Angry Birds: The Development-Biodiversity Tradeoff in India's Tropical Forests

Raahil Madhok*

This Version: September 21, 2021
Latest Version: click here

Abstract

The tropics hold two-thirds of global biodiversity yet experience more deforestation than any other biome. One exception is India, which recorded stable forest cover amidst rapid economic growth. This paper shows that, despite seeming resilience, India's forests face a substantial development-biodiversity tradeoff. I show this using novel data from 2015-18 on infrastructure encroachments and millions of bird sightings recorded through a popular app. My identification strategy exploits within-observer travel across space and time, overcoming endogeneity from pooling observers with different abilities. I find that development, especially village resettlements, triggers a 3% decline in observed avian diversity, and accounts for over half of India's species loss over four years. I also demonstrate how to remove learning biases in observational data, and provide evidence of species resilience to disturbance in intact forests but not fragmented ones. Lastly, I show that slowing down construction, combined with forest regeneration, can neutralize the tradeoff in the medium run.

Keywords: sustainable development, economic development, infrastructure, biodiversity, conservation, ecological impacts, citizen science.
JEL Codes: Q01, Q56, Q57, Q20, O13.

---

*University of British Columbia, Food and Resource Economics Group, Faculty of Land and Food Systems; madhokr@mail.ubc.com. For helpful feedback and comments, I thank Sumeet Gulati, Patrick Baylis, Frederik Noack, Mook Bangalore, Teevrat Garg, Ryan Abman, Sonja Kolstoe, Sarah Jacobson, Subhrendu Pattanayak, the Wildlife and Conservation Economics Lab at UBC, and workshop participants at Columbia (SIPA Sustainable Development Workshop), UBC CFREE, MIT/Sloan Energy Camp, Canadian Resource and Environmental Economics Association Early Career Workshop, Association of Environmental and Resource Economics Summer Conference, and Camp Resources. I am especially grateful for the SSHRC CGS Doctoral Fellowship for generous funding. All errors are my own.

1 Introduction

Large-scale infrastructure investment is an emblematic feature of economic development. Global infrastructure spending totalled $US 2.3 trillion in 2015 alone, of which 60% was in Asia (Oxford Economics, 2017). Although critical for economic progress, infrastructure expansion narrows the frontier between human activity and fragile ecosystems. Economists have long sought to quantify the environmental costs of development (Grossman and Krueger, 1995; Dasgupta et al., 2002; Copeland and Taylor, 2004), however there is little credible evidence on associated changes in biodiversity beyond global simulations (Newbold et al., 2016). A deeper understanding of the development-biodiversity tradeoff can assist countries to integrate conservation into development planning and achieve win-win outcomes.

This paper provides a first step towards understanding the development-biodiversity tradeoff in the tropics. The tropics harbour two-thirds of Earth's biodiversity (FAO and UNEP, 2020), yet experienced over half of its deforestation since 2000 (Pacheco et al., 2021). India notably avoided widescale forest loss—mainly through concerted tree-planting—despite recording rapid economic growth (Forest Survey of India, 2019). However, this does not negate a tradeoff: tree cover abstracts from nuanced forest ecology whereby, even if development leaves forests unscathed, important inhabiting species may become threatened and require policy attention. Elusive measurement of such species has led to biodiversity being overlooked in studies of anthropogenic drivers of deforestation (Foster and Rosenzweig, 2003; Burgess et al., 2012).

I uncover the development-biodiversity tradeoff in India's forests between 2015-2018. This constitutes a unique laboratory 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 their sightings on species-specific (e.g. eBird, eButterfly) or general (e.g. iNaturalist) platforms launched over the last 5-10 years. Their geocoded uploads serve as a new, high-resolution repository of biodiversity data unmatched in the literature. Third, despite stable forest cover on net, small encroachments for infrastructure (hereafter, forest infrastructure) are commonplace in India and account for 20% of annual deforestation. The Forest Act (1980) mandates rigorous environmental review of such projects before construction. This process was fast-tracked and digitized in 2014, jeopardizing pristine forests while also unlocking a new data source for estimating the development-biodiversity tradeoff at a local level.

To measure development, I extract key data from thousands of forest infrastructure application by firms that passed environmental review. Each observation describes a forest patch diverted for construction and uniquely bundles both infrastructure development and deforestation into a single variable. For my purposes, this new data improves on conventional satellite measures because the latter cannot distinguish the source of deforestation. Moreover, pixel values are annual aggregates, which masks variation in deforestation throughout the year. In contrast, my data specifically captures development-driven deforestation and features sharp monthly changes to the landscape as projects are rolled out.

To measure biodiversity, I obtain hundreds of thousands of geocoded birdwatching diaries from eBird, the world's largest crowd-sourced platform for wildlife sightings (Sullivan et al., 2009). Birds are a credible indicator species for ecosystem health, sensitive to environmental change, and documented with high spatial precision (Morrison, 1986; Gregory et al., 2003). I track nearly 10 million observations by eBird users to assemble a biodiversity dataset with with unparalleled spatiotemporal resolution. These observations span 95% of Indian districts from the Himalayas to the Western Ghats.

The fine-grained data repository permits a research design based on high-dimensional fixed effects, which I employ to estimate the impact of development on bird species diversity (hereafter, species diversity) in a typical Indian district. The data also allows me to characterize heterogeneity in a way that has eluded previous studies. I disaggregate estimates down to the project category to show which types of infrastructure are the least and most harmful. I also stratify districts by initial forest fragmentation level to reveal whether infrastructure has differential effects in pristine or already-fragmented habitats. These estimates can help policy makers optimize the allocation of conservation budgets by choosing the location and infrastructure mix with minimal harm to local ecosystems.

