# Infrastructure, Institutions, and the Conservation of Biodiversity in India

Raahil Madhok*
University of Minnesota

June 4, 2024

**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. For helpful comments, I thank Sumeet Gulati, Patrick Baylis, Frederik Noack, Siwan Anderson, Eyal Frank, Ryan Abman, Teevrat Garg, Meredith Fowlie, Sonja Kolstoe, Bianca Cecato, Tatiana Zarate, and the Wildlife and Conservation Economics Lab at UBC. I also thank seminar audiences at the World Bank, NBER Summer Institute (EEE), Vancouver School of Economics, Columbia IPWSD, CREEA, AERE, WEAI, PacDev, and Camp Resources. Veena Vinod provided excellent research assistance. I am grateful for the SSHRC CGS Doctoral Fellowship for generous funding. All errors are my own.

# 1 Introduction

Global infrastructure spending totalled $US 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 widescale 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. Their geocoded uploads serve as a new, high-resolution biodiversity repository unmatched in the literature. Third, India publicly reports forest encroachments by infrastructure. Deforestation for building roads, mines an 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, which I use to estimate the impact of infrastructure development on bird species diversity (hereafter, species diversity) in a typical Indian district. In addition, I decompose estimates by project category, ownership (public/private), and shape 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 both birds and users is the main threat 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. These tests help rule out concerns of endogenous mobility.

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, common and vulnerable species are the most sensitive to infrastructure-driven deforestation. I manually match bird taxonomies with their IUCN Red List status and count the number of times each user observes a common, vulnerable, or endangered species. Poisson regression estimates show a sharp decline in the abundance of common and vulnerable species following infrastructure expansion.

Lastly, 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 supports earmarking degraded landscapes for protection, advancing the debate about how to target conservation action (Betts et al., 2017).

The results are robust to a variety of stress tests and alternative samples. Estimates are stable when accounting for individual-level and sub-state seasonality as well as more flexible specifications of learning. To show that results are not driven by a changing user base, I show that estimates are similar on the subsample of users who signed up in 2015. Estimates are also similar when dropping users' home districts, suggesting that estimates are not biased by different behaviour at home and away. I also test robustness to more intricate diversity indices. Estimates remain negative but become noisy due to low-quality bird counts. Lastly, to show that estimates are unbiased by the switch to "balcony birdwatching" during COVID-19, I document stable estimates when dropping the year 2020.

The infrastructure-biodiversity tradeoff is also apparent under an alternative, widely implemented, identification strategy 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 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 and 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, 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 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. These findings are crucial given India's prioritization of Northeastern states—which have the higher forest cover and largest tribal populations in the country—for industrialization (Nayak et al., 2020).

**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, I document the extent to which infrastructure development erodes biodiversity.

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. The ecology literature has long documented anthropogenic pressures on ecosystems. In these studies, field workers often count species in transects with different levels of human activity. This yields accurate data but limits analysis to cross-sectional comparisons (Reis et al., 2012; Stephens et al., 2004). Although citizen science dramatically improves data coverage, most interest in the ecology literature has been in identifying endogeneity i.e., from seasonality, site choice, and detection ability, rather than conducting quasi-experiments (Callaghan et al., 2019; Kelling et al., 2019). I advance this literature by accounting for documented as well as undocumented biases, especially within-user learning, to arrive at quasi-experimental estimates of the infrastructure-biodiversity tradeoff.

The third contribution of this paper is to extend research at the intersection of political economy and conservation. I do this by showing empirically that inclusive institutions matter for natural resource 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 the impor-

[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), Noack et al. (2021) and Noack et al. (2019). The first studies pollution effects on bird abundance, whereas I focus on effects of habitat loss. The second studies the impact of farm size on bird species diversity. The third studies the impact of plant species diversity on income.

[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 infrastructure policy and resource depletion. Neither consider biodiversity per se. Noack et al. (2021) shows that different land

tance of institutions in moderating economy-environment tradeoffs, yet few have tested the claim credibly (Börner et al., 2020). I advance this literature by credibly estimating of the role of institutions in reducing species loss.

I am also able to fill a gap in this literature by pinning down mechanisms. Duflo and Pande (2007) use the same institutional data to show that the poverty impact of dams is muted in inclusive districts, and argue that the poor are better able to obtain compensation in these districts. Lee (2019) confirm that inclusive districts have better contemporary state capacity. My results point to a mechanism with "teeth", namely, higher rates of informed consent in inclusive districts. My paper therefore ties together this literature and shows that grassroots institutions are crucial for successful sustainable development.

The next section provides background on infrastructure-driven deforestation in India. Section 3 introduces the construction permit and citizen science data. Section 4 presents stylized facts to motivate the research design, described in Section 5. 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.

# 2 Background

**Forest Act (1980) Regulates Construction in Forests:** India's Forest (Conservation) Act (1980) protects forests from "conversion to non-forest uses" (MoEFCC, 1980). Infrastructure is among the main regulated activities because it fragments important habitats. For example, roads split contiguous forests into smaller patches, restricting wildlife movement and gene flow (Riley et al., 2006). Infrastructure also generates noise pollution, changes soil properties, and facilitates invasive species dispersal (Laurance et al., 2009).

When non-forest sites are unfeasible, the Act permits infrastructure encroachments pending a rigorous environmental review process. 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 in government records irrespective of ownership (public/private land), 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.

institutions in Germany led to differences in farm size which, in turn, impacts bird diversity.

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

**Informed Consent Required Since 2006:** Infrastructure encroachment is often at the expense of India's Scheduled Tribes. Numbering at 200 million, tribes are forest-dependent and conserve biodiversity through traditional knowledge. Yet, they have been excluded from development decisions since the Act historically gave powers only to the State. In 2006, the landmark Forest Rights Act (FRA) democratized forest governance by granting tribes forest management rights. These rights also prevent firms from diverting forests without informed consent from the Gram Sabha (village council)[5]. Despite its promise, the FRA is often overlooked (Menon, 2016). Empirically, weak implementation provides variation to study the merits of grassroots institutions (Section 7).

**Project Approval is Granted via Forest Clearances:** 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 environmental review. Stage-II is granted after FRA compliance and payment into a tree-planting fund. As discussed below, these funds are rarely disbursed and pose little threat to my research design.

To receive stage-I approval, the firm first submits a proposal to the District Forest Office (DFO). The DFO commissions a site inspection report, which is then forwarded to the State Forest Department. At this point, approval is granted to small projects (0-5 ha., except mining). Medium projects (5-40 ha.) are forwarded to the Regional Office (RO) and large projects ($> 40$ ha.) to the Ministry of Environment Forests and Climate Change (MoEFCC). The FAC[6] then rules on stage-I, and environmental review is complete.

To receive stage-II approval, firms pay a state-fixed rate into a compensatory afforestation fund. Despite potential to offset my findings, the fund is fraught with issues and tree-planting is rarely carried out. A recent audit found that just 7% of land secured for afforestation between 2006-12 had been planted in 2013 (MoEFCC, 2013). Other studies either find no impact of plantations on satellite forest cover (Coleman et al., 2021), or no existence of plantations whatsoever during field visits (World Rainforest Movement, 2019). Sanctioned afforestation thus poses minimal threat to my research design.

Another stage-II requirement is for the DFO to verify ST rights and submit evidence of Gram Sabha consent to the MoEFCC. After fundraising and FRA compliance, the central FAC makes their decision and, if approved, the firm begins deforestation. My data consists of firms with stage-II approval.

Whether *approved* deforestation can be taken to mean *actual* deforestation is an empirical question. In Section 4.1 I verify that stage-II project approvals trigger satellite-

[5] FRA guildelines: `http://forestsclearance.nic.in/writereaddata/public_display/schemes/981969732$3rdAugust2009.pdf`

[6] The regional FAC consists of senior RO and DFO officers as well as non-government forestry experts.

observed deforestation. References to approved project area hereafter can thus be interpreted as actual area of forest diversion.

**Projects Take Several Years to Complete:** Projects take many years to complete after permit approval. 60% of Indian projects experience time overruns, with dispute settlement between owners and contractors reported as the main delay (Salunkhe and Patil, 2014; Prasad et al., 2019). Disputes take around 6.5 years to resolve (Construction Federation of India, 2015). Given that my study period is 6 years, and that other delays may occur, it is unlikely that projects in my sample will complete construction during the study period.

Project non-completion is an advantage from an inference standpoint. It ensures that estimates capture species loss from habitat loss during construction, not through broader economic effects of industry activity once the project is operational.

**Forest Clearance was Fast-tracked and Digitized in 2014:** In 2014, prime minister candidate Modi promised to speed up the FC process, which the Environment Minister referred to as a "roadblock to growth". Delivering on this after election, the annual rate of project approvals jumped 60% compared to the previous 40 years (authors calculation). The share of rejected proposals also declined post-2014. Faster approvals materialized through easing norms, exempting certain projects from FRA compliance, and changing project size thresholds for state versus central review. Another cornerstone of the fast-track initiative is the digitization. An online portal automates much of the review process and reduces turnaround time. For researchers, an added benefit is data availability and process standardization, which reduces variation from state-level bureaucratic differences.

# 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.

## 3.1 Forest Clearances

**Data Collection:** Forests cover 22% of India (Figure 1A)(Forest Survey of India, 2019). As previously noted, these forests are being degraded by infrastructure encroachment. Administrative data on infrastructure-driven deforestation rarely exist, and previous work mainly relies on satellite data. Yet satellites have difficulty distinguishing anthropogenic

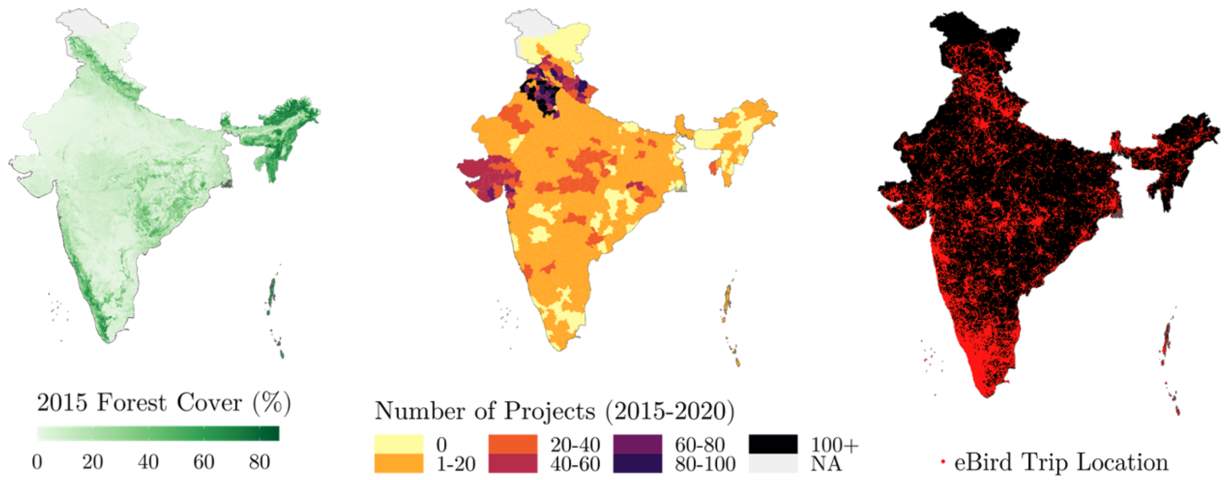


Figure 1: Infrastructure Encroachments and eBird Activity

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.

intrusions from natural sources (e.g. forest fires). Moreover, the best satellites report annual data, which mask within-year encroachments and their ecosystem impacts.

I construct a dataset of monthly infrastructure encroachments using newly digitized FC proposals approved between 2015-2020. Proposals submitted post-digitization and approved during the study period (N=6,597) were scraped (the digital subsample) [7]. 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. Figure A1 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.), the main variables for the analysis. 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[8]), and shape (linear, nonlinear). Linear projects are contiguous in terms of land displaced, such as roads. Nonlinear projects are non-contiguous, such as mines and dams. Digital applications also report whether a cost-benefit analysis was commissioned

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

[8] "Neither" are mainly joint public-private partnerships.

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 for 8,329 approved projects. Total area (ha.) is the summed deforestation area of all projects in the sample for the particular category.

and whether informed consent was obtained. This enables analysis of how local institutions mediate ecosystem impacts (Section 7). Appendix B1.1 provides more data details.

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 resettlement projects 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—which are those not in the listed categories—are most common, but reflect small patches (see Appendix B1.1 for more details)[9]. Transportation is the only category that is both numerous and accounts for a large (20%) share of total deforestation. Appendix B1.2 provides statistics by project ownership and shape.

**Panel Data Structure:** Project data are aggregated to the district and year-month level, both overall and category-wise (e.g. total deforestation in January 2018 for electricity projects in Delhi). For multi-district projects ($<$ 4% of sample) such as roads, the section falling into each district is listed as a separate row, and thus aggregation is over project *patches* in a district-year-month (Appendix B1.1). Aggregation is at the district level, firstly, because the district is the only consistent location identifier. Second, districts are the administrative unit immediately 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 (cream color in Figure 1B). This is reasonable since all projects undergo the FC process, and the full sample contains the universe of approvals.

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

## 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. Representative, here, means checklists best reflecting local species diversity at the user's location. The manual suggests keeping complete, stationary and travelling protocols (99% of sample) as well as lists collected in $<$ 5hrs and with group size $\leq$ 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 640). Figure 1C plots the trips: areas with high forest cover and development activity are popular among eBirders. 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.

**Summary Statistics and Sample Representativity:** Over 1,600 trips by 100 users are recorded in the average district during the study period (Table A1). Users are quite active: they typically record in four districts, two states, and six periods over the study period. This within-user variation is the cornerstone of my empirical strategy (Section 5). About 23 species are recorded during the average trip. Importantly, these species are observed while traversing 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 their data uploads are a reasonable guide to local biodiversity[10].

**Final Sample: Outcomes and Aggregation:** My main outcome measure is species richness, the number of unique species observed on a trip. Richness indicates conservation value, proxies the number and stability of ecosystem services, and is a widely used bio-

[10]Although there are no ground-truthing studies in India, several studies use a similar sample selection procedure to show that eBird data mirrors observations from nearby bird census sites in North America (Horns et al., 2018; Munson et al., 2010) and Australia (Callaghan et al., 2018). Ground-truthing the sample selection procedure, paired with wide spatial coverage of Indian eBird users, builds confidence that eBird data can deliver reliable statistical inferences about factors affecting biodiversity.

diversity metric (Fleishman et al., 2006). I also compute abundance metrics that count species according to their conservation importance. To do this, I manually code individual taxa reported in eBird with their corresponding IUCN Red List status: least concern, near-threatened, vulnerable, endangered, or critically endangered[11].

Despite having 1 million trips, the final panel aggregates users' trips in each district-year-month, which keeps relevant variation and reduces noise. Otherwise there would be no variation in development across trips in the same month and location since deforestation is district-monthly. Aggregation retains information from all 17,634 users but reduces the effective sample to 161,896, where each observation is at the user-district-month level. The outcome is mean species richness across users' trips in a district-year-month. I count of the number of trips underlying this mean, truncate at the 99th percentile to exclude outliers[12], and use these as regression weights in a robustness check. During aggregation, I also compute the *number of times* that each user recorded species in each IUCN category in a district and year-month. This enables investigation of whether changes in species diversity are being driven by common, vulnerable, or endangered species.

One concern is that aggregating users' trips in a district masks within-district sorting. I design a test for sorting in Section 5.3.3 and show that users do not shuffle within districts following project approval. This suggests they do not birdwatch near construction sites in the first place, though I am unable to verify this since projects are not geocoded.

**Who uses eBird?** To estimate the infrastructure-biodiversity tradeoff, it is important that eBirders collect representative data moreso than themselves be representative of the population. The latter is implausible since users have smartphones and the privilege to engage in recreation. Yet citizen science is becoming commonplace in conservation studies, making it important to frame a deeper discussion about who these citizen scientists are.

Since eBird does not record demographics, I characterize users by matching their approximate home locations to the nationally representative Demographic and Health (DHS) survey. This method, first proposed by Blanchard et al. (2023), reveals how residents of users' home locations compare to the typical Indian. Perhaps unsurprisingly, I find that users are from more urban and better-off places. Yet despite statistical differences, they are not atypical along either of these dimensions. Appendix B3 elaborates and provides supporting results (Figures B1-B2, Table B5).

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

[12] As an example outlier, the maximum is 3779 trips in a district-month by one user i.e., 126 trips *per day*.

