpool, Ademola Ajagbe, Leon A Bennun, Gill Bunting, Ian J Burfield, Stuart HM Butchart, et al. 2019. Important bird and biodiversity areas (ibas): the development and characteristics of a global inventory of key sites for biodiversity. *Bird Conservation International* 29 (2): 177–198.

Dubey, Sanghamitra, Tushar Dash, and Geetanjoy Sahu. 2017. Political economy of community forest rights. *Economic and Political Weekly* 52.

Duflo, Esther, and Rohini Pande. 2007. Dams. *The Quarterly Journal of Economics* 122 (2): 601–646.

eBird Basic Dataset. 2019. Version ebd_in_201401_201904_relapr-2019.

Elhorst, JP, and SH Vega. 2015. The slx model. *Journal of Regional Science* 55 (3): 339–363.

Elvidge, Christopher D, Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu, and Tilottama Ghosh. 2017. Viirs night-time lights. *International Journal of Remote Sensing* 38 (21): 5860–5879.

FAO, and UNEP. 2020. State of the world's forests: Forests, biodiversity and people. Technical report, FAO.

Fitzpatrick, Matthew C, Evan L Preisser, Aaron M Ellison, and Joseph S Elkinton. 2009. Observer bias and the detection of low-density populations. *Ecological applications* 19 (7): 1673–1679.

Fleishman, Erica, Reed F Noss, and Barry R Noon. 2006. Utility and limitations of species richness metrics for conservation planning. *Ecological indicators* 6 (3): 543–553.

Forest Survey of India. 2019. India state of forest report. Technical report, Ministry of Environment, Forests and Climate Change. Government of India.

Foster, Andrew D, and Mark R Rosenzweig. 2003. Economic growth and the rise of forests. *The Quarterly Journal of Economics* 118 (2): 601–637.

Fraixedas, Sara, Andreas Lindén, Markus Piha, Mar Cabeza, Richard Gregory, and Aleksi Lehikoinen. 2020. A state-of-the-art review on birds as indicators of biodiversity: Advances, challenges, and future directions. *Ecological Indicators* 118: 106728.

Frank, Eyal G, and Wolfram Schlenker. 2016. Balancing economic and ecological goals. *Science* 353 (6300): 651–652.

Garg, Teevrat, and Ajay Shenoy. 2021. The ecological impact of place-based economic policies. *American Journal of Agricultural Economics* .

Grossman, Gene M, and Alan B Krueger. 1995. Economic growth and the environment. *The Quarterly Journal of Economics* 110 (2): 353–377.

Hagerty, Nick. 2021. Adaptation to surface water scarcity in irrigated agriculture. *Working paper* .

Hanna, Rema, and Paulina Oliva. 2015. The effect of pollution on labor supply: Evidence from a natural experiment in mexico city. *Journal of Public Economics* 122: 68–79.

Hanski, Ilkka. 1998. Metapopulation dynamics. *Nature* 396 (6706): 41–49.

Henderson, J Vernon, Adam Storeygard, and David N Weil. 2012. Measuring economic growth from outer space. *American Economic Review* 102 (2): 994–1028.

Hoffmann, Lars, Gebhard Günther, Dan Li, Olaf Stein, Xue Wu, Sabine Griessbach, Yi Heng, et al. 2019. From era-interim to era5: the considerable impact of ecmwf's next-generation reanalysis on lagrangian transport simulations. *Atmospheric Chemistry and Physics* 19 (5): 3097–3124.

Horns, Joshua J, Frederick R Adler, and Çağan H Şekercioğlu. 2018. Using opportunistic citizen science data to estimate avian population trends. *Biological conservation* 221: 151–159.

Huffman, George J, David T Bolvin, Dan Braithwaite, Kuolin Hsu, Robert Joyce, Pingping Xie, and Soo-Hyun Yoo. 2019. Nasa global precipitation measurement (gpm) integrated multi-satellite retrievals for gpm (imerg). *Algorithm Theoretical Basis Document (ATBD) Version* 4: 26.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jayadevan, Praveen, Rajah Jayapal, and Aasheesh Pittie. 2016. A checklist of the birds of india. *Indian BIRDS* 11: 113–172.

Johnston, Alison, Daniel Fink, Wesley M Hochachka, and Steve Kelling. 2018. Estimates of observer expertise improve species distributions from citizen science data. *Methods in Ecology and Evolution* 9 (1): 88–97.

Kaczan, David J. 2020. Can roads contribute to forest transitions? *World Development* 129: 104898.

Kelling, Steve, Alison Johnston, Aletta Bonn, Daniel Fink, Viviana Ruiz-Gutierrez, Rick Bonney, Miguel Fernandez, et al. 2019. Using semistructured surveys to improve citizen science data for monitoring biodiversity. *BioScience* 69 (3): 170–179.

Kelling, Steve, Alison Johnston, Wesley M Hochachka, Marshall Iliff, Daniel Fink, Jeff Gerbracht, Carl Lagoze, et al. 2015. Can observation skills of citizen scientists be estimated using species accumulation curves? *PloS one* 10 (10): e0139600.

Khemani, Stuti. 2004. Political cycles in a developing economy: effect of elections in the indian states. *Journal of development Economics* 73 (1): 125–154.

Kolstoe, Sonja, and Trudy Ann Cameron. 2017. The non-marketo tht value of birding sites and the marginal value of additional species: biodiversity in a random utility model of site choice by ebird members. *Ecological economics* 137: 1–12.

Lal, Apoorva, Saad Gulzar, and Ben Pasquale. 2021. Representation and forest conservation: Evidence from india?s scheduled areas.

Lee, Alexander. 2019. Land, state capacity, and colonialism: Evidence from india. *Comparative Political Studies* 52 (3): 412–444.