Despite the power of citizen science, its opportunistic nature yields more endogeneity than typical administrative datasets. For example, the Siberian bird migration to India in winter, and lull in birdwatching activity during monsoons, induces stark seasonality. Users also disproportionately visit more biodiverse locations, especially districts across the Western Ghats. Lastly, users possess a range of abilities, complicating inference from cross-user comparisons. I employ time fixed effects to address seasonality, district fixed effects to address site choice, and individual fixed effects to compare trips *within* the same user. Even with the ability bias removed, within-user residuals still trend upward due to learning. I use user-by-year fixed effects to to address this and show that this method outperforms others in the literature.

My analysis yields three key findings that inform conservation policy. First, the global development-biodiversity tradeoff established from simulations is visible at the local level using more disaggregated data and quasi-experimental methods. One square kilometre of forest infrastructure reduces species diversity by 3%, 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. In aggregate, over half of the observed decline in species diversity over the study period can be attributed to development in India's forests.

Second, village resettlements and "other" projects drive the bulk of species loss. Resettlements involve relocating entire communities either due to displacement from other projects or because they live in a protected area. These projects are much larger than average because they involve constructing settlements for thousands of households. "Other" projects include ecotourist resorts, petrol stations, and police camps. Surprisingly, several other categories increase species diversity. Mining in particular has a large positive impact possibly because mines open up dense forest edges where differentially adapted species congregate (called edge effects).

Third, species are resilient to encroachments in intact forests, but not in fragmented ones. My results show that a marginal intrusion has no impact in intact forests, but reduces species diversity by up to 20% in patchy districts. This advances a long-standing debate in ecology about which landscape types merit prioritization (Betts et al., 2017).

Translating my results into comprehensive biodiversity projections is out of scope because reliable projections necessitate models of species ranges, species-area functions, and complex ecosystem dynamics. Instead, I estimate species responses to different project review schemes. My calculations show that slowing the fast-track policy yields less per-period fragmentation until intrusions are small enough that species are tolerant to disturbance. However, the cumulation of marginal intrusions has a net negative effect in the medium run. Thus, I argue that afforestation is paramount to offset erstwhile encroachments and produce a connected landscape that preserves biodiversity. Although limited in scope, these results are important given India's plans to prioritize Northeastern states, which have the highest forest cover in the country, for industrialization (Nayak et al., 2020).

Two more limitations should be noted. First, the estimated development-biodiversity tradeoff may be overstated due to approved projects that were never constructed. I am unable to ground-verify each project with satellite imagery, so I assume deforestation work begins within the month. Many studies using permitting data share this limitation, and the one-month assumption has been invoked in other infrastructure studies in India (Aggarwal, 2018). Indeed, I find little evidence of delayed construction through an event study and distributed lag specification. Second, my species diversity measure is imperfect, since eBird users are not in the business of collecting representative data. Despite sample restrictions to improve representativity, and controlling for effort and learning, measurement bias may persist. All research designs using citizen science data share this limitation.

My main contribution is to document the extent of biodiversity loss from development. In doing so, I am able to bridge the economics and ecology literatures. Most economic studies that quantify infrastructure externalities estimate pollution costs (Currie et al., 2015; Hanna and Oliva, 2015), a handful estimate forest loss and, to my knowledge, none explicitly examine biodiversity[1]. Most related, Asher et al. (2020) and Garg and Shenoy (2021) find surprisingly little effect of infrastructure on forest cover in India. Baehr et al. (2021) also find muted effects in Cambodia. While one may conclude that ecosystems are resilient to development, my results indicate otherwise. Using more detailed species-level data, I show that development *does* trigger species loss, and am among the first to show this in the environmental economics literature.

My findings are also relevant for the ecology literature, which has long documented anthropogenic pressures on ecosystems. In these studies, field workers often count species in transects with differing levels of human activity. This yields accurate data but limits study designs to cross-sectional comparisons (Reis et al., 2012; Stephens et al., 2004). Although citizen science dramatically improves data coverage, its low quality has garnered more interest in identify-

[1] An exception is Liang et al. (2020), who find reduced bird abundances from U.S. industrial emissions, but the effects of land use change from infrastructure per se is unexplored.

ing endogeneity—particularly 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 both documented biases as well as undocumented ones, especially within-user learning, to arrive at quasi-experimental estimates of the development-biodiversity tradeoff[2]. These estimates can be generalized beyond existing cross-sectional estimates due to the national-scale panel nature of my data.

The rest of this paper is organized as follows. The next section provides background on infrastructure-driven deforestation in India. Section 3 describes the administrative and citizen science data. Section 4 and 5 presents the empirical strategy and results, respectively. Section 6 discusses implications for conservation policy, and Section 7 concludes.

2 Background

India's Forest Act (1980) Regulates Construction in Forests. India's Forest (Conservation) Act (1980) protects its forests from "conversion to non-forest uses" (MoEFCC, 1980). Forest 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 since many fauna avoid clearings as narrow as 30 metres (Riley et al., 2006; Benítez-López et al., 2010). These projects also generate noise pollution, change soil properties, and facilitate invasive species dispersal (Laurance et al., 2009).