### 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 and duration, protocol, group size, and spatial coverage. Duration (minutes) and distance (km) are recorded by eBird. Protocol equals one if the user is moving and zero if stationary. Group size is the birdwatching party size[13]. 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).

**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 A2 regresses log nightlights on project permits at the district-year level and reveals no correlation even up to two years later. Point estimates are near zero. This implies 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).

# 4 Empirical Patterns

I next present four stylized facts that make the data ideal for studying the infrastructure-biodiversity tradeoff in India. The first verifies that project approvals trigger real defor-

[13]If multiple eBird users birdwatching together, only one needs to record a checklist. This is shared at the end, and the others can edit it to include additional species they observed.

Table 2: Correlation between approved and actual deforestation

| | Log-Log | Log-Log | Linear-Linear |
|---|---|---|---|
| | (1) | (2) | (3) |
| Log(Infrastructure+1) | -0.024** | -0.024** | |
| | (0.010) | (0.010) | |
| Infrastructure ($km^2$) | | | 1.033 |
| | | | (1.214) |
| Nightlights | No | Yes | Yes |
| District FEs | ✓ | ✓ | ✓ |
| State × Year FEs | ✓ | ✓ | ✓ |
| Observations | 3822 | 3822 | 3822 |
| $R^2$ | 0.988 | 0.988 | 0.985 |

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

estation. The second and third illustrate shortcomings and remedies when 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

Using deforestation permits to study biodiversity impacts requires an assumption that authorized deforestation equates to actual deforestation. I test this assumption using validation data that measures forest cover from outer space. Annual forest cover (% of a pixel) is from the VCF satellite product and converted to $km^2$ (Townshend et al., 2017)[14]. Since validation data is annual, I estimate the following equation on aggregated data:

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

where $d$ indexes districts and $t$ indexes years. $ForestCover$ is *actual* forest cover and $Infrastructure$ is the cumulative area of *approved* deforestation. $X'_{dsy}$ is a control for night-light intensity, which helps disentangle infrastructure from other anthropogenic drivers of forest loss. $\gamma_d$ and $\theta_{st}$ are district and state-year fixed effects, respectively. $\beta < 0$ tests whether approved deforestation translates into actual deforestation.

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

I specify a log model as the main specification since $Infrastructure_{dst}$ is an aggregate over *differently sized* projects approved in a district. Small encroachments are unlikely to be captured by the satellite, introducing noise at low values of $Infrastructure$. Also, if small projects are sited in highly forested areas (and overlooked by the satellite), then $\hat{\beta}$ 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.

Forest cover declines as districts approve more projects (Table 2). A 1% increase in approved deforestation results in a .02% decline in actual forest cover (column 1). This is a reasonable magnitude since infrastructure is just one of many sources of deforestation in a district. The result is robust to including nightlights as a covariate (column 2). The coefficient becomes positive and noisy under a linear specification (column 3). This is likely misspecified and coincides with the explanation in the previous paragraph.

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

Citizen science is revolutionizing biodiversity monitoring. Yet, loose restrictions on when, where, and by whom data are collected yields more endogeneity than most administrative data. eBird records details on the observation process, helping ease such concerns.

**Seasonality:** The seasonality bias arises from the ability to record trips at any time. Figure 2A shows stark seasonality in collective species richness (left axis), with a peak in winter when Siberian birds migrate to India, and a trough during lulls in activity (right axis) during monsoons. I address seasonality by exploiting within-month variation so that all time-invariant differences across months, such as seasonal species fluctuations, are eliminated. I do this separately by state since migratory patterns vary regionally. I also test cases where seasonality is assumed at the sub-state level. Neither approach addresses sub-state seasonality from climate-induced shifts in bird migration. This requires district-yearmonth fixed effects, which is the level of treatment. This concern is minimal, though, since climate adaptation occurs over time spans longer than my study period.

**Site Selection:** The site selection bias arises from the ability to record species from anywhere, triggering a convergence of 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 biodiverse 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.

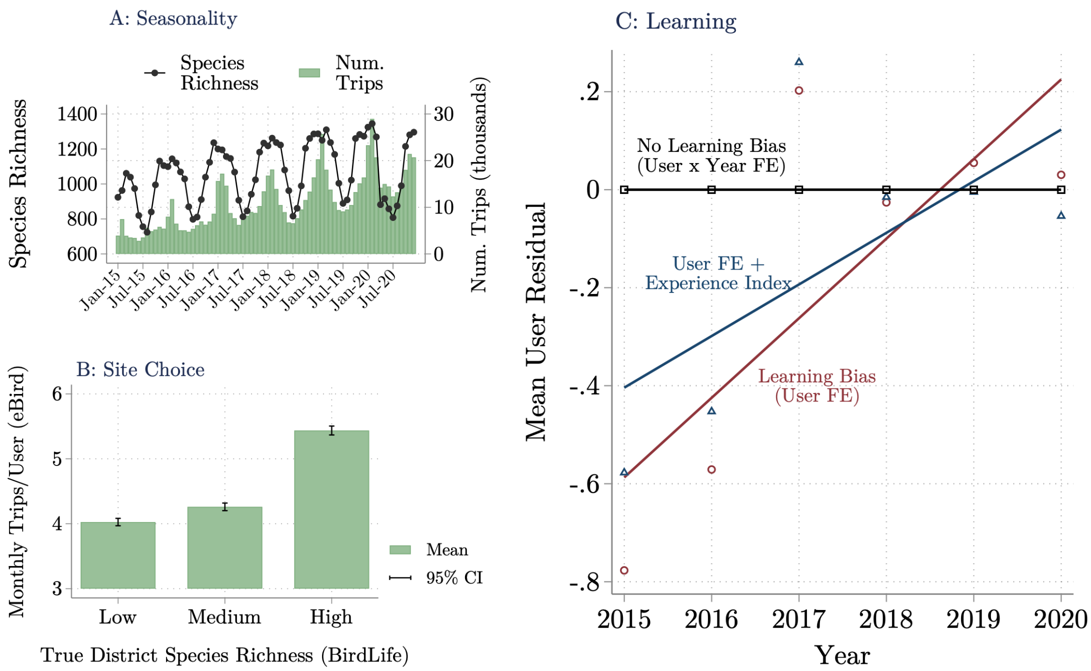


Figure 2: Biases in Citizen Science

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 heterogeneous abilities and learning rates (Farmer et al., 2014; Fitzpatrick et al., 2009). To account for low-ability users potentially misidentifying certain species, Kelling et al. (2015) construct a fixed ability score for eBird users based on a random effects model. For causal inference, this requires a strong orthogonality assumption between the score and other unobserved user attributes. Instead, I relax this assumption by comparing species richness across time and space *within the same user*, making the ability score superfluous.

Red circles in Figure 2C show species richness residualized on user, state-month, and district fixed effects. An upward trend remains, evidence of a potential user-specific learning curve. Blue 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 not only driven by short term experience, but also by longer-term unobservables (see gradual upward trend in Figure 2A). For example, a novice may detect the same common species

Table 3: Variation in Species Richness Under Various Fixed Effects

| | $1 - R^2$ | $\sigma_\epsilon$ |
|---|---|---|
| | (1) | (2) |
| District FE | 0.825 | 16.998 |
| District + State-Month + Year FE | 0.806 | 16.798 |
| User + District + State-Month + Year FE | 0.515 | 13.418 |
| User-Year + District + State-Month FE | 0.441 | 12.401 |

Note: This table summarizes regressions of species richness on sets of fixed effects (rows). Data are at the user-district-month level. Column 1 reports $1 - R^2$ i.e., the fraction of variation not explained by the fixed effects. Column 2 is the standard deviation of the residuals (units = number of species).

month-to-month, gradually listing rarer species after learning their songs.

My solution hinges on restricting variation from within-user to within-user-by-year. This has three advantages. First, user-specific annual trends are removed, including accumulated trips, number of months per year of activity, and other longer-term learning indicators. Second, it is agnostic about the shape of the learning curve. A line of best fit (Figure 2C) is one possibility, but the true shape is unknown. Lastly, it allows for differential learning rates between users. Partialling out user-by-year fixed effects mechanically flattens the trend (black line), implying that residual variation has been stripped of the learning bias. This is the variation that I use to estimate the impact of infrastructure encroachment on biodiversity 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 A1 showed that users visit multiple districts and states over the study period. Figure A2 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 3 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 to quantify the infrastructure-biodiversity tradeoff. Development projects fragment district forests, and eBird users venture to these districts to record birds. My specifications compare species diversity *within* users' trips as they travel for birdwatching. 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:

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

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 $m$. I also estimate versions where the outcome is a count of common, vulnerable, and endangered species observed by user $i$. $Infrastructure_{dsym}$ 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 within the district. User-by-year fixed effects, $\phi_{iy}$, absorb cross-user differences in ability and account for individual-specific learning. District fixed effects, $\gamma_d$, absorb mean checklist length and other other fixed district attributes. State-by-month fixed effects, $\theta_{sm}$, control for state-specific seasonality.

**Identifying Variation:** $\phi_{iy}$ 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 identification is from cross-sectional comparisons. In general, users are more active than these limiting cases (Figure A2) 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, $\beta_1$ 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.

**TWFE Estimator:** $\beta_1$ is estimated via TWFE. To understand why this estimator applies to my context, note that some districts receive continuous "doses" of infrastructure in each period, whereas others remain untreated. $\beta_1$ thus represents a dose response to infrastructure at the district level, rather than the average effect across users, since changes in species richness are not private to user $i$; it is best interpreted as the slope of such a dose-response curve, where users simply report different levels of response. Callaway et al. (2024) show that if the dose-response function is heterogeneous, then a stronger parallel trends assumption is required: the average change in species richness across districts at a given level of infrastructure is the same as all districts would experience, on average, if they all experienced that dose. If this holds, and the dose is normally distributed, then a TWFE estimator recovers a weighted average of slopes that approximates the average causal response (Callaway et al., 2024). Indeed, Figure A3 shows that conditional on fixed effects, $Infrastructure_{dsym}$ is slightly more peaked but otherwise resembles a normal distribution. I discuss other identification assumptions in Section 5.3.

**Clustering:** Standard error clustering is an experimental design issue in this analysis, which leaves the choice of cluster somewhat subjective (Abadie et al., 2017). I cluster at the biome level 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 $\epsilon_{idsym}$ across time and space within a biome. Maps of India's 12 biomes are obtained from the Nature Conservancy[16] and plotted in Figure A4. For districts spanned by multiple 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 biodiver-

[15] The fact that 10 users are active in the typical district-month, together covering 20% of district area (Table A1), helps ensure that the local species pool is reported, even if one user misses a species.

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

sity 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

**Decomposed Specification:** 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:

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

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. $\beta_{1k}$ measures 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).

**Treatment Heterogeneity:** There is debate among conservationists about whether biodiversity is better conserved by protecting intact or already-fragmented landscapes. I help resolve this by investigating whether the same infrastructure intrusion has different effects by baseline ecosystem quality. I estimate treatment heterogeneity as follows:

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

where $EQ_d$ is a fixed measure of ecosystem quality in district $d$. It is measured with pre-period 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 remote, 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. $\beta_2$ reveals whether the infrastructure-biodiversity tradeoff is accentuated or muted in more pristine landscapes, independent of project type.

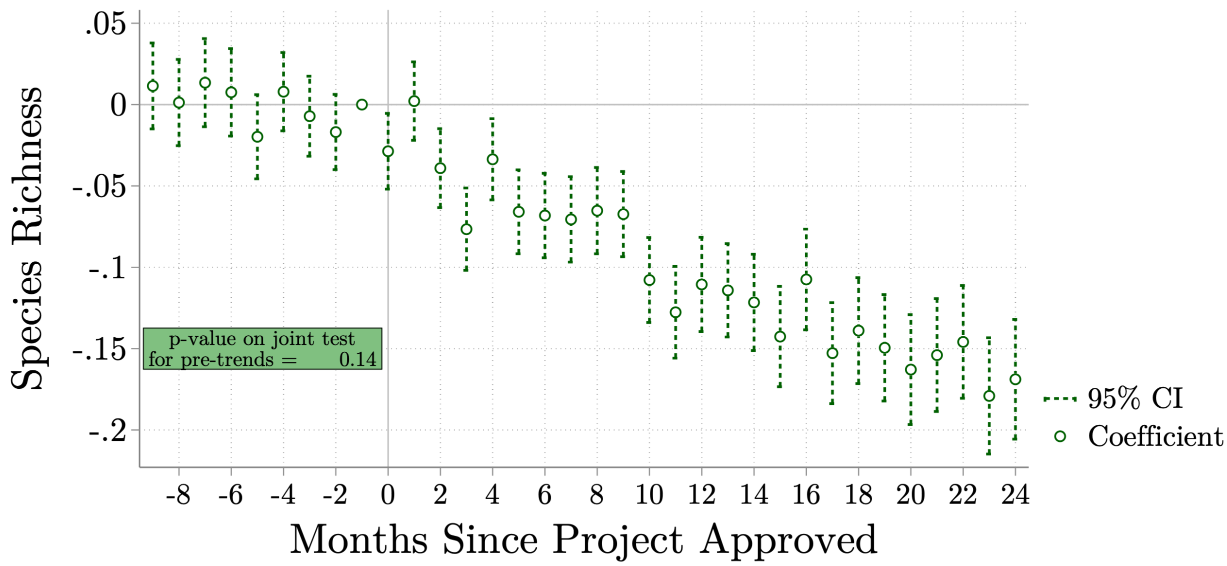


Figure 3: Event Study Results

Note: Data are a stacked project-district-month panel. Dark green circles are coefficients from an event study regression (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 omitted period is -1. Dotted lines are 95% confidence intervals. 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 project.

## 5.3 Identifying Assumptions

### 5.3.1 Assumption I: Parallel Trends

Without an established test for the strong parallel trends assumption (Callaway et al., 2024), I instead test the standard version: 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. Two challenges arise in testing this. First, treatment is continuous and comprises multiple project approvals in a district. Second, treatment is cumulative i.e., the post-period of one project in a district is the pre-period of the next one. I address these challenges by stacking project data into a balanced project-district-month panel, which enables estimation of 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 $\tau$ months relative to the date of project approval. Lack of pre-trends are indicated by $\beta_{1\tau} = 0 \; \forall \, \tau < 0$ (individual test), or by failure to reject the null that pre-period coefficients are jointly zero (joint test).

Figure 3 displays estimates of Equation 5. Visually, coefficients fluctuate tightly around zero during the pre-period and show minimal trend. Formally, the p-value for a joint significance test implies that pre-period coefficients are jointly indistinguishable from zero. Yet two months after approval, coefficients turn sharply negative and continue downward. Section 6.5 elaborates on the persistence of species loss. Overall, lack of pre-trends support the parallel trends assumption and suggests that project siting is generally unpredictable. For example, if projects were selectively approved in high-growth districts, then species diversity would trend downwards prior to project approval. Thus, while the FAC may approve projects based on certain district characteristics, these characteristics appear unrelated to changes in local species diversity around the time of approval.

#### 5.3.2 Assumption II: No Spatial Spillovers

$\beta_1$ 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 since habitat loss triggers species dispersal to other districts, introducing spatial links unmodeled in Equation 2. The severity of the violation is unclear: it is zero if species relocation is random, since spillovers would be orthogonal to infrastructure development. It is positive if species disperse to less fragmented districts. However positive bias converges to zero if dislocated species are already found in destination districts.

I address SUTVA by modelling spillovers with a spatial matrix, $\mathbf{W}$. This transforms $Infrastructure_{dsym}$ into a spatially-weighted average of development within a pre-specified radius around district $d$, known as a "spatial lag of X" (SLX) (Elhorst and Vega, 2015):

$$SLX_{dsym} = (I_T \otimes \mathbf{W}_D) \cdot Infrastructure_{dsym} \tag{6}$$

$\mathbf{W}_D$ is a symmetric $D \times D$ matrix where $D$ is the number of districts in India. $I_T$ is a $T \times T$ identity matrix where $T$ is the number of year-months in the study period. The kronecker product signifies that $\mathbf{W}_D$ is applied to the infrastructure variable in each period and then stacked into a panel of spatially lagged infrastructure development.