Liang, Yuanning, Ivan Rudik, Eric Yongchen Zou, Alison Johnston, Amanda D Rodewald, and Catherine L Kling. 2020. Conservation cobenefits from air pollution regulation: Evidence from birds. *Proceedings of the National Academy of Sciences* 117 (49): 30900–30906.

Liang, Yuanning, Ivan J Rudik, and Eric Zou. 2021. Economic production and biodiversity in the united states. Technical report, National Bureau of Economic Research.

Libecap, Gary D. 2011. Institutional path dependence in climate adaptation: Coman's" some unsettled problems of irrigation". *American Economic Review* 101 (1): 64–80.

Madhok, Raahil, and Sumeet Gulati. 2022. Ruling the roost: Avian species reclaim urban habitat during india's covid-19 lockdown. *Biological Conservation* : 109597.

Magurran, Anne E. 2013. *Measuring biological diversity*. John Wiley & Sons.

Meiyappan, Prasanth, Parth S Roy, Yeshu Sharma, Reshma M Ramachandran, Pawan K Joshi, Ruth S DeFries, and Atul K Jain. 2017. Dynamics and determinants of land change in india: integrating satellite data with village socioeconomics. *Regional environmental change* 17 (3): 753–766.

Ministry of Tribal Affairs. 2022. Monthly update on status of implementation of the scheduled tribes and other traditional forest dweller (recognition of forest rights) act, 2006.

MoEFCC. 1980. Forest (conservation) act, 1980.

———. 2003. Forest (conservation) rules, 2003, ministry of environment and forests notification.

———. 2013. Report of the comptroller and auditor general of india on compensatory afforestation in india.

Morrison, Michael L. 1986. Bird populations as indicators of environmental change. In *Current ornithology*. Springer, 429–451.

Munson, M Arthur, Rich Caruana, Daniel Fink, Wesley M Hochachka, Marshall Iliff, Kenneth V Rosenberg, Daniel Sheldon, et al. 2010. A method for measuring the relative information content of data from different monitoring protocols. *Methods in Ecology and Evolution* 1 (3): 263–273.

Noack, Frederik, Christian Levers, Johannes Kamp, and Ashley Larsen. 2021. A bird's eye view on farm size and biodiversity. *Mimeo* .

Noack, Frederik, Marie-Catherine Riekhof, and Salvatore Di Falco. 2019. Droughts, biodiversity, and rural incomes in the tropics. *Journal of the Association of Environmental and Resource Economists* 6 (4): 823–852.

Nunn, Nathan. 2009. The importance of history for economic development. *Annual Review of Economics* 1 (1): 65–92.

Odum, Eugene P. 1969. The strategy of ecosystem development: An understanding of ecological succession provides a basis for resolving man's conflict with nature. *science* 164 (3877): 262–270.

Oxford Economics. 2017. Global infrastructure outlook. Technical report, Global Infrastructure Hub.

Pacheco, Pablo, Karen Mo, Nigel Dudley, Aurelie Shapiro, Naikoa aguilar amuchastegui, Pui-Yu Ling, Christa Anderson, et al. 2021. Deforestation fronts: Drivers and responses in a changing world. Technical report, WWF.

Prasad, KV, V Vasugi, R Venkatesan, and Nikhil S Bhat. 2019. Critical causes of time overrun in indian construction projects and mitigation measures. *International Journal of Construction Education and Research* 15 (3): 216–238.

Raleigh, Clionadh, Rew Linke, Håvard Hegre, and Joakim Karlsen. 2010. Introducing acled: An armed conflict location and event dataset. *Journal of peace research* 47 (5): 651–660.

Salunkhe, Ashwini Arun, and Rahul S Patil. 2014. Effect of construction delays on project time overrun: Indian scenario. *Int. J. Res. Eng. Technol* 3 (1): 543–547.

Strimas-Mackey, M., W.M. Hochachka, V. Ruiz-Guiterrez, O.J. Robinson, E.T. Millet, T. Auer, S. Kelling, et al. 2020. *Best Practices for Using eBird Data. Version 1.0*. Cornell Lab of Ornithology, Ithaca, NY.

Sullivan, Brian L, Christopher L Wood, Marshall J Iliff, Rick E Bonney, Daniel Fink, and Steve Kelling. 2009. ebird: A citizen-based bird observation network in the biological sciences. *Biological conservation* 142 (10): 2282–2292.

Townshend, John, Matthew Hansen, Mark Carroll, Charlene DiMiceli, Robert Sohlberg, and Chengquan Huang. 2017. Annual global automated modis vegetation continuous fields (mod44b) at 250 m spatial resolution for data years beginning day 65, 2000?2014, collection 5 percent tree cover, version 6. Technical report, University of Maryland, College Park.

Trippi, Michael H, and Susan J Tewalt. 2011. Geographic information system (gis) representation of coal-bearing areas in india and bangladesh. Technical report.

Tsuda, Shunsuke, Yoshito Takasaki, and Mari Tanaka. 2023. Human and nature: economies of density and conservation in the amazon rainforest. *Working paper* .

Venkataraman, K., and C. Sivaperuman. 2018. In *Biodiversity Hotspots in India*.

Weitzman, Martin L. 1992. On diversity. *The Quarterly Journal of Economics* 107 (2): 363–405.

———. 1993. What to preserve? an application of diversity theory to crane conservation. *The Quarterly Journal of Economics* 108 (1): 157–183.

Wooldridge, Jeffrey M. 1999. Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics* 90 (1): 77–97.

World Rainforest Movement. 2019. Compensating for forest loss or advancing forest destruction? a study of compensatory afforestation in india.