When non-forest sites are unfeasible, the Act permits forest infrastructure pending an application and review process. It also sets up a forest advisory committee (FAC) to rule on construction proposals. Projects involving *any* amount of deforestation undergo the review process.

Despite the Act's intent, huge swathes of India's forests have been transferred to public and private firms. Between 1985-2014, approximately 4000 km^2 was clearcut for 23,000 infrastructure projects. Overall deforestation during this period averaged 835km^2 per year (Meiyappan et al., 2017)[3], implying that infrastructure intrusions accounted for 20% of annual deforestation in the three decades preceding this study.

Project Approval is Granted via Forest Clearances. The step-by-step journey of a project proposal was delineated in a 2003 Amendment (MoEFCC, 2003) and is known as the Forest Clearance (FC) Process[4]. There are two stages: stage-I approval is granted after environmental review. Stage-II is granted after payment into an offsetting tree-planting fund.

To receive stage-I approval, the firm first submits an application to the District Forest Office (DFO). The DFO may commission a site inspection report, which typically includes stipulations

[2] An exception is Noack et al. (2021), who use eBird to estimate the causal impact of farm size on species diversity, controlling for learning.

[3] Forest loss between 1985-2005 was 18,000km^2 (Meiyappan et al., 2017). Between 2006-2014, it was 6223km^2 (Global Forest Watch). Therefore, the annual deforestation rate between 1985-2014 is: $(18000 + 6223)/29 = 835km^2/year$.

[4] Terminology is important here. Forest Clearance refers to a firm receiving "clearance" to initiate construction, not clear-cutting forests, although the latter follows.

to change project size, area, or move location altogether. The report is forwarded to the State Forest Department, which can add more stipulations. 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.) are directly forwarded to the Ministry of Environment Forests and Climate Change (MoEFCC). The FAC meets[5] to rule on stage-I, indicating that environmental review is complete.

My data consists of firms with stage-II approval. To receive stage-II approval, the firm must first pay into a compensatory afforestation fund. The money is used for afforestation on contiguous non-forest land or, if unavailable, land elsewhere in the state. Rates are fixed by the state and tree-planting begins within 1-2 years (MoEFCC, 2013). After payment, the central FAC makes their decision and, if approved, the firm can begin deforestation. The MoEFCC may permit deforestation before stage-II in special cases, and premature deforestation without approval has also been reported. I test for this behaviour in the robustness checks.

The Forest Clearance process was Fast-Tracked in 2014. In 2014, then-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 stage-I and II approvals jumped 60% compared to the previous 40 years (authors calculation). The share of rejected proposals also considerably declined post-2014. Faster approvals became possible through easing norms (e.g. diluting no development zones), eliminating public hearings, and changing thresholds for MoEFCC review. Pre-2014, projects larger than 15 ha. were reviewed by the MoEFCC, whereas the new 40 ha. threshold implies that the majority of projects are approved without Ministry involvement.

Another cornerstone of the fast-track initiative is the digitization of the FC process. An online portal automates each stage of the decision process and enables a reduction in turnaround time. For research purposes, an added benefit is process standardization, which reduces variation from state-level bureaucratic differences.

3 Data

I estimate the development-biodiversity tradeoff by drawing on several new datasets. I use FC infrastructure permits as a measure of development in India's forests. Data on species diversity are from eBird, a digital diary for birdwatchers. Its individual level nature and vast coverage allows for a saturated model that reduces important selection biases. I combine these with multiple high-resolution satellite datasets to control for weather and forest characteristics as well as to explore treatment heterogeneity. The final panel covers nearly all of India from 2015 to 2018. This section describes the data, panel construction, and methods for reducing endogeneity.

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

3.1 Data Description

Forest Infrastructure. This paper studies the encroachment of economic activity into India's forests, which cover 22% of the country (Forest Survey of India, 2019). Administrative data on development-driven deforestation rarely exist, and previous work largely relies on satellite data. I choose not to use satellites due to difficulty distinguishing anthropogenic intrusions from natural sources (e.g. forest fires). Moreover, satellite data report annual aggregates, which mask encroachments scattered throughout the year with potentially large impacts on local ecosystems.

I construct a new dataset of infrastructure encroachments using digitized FC applications. I scraped all project proposals submitted through the online portal[6] since its creation in 2014 and assembled a district-monthly panel spanning five years. The sample frame includes nearly 20,000 proposals at various stages, each with information on: project category (road, mine, etc.), forest and non-forest area requested for diversion, and the application stage. I homogenize similar project categories to simplify the analysis. A district-wise breakup of forest diversion is provided for projects spanning multiple districts (e.g. transmission lines). I geocode district names using the 2011 Census shapefiles to avoid issues with district splitting and to match across other spatial datasets. Section B1 provides more details on data preparation.

The analysis sample consists of 2,770 stage-II approved projects. Three "mega-projects" in Telangana severely skew the project size distribution and were dropped (see section B1 for details of these projects). I assume deforestation work begins in the month that the FC permit is awarded. In the robustness checks, I consider stage-I approvals, extend the window between approval and construction, and add back the outliers.

I aggregate project-level data to district and year-month level, both overall and category-wise (e.g. total deforestation in January 2018 for electricity projects in Delhi). I do this for two reasons. First, projects are not geocoded and only identify the district. Second, the district is the administrative planning unit immediately below the Indian state (similar to a United States county) and forms a natural unit for local policy implementation. States wield significant decision-making power when reviewing project proposals in their districts in consultation with the DFO and MoEFCC. I balance the panel by assuming that districts and time periods not in the portal had zero stage-II approvals. This is reasonable since the portal contains all applications submitted during my study period. This improves sample size and power, but also may lead to overstated effects since pre-2014 submissions (not in my sample) may have been approved between 2015-2018. The latter point is a study limitation since I am unable to collect data from non-digitized applications.