I delineate several boundaries over which spillovers can materialize. Most realistically, I allow disturbances to disperse birds across districts *in the same biome*, but less so to further away districts. This is modelled by $w_{dj} \in \mathbf{W}_D = \frac{1}{distance_{dj}}$, where $distance_{dj}$ is distance from district $d$ to $j$ if in the same biome and zero otherwise. Since some species exist in multiple biomes, I test robustness to spillovers within arbitrary, potentially biome-spanning, circles of radii up to 500km. Finally, I add $SLX_{dsym}$ to Equation 2. The coefficient captures changes to species diversity in district $d$ when other districts in the biome become relatively more fragmented. Conditional on this, $\beta_1$ 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. Since mobility is a key source of variation, $\beta_1$ is unbiased only if unobserved determinants of species diversity are conditionally uncorrelated with site choice. Users may exhibit fixed, heterogeneous preferences over nature (e.g. enjoying hiking) that affect site choice. This does not bias $\beta_1$ since they are absorbed by user fixed effects. The orthogonality assumption is only violated if mobility reflects systematic *changes* in unobserved preferences. Suppose deforestation causes sorting towards more pristine districts. District fixed effects account for overall biodiversity, user fixed effects account for static preferences, but neither accounts for deviations from within-user species diversity as they sort across districts. These deviations are non-random if sorting correlates with project development.

I test for *cross-district* sorting on aggregated data with the following equation:

$$Users_{dsym} = \alpha + \beta_1 Infrastructure_{dsym} + \beta_2 SLX_{dsym} + \Gamma X'_{dsym} + \gamma_d + \theta_{sm} + \mu_y + \epsilon_{dsym} \quad (7)$$

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 fourth term is the spatial lag defined in Equation 6, except $w_{dj} = \frac{1}{distance_{dj}}$ varies across the full support of $\mathbf{W}_D$ (instead of only within-biome). This model choice enable users to sort anywhere in India following project construction in district $d$, but with lower probability towards further destinations. I also test specifications that assign zero sorting probabilities to districts beyond 100km, 200km, and 500km. 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 7 with spatial coverage as the outcome and omitting $SLX_{dsym}$. Spatial coverage is the percent of district grid cells (10km resolution) "birdwatched" by the average user in district $d$ at time $ym$. 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, corroborating the orthogonality assumption and improving confidence in the TWFE design.

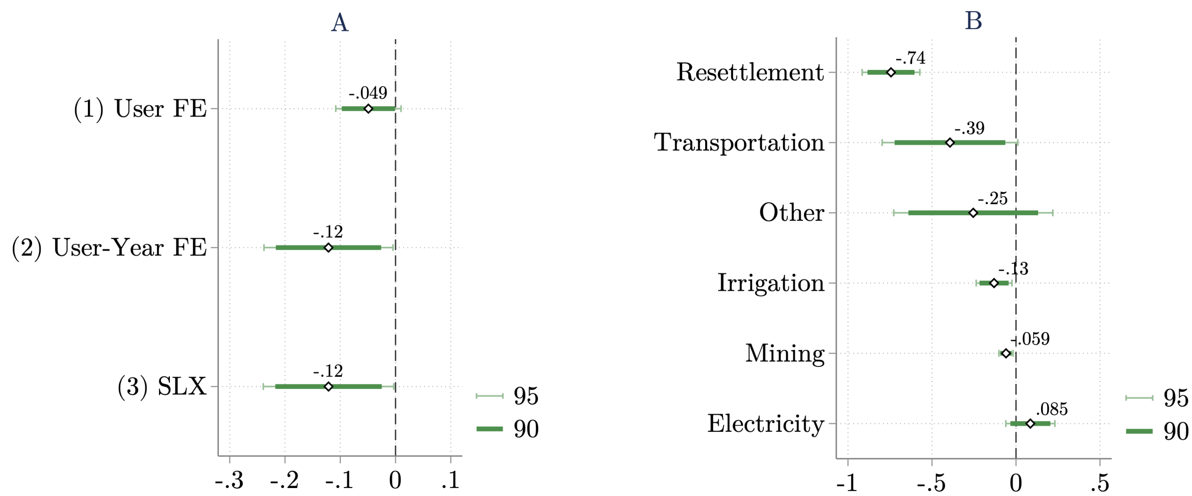


Figure 4: Estimates of the Infrastructure-Biodiversity Tradeoff in India

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 fixed effects for user, district, state-month, and year. Specification (2) includes user-by-year, district, and state-month fixed effects. Specification (3) controls for spatial spillovers within the biome (Section 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 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 (see Table A3 for tabulated estimates). Specifications (1) and (2) estimate Equation 2 with and without the learning curve, respectively. The main coefficient ($\beta_1$) 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 forest infrastructure 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* occupied 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. Column 3 of Table A3 shows that, conditional on the direct effect, the spillover coefficient is positive but insignificant. Table A4 tests robustness to allowing spillovers to materialize over different distances, ranging from 100km to nationwide. 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 A3 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 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 A5 tests for cross-district sorting by estimating Equation 7 on the same user sample that identified species impacts in Figure 4A. The outcome is log number of users. 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 7. The coefficient is negative and insignificant, suggesting that eBird users continue visiting the same birdwatching locations as districts undergo development. This is an indication 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[17]. 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 $1km^2$ 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 $1km^2$ 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 may explain the noisy impact of electricity projects, which have the second highest noise-

[17]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.

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 miscount the species pool, despite my attempts to prevent this (Section 3.2).

Table A6 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 B2 present additional results by project ownership and shape (Table B4).

## 6.3 Estimates by IUCN Threat Status

Having established that infrastructure drives species loss, I next investigate which species are under threat. Table 4 shows estimates of Equation 2 with the outcome measured as counts of common, vulnerable, and endangered species[18]. These estimates 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.

These results suggest that the decline in species *diversity* in the main estimates (Fig-

[18]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 4: 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 group size, and log spatial coverage. Standard errors clustered by biome.

ure 4A) is driven by lower *abundance* of common and vulnerable species. The negative estimates in columns 3 and 4 are particularly concerning since these species are already vulnerable according to the IUCN. Given that site monitoring is required during project review (Section 2), my findings highlight the need to monitor vulnerable taxa to prevent these species from being further down-listed.

### 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 5 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[19] (Column 1). To account for potential selection of 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 dis-

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

Table 5: 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 ($km^2$) × Baseline Forest Cover | 0.065* | 0.056 | | |
| | (0.030) | (0.032) | | |
| Infrastructure ($km^2$) × Baseline Species Richness | | | 0.052** | 0.046** |
| | | | (0.020) | (0.019) |
| Infrastructure × Category Shares | No | Yes | No | Yes |
| Controls | Yes | Yes | Yes | Yes |
| User x Year FEs | ✓ | ✓ | ✓ | ✓ |
| District FEs | ✓ | ✓ | ✓ | ✓ |
| State × 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 also at the district level and measured 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 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.

tribution. 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). 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).

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

My estimates reflect how species respond to habitat loss within the month. This overlooks lagged effects, either if it takes time for species diversity to equilibrate (Odum, 1969), or, if there is simply 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

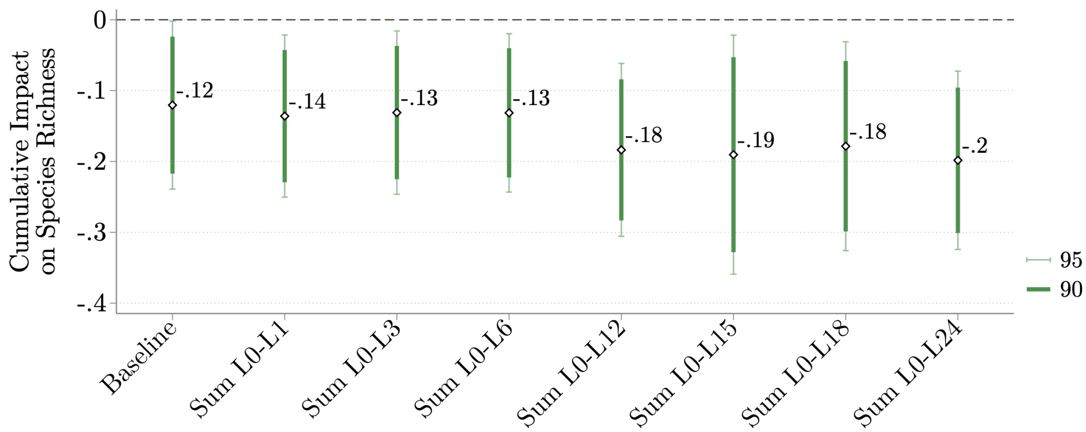


Figure 5: Cumulative Dynamic Lag Results

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 group size, and log spatial coverage. Standard errors clustered by biome.

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. 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 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 feature a U-shape as species return after the disturbance dies down. However, in this context, once the forest patch is replaced by infrastructure, there is little room for regrowth. Second, the dynamic results also highlight the ineffectiveness of compensatory afforestation requirements. Compensatory afforestation would also be indicated by 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 demonstrates robustness to: alternative forms of seasonality (Table A7), different data samples, different diversity metrics (Table A8) and spatial clustering (Table A9). The next section shows robustness to an alternative research design.

**Alternative Fixed Effects: Seasonality, Skill, and Location:** Table A7 tests robustness to alternative fixed effects. Columns 1-4 specify alternative forms of seasonality. Column 1 includes user-by-month fixed effects, which relies on comparisons across districts and years within a user-month. This accounts for seasonality exhibiting an individual component i.e., winter migratory species mainly reported by experts (Johnston et al., 2018). Columns 2 and 3 include district-month fixed effects, which accounts for sub-state seasonality[20]. Column 4 uses biome-by-month fixed effects in case seasonality is biome- rather than state-specific. All four estimates are strikingly similar to the baseline estimate, suggesting that biases from individual and regional seasonality are negligible.

Column 5 tests another way of accounting for skill. User-year fixed effects do not account for within-year learning, nor learning among users active for less than one year. Column 5 therefore adds fixed effects for cumulative number of trips per user. Column 6 adds time-of-day fixed effects to account for different species availability and user activity throughout the day[21]. Estimates are virtually unchanged from the baseline finding.

Column 7 tests robustness to sorting by dropping the user fixed effect and adding a fine-grained location fixed effect. Throughout the paper, user fixed effects remove biases from user heterogeneity, but requires a no-sorting assumption for identification. The reverse is to track species diversity in a fixed location as it develops, which obviates this assumption, but then pools checklists from heterogeneous users. I implement this with $10km^2$ cell fixed effects, which guards against sorting even within districts. Raw eBird data are aggregated to the user-cell-month level. Species loss is still observed, but estimates are noisy. This is likely from pooling heterogenous users within cells. Although effort covariates remove some heterogeneity, differences in fixed unobservables (e.g. ability) inflate the error. My chosen design, with user fixed effects, solves this issue and, coupled with evidence of minimal sorting (Section 6.1), delivers credible estimates.

**Sample Restrictions: fixed user base, home vs. away, and non-COVID years:** Table A8 columns 1-4 show estimates from alternative samples. Whereas the main specification

[20]Column 2 of Table A7 uses user-year and district-month fixed effects. This is likely too saturated to yield precise estimates. Column 3 includes user and district-month fixed effects as a compromise.

[21]Time-of-day categorizes mean hour-of-day for user trips in a district-month. Categories are: morning: 6am-12pm; afternoon: 12pm-6pm; evening: 6pm-12am; night: 12am-6am.

accounts for within-user behaviour changes, it does not account for a changing user base. Column 1 thus fixes the sample to 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 B3 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.

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.

**Outliers: Dropping mega-projects and IHS:** 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[22] 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 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 (arcsinh) transformation[23] to $Infrastructure_{dsym}$ (Bellemare and Wichman, 2020). There are two advantages: first, $arcsinh(x)$ is defined at $x = 0$, which is common in districts with no forest or no projects (Figure 1B). Second, it mimics the natural log and thus reduces the influence of outliers. 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: Shannon and Simpson Index:** Species richness has been criticized for its simplicity. Somewhere with one pigeon and 99 crows, and another

[22]Lift irrigation is where water is transported by pumps rather than by exploiting natural flow.
[23]This uses the function $arcsinh(x) = ln(x + (x^2 + 1)^{1/2})$.

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. A limitation of implementing these indices is that bird counts are notoriously imprecise given difficulties with 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. 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 A7 shows 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.

**Standard Error Clustering:** Table A9 shows the baseline estimates adjusted for alternative clustering. Column 1 replicates the baseline, whereby unobservable biophysical determinants of species diversity are assumed to be correlated within biomes, even though treatment varies by district. Columns 2 and 3 show that estimates are quite similar when clustering by district or state, respectively. Clustering by state is a compromise between large biome clusters (Figure A4) and smaller districts. Columns 4-7 investigate spatial correlation more systematically by implementing Conley (1999) standard errors for several choices of the kernel cutoff distance. Reassuringly, precision remains similar, even when allowing for long distance spatial correlation up to 1000km.

**Robustness by Project Category:** Tables A10 and A11 report the same robustness checks with infrastructure decomposed by project category. Coefficients on most categories remain negative across the most stress tests. The negative effect of irrigation, transportation, mining, and resettlement projects is robust to alternative fixed effects and most other robustness tests. Electricity and Other projects continue to have no impact.

### 6.7 Robustness: Instrumental Variable Estimates

Despite the robustness checks, parallel trends, and no evidence of endogenous sorting, causal interpretability of estimates may remain in question since infrastructure is non-random. Next, I show that results are robust to an IV design based on close races between incumbent and runner-ups in State elections (Appendix B4 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.

First-stage variation derives from the idea that incumbents target resources toward their supporters (Dixit and Londregan, 1996)[24]. This, however, leads to a bundled estimate of predicted projects: if voters prefer less deforestation, then incumbent winners may *reduce* project approvals as a reward. At the same time, incumbents have been shown to increase public investment overall (Khemani, 2004), which may involve approving *more* projects. I control for nightlights to partially capture the investment channel, which leaves identification to rely on the assumption that, conditional on controls and fixed effects, district-level incumbent strength affects local biodiversity only by sanctioning forest diversion for infrastructure. I acknowledge 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 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 just barely lost:

$$\text{First Stage:} \quad Infrastructure_{dsym} = \mu_1 IC_{dsy} + \mu_2 C_{dsy} + \mu_3 (f(M_{dsy}) \cdot I_{dsy}) + \mu_4 f(M_{dsy}) + \mu_5 I_{dsy} + \mu_6 E_{sy} + \Gamma X'_{idsym} + \phi_i + \gamma_d + \theta_{sm} + \nu_y + \epsilon_{idsym} \tag{8}$$

$$\text{Second Stage:} \quad SR_{idsym} = \beta_1 \widehat{Infrastructure}_{dsym} + \beta_2 C_{dsy} + \beta_3 (f(M_{dsy}) \cdot I_{dsy}) + \beta_4 f(M_{dsy}) + \beta_5 I_{dsy} + \beta_6 E_{sy} + \Gamma X'_{idsym} + \phi_i + \gamma_d + \theta_{sm} + \nu_y + \epsilon_{idsym} \tag{9}$$

[24] The literature distinguishes patronage i.e., awarding incumbent-supporting areas irrespective of political goals, and tactical redistribution, which is to achieve political goals. I am agnostic about the motivations.

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 most recent election. Elections are close if the win margin is $< 2$ percentage points (p.p.). Note that while the *outcome* of close elections is random, its *existence* may not be. I therefore control for $C_{dsy}$, the share of close-election constituencies in the district where the incumbent ran, to disentangle the election outcome from the factors that led to a close election in the first place. $M_{dsy}$ is the mean win margin (close or not), which I include 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 demanding here since eBird users are unlikely to traverse many closely contested districts in a given year. I thus include user and year fixed effects separately.

Figure A5 plots the first stage (Panel A) and reduced form (Panel B)[25]. A negative margin means that the incumbent lost. The first stage shows a sharp discontinuous decrease in forest encroachments when incumbents win, accompanied by an increase in species richness. Table A12 Column 1 presents 2SLS estimates of Equation 9. Columns 2 and 3 vary the close election bandwidth whereas columns 4 and 5 test sensitivity to using a second- and third-order polynomial in the win margin. Coefficient estimates are negative and significant across the board and first-stage F-statistics are near rule-of-thumb levels; the infrastructure-biodiversity tradeoff is thus robust to this IV design. When user- and year-fixed effects are interacted (column 6), coefficient sign and magnitude remain very similar but precision declines (p=0.11) due to the demanding specification.

Coefficient size implies larger species loss compared to the TWFE estimates. However, a direct comparison cannot be made since the 2SLS and OLS estimates apply to different populations[26]. More important is that the two designs agree in terms of coefficient sign and statistical significance. This helps build confidence that my main findings capture a robust relationship between infrastructure development biodiversity loss.

[25]Figure A6 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).

[26]2SLS measures the LATE for the subsample that took up the treatment (the "compliers"). OLS is an ATE for the compliers and the subsample that would have taken up the treatment regardless ("always-takers") and those who would have never taken up the treatment ("never-takers"). Since projects are not random, I cannot distinguish these subsamples, in which case directly comparing LATE and ATE is misleading.

# 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 start 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 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 this paper, I re-conceptualize *raiyatwari* and *zamindari* districts as inclusive and extractive, respectively. If disaffected groups are better able to engage in the development process and protect their livelihoods in inclusive districts, then the adverse environmental impacts of infrastructure development should be smaller in these districts. I provide evidence for this mechanism in Section 7.3.

## 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:

$$\begin{aligned} SR_{idsym} = \alpha + \beta_1 \cdot Infrastructure_{dsym} + \beta_2(Infrastructure_{dsym} \cdot Inclusive_d) \\ + \Omega(Infrastructure_{dsym} \cdot X'_d) + \Gamma X'_{idsym} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idsym} \end{aligned} \tag{10}$$

[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).

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 $Infrastructure$. $\Omega$ 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 $\beta_1$ and $\beta_2$, which captures the main infrastructure-biodiversity tradeoff, and any moderation of the tradeoff depending on the type of local institution. 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, 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 $\beta_2$ is biased. The interaction coefficient $\Omega$ 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 10 are in Table 6. 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 $\beta_2$ (Section 6.4, Table 5).

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 $\beta_2$ 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 A13 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 substate 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

Table 6: The Impact of Infrastructure on Biodiversity by Institutional Type

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $\beta_1$: Infrastructure ($km^2$) | -0.551$^{***}$ | -0.447$^{***}$ | -0.434$^{***}$ | -0.394$^{***}$ | -0.551$^{**}$ |
| | (0.067) | (0.051) | (0.073) | (0.082) | (0.214) |
| $\beta_2$: Infrastructure ($km^2$) | 0.434$^{**}$ | 0.340$^{*}$ | 0.315$^{*}$ | 0.421$^{**}$ | 0.434$^{***}$ |
| $\times$ Inclusive (=1) | (0.116) | (0.144) | (0.134) | (0.129) | (0.142) |
| Infrastructure $\times$ Tribal Share | Yes | Yes | Yes | Yes | Yes |
| Infrastructure $\times$ Baseline Forest | Yes | Yes | Yes | Yes | Yes |
| Infrastructure $\times$ High-Activity | No | No | No | Yes | No |
| $\beta_1 + \beta_2$ | -0.116 | -0.107 | -0.119 | 0.027 | -0.116 |
| | (0.159) | (0.151) | (0.201) | (0.176) | (0.257) |
| User $\times$ 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 group size, and log spatial coverage.

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 impor-

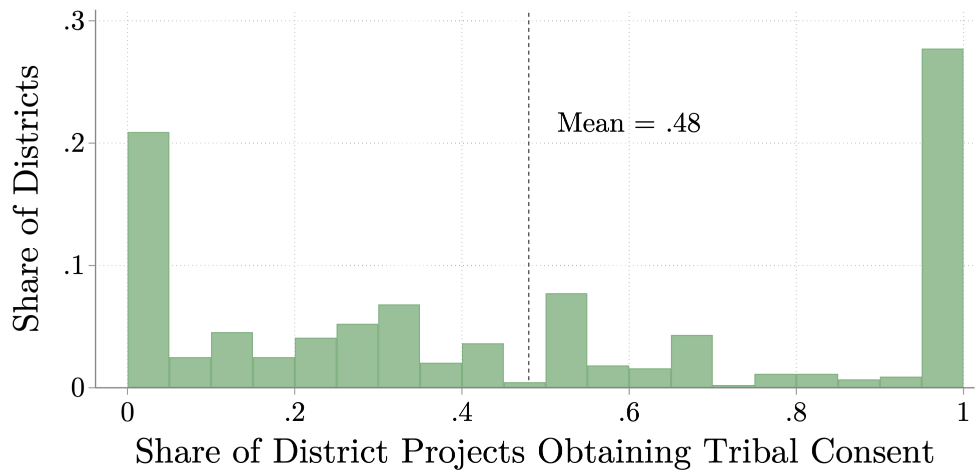


Figure 6: Enforcement of Forest Rights Act (2006)

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).

tant 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). If the policy is binding, however, then there would be no variation. Recent reports indicate that FRA enforcement is weak, often bypassing consent altogether (Dubey et al., 2017). The project sample reports whether consent was obtained (Figure A8). Figure 6 shows the distribution of projects obtaining Gram Sabha consent. The lack of right-tail bunching is evidence of imperfect compliance. There are districts where inclusive development is always, sometimes, and never observed. I exploit this fact to study if inclusive institutions, as defined by Banerjee and Iyer (2005), 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.

[28] Value judgment is used for projects $>$ 20 ha., which is more than 90% of my projects.

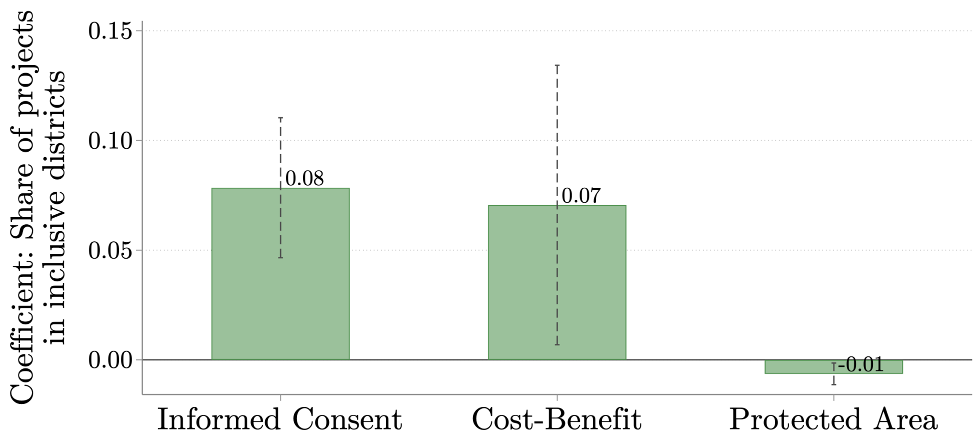


Figure 7: Mechanisms: Informed Consent, Cost-Benefit Analyses, and Project Placement

Note: Data are at the project level. Bars represent coefficients from Equation (11). 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 commissioned during project review. Protected Area equals one if the project is sited in or near one. All specifications control for: project size, district tribal population share, baseline forest cover, and district area. Grey bars represent 95% confidence intervals.

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

$$Y_{pdsym} = \alpha + \beta_1 \cdot Inclusive_d + \Gamma X'_{pdsym} + \theta_{sm} + \epsilon_{pdsym} \tag{11}$$

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 10. $X'_{pdsym}$ is a set of covariates including project size, tribal population share, baseline forest cover, and district size. $\theta_{sm}$ are state-month fixed effects. $\beta_1$ reveals the proportion of projects with each feature in inclusive versus extractive districts.

**Results:** Projects in inclusive districts are 8 p.p. more likely to obtain informed consent from tribal groups compared to projects in extractive districts in the same state (Figure 7; Table A14 for table). Forest officers in inclusive districts are also 7 p.p. more likely to commission extra cost-benefit reports during project review. Lastly, projects in inclusive districts are 1 p.p. less likely to be sited near a protected area. These are three important mechanisms driving the smaller infrastructure-biodiversity tradeoff in Table 6.

Official guidelines here: http://forestsclearance.nic.in/writereaddata/Addinfo/0_0_7111512571261CostBenefitAnalysisGuidelinesforforestlanddiversion-2017.pdf

The previous results do not appear to be driven by sorting between districts. If sorting were present, coefficients could reflect selection of conservation friendly projects into certain districts, even if institutions had no effect. To help rule this out, Table A15 estimates the project category distribution by institutional type using the same data as Equation 11. Values are coefficients from regressing a project category indicator on $Inclusive_d$. Column 1 is a raw correlation, column 2 adds $X'_{pdsym}$ from Equation 11, while columns 3 and 4 add fixed effects. Overall, project categories are well balanced across institutional types. One exception are electricity projects, which are 7 p.p. more common in inclusive districts under state fixed effects. Recall that these were the only projects with a positive (albeit insignificant) effect on species diversity (Figure 4B). These features of electricity projects may partially help explain the offsetting effect of inclusive institutions in Table 6.

Overall, the balance table implies that inclusive institutions have "teeth", as observed through higher rates of informed consent and a smaller ecological footprint of projects in these districts (Table 6). 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


**Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge**, "When should you adjust standard errors for clustering?," Technical Report, National Bureau of Economic Research 2017.

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

**Anukriti, S, Bilge Erten, and Priya Mukherjee**, "Women's Political Representation and Intimate Partner Violence," 2022.

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

**_ , Tobias Lunt, Ryu Matsuura, and Paul Novosad**, "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*, 2021, *35* (4), 845–871.

**Baehr, Christian, Ariel BenYishay, and Bradley Parks**, "Linking Local Infrastructure Development and Deforestation: Evidence from Satellite and Administrative Data," *Journal of the Association of Environmental and Resource Economists*, 2021, *8* (2), 375–409.

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

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

**Betts, Matthew G, Christopher Wolf, William J Ripple, Ben Phalan, Kimberley A Millers, Adam Duarte, Stuart HM Butchart, and Taal Levi**, "Global forest loss disproportionately erodes biodiversity in intact landscapes," *Nature*, 2017, *547* (7664), 441–444.

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

**Blanchard, Paul, Douglas Gollin, and Martina Kirchberger**, "Perpetual Motion: Human Mobility and Spatial Frictions in Three African Countries," *CEPR Discussion Paper No. DP16661*, 2023.

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

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

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

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

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

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

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

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

**Coleman, Eric A, Bill Schultz, Vijay Ramprasad, Harry Fischer, Pushpendra Rana, Anthony M Filippi, Burak Güneralp, Andong Ma, Claudia Rodriguez Solorzano, Vijay Guleria et al.**, "Limited effects of tree planting on forest canopy cover and rural livelihoods in Northern India," *Nature Sustainability*, 2021, *4* (11), 997–1004.

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

**Construction Federation of India**, "India Construction Sector ? Moving Forward," 2015.

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

**Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker**, "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings," *American Economic Review*, February 2015, *105* (2), 678–709.

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

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

**Dubey, Sanghamitra, Tushar Dash, and Geetanjoy Sahu**, "Political economy of Community forest rights," *Economic and Political Weekly*, 06 2017, *52.*

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

**eBird Basic Dataset**, "Version ebd_IN_201401_201904_relApr-2019," 2019.

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

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

**FAO and UNEP**, "State of the World's Forests: Forests, Biodiversity and People," Technical Report, FAO 2020.

**Farmer, Robert G, Marty L Leonard, Joanna E Mills Flemming, and Sean C Anderson**, "Observer aging and long-term avian survey data quality," *Ecology and Evolution*, 2014, *4* (12), 2563–2576.

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

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

**Forest Survey of India**, "India State of Forest Report," Technical Report, Ministry of Environment, Forests and Climate Change. Government of India 01 2019.

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

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

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

**Garg, Teevrat and Ajay Shenoy**, "The Ecological Impact of Place-Based Economic Policies," *American Journal of Agricultural Economics*, 2021.

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

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

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

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

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

**Hoffmann, Lars, Gebhard Günther, Dan Li, Olaf Stein, Xue Wu, Sabine Griessbach, Yi Heng, Paul Konopka, Rolf Müller, Bärbel Vogel et al.**, "From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations," *Atmospheric Chemistry and Physics*, 2019, *19* (5), 3097–3124.

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

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

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

**Jayadevan, Praveen, Rajah Jayapal, and Aasheesh Pittie**, "A checklist of the birds of India," *Indian BIRDS*, 07 2016, *11*, 113–172.

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

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

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

**_ , _ , Wesley M Hochachka, Marshall Iliff, Daniel Fink, Jeff Gerbracht, Carl Lagoze, Frank A La Sorte, Travis Moore, Andrea Wiggins et al.**, "Can observation skills of citizen scientists be estimated using species accumulation curves?," *PloS one*, 2015, *10* (10), e0139600.

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

**Kolstoe, Sonja and Trudy Ann Cameron**, "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*, 2017, *137*, 1–12.

**Lal, Apoorva, Saad Gulzar, and Ben Pasquale**, "Representation and Forest Conservation: Evidence from India?s Scheduled Areas," Feb 2021.

**Laurance, William F, Miriam Goosem, and Susan GW Laurance**, "Impacts of roads and linear clearings on tropical forests," *Trends in ecology & evolution*, 2009, *24* (12), 659–669.

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

**Liang, Yuanning, Ivan J Rudik, and Eric Zou**, "Economic Production and Biodiversity in the United States," Technical Report, National Bureau of Economic Research 2021.

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

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

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

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

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

**Menon, Meena**, "The unmaking of the forest rights act," *Economic and Political Weekly*, 2016, pp. 15–18.

**Ministry of Tribal Affairs**, "Monthly Update on Status of Implementation of the Scheduled Tribes and Other Traditional Forest Dweller (Recognition of Forest Rights) Act, 2006," 2022.

**MoEFCC**, "Forest (Conservation) Act, 1980," 1980.

_ , "Forest (Conservation) Rules, 2003, Ministry of Environment and Forests Notification," 2003.

_ , "Report of the Comptroller and Auditor General of India on Compensatory Afforestation in India," 2013.

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

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

**Nayak, Rajat, Krithi K Karanth, Trishna Dutta, Ruth Defries, K Ullas Karanth, and Srinivas Vaidyanathan**, "Bits and pieces: Forest fragmentation by linear intrusions in India," *Land Use Policy*, 2020, *99*, 104619.

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

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

**Nunn, Nathan**, "The Importance of History for Economic Development," *Annual Review of Economics*, 2009, *1* (1), 65–92.

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

**Oxford Economics**, "Global Infrastructure Outlook," Technical Report, Global Infrastructure Hub 2017.

**Pacheco, Pablo, Karen Mo, Nigel Dudley, Aurelie Shapiro, Naikoa aguilar amuchastegui, Pui-Yu Ling, Christa Anderson, and Alexander Marx**, "Deforestation fronts: Drivers and responses in a changing world," Technical Report, WWF 01 2021.

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

**Reis, Elivânia, Germán Manuel López-Iborra, and Renato Torres Pinheiro**, "Changes in bird species richness through different levels of urbanization: Implications for biodiversity conservation and garden design in Central Brazil," *Landscape and urban planning*, 2012, *107* (1), 31–42.

**Riley, Seth, John Pollinger, Raymond Sauvajot, Eric York, Cassity Bromley, Todd Fuller, and Robert Wayne**, "A southern California freeway is a physical and social barrier to gene flow in carnivores," *Molecular Ecology*, 07 2006, *15*, 1733–41.

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

**Stephens, Scott E, David N Koons, Jay J Rotella, and David W Willey**, "Effects of habitat fragmentation on avian nesting success: a review of the evidence at multiple spatial scales," *Biological conservation*, 2004, *115* (1), 101–110.

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

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

**Townshend, John, Matthew Hansen, Mark Carroll, Charlene DiMiceli, Robert Sohlberg, and Chengquan Huang**, "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 2017.

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

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

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

_, "What to preserve? An application of diversity theory to crane conservation," *The Quarterly Journal of Economics*, 1993, *108* (1), 157–183.

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

**World Rainforest Movement**, "Compensating for Forest Loss or Advancing Forest Destruction? A Study of Compensatory Afforestation in India," 2019.

# A Supplementary Tables and Figures

## A1 Tables

Table A1: eBird Summary Statistics (2015-2020)

| | Mean | Std. Dev. | Obs. |
|---|---|---|---|
| *District* | | | |
| Num. Users | 109.01 | 193.91 | 628 |
| Num. Trips | 1671.86 | 5497.55 | 628 |
| Coverage (%) | 52.88 | 32.26 | 628 |
| *User* | | | |
| Num. Districts | 3.99 | 7.45 | 16899 |
| Num. States | 1.93 | 2.21 | 16899 |
| Num. Year-months | 6.41 | 11.32 | 16899 |
| *User-District-Time* | | | |
| Species Richness | 23.39 | 18.72 | 173813 |
| Duration (min) | 85.51 | 70.70 | 173813 |
| Distance (km) | 3.06 | 6.02 | 173813 |
| Coverage (%) | 9.58 | 16.98 | 173813 |
| *District-Time* | | | |
| Rainfall (mm) | 0.34 | 0.82 | 21750 |
| Temperature ($^\circ$ C) | 23.30 | 7.22 | 21750 |
| Nightlights (radiance) | 2.61 | 7.28 | 21750 |
| Coverage (%) | 18.89 | 25.93 | 21750 |
| Num. Users per District-Yearmonth | 8.07 | 14.60 | 21750 |

Note: District variables reflect total eBird activity in a district during the study period. User variables describe number of locations and time-periods in which the user is active. Variables at the user-district-time level are means over users' trips in a district-month. Coverage is measured as the percentage of district cells traversed by a user on a 10km grid. Remaining covariate details are explained in section 3.