Species Diversity. eBird entered the Indian market in 2014 and requires only a smartphone. On each trip, users provide a detailed taxonomy of bird sightings called a checklist, and the app automatically records the location and date. Every checklist is vetted by ornithology experts

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

(Sullivan et al., 2009). eBird is among the best citizen science products for research because it records information about the observation process, including: GPS coordinates, trip duration, protocol (e.g. stationary or travelling), and whether all detected species are recorded, called a complete checklist, or only highlights. I use this information to select users with checklists that best represent species diversity on the ground as well as to control for observer effort.

My sample frame consists of the eBird Basic Dataset for India (eBird Basic Dataset, 2019), which includes nearly 730,000 trips between 2015-2018 that passed the vetting process. The collective checklists comprise nearly 13 million trip-species observations. To identify representative users, I first restrict the sample to stationary, traveling, banding, and random trips (see Table A1 for a full breakup). Banding involves capture-and-release methods for large scale research endeavours. Random trips involve periodically finding a new location by walking 3 to 5 miles in a random direction from the previous location. Excluded protocols reflect efforts to find specific species. Second, I remove duplicate observations among birdwatching groups where a "leader" copies their checklist into group member's accounts. Third, I remove checklists recording a single species, where birdwatching is unlikely the main activity (e.g. observing a hummingbird through the window while cooking). Lastly, I use trip coordinates to identify districts as per 2011 boundaries. This provides a matching key across datasets and also reveals off-coast boating trips, which I drop. The sample selection process leaves approximately 8.5 million observations, 486,000 trips, 10,000 users, and 608 districts (out of 640).

My outcome variable is species richness, the number of unique species observed during a trip. Species richness indicates conservation value, proxies the number and stability of ecosystem services, and is a widely used biodiversity metric (Fleishman et al., 2006). Alternative metrics integrate abundance weights, however eBird species counts involve considerable measurement error; 11% of checklists omit counts altogether. Nevertheless, I estimate results with two abundance-weighted diversity indices in the robustness checks.

To construct a panel, I aggregate mean species richness across users' trips in each district and time period. This keeps the 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. I keep count of the number of trips over which the mean is computed and use this as a regression weight. I truncate this variable at the 99th percentile to exclude outliers[7].

Covariates. The first set of covariates are environmental and include: temperature, rainfall, and forest cover. Controlling for weather is important because it affects species detection. Monthly temperature is from the ERA5 reanalysis product on a $0.125° \times 0.125°$ grid (Hoffmann et al., 2019). Rainfall is from the NASA-operated GPM Level 3 product on a $0.1° \times 0.1°$ grid (Huffman et al., 2019). Forest cover is an important covariate because it partially determines project placement and because checklists may be longer in forested districts compared to sparse ones.

[7] As an example of one outlier, the maximum is 890 trips per month in the same district by a single user. This amounts to approximately 30 birdwatching trips *per day*.

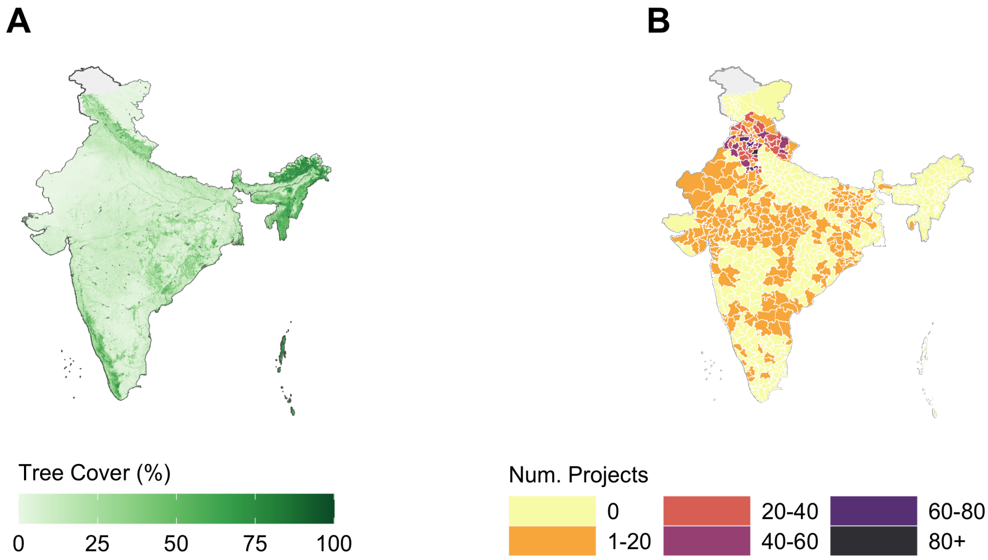

Figure 1: Spatial Distribution of Forest Cover and Forest Infrastructure

Note: Panel A shows a heat map of 2015 forest cover. Pixels are at $250m \times 250m$ resolution and shaded by percent forest cover. Panel B maps the number of development projects requiring deforestation that were approved for construction between 2015-2018.