Table A2: Correlation Between Infrastructure Permits and Nightlights

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Infrastructure ($km^2$) | -0.0008 | -0.0006 | -0.0006 | -0.0000 | -0.0004 |
| | (0.0006) | (0.0005) | (0.0005) | (0.0004) | (0.0003) |
| Infrastructure (t-1) | | | | -0.0010 | -0.0010* |
| | | | | (0.0006) | (0.0006) |
| Infrastructure (t-2) | | | | | 0.0006 |
| | | | | | (0.0007) |
| Weather Controls | No | No | Yes | Yes | Yes |
| District FEs | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FEs | ✓ | | | | |
| State $\times$ Year FEs | | ✓ | ✓ | ✓ | ✓ |
| Observations | 3840 | 3822 | 3822 | 3185 | 2548 |
| $R^2$ | 0.993 | 0.996 | 0.996 | 0.997 | 0.998 |

Note: Data are at the district-year level. The outcome is log of mean nightlight intensity across gridcells in a district. The explanatory variable is cumulative forest area approved for deforestation to build infrastructure. Columns 4 and 5 include lags of infrastructure. Weather controls include temperature and rainfall. Standard errors clustered by district.

Table A3: Impact of Forest Infrastructure on Species Diversity

| | Main Estimates | | | Sensitivity | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Infrastructure ($km^2$) | -0.049* | -0.121** | -0.121** | -0.105 | -0.120** | -0.110* |
| | (0.027) | (0.053) | (0.054) | (0.062) | (0.054) | (0.059) |
| Infrastructure (district j $\neq$ d) (Standard Deviations) | | | 0.232 | | | |
| | | | (0.251) | | | |
| Non-forest Land Diversion ($km^2$) | | | | | | -0.046 |
| | | | | | | (0.056) |
| Weather Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Behaviour Controls | Yes | Yes | Yes | No | Yes | Yes |
| General Economic Trends | Yes | Yes | Yes | No | No | Yes |
| Outcome Mean | 23.672 | 23.748 | 23.748 | 23.748 | 23.748 | 23.748 |
| Coeff. Equality (p-val) | | | | | | 0.540 |
| User FEs | ✓ | | | | | |
| User $\times$ Year FEs | | ✓ | ✓ | ✓ | ✓ | ✓ |
| District FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| State $\times$ Month FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FEs | ✓ | | | | | |
| Observations | 167258 | 161902 | 161902 | 161896 | 161896 | 161896 |
| $R^2$ | 0.635 | 0.690 | 0.690 | 0.559 | 0.690 | 0.690 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Columns 1-3 are the same as in Figure 4A. Columns 4-5 successively add controls. Weather controls include temperature and rainfall. Behaviour controls include: traveling trips, log duration, log distance, log group size, and log spatial coverage. General economic trends are measured by nightlights. Column 6 adds cumulative non-forest land diversion, which is available only the digital subsample of project proposals. "Coeff. Equality" is the p-value on the test for equality between the infrastructure and non-land forest diversion coefficients. Standard errors clustered by biome.

Table A4: Robustness—Spatial Spillovers

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Infrastructure (Standard Deviations) | -0.396** | -0.402** | -0.408** | -0.410** |
| | (0.173) | (0.178) | (0.177) | (0.182) |
| Infrastructure (district j $\neq$ d) (Standard Deviations) | 0.025 | 0.081 | 0.399 | 0.441 |
| | (0.473) | (0.258) | (0.647) | (0.421) |
| Distance Cutoff | 100km | 200km | 500km | None |
| User $\times$ 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 user's trips in a district-month. All infrastructure variables are standardized for comparability. In all columns Infrastructure (row 1) is cumulative area of forest occupied by infrastructure in district $d$ during a year-month. In column 1, infrastructure in other districts $j$ describes cumulative encroachment area in other districts within 100km of the focal district $d$. In each time period, $Infrastructure_{dsym}$ is multiplied by a $N \times N$ (where $N$ is the number of districts in India) dimensional weight matrix $W$ with elements $w_{dj} = 1/distance_{dj}$ for districts $j$ within 100km of $d$ and zero otherwise. Columns 2 and 3 extend the distance cutoff to 200km and 500km, respectively. Column 4 applies the inverse distance weight to all districts. Section 5.3.2 elaborates the procedure. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log group size, and log spatial coverage. Standard errors clustered by biome.

Table A5: Tests of Endogenous Sorting

| | Across-District | | | Within-District |
|---|---|---|---|---|
| | (1) Num. Users | (2) Num. Users | (3) Num. Users | (4) % District Area |
| Infrastructure (Standard Deviations) | 0.010 | 0.010 | 0.008 | -0.007 |
| | (0.029) | (0.029) | (0.028) | (0.011) |
| Infrastructure (district $j \neq d$) (Standard Deviations) | -0.023 | -0.017 | -0.013 | |
| | (0.020) | (0.024) | (0.029) | |
| Controls | Yes | Yes | Yes | Yes |
| Data Aggregation | District | District | District | District |
| Distance Cutoff | 100km | 200km | 500km | |
| District FEs | ✓ | ✓ | ✓ | ✓ |
| State × Month FEs | ✓ | ✓ | ✓ | ✓ |
| Year FEs | ✓ | ✓ | ✓ | ✓ |
| Observations | 21256 | 21256 | 21256 | 21256 |
| $R^2$ | 0.808 | 0.808 | 0.808 | 0.976 |

Note: $^{*}$ $p < .1$, $^{**}$ $p < .05$, $^{***}$ $p < .01$. Data are aggregated at the district-year-month level. The outcome in columns 1-3 is log number of users in a district. The outcome in column 4 is % of district grid cells traversed by the average user. All infrastructure variables are standardized for comparability. Infrastructure (row 1) is cumulative area of forest occupied by infrastructure in district $d$ during a year-month. In column 1, infrastructure (district $j \neq d$) is inverse-distance weighted infrastructure in districts $j$ within 100km of $d$. In columns 2 and 3, the distance cutoff is extended to 200km and 500km, respectively. Controls are the same as the main specification. Experience, duration, distance, group size, and % traveling trips are aggregated to district means and logged. Standard errors clustered by biome.

Table A6: Impact of Forest Infrastructure on Species Diversity by Category

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Electricity | 0.088 | 0.090 | 0.085 | 0.091 |
| | (0.066) | (0.066) | (0.066) | (0.066) |
| Irrigation | -0.105* | -0.130** | -0.131** | -0.121** |
| | (0.052) | (0.049) | (0.048) | (0.042) |
| Mining | -0.061 | -0.059*** | -0.059*** | -0.111*** |
| | (0.035) | (0.019) | (0.019) | (0.021) |
| Other | -0.165 | -0.257 | -0.254 | -0.264 |
| | (0.229) | (0.215) | (0.215) | (0.210) |
| Resettlement | -1.100*** | -0.745*** | -0.744*** | -0.728*** |
| | (0.062) | (0.080) | (0.078) | (0.088) |
| Transportation | -0.265 | -0.392* | -0.393* | -0.444** |
| | (0.305) | (0.181) | (0.183) | (0.188) |
| Weather Controls | Yes | Yes | Yes | Yes |
| Behaviour Controls | No | Yes | Yes | Yes |
| General Economic Trends | No | No | Yes | Yes |
| Outcome Mean | 23.748 | 23.748 | 23.748 | 23.983 |
| Sample | Full | Full | Full | High-Activity |
| User x Year FEs | ✓ | ✓ | ✓ | ✓ |
| District FEs | ✓ | ✓ | ✓ | ✓ |
| State × Month FEs | ✓ | ✓ | ✓ | ✓ |
| Observations | 161896 | 161896 | 161896 | 150011 |
| $R^2$ | 0.559 | 0.690 | 0.690 | 0.687 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all specifications is mean species richness across users' trips in a district-month. Rows denote cumulative area of infrastructure encroachments by a particular category in a district-month. Column 1-3 successively add controls. Weather controls include temperature and rainfall. Behaviour controls include: traveling trips, log duration, log distance, log group size, and log spatial coverage. General economic trends is measured by nightlights. Column 3 is the same as Figure 4B. Column 4 restricts the sample to districts with high eBird usage, measured as districts with above-median numbers of users, recording above-median trips per user. Standard errors clustered by biome.

Table A7: Robustness—Alternative Fixed Effects

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Infrastructure ($km^2$) | -0.125** | -0.114 | -0.080** | -0.108* | -0.117* | -0.119** | -0.049 |
| | (0.041) | (0.064) | (0.032) | (0.051) | (0.054) | (0.054) | (0.056) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| User x Year FEs | | ✓ | | ✓ | ✓ | ✓ | |
| User × Month FEs | ✓ | | | | | | |
| District FEs | ✓ | | | ✓ | ✓ | ✓ | |
| District × Month FEs | | ✓ | ✓ | | | | |
| State × Year FEs | ✓ | | ✓ | | | | |
| State × Month FEs | | | | | ✓ | ✓ | ✓ |
| Biome × Month | | | | ✓ | | | |
| Experience FEs | | | | | ✓ | | |
| Cell FEs | | | | | | | ✓ |
| Year FEs | | | | | | | ✓ |
| Time-of-Day FEs | | | | | | ✓ | |
| Observations | 143394 | 161087 | 166446 | 161909 | 161563 | 161665 | 282428 |
| $R^2$ | 0.702 | 0.706 | 0.654 | 0.688 | 0.694 | 0.690 | 0.542 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all columns is mean species richness across users' trips in a district-month. The estimating equation is the same as Equation 2 with different fixed effects described in the footer. Experience fixed effects indicate cumulative number of trips taken by a user. Time-of-day fixed effects indicate whether users' trips in a district-month were recorded on average during morning, afternoon, evening, or night. In column 7, data are at the user-grid-cell level. Variation in observations across columns 1-6 is from dropping singletons. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log group size, and log spatial coverage. Standard errors clustered by biome.

Table A8: Robustness—Other

| | (1) SR | (2) SR | (3) SR | (4) SR | (5) SR | (6) SR | (7) Shannon | (8) Simpson | (9) SR |
|---|---|---|---|---|---|---|---|---|---|
| Infrastructure | -0.118** | -0.093* | -0.124** | -0.111*** | -0.246** | -0.605*** | -0.033 | -0.036 | -0.112* |
| ($km^2$) | (0.053) | (0.050) | (0.054) | (0.010) | (0.082) | (0.138) | (0.276) | (0.060) | (0.057) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Infrastructure Unit | $km^2$ | $km^2$ | $km^2$ | $km^2$ | $km^2$ | IHS | $km^2$ | $km^2$ | $km^2$ |
| Sample Restriction | 2015 users | Away | Active | non-COVID | Truncate | None | None | None | None |
| User x Year FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| District FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| State x Month FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Weights | None | None | None | None | None | None | None | None | Trips |
| Observations | 82339 | 123280 | 150011 | 123416 | 161896 | 161896 | 161896 | 157766 | 161896 |
| $R^2$ | 0.684 | 0.685 | 0.687 | 0.690 | 0.690 | 0.690 | 0.645 | 0.445 | 0.762 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all columns (except 7 and 8) is mean species richness (SR) across users' trips in a district-month. Column 1 is estimated on users who signed up for eBird in 2015. Column 2 drops observations in users' home districts. Column 3 restricts the sample to districts with above-median numbers of users who record an above-median number of trips per user. Column 4 drops the year 2020. Column 5 drops the three largest projects. Column 6 uses the inverse hyperbolic sine of the explanatory variable. Columns 7 and 8 show results with two alternative species diversity metrics. Column 9 weights the regression with number of trips underlying the outcome. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log group size, and log spatial coverage. Standard errors clustered by biome.

Table A9: Robustness—Alternative Standard Errors

| | Standard Error Boundary | | | Conley Spatial Error Cutoff | | | |
|---|---|---|---|---|---|---|---|
| | Biome | District | State | 100km | 200km | 500km | 1000km |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Infrastructure ($km^2$) | -0.122** | -0.122** | -0.122* | -0.122* | -0.122* | -0.122* | -0.122** |
| | (0.051) | (0.060) | (0.062) | (0.064) | (0.069) | (0.066) | (0.054) |
| User × Year | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| District | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| State × Month | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Observations | 161,907 | 161,907 | 161,907 | 161,907 | 161,907 | 161,907 | 161,907 |
| $R^2$ | 0.694 | 0.694 | 0.694 | 0.694 | 0.694 | 0.694 | 0.694 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Coefficient estimates and standard errors from baseline specification with alternative error clustering. Column 1 replicates the main estimate with clustering at the biome level. In columns 2-3, standard errors are clustered by district and state, respectively. Columns 4-7 implement Conley (1999) standard errors for four different values of the kernel cut off distance (in km). The R software was used to compute Conley errors; observations differ slightly from the main results due to differences in the way R drops singletons.

Table A10: Category-Wise Robustness—Alternative Fixed Effects

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Electricity | -0.051 | 0.147* | 0.036 | 0.095 | 0.090 | 0.037 |
| | (0.090) | (0.071) | (0.098) | (0.066) | (0.065) | (0.051) |
| Irrigation | -0.026 | -0.158** | -0.014 | -0.127** | -0.126** | -0.062 |
| | (0.046) | (0.064) | (0.046) | (0.050) | (0.051) | (0.109) |
| Mining | -0.205** | -0.051 | -0.257*** | -0.055** | -0.056** | -0.154 |
| | (0.067) | (0.047) | (0.046) | (0.022) | (0.018) | (0.259) |
| Other | -0.236 | -0.255 | -0.152 | -0.246 | -0.256 | -0.276 |
| | (0.165) | (0.251) | (0.234) | (0.225) | (0.219) | (0.258) |
| Resettlement | -0.778*** | -0.530*** | -0.447*** | -0.761*** | -0.756*** | -0.159*** |
| | (0.166) | (0.065) | (0.060) | (0.084) | (0.077) | (0.036) |
| Transportation | -0.462 | -0.431*** | -0.513 | -0.414** | -0.398* | 0.159 |
| | (0.284) | (0.136) | (0.299) | (0.167) | (0.186) | (0.233) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| User x Year FEs | | ✓ | | ✓ | ✓ | |
| User × Month FEs | ✓ | | | | | |
| Experience FEs | | | | ✓ | | |
| District FEs | ✓ | | | ✓ | ✓ | |
| District × Month FEs | | ✓ | ✓ | | | |
| Cell FEs | | | | | | ✓ |
| State × Month FEs | | | | ✓ | ✓ | ✓ |
| State × Year FEs | ✓ | | ✓ | | | |
| Year FEs | | | | | | ✓ |
| Time-of-Day FEs | | | | | ✓ | |
| Observations | 143384 | 161029 | 166409 | 161557 | 161665 | 282427 |
| $R^2$ | 0.702 | 0.706 | 0.654 | 0.694 | 0.690 | 0.542 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all columns is mean species richness across users' trips in a district-month. Coefficients describe the marginal impact of infrastructure encroachment by projects of a given category. The estimating equation is the same as Equation 3 with different fixed effects described in the footer. Experience fixed effects indicate cumulative number of trips taken by a user. Time-of-day fixed effects indicate whether users' typical trips in a district-month were recorded during morning, afternoon, evening, or night. In column 6, data are at the user-grid-cell level. Variation in observations across columns 1-5 is from dropping singletons. All regressions control for: temperature, rainfall, traveling trips, log night-lights, log duration, log distance, log group size, and log spatial coverage. Standard errors clustered by biome.

Table A11: Category-Wise Robustness—Other

| | (1) SR | (2) SR | (3) SR | (4) SR | (5) SR | (6) SR | (7) Shannon | (8) Simpson | (9) SR |
|---|---|---|---|---|---|---|---|---|---|
| Electricity | -0.038 | 0.078 | 0.091 | 0.197 | -0.224 | 0.132 | -0.132 | -0.099 | 0.026 |
| | (0.145) | (0.076) | (0.066) | (0.438) | (0.854) | (0.926) | (0.287) | (0.076) | (0.029) |
| Irrigation | 0.042 | -0.049 | -0.121** | -0.056* | -0.164 | -0.580** | -0.034 | -0.045 | -0.175*** |
| | (0.051) | (0.062) | (0.042) | (0.027) | (0.142) | (0.214) | (0.214) | (0.048) | (0.049) |
| Mining | -0.055 | -0.023 | -0.111*** | -0.102*** | -0.031 | -0.332 | -0.333* | 0.028 | 0.077 |
| | (0.046) | (0.091) | (0.021) | (0.027) | (0.031) | (0.403) | (0.178) | (0.088) | (0.132) |
| Other | -0.330 | -0.230 | -0.264 | -0.209 | -0.257 | -0.646 | 0.024 | -0.072 | -0.323*** |
| | (0.204) | (0.228) | (0.210) | (0.224) | (0.209) | (0.462) | (0.853) | (0.176) | (0.068) |
| Resettlement | -0.875*** | -0.951*** | -0.728*** | -0.565*** | -0.750*** | -1.682*** | 2.362*** | 0.348** | -0.563*** |
| | (0.097) | (0.125) | (0.088) | (0.095) | (0.084) | (0.307) | (0.162) | (0.124) | (0.066) |
| Transportation | -0.551*** | -0.502** | -0.444** | -0.241 | -0.381** | -0.882*** | -0.783 | 0.073 | -0.137 |
| | (0.125) | (0.168) | (0.188) | (0.140) | (0.166) | (0.203) | (1.449) | (0.270) | (0.159) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Infrastructure Unit | $km^2$ | $km^2$ | $km^2$ | $km^2$ | $km^2$ | IHS | $km^2$ | $km^2$ | $km^2$ |
| Sample Restriction | 2015 users | Away | Active | non-COVID | Truncate | None | None | None | None |
| User x Year FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| District FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| State x Month FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Weights | None | None | None | None | None | None | None | None | Trips |
| Observations | 82339 | 123280 | 150011 | 123416 | 161896 | 161896 | 161896 | 157766 | 161896 |
| $R^2$ | 0.684 | 0.685 | 0.687 | 0.690 | 0.690 | 0.690 | 0.645 | 0.445 | 0.762 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all columns (except 7 and 8) is mean species richness (SR) across users' trips in a district-month. Coefficients show the marginal impact of infrastructure encroachment by projects of a given category. Column 1 is estimated on users who signed up for eBird in 2015. Column 2 drops observations in users' home districts. Column 3 restricts the sample to districts with above-median numbers of users who record an above-median number of trips per user. Column 4 drops the year 2020. Column 5 drops the three largest projects. Column 6 uses the inverse hyperbolic sine of the explanatory variable. Columns 7 and 8 show results with two alternative species diversity metrics. Column 9 weights the regression with number of trips underlying the outcome. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log group size, and log spatial coverage. Standard errors clustered by biome.

Table A12: Robustness: Close Election Design

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Infrastructure ($km^2$) | -1.850** | -2.695** | -1.308 | -2.303** | -2.044** | -2.468 |
| | (0.713) | (1.171) | (0.779) | (1.005) | (0.801) | (1.400) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| First Stage F-Statistic | 8.842 | 9.043 | 8.981 | 6.414 | 8.442 | 5.234 |
| Bandwidth | 2 | 3 | 5 | 2 | 2 | 2 |
| Polynomial Order | 1 | 1 | 1 | 2 | 3 | 1 |
| User FEs | ✓ | ✓ | ✓ | ✓ | ✓ | |
| User x Year FEs | | | | | | ✓ |
| District FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| State $\times$ Month FEs | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FEs | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Observations | 134448 | 134448 | 134448 | 134448 | 134448 | 129704 |

Note: $^{*}$ $p < .1$, $^{**}$ $p < .05$, $^{***}$ $p < .01$. Data are at the user-district-month level as in the main specifications. The outcome is mean species richness across user's trips in a district-month. Coefficients are 2SLS estimates as specified in Equation 9. Infrastructure is instrumented with the fraction of constituencies in a district where the incumbent won in a close race during the most recent state election. All regressions control for the same user- and district-level covariates as Equation 2 as well as the fraction of close-election district constituencies where the incumbent ran, election year, and the interaction of victory margin with an indicator for whether any incumbent ran in the district. Column 1 defines close election as a win margin of 2 percent. Columns 2-3 expand the win margin to 3 and 5 percent, respectively. Columns 4 and 5 uses a second- and third -rder polynomial in the win margin. Column 6 uses user-by-year fixed effects.

Table A13: Robustness: Treatment Heterogeneity by Institutional Type

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Infrastructure | -0.519** | -0.510*** | -0.548*** | -0.558*** | -0.236 | -0.607*** | -0.551*** | -1.227*** | -0.551** |
| | (0.116) | (0.098) | (0.037) | (0.083) | (0.265) | (0.079) | (0.067) | (0.202) | (0.189) |
| Infrastructure | 0.037 | 0.327*** | 0.435** | 0.421** | 0.377** | 0.454* | 0.434** | 1.529*** | 0.434** |
| × Inclusive (=1) | (0.136) | (0.032) | (0.132) | (0.114) | (0.125) | (0.191) | (0.116) | (0.329) | (0.149) |
| Infra. × Tribal Share | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Infra. × Baseline Forest | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Unit | $km^2$ | $km^2$ | $km^2$ | $km^2$ | $km^2$ | $km^2$ | $km^2$ | IHS | $km^2$ |
| Sample Restriction | None | None | None | None | 2015 users | preCOVID | Truncate | None | None |
| User × Month FEs | ✓ | | | | | | | | |
| User × Year FEs | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| District FEs | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| District × Month FEs | | ✓ | | | | | | | |
| State x Month FEs | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| State × Year Fes | ✓ | | | | | | | | |
| Experience FEs | | | ✓ | | | | | | |
| Time-of-day FEs | | | | ✓ | | | | | |
| Clustering | Biome | Biome | Biome | Biome | Biome | Biome | Biome | Biome | State |
| Observations | 47609 | 58587 | 58204 | 58678 | 29208 | 43788 | 58760 | 58760 | 58760 |
| $R^2$ | 0.713 | 0.719 | 0.713 | 0.705 | 0.691 | 0.698 | 0.704 | 0.704 | 0.704 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome in all specifications is mean species richness across users' trips in a district-month. Inclusive (=1) means the district has historically inclusive institutions. Sample restricted to 163 districts in Banerjee and Iyer (2005) and aggregated to 1991 boundaries. ST share is the fraction of district population belonging to a tribal group as measured in 2011. Experience fixed effects indicate cumulative number of trips taken by a user. Time-of-day fixed effects indicate whether users' typical trips in a district-month were recorded during morning, afternoon, evening, or night. Column 5 is estimated on users who signed up for eBird in 2015. Column 6 drops the year 2020. Column 7 drops the three largest projects. Column 8 uses the inverse hyperbolic sine of the explanatory variable. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log group size, and log spatial coverage. Fixed effects and clustering described in the footer.

Table A14: Mechanisms: Institutions and Sustainable Infrastructure

| | (1) Informed Consent | (2) Cost-Benefit | (3) Protected Area |
|---|---|---|---|
| Inclusive (=1) | 0.078*** | 0.071** | -0.006** |
| | (0.015) | (0.029) | (0.002) |
| Controls | Yes | Yes | Yes |
| Outcome Mean | 0.234 | 0.156 | 0.007 |
| State × Time FEs | ✓ | ✓ | ✓ |
| N | 2275 | 2275 | 2270 |
| $R^2$ | 0.541 | 0.510 | 0.237 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the project level for the digital subsample. Inclusive (=1) means the district has historically inclusive institutions. Sample restricted to 163 districts in Banerjee and Iyer (2005) under British Rule. Informed consent indicates whether the Gram Sabha was consulted and the FRA followed. Cost-Benefit Analyses are reports commissioned during project review. Protected Area equals one if the project is sited in or near one. All specifications include controls for: project size, district population share belonging to a tribal group as measured in 2011, baseline forest cover, and district area.

Table A15: Balance Table: Project Category Distribution by Institutional Type

| | (1) No Controls or FEs | (2) Controls | (3) State FEs | (4) State + Year FEs |
|---|---|---|---|---|
| Electricity | -0.016 | 0.039 | 0.066** | 0.071** |
| | (0.023) | (0.038) | (0.029) | (0.029) |
| Irrigation | -0.049 | -0.019 | -0.012 | -0.005 |
| | 0.034 | 0.032 | 0.020 | 0.020 |
| Mining | -0.020 | 0.010 | -0.018* | -0.017* |
| | 0.015 | 0.011 | 0.010 | 0.010 |
| Other | 0.000 | -0.081 | -0.027 | -0.028 |
| | 0.048 | 0.049 | 0.053 | 0.050 |
| Resettlement | -0.020 | -0.016 | -0.001 | -0.001 |
| | 0.012 | 0.010 | 0.008 | 0.008 |
| Transportation | 0.104* | 0.067 | -0.008 | -0.019 |
| | 0.056 | 0.038 | 0.049 | 0.046 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Values describe the *difference* in project shares of each category between districts with inclusive and extractive institutions. For example, the first cell implies that inclusive districts have 1.6p.p less electricity projects than extractive districts. Values in each cell are from separate project-level regressions of an indicator for that category on an indicator for whether its district of approval is inclusive. Column 1 includes no other controls or fixed effects and describes the difference in mean project shares. Column 2 adds controls for: project size, district population share belonging to a tribal group as measured in 2011, baseline forest cover, and district area. Column 3 successively adds state fixed effects, and column 4 adds state and year fixed effects.

## A2 Figures

**F. No.8-43/2005-FC**
Government of India
Ministry of Environment & Forests
(FC Division)

Indira Paryavaran Bhawan,
Aliganj, Jorbagh Road,
New Delhi – 110003

**Dated: 15th September, 2016**

To,

The Principal Secretary (Forests),
Government of Rajasthan,
Jaipur.

**Subject: Diversion of 185 hectares of forest land for Lahasi Medium Irrigation Project in Baran district of Rajasthan.**

Sir,

I am directed to refer to the State Government of Rajasthan's letter No. P.1 (30) Van/2005 dated 25.04.2005 on the subject cited above seeking prior approval of the Central Government under the Forest (Conservation) Act, 1980. After careful consideration of the proposal by the Forest Advisory Committee (FAC) constituted under Section-3 of the said Act, **In-principle** approval was granted vide this Ministry's letter of even number dated 08.02.2006 subject to fulfilment of certain conditions. The State Government has furnished compliance report in respect of the conditions stipulated in the approval and has requested the Central Government to grant final approval.

In this connection, I am directed to say that on the basis of the compliance report furnished by the Government of Rajasthan vide their letters no. F.14 ()2005/FCA/APPCF/7943 dated 16.10.2014,F.14()2005/FCA/APPCF/9067 dated 29.12.2014 and F.14()2005/ FCA/APPCF/542 dated 03.03.2016, **Final/Stage-II approval** of the Central Government is hereby granted under Section-2 of the Forest (Conservation) Act, 1980 for diversion of 185 hectares of forest land for Lahasi Medium Irrigation Project in Baran district of Rajasthan in favour of Water Resources Department, Govt. of Rajasthan subject to fulfilment of the following conditions:

1. Legal status of the diverted forest land shall remain unchanged.
2. Compensatory afforestation shall be raised and maintained by the State Forest Department at the project cost.
3. Non-forest land to be transferred and mutated in favour of the State Forest Department for raising Compensatory Afforestation shall be notified as reserved Forest under Section-4 or Protected Forest under Section-29 of the Indian Forest Act, 1927 or under the relevant Section(s) of the local Forest Act. The Nodal officer must report compliance within a period of 6 months from the date of grant of final approval and send a copy of the notification declaring the non-forest land under Section 4 or Section 29 of the Indian Forest Act, 1927, or under the relevant section of the local Forest Act as the case may be, to this Ministry for information and record.
4. The State Government and the User Agency shall ensure implementation of approved R&R plan.
5. The project area shall be demarcated on ground at the project cost using four Feet high RCC pillars with each pillar inscribed with serial No. forward and backward bearing, distance between two adjacent pillars and GPS Co-Ordinates.
6. The tree felling in the forest area, so diverted, shall only be as per the actual requirement and

Figure A1: Example Approval Letter

Note: Scanned letter from Principal Secretary of State Forest Ministry approving proposal for deforestation of 185 ha. for irrigation project. Encroachment area, district, and state are extracted for construction of manual sample. Additional fields from the actual proposal provided for the digital sample.

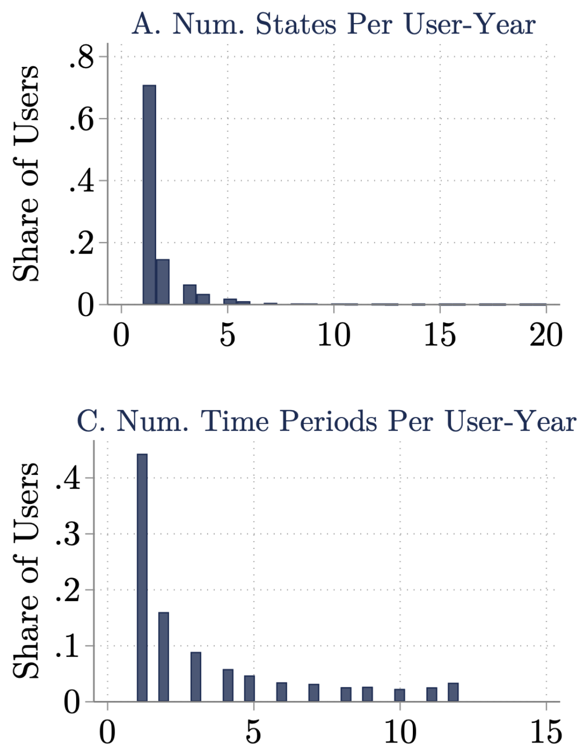


Figure A2: Within-User Distribution of Spatiotemporal Activity

Note: Distributions are based on aggregating eBird data to the user level (N=17,634 users). Panel A is the distribution of total number of states traversed for eBirding per user during a year. Panel B shows the same for total number of districts traversed per user-year. In Panel C, x-axis denotes number of months in the year in which users recorded trips.

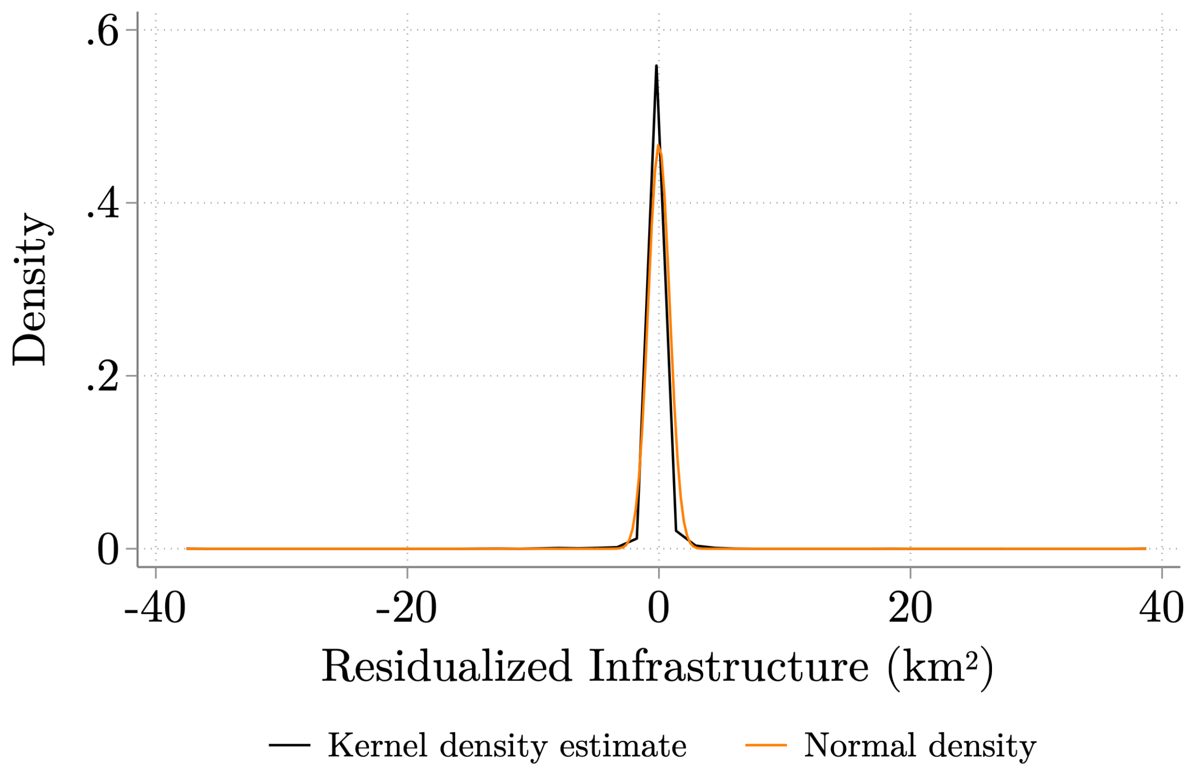


Figure A3: Density of Residualized Infrastructure

Note: Kernel density estimates using the Silverman bandwidth. Residualized infrastructure is obtained by computing residuals after regressing cumulative infrastructure (in $km^2$) on fixed effects from Equation 2.