Annual forest cover is from the VCF satellite product, which measures the percentage of a pixel under forest cover on a 250×250m grid (Townshend et al., 2017). To compute the covariates, I extract mean monthly rainfall (mm) and temperature (°C), and mean yearly forest cover, over all cells within a district weighted by cell overlap fraction.

The second set of covariates captures observer effort and includes: trip duration, site accessibility, and checklist representativity. Duration (minutes) is automatically recorded by eBird. Access refers to projects opening up previously inaccessible forest patches. Some projects involve approach roads for shuttling supplies, which may draw an influx of birders to new sites and upward bias my estimates. I measure accessibility by first overlaying each month's trip coordinates onto a 5km x 5km grid and then computing the fraction of district cells traversed by eBirders. To my knowledge, this is the first study that measures and controls for access in this way. Lastly, although 95% of checklists are complete, I control for checklist representativity as an extra safeguard. Panel aggregation transforms the completeness indicator into a continuous measure (on a scale of 0 to 1) of trip representativity in a given district-month. I include this as a covariate because some incomplete checklists may still accurately represent local species diversity. In the robustness checks I drop incomplete lists altogether.

3.2 Summary Statistics

Figure 1 illustrates the spatial distribution of forest cover and forest infrastructure. Dense forests are concentrated in North and Northeast India as well as the Western Ghats (Panel A). Areas with 10-30% forest cover are found nationwide. Apart from the Northeast, most forested areas have been fragmented by infrastructure (Panel B). The dense forests of Punjab, Haryana, Himachal Pradesh and Uttarakhand in Northern India suffered the most encroachment.

The average project is 4.76 ha., roughly seven soccer pitches, and a total of 13,000 ha. was approved for deforestation during the study period (Table A2). Transportation accounted for nearly 40% of this total and is also the largest category share. In contrast, resettlement projects make up the fewest projects, although average size vastly exceeds most other categories.

Table A3 summarizes the outcome and covariates in districts with and without forest infrastructure. Users record three more species per trip, on average, in districts with infrastructure than districts without. Of course, this naive comparison bundles other time-varying and static differences. Tree cover, birdwatching activity, and trip length are also higher in deforestation districts. Deforestation districts also have substantially lower population density than non-deforestation districts, leading to higher species availability through less human disturbance. The next section demonstrates how to correct such biases.

3.3 Mitigating Selection Bias in Citizen Science

The promise of citizen science lies in its dual aim of large-scale data collection and community engagement. However, loose restrictions on when, where, and by whom data are collected yields more endogeneity than typical administrative datasets. Fortunately, eBird records details about the observation process, which I leverage to mitigate data quality concerns.

Seasonality. The seasonality bias arises from the ability to record trips at any time[8]. Figure 2A demonstrates 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 birdwatching 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.

Site Selection. The site selection bias arises from the ability to record species from anywhere[9]. This triggers a convergence of activity in more "attractive" districts. Figure 2B shows that eBird users are more active in districts with higher actual species diversity (not eBird-reported) compared with less diverse ones. Users will populate longer checklists in the former districts since the species pool is larger. Previous studies account for this with a variety of fixed landscape covariates (Kelling et al., 2015), but species richness may still vary endogenously. I take a stricter approach with district fixed effects that rules out site selection biases more confidently.

Learning Bias. The learning bias arises from the wide range of user abilities (Farmer et al., 2014; Fitzpatrick et al., 2009). I decompose this into a fixed and variable component, called innate ability and the learning curve, respectively. Low-ability users might misidentify species

[8] In contrast, other surveys like the North American Breeding Bird Survey only operate in specific months

[9] In contrast, the North American Breeding Bird Survey mandates specific routes

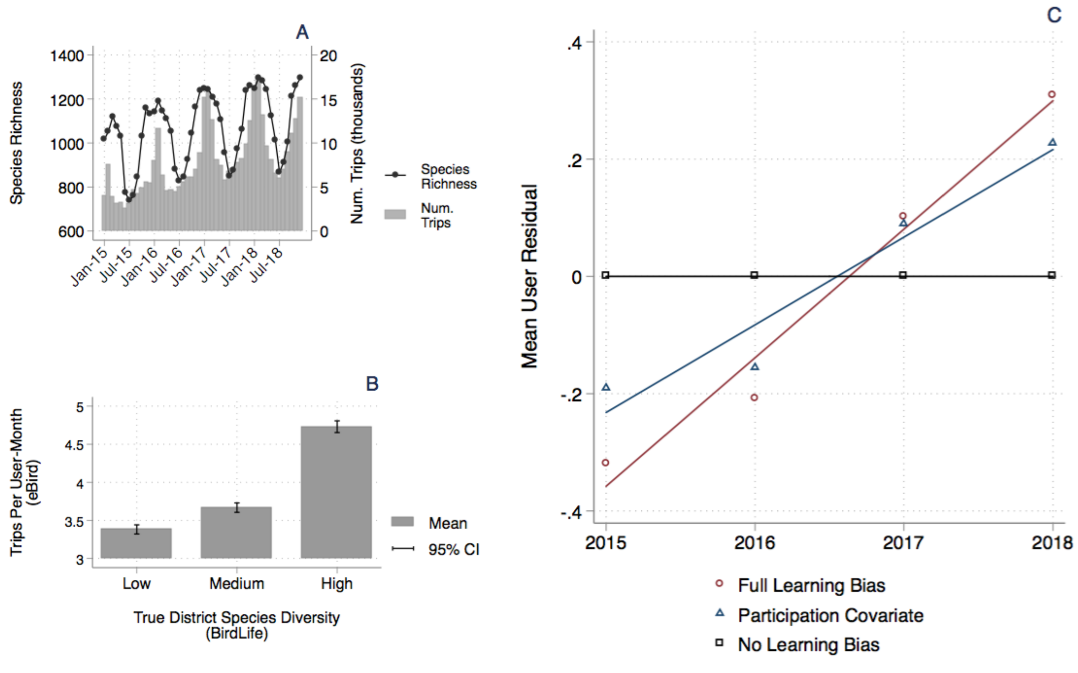

Figure 2: Biases in Citizen Science

Note: The left y-axis of panel A shows total species richness across the study sample. The right y-axis shows the total number of trips across all users. Panel B shows mean number of trips per user-month in three quantiles of *actual* species richness, as per historic range maps from BirdLife International (2018). Red circles in Panel C shows the average residual variation in species richness per user from regressing species richness on user, district, state-month, and year fixed effects. Blue triangles show the same with a control for experience. Black squares partial out user-year, district, and state-month fixed effects.

(false positives) or overlook them (false negatives), whereas the opposite is true for high-ability users. Volunteer training or ability scores are typically used to reduce inter-observer variability. Kelling et al. (2015) construct a fixed ability score for eBird users based on predictions from a random effects model. For causal inference, however, this requires a strong orthogonality assumption between the score and all other unobserved user attributes. I relax this assumption by simply comparing species richness across time and space *within the same user*, which makes the ability score superfluous. Figure A1 verifies that there is sufficient variation to do this.

Red circles in Figure 2C show species richness residualized on user, state-month, and district fixed effects. Notably, a steep upward trend remains, which is illustrative evidence of the learning bias. To account for this, I first replicate Kelling et al. (2015) and control for a participation index. The index increments with each month of experience and assumes constant returns. The learning curve flattens as expected, but is not fully absorbed (blue triangles). This is novel evidence that learning is driven not only 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 month-to-month, gradually listing rarer species over longer time spans 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 number of trips accumulated, number of months per year of birdwatching, 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 the residual variation has been stripped of the learning bias.

4 Empirical Strategy

Main Specification. My analysis leverages panel fixed effects to quantify the development-biodiversity tradeoff across India. Infrastructure projects periodically began fragmenting district forests throughout my study period. At the same time, eBird users ventured to these same districts to record birds. My specifications compare observed species diversity *within* a user's trips as they travel for birdwatching. The identification 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 forest intrusions. I implement the following estimating equation:

$$SR_{idt} = \alpha + \beta_1[Infrastructure]_{dt} + \beta_2[X]_{idt} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idt} \tag{1}$$

where SR_{idt} is mean species richness observed by user i across their trips in district d during time t (year-month). $Infrastructure_{dt}$ is the cumulative area of forest infrastructure in the same district and time period. X_{idt} is a vector of mainly weather and behavioural covariates described in section 3. Importantly, X_{idt} includes a measure of site accessibility that accounts for projects simultaneously opening up previously inaccessible parts of the forest. It also includes the cumulative area of non-forest land diverted for the same projects. User-by-year fixed effects, ϕ_{iy}, absorb cross-user differences in ability and account for individual-specific learning. District fixed effects, γ_d, absorb mean checklist length and other other fixed district attributes. State-by-month fixed effects, θ_{sm}, control for state-specific seasonality. Conditional on covariates and fixed effects, β_1 identifies the impact of development on species diversity by leveraging remaining variation across users' trips in different locations and months within the year.

The specification is weighted by the number of trips underlying the outcome, since SR_{idt} is an aggregate. The cumulative distribution of the weighting variable (Figure A2) shows that 10% of observations are a mean over more than 10 trips (precisely measured) and 90% are over fewer than 10 trips (less precisely measured). Regression weights ensure that observations influence β_1 in proportion to their measurement precision rather than be treated equally.

I view standard error clustering mainly as an experimental design issue (Abadie et al., 2017), which leaves the choice of cluster somewhat subjective. I cluster at the biome level to allow for arbitrary correlation of ϵ_{idt} across time and space within a biome[10]. Biomes form the most appropriate cluster because they delineate biological communities with shared eco-climatic characteristics that are unobserved in my model. In contrast, unobservable components of biodiversity are unlikely to adhere to district or state boundaries. I still report results from district level clustering in a robustness check since $Infrastructure_{dt}$ varies at the district level. Clusters are not interacted with time to account for serial correlation (Bertrand et al., 2004).

Threats to Validity. The research design rules out several alternative mechanisms that threaten identification. First, it is unlikely that other infrastructure programs unfold with the same sharp timing as project approvals. For example, the Pradhan Mantri Gram Sadak Yojana road program was rolled out during my study period to provide paved roads to unconnected villages. However this, along with other development programmes, affects districts with forest infrastructure as well as those without (or yet to be treated). The threat is thus limited to factors differentially affecting districts with forest infrastructure and with the same timing as project approvals. This threat is minimal given the targeted nature of cross-cutting development programmes based on rules unrelated to forest cover (Asher et al., 2020; Burlig and Preonas, 2016).

A second threat to identification arises if projects are selectively approved in districts with trending biodiversity. Equation (1) is identified under the assumption that, in the absence of forest infrastructure, districts receiving and not receiving projects would have experienced similar changes in species diversity. Since biodiversity tends to change over longer periods, the sharp timing of project approvals again guards against these slower changes. To be sure, I estimate an event study and find little evidence of pre-trends. While the FAC may approve projects in districts with certain characteristics, these characteristics are generally orthogonal to local biodiversity around the date of approval.

The final threat to identification is spatial spillovers. If projects are spatially correlated, then β_1 may be biased by two opposing forces; biased downward if species in district d are threatened by development elsewhere, and biased upward if development elsewhere drives species immigration into district d. This threat is arguably minimal in my context. Industrial clusters are unlikely to form *in forests* because of Forest Act (1980) stipulations, making other-district forest infrastructure plausibly orthogonal to that in district d. An exception is spatial correlation from single projects spanning multiple districts. However Table A4 shows that less than 4% of projects spill into adjacent districts. In the robustness checks I verify that, as expected, β_1 is robust to a variety of spatial dependence structures.

Decomposed Specification. I decompose $Infrastructure_{dt}$ in equation (1) into individual project

[10] I obtained biome shapefiles from the Nature Conservancy (accessed from `http://worldmap.harvard.edu/data/geonode:wwf_terr_ecos_oRn`). For districts covered by multiple biomes, I choose the biome with the largest overlapping area as the cluster.

categories. This not only reveals particularly harmful categories, but also those with negligible, or even positive, impacts on species diversity. I estimate the following specification:

$$SR_{idt} = \alpha + \sum_{k=1}^{6} \beta_{1k}[Infrastructure]_{kdt} + \beta_2[X]_{idt} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idt} \quad (2)$$

where the term under summation is cumulative forest area diverted for category k projects in district d at time t. Non-forest area diverted for corresponding categories is included as a covariate. Remaining terms are defined as in equation (1). β_{1k} reveals the differential impacts of e.g. irrigation versus transportation projects, and are useful for policymakers tasked with allocating the right mix of projects that balance development and conservation.

Treatment Heterogeneity. There is debate among conservationists about whether biodiversity is better conserved by protecting intact or already-fragmented landscapes. While both approaches are valuable, some species may be surprisingly resilient to landscape modification up to a threshold. I estimate treatment heterogeneity over the baseline "patchiness" distribution, which helps resolve the debate in addition to non-parametrically identifying habitat thresholds:

$$SR_{idt} = \alpha + \sum_{j=1}^{5} \beta_{1j} \left(\mathbb{1}\{F_{jd}\} \times Infrastructure_{dt}\right) + \beta_2[X]_{idt} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idt} \quad (3)$$

where $F_{jd} = 1 \; \forall \; ForestCover_{jd} \in [j, j+1)$ in baseline year 2015 for district forest cover quintiles $j = \{1, ..., 5\}$. A separate coefficient, β_{1j}, is estimated for the effect of a one km^2 intrusion in each quintile of the initial forest cover distribution. It also enables identification of non-linearities without imposing any modelling restrictions, which is important for determining critical habitat thresholds.

5 Results

This section presents three important pieces of evidence on the development-biodiversity trade-off. First, I find a statistically significant decline in local species richness following the development of infrastructure in India's forests, across a variety of specifications and robustness checks. Resettlement and "other" projects are particularly harmful. Second, species are resilient to habitat encroachments in relatively intact forests, but not in fragmented forests. Lastly, I show that species loss persists over the medium run.

5.1 Development Reduces Biodiversity

Figure 3A shows specifications with and without the learning curve. The former includes user along with district, state-by-month and year fixed effects. The latter estimates equation (1),

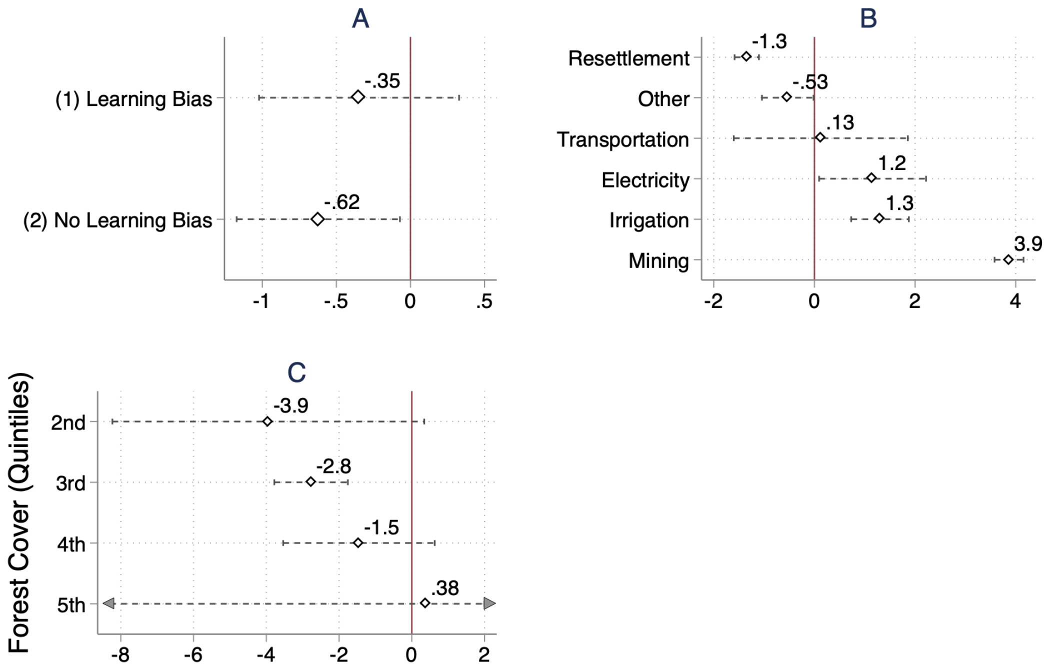

Figure 3: Impact of Forest Infrastructure on Species Richness

Note: In panel A, specification (1) includes fixed effects for user, district, state-month, and year. Specification (2) includes user-by-year district, and state-month fixed effects. Panel B is a single regression where deforestation is decomposed into project categories. Panel C is from a single regression where forest infrastructure is interacted with quintiles of baseline forest cover. The coefficient on the interaction is shown for each fragmentation quintile (y-axis). 95% confidence intervals are shown by the dashed bars. All regressions are weighted by the number of user-trips in a district-time period and include controls for: non-forest area, percent tree cover, access, temperature, rainfall, and reporting type. Standard errors clustered by biome.

with user-by-year fixed effects to absorb the learning curve. The difference between the two represents a straightforward test of the learning bias in citizen science, which has rarely been quantified (Kelling et al., 2015).

I begin with the question of whether a development-biodiversity tradeoff exists. The main coefficient in both specifications is negative, indicating that development *reduces* species diversity (see Table A5 for tabulated results). Precision is weak in specification (1) arguably because within-user learning offsets the species decline (see Figure 2), yielding an overall impact statistically indistinguishable from zero. In specification (2), my preferred specification, absorbing the learning curve yields a steep decline in species richness. Comparing both specifications, the learning curve erases nearly half of the infrastructure impact if unaccounted for.

The main coefficient indicates that an additional km^2 of forest infrastructure in a district reduces observed species richness per month by 0.62 species, equivalent to 3% of the average checklist. To put this in perspective, this implies that forest infrastructure accounted for approximately 60% of India's observed species loss between 2015-2018[11]. In contrast, diversion of non-forest land for the same projects has no impact on species diversity (Table A5). This suggests that habitat loss from deforestation is a key mechanism driving species loss. It also implies that a redistribution of economic activity away from fragile forestland can blunt biodiversity decline.

5.2 Effects are Driven by Two Project Categories

In a novel addition to the literature, Figure 3B shows decomposed estimates of the tradeoff (equation (2)). Species diversity declines are mainly triggered by resettlement and "other" projects, which include: petrol pumps, police camps, and communications towers. Resettlements are the largest projects but make up the smallest share due to high costs of relocating communities. An example is the diversion of 2.85 km^2 of forest in the Betul district of Madhya Pradesh to relocate a nearby village within the Satpura Tiger Reserve. The project was approved in April 2017 and includes construction of housing, playgrounds, and roads[12].

Counterintuitively, the remaining categories *increase* species diversity, although the point estimate is imprecise for transportation. One explanation is that roads often carve through contiguous forests and block animal migration corridors. Resulting declines in species diversity may then be offset by reduced threats from predators. Electricity and irrigation projects, in contrast, have less of a barrier effect. However, both have large water requirements, which may draw species otherwise difficult to spot, plausibly explaining the positive and significant coefficients. Indeed, half of the irrigation projects in my sample are drinking water wells, and 10% of electricity projects involve dams or reservoirs.

The positive impact of mining is the largest in magnitude. To understand why, note that Indian mines are often located in remote, densely forested districts and are impressively large—the second largest in my sample. Excavation opens long perimeters of new forest "edges" where only certain species proliferate (Murcia, 1995)[13]. The changing species pool near mines will, in turn, be reflected in eBird. Many mines also require supply roads. The compounded barrier *and* edge effects may explain the higher magnitude and precision of the mining coefficient. Finally, despite the similar size of resettlement and mining projects, settlements usually encroach into green belts with a pre-existing edge. The coefficient thus avoids offsetting edge effects and mainly captures species declines from construction-related habitat loss per se.

[11] The average district had 0.15 km^2 of forest infrastructure between 2015-2018, implying a monthly loss of 0.09 species (0.15 $\times$ -0.62). During the same period, mean species richness declined by 0.15, suggesting infrastructure accounted for 0.09/0.15=60% of species diversity loss over four years

[12] The site inspection report for this project can be found at http://forestsclearance.nic.in/writereaddata/SIR/06022017561SBScan_02-06-2017_1501.pdf

[13] "Edge effects" refer to sharp changes in species abundance and diversity at the border between habitats. Edge species experience more sun, temperature extremes, and lower humidity than those in the forest interior (Murcia, 1995).

5.3 Species are Resilient in Intact Forests

To what extent should resources be allocated towards protecting fragmented landscapes versus largely intact ones? This question is important given the scarce conservation budgets of most countries. My results suggest greater returns from conserving fragmented landscapes. Figure 3C illustrates treatment heterogeneity over the forest cover distribution (equation (5)). The y-axis is the interaction of cumulative project construction with quintiles of baseline forest cover. The first quintile (low forest cover) is omitted so that all coefficients are interpreted relative to intrusions into degraded forests.

A 1 km^2 intrusion poses a larger threat to species in districts with higher baseline fragmentation. Put differently, species are more resilient in intact landscapes compared to fragmented ones. My result corroborates existing theory from landscape ecology (Hanski, 1998) and is among the first empirical tests. An exception is Betts et al. (2017), who find the opposite result: that species are more threatened in intact landscapes. They interact time-varying forest loss with an indicator for baseline forest cover above 90%, whereas I use a continuous measure of baseline forest cover that includes substantially