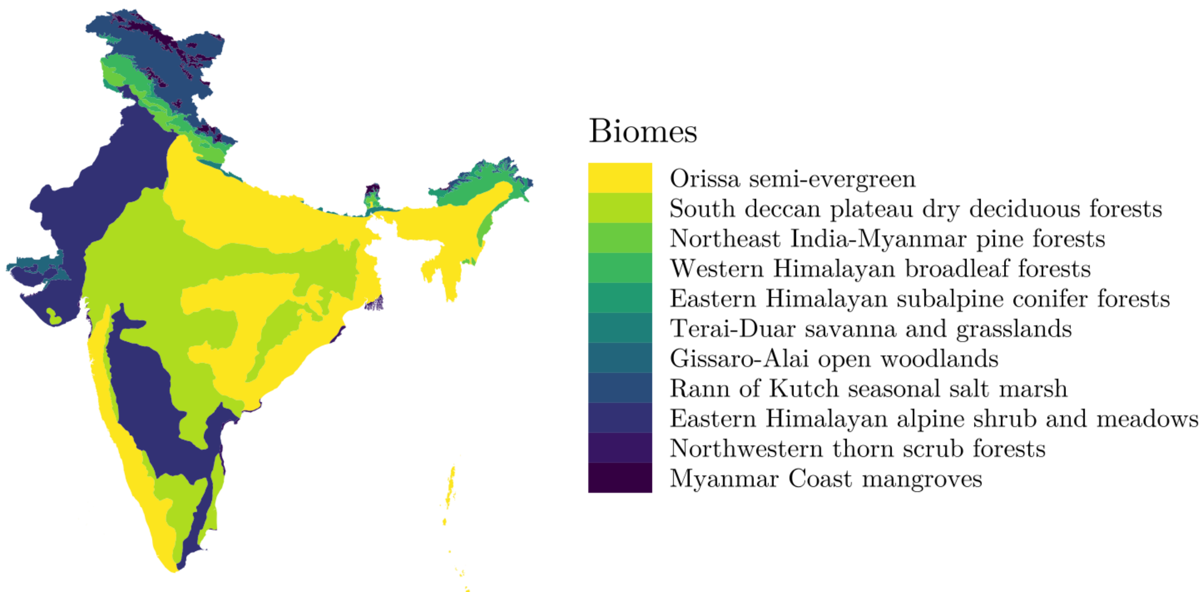


Figure A4: Biomes of India

Note: Data obtained from the Nature Conservancy Terrestrial Ecoregion shapefiles.

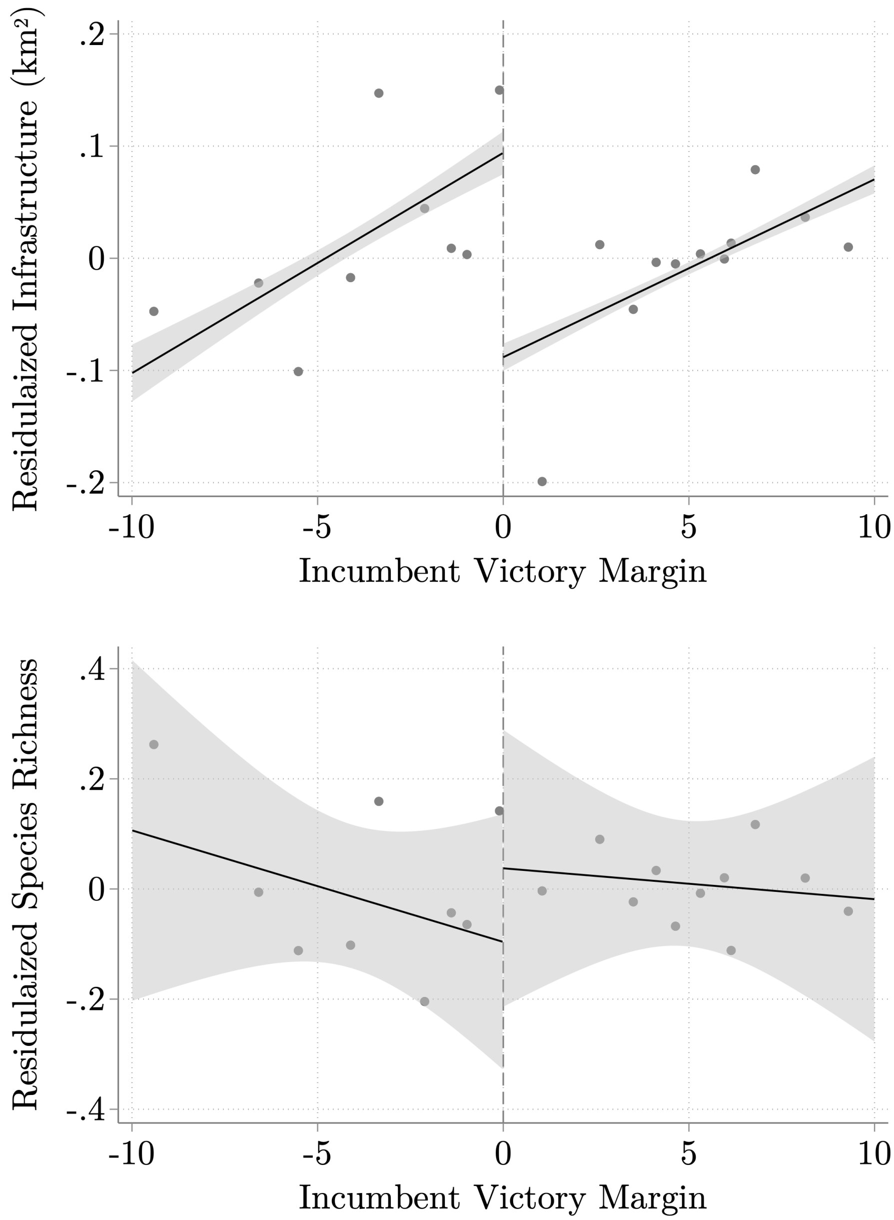


Figure A5: Margin of Victory, Infrastructure Project Approvals, and Species Richness

Note: Panel A (top) shows the first stage (Equation 8) and Panel B (bottom) shows the reduced form. Each figure plots a binscatter of incumbent win margin in close elections (x-axis) against the outcome (y-axis) residualized on user, district, state-month and year fixed effects as well as the same set of covariates as the main TWFE specification. Points to the left of zero denote districts where incumbents lost in close elections. A linear fit is generated separately for each side of zero, with 95% confidence intervals displayed.

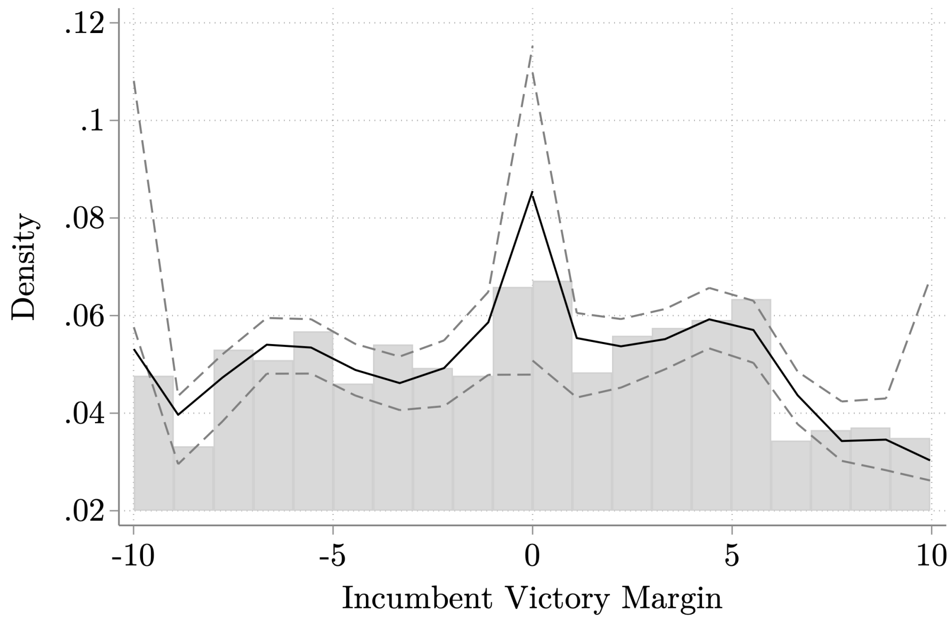


Figure A6: Density Discontinuity Test for Manipulation

Note: The figure plots a density test from Cattaneo et al. (2020). The black line traces the density of observations in each margin-of-victory bin. Dashed lines are 95% confidence intervals around the local linear density estimates. Grey bars are a histogram of victory margin.

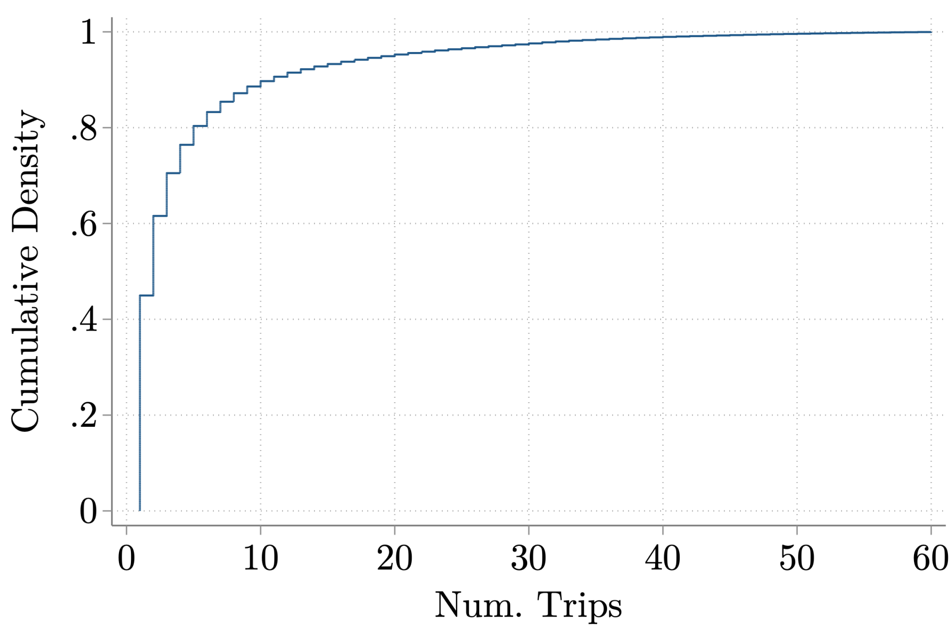


Figure A7: Cumulative Distribution of Weighting Variable

Note: The x-axis measures the total number of birdwatching trips taken by a user in a district-time period. Data are truncated at the 95th percentile to remove outliers. The y-axis is the percentage of observations with values less than the x-axis value.

**FORM-I**

**Government of Odisha**
**Office of the District Collector, Khordha.**

Date. 23.10.2019

No. 2118

**TO WHOM IT MAY CONCERN**

In compliance of the Ministry of Environment & Forests (MoEF), Government of India's letter No. 11-9/98-FC (pt) dated 03rd August 2009 wherein the MoEF issued guidelines on submission of evidences for having initiated & completed the process of settlement of rights under the Scheduled Tribes and Other Traditional Forest Dwellers (Recognition of Forest Rights) Act 2006 on the forest land proposed for diversion, It is certified that Ac. 26.150 of forest land to be diverted in favour of AIIMS, Bhubaneswar.**(I.e. for development of public health facilities)** in Khordha District falls within the jurisdiction in Bhubaneswar Tahasil.

It is further certified that:

(a) *The complete process for identification and settlement of rights under the FRA has been carried out for the entire Ac.26.150 of forest area proposed for diversion. A copy of records of all consultations and meetings of the GramaSabha/Sub-Division Level Committee and the District Level Committee are enclosed as annexure 1 to annexure 3.*

(b) *The diversion of forest land for facilities managed by the Government as required under section 3 (2) of the FRA have been completed.*

(c) *The proposal does not involve recognized right of Primitive Tribal Groups and Pre-agricultural communities.*

Signature

23/10/19

Encl: As above

(Sitansu Kumar Rout)
DISTRICT COLLECTOR

Figure A8: Example Letter of Informed Consent

Note: Figure shows scanned letter from Principal Secretary of State Forest Ministry (Rajasthan) approving proposal for deforestation of 185 hectares for irrigation project. Encroachment area, district, and state are extracted for construction of manual sample. Additional fields from the actual proposal/application itself is only provided for the digital sample.

# B Online Appendix

## B1 Infrastructure Sample

This Appendix provides supplementary details about the infrastructure data sample. It also presents additional summary statistics about projects by ownership type and shape.

### B1.1 Infrastructure Sample Construction

Table B1: Procedure for Project Categorization

| Raw Entry | Recategorized Entry |
|---|---|
| hydel, sub station, thermal, transmission line, village electricity, wind power, solar power | Electricity |
| road, railway | Transportation |
| canal, irrigation, drinking water | Irrigation |
| forest village conversion, rehabilitation | Resettlement |
| mining, quarrying, borehole prospecting | Mining |
| Everything else | Other |

Note: The left column is the verbatim project categories as entered in application by the firm. The right column describes the synthesized categories for the purpose of simplifying project types.

**Sample Construction and Digitization.** The project sample consists of projects approved between 2015-2020. Applications submitted after 2014 (N=6,597) were scraped from the online portal (the digital subsample). Applications submitted before 2014 but approved afterwards (N=1,732) were manually digitized (the manual subsample). The application itself was not available for the manual subsample, but a PDF of the approval letter listing project size, location etc. was available in the portal.

The manual subsample was digitized as follows. First, a PDF of each approval letter was downloaded. The district of each project was extracted from the subject header (see Figure A1) and cross-checked online. If only the village was given, the district was identified on Google. Second, project size (hectares approved for forest diversion) was also extracted from the letter. For projects that span multiple districts (e.g. roads), a separate document called "Form A" (also available in the portal) was downloaded to identify hectares per district. 26 multi-district projects did not specify a district-wise breakdown, in which case total project size was divided equally across districts. Lastly, project category was extracted from the letter. In some cases it was taken from the Form A document which includes a detailed project description[29] .

**Project Categorization.** Verbatim project categories often refer to the same type of project. To simplify the analysis, I re-categorize projects according to Table B1[30]. Because of inconsistencies in raw data entry by firms, I manually examine and reassign projects mistakenly categorized as “other”. “Other” project descriptions with the word “power”, “substation”, and “kv” are placed in the Electricity category. “Other” projects with the word “resettle”, “relocate”, and “pattayam”[31] are placed in the Resettlement category.

**District Splitting.** I address the problem of district splitting and name changes. In the application, district name is an unstructured string with many spelling inconsistencies. To form a consistent matching key, I link the official district names and codes from the 2011 Census. To do this, I first manually rename district names to their original 2011 census name if there was a name change after 2011. Second, in the case of a district split after 2011, I rename the “child” districts to the name of the 2011 “parent” district. These two steps produce a single geographical unit that can be tracked consistently across years. Third, I perform a fuzzy match between district names in the application and the 2011 key using the Levensthein distance between the district strings. This algorithm identifies the official census code for 98% of districts in the application sample. The remaining 2% (8 districts) represent districts with more complex redrawing procedures and are dropped.

**Panel Aggregation.** I transform project-level data into a district-monthly panel in three steps. First, since each project can span multiple districts (e.g. roads), I reshape the data from project to project-district level. Each row contains the project component area falling into a specific district. The sum of rows for a single project equals total project size. Second, I aggregate to the total forest area diverted in each district and year-month, overall and category-wise (e.g. total deforestation in January 2018 for electricity projects in Delhi). Third, I balance the panel by assuming districts and time periods not in the portal had zero stage-II approvals.

[29]The category of each project in the manual subsample was available digitally, and scraped, but the majority were listed as “Other”. I manually categorized them based on the subject header (see Figure A1)

[30]There are 89 projects categorized as “industry” which together make $< 0.1\%$ of total area cleared during the study period. For this reason, I include industrial projects in the “other” category.

[31]*Pattayam* means deed and refers to a scheme providing land to the landless, typically tribal families.

### B1.2 Summary Statistics by Project Ownership and Shape

Table B2: Summary Statistics of Projects by Ownership and Shape

| | Num. Projects | Mean Size (ha.) | SD (ha.) | Total Area (ha.) |
|---|---|---|---|---|
| *Panel A: Ownership* | | | | |
| Public | 4,666 | 9.6 | 89.9 | 44,861.3 |
| Private | 1,549 | 1.9 | 15.2 | 2,910.6 |
| Neither | 382 | 2.7 | 24.9 | 1,019.8 |
| *Panel B: Shape* | | | | |
| Linear | 5,768 | 4.8 | 28.2 | 27,472.6 |
| Nonlinear | 829 | 25.7 | 201.0 | 21,319.1 |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Data are arranged at the project level for 6,597 approved projects that reported ownership. Panel A splits projects by ownership type. Panel B splits projects by shape. Total area (ha.) is the summed deforestation area of all projects in the sample for the particular project group.

Table B3: Percent of Projects in Each Category by Ownership and Shape

| | Ownership (%) | | | Shape (%) | |
|---|---|---|---|---|---|
| | Public | Private | Neither | Linear | Nonlinear |
| | (1) | (2) | (3) | (4) | (5) |
| Electricity | 81.48 | 15.86 | 2.67 | 86.03 | 13.97 |
| Irrigation | 95.34 | 2.48 | 2.17 | 72.05 | 27.95 |
| Mining | 49.15 | 35.59 | 15.25 | 0.00 | 100.00 |
| Other | 58.07 | 33.53 | 8.40 | 85.04 | 14.96 |
| Resettlement | 100.00 | 0.00 | 0.00 | 5.00 | 95.00 |
| Transportation | 88.41 | 9.54 | 2.05 | 99.03 | 0.97 |

Note: Data are arranged at the project level for 6,597 approved projects that reported ownership and shape, prior to aggregating to the district level. Cell values denote row percentages *within each group* (i.e. % of projects in each category falling under different ownership types). Thus, the row sum of columns 1-3 equals 100, and same for columns 4-5.

Table B2 shows summary statistics by ownership and shape. Over 70% of projects are publicly owned (Panel A). These are about five times larger than non-public projects. Grouped by shape, only 10% of projects are nonlinear, but these are five times larger than linear projects (Panel B). Table B3 shows the project distribution by ownership, shape, and category. Public projects are largest because resettlement and irrigation—the second and third largest category—are almost all publicly owned (column 1). Mining and "other" feature a more even public-private split than any other category. Nonlinear projects are largest because nearly all mining and resettlement projects are non-linear (column 5).

Table B4: Impact of Infrastructure on Species Diversity by Project Ownership and Shape

| | (1) | (2) | (3) |
|---|---|---|---|
| *Panel A: Ownership* | | | |
| Public | -0.197*** | -0.203*** | -0.204*** |
| | (0.047) | (0.048) | (0.048) |
| Private | -0.806 | -0.299 | -0.297 |
| | (0.925) | (0.737) | (0.735) |
| Neither | 2.911 | 0.852 | 0.851 |
| | (1.854) | (0.667) | (0.669) |
| *Panel B: Shape* | | | |
| Linear | -0.260 | -0.345* | -0.344* |
| | (0.193) | (0.168) | (0.168) |
| Nonlinear | -0.177*** | -0.170*** | -0.171*** |
| | (0.043) | (0.036) | (0.035) |
| Weather Controls | Yes | Yes | Yes |
| Behaviour Controls | No | Yes | Yes |
| General Economic Trends | No | No | Yes |
| Outcome Mean | 23.748 | 23.748 | 23.748 |
| User x Year FEs | ✓ | ✓ | ✓ |
| District FEs | ✓ | ✓ | ✓ |
| State $\times$ Month FEs | ✓ | ✓ | ✓ |
| Observations | 161896 | 161896 | 161896 |
| $R^2$ | 0.559 | 0.690 | 0.690 |

Note: The outcome in all specifications is mean species richness across users' trips in a district-month. In Panel A, rows denote cumulative area of infrastructure encroachments by projects of a particular ownership type in a district-month. Panel B reports the same by project shape. Column 1-3 successively add controls. Weather controls include temperature and rainfall. Behaviour controls include: traveling trips, log duration, log distance, log group size, and log spatial coverage. General economic trends is measured by nightlights. Standard errors clustered by biome.

## B2 Additional Results

### B2.1 Impacts by Project Ownership and Shape

Table B4 groups estimates by project ownership and shape. Public projects are the main threat to biodiversity (Panel A), mainly because resettlement, transport, and irrigation projects—the three most harmful categories—are almost all publicly owned (Table B3). In contrast, the effect of private projects is noisy, likely because "other" projects are largely privately owned (Figure 4B). Similar logic helps interpret the estimates by project shape

(Panel B). Nonlinear projects have a robust negative impact on species diversity. The magnitude is a combination of the small mining and large resettlement coefficients in Figure 4B, which are almost all nonlinear (Table B3). The tradeoff for linear projects is weaker but the magnitude is twice as large. The large magnitude is a combination of the transportation, electricity, and irrigation coefficients, which are mostly linear (Table B3). Wide standard errors likely come from "other" projects, which are also mostly linear.

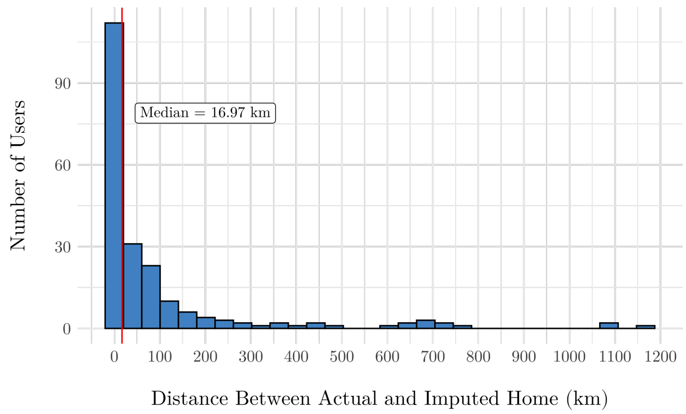


Figure B1: Distance between real and imputed home locations

Note: Data are from 210 eBird users who volunteered their actual home locations. Distance is the straight-line distance between their actual home and the centre of their trips (imputed home).

## B3 User Demographics

eBird does not release data on user demographics. This appendix describes a method for inferring demographics when official data is unavailable. First, I impute user home locations as the gravitational centre of their trips. Second, I compare the distribution of user home locations to the general population to see whether they are rural or urban. Lastly, I characterize users more precisely by studying respondents from a large household survey who live near eBird users. The last two steps are inspired by Blanchard et al. (2023).

**User Home Locations.** Home is defined as the gravitational centre of users' trips. I start with the full sample frame (all protocols) and find the centroid of users' trips. Since trips far from the main cluster (e.g. trips during vacation) warp the centroid, I drop outliers and then re-compute home. Outliers are identified by computing the straight-line distance from home to each destination, and then dropping those with distances < Q1-1.5*IQR or above Q3+1.5*IQR, where Q1 and Q3 are the first and third quartiles, respectively.

This method produces a fairly accurate approximation of home location. 210 users volunteered real home locations, which I use for corroboration. I compute the straight-line distance (in km) between their real and imputed home. Figure B1 shows that the median difference is only 17km, which suggests considerable accuracy of the imputation.

I acknowledge this check is based on a selected sample. However, a similar imputation is applied for estimating eBird travel costs by Kolstoe and Cameron (2017), who received special access to eBird member profiles (including home address) for a much larger sample. Their results are robust to using imputed and real home locations.

**Where Do Users Live?** To visualize how representative users are in terms of where they

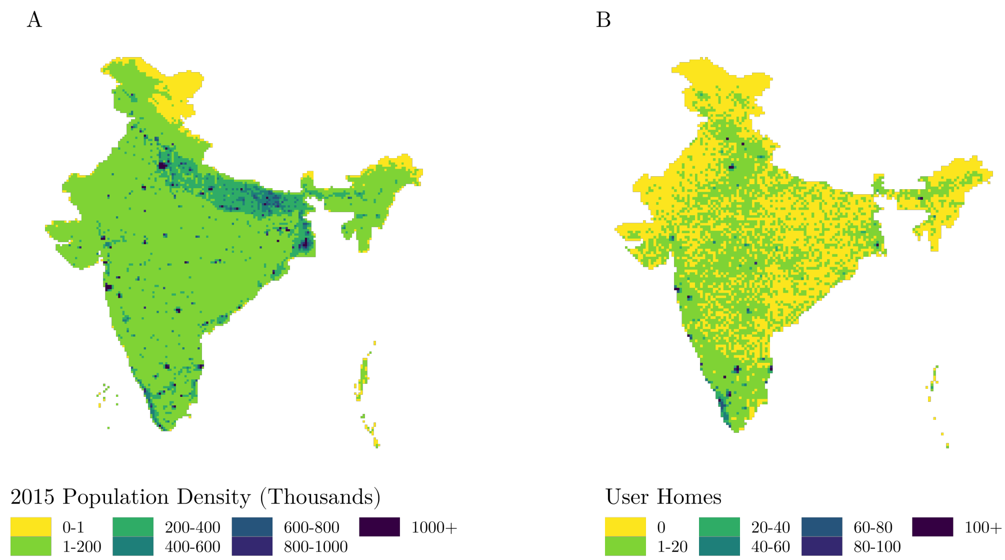


Figure B2: Population density of overall population and eBird users

Note: Both maps are at 20km resolution. Panel A) is the total population count in a cell from WorldPop. The method used for calculating population counts is described in the manual: `https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained`. Panel B) shows the count of eBird user home locations in each cell.

live, I map imputed homes of all 17,634 users against gridded population density data for India. User density is mapped by constructing a $20 \times 20$km resolution grid and counting the number of user homes in each cell. Population density for 2015 is obtained from WorldPop[32]. Data are at 1km resolution and aggregated to 20km for consistency.

Dark hotspots in panel A of Figure B2 are India's largest cities. Many of these cities are also home to the highest density of eBird users (Panel B). While it may be unsurprising that eBird users live in big cities, Panel B also shows many remote eBird users (green).

To assess representativity, I compare the fraction of users living in "mega-cities" with more than 1 million population to that of the overall population. City polygons are obtained from the Global Rural-Urban Mapping Project (GRUMP), and I add a 3km buffer to include suburbs. Overlapping boundaries are dissolved into a single region. Extracting WorldPop counts over these polygons reveals that 27% of the Indian population live in megacities. The equivalent number for eBird users is 43%.

**Location Profiles from the DHS 2015-16 Survey**. As a last step to characterize eBird users, I draw on the DHS, a nationally representative household survey of 600,000 households. Households are grouped into georeferenced clusters, usually a village or town. There are 28,395 clusters with available coordinates. My goal is to identify clusters comparable to

[32]Data accessed from: `https://www.worldpop.org/`. I use the 1km resolution unconstrained mosaic.

Table B5: T-test for equality of means between matched eBird and DHS samples

| Variable | All | Urban | Rural |
|---|---|---|---|
| HH Size | -0.375*** | -0.193*** | -0.341*** |
| Cellphone (=1) | 0.053*** | 0.013*** | 0.040*** |
| Fridge (=1) | 0.253*** | 0.089*** | 0.181*** |
| Car (=1) | 0.089*** | 0.056*** | 0.067*** |
| Sep. Kitchen (=1) | 0.139*** | 0.035*** | 0.190*** |
| Colour TV (=1) | 0.212*** | 0.045*** | 0.190*** |
| Internet (=1) | 0.110*** | 0.062*** | 0.041*** |
| Washing Machine (=1) | 0.191*** | 0.109*** | 0.101*** |
| Flush Toilet (=1) | 0.271*** | 0.053*** | 0.216*** |

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Difference in characteristics between DHS respondents living in similar locations as eBird users compared to the overall DHS survey. The former dataset is weighted by the number of users to which each cluster is matched and the latter uses DHS survey weights. Robust standard errors are bootstrapped.

where eBird users live. One challenge is that cluster centroids are displaced to ensure privacy. Urban and rural clusters are displaced by up to 2 and 10km, respectively.

I start by defining eBird users as urban if they live in cities and rural if not. Cities are defined by GRUMP polygons (see above). Next, I identify DHS clusters within 5km of urban user homes and 10km of rural user homes to account for displacements. This may generate mismatched pairs if, for example, a user living in a Delhi suburb is matched to a nearby rural cluster as well as urban clusters inside Delhi. Therefore, I only keep matches if the population density of the DHS cluster is within 25% of that in the user's home location, both calculated over a 5km buffer. This method matches 61% of users with at least one comparable DHS cluster. Note that the same cluster can match to several nearby users, resulting in duplicates. This is equivalent to a weighted dataset of unique DHS respondents with weights equal to the number of users to which the cluster is matched (Blanchard et al., 2023). I call this the "matched eBird" sample.

This procedure presents a new way to assess whether eBird users live in locations that are statistically similar to the average population. As such, the citizen scientists of India are characterized in a data-poor context. I compare users along several DHS wealth indicators, including household size and ownership of physical assets. T-tests for equality of means are conducted between the matched eBird sample and the overall DHS sample, with bootstrapped standard errors robust to heteroskedasticity. Survey weights are used for the overall sample and the number of matched users for the matched eBird sample.

Table B5 shows the results. Overall, there are statistically significant differences in wealth between the matched and overall sample, indicating that eBird users live in nonrepresentative locations. Compared to the overall population, eBird users live in places with smaller household sizes and better access to amenities such as a fridge, car, separate kitchen, and flush toilets. These differences persist even within rural and urban subsamples. Put differently, the urban locations where eBird users live are wealthier than the average urban location. Yet it should also be noted that these wealth differences are quantitatively small. Thus, while eBird users live in places that are not nationally representative, these places are not markedly atypical either.

## B4 Close Election Design: Context and Data

Section 6.7 tests robustness of the infrastructure-biodiversity tradeoff to an IV strategy based on close elections between incumbents and runner ups. The fraction of constituencies in a district where the incumbent won in close elections is used as an instrument for infrastructure approvals. Identification is based on comparisons of eBird observations among users travelling between places where incumbents just barely won to where they just barely lost. This appendix outlines the political context and data details.

**Context:** India has a federal structure with national and state assemblies. States are partitioned into administrative districts, which are politically significant units since States appoint several district officials, including a District Forest Officer. Districts are further split into single-member State Assembly constituencies with leaders elected through a simple majority voting rule. The constitution requires state elections every five years, although elections are not synchronized across states. One limitation is that my 6-year panel is shorter than ideal for estimating the impact of elections on infrastructure and biodiversity. However, this drawback is partially mitigated by the staggered nature of state elections. There are 32 statewide elections across 30 states during the study period.

**Data:** Election data are from the Trivedi Center for Political Data and distributed through the Socioeconomic High-Resolution Rural-Urban Geographic Dataset on India (SHRUG) (Asher et al., 2021). Both winner-level and candidate-level data are available at the constituency level. The main data include candidate party, election year, and vote share. First, in each election year, I use the winner-level data to identify the winner party in the previous election. Next, I use the candidate-level data to identify incumbent candidates based on whether their party is the same as the previous election winner. 94% of constituencies had an incumbent go up for re-election. Lastly, I compute the win margin as the difference in vote shares between the winning candidate (highest vote share) and runner up (second highest vote share). Elections are quite competitive: half of elections in my sample were decided by margins $< 10\%$. In the main analysis, I classify "close" elections as those decided by margins within 2 percent.

Election data are at the constituency-year level whereas the eBird panel is at the user-district-month level. I use the crosswalk provided in the SHRUG to link constituencies to districts. There are an average of 6 constituencies in a district. I aggregate win margins and close-election dummies to the district level. The latter creates measures of "winning party strength" i.e., the fraction of constituencies with close elections and with incumbent winners. This strategy follows on previous studies that have studied electoral impacts by aggregating over constituencies (Anukriti et al., 2022; Cole, 2009; Clots-Figueras, 2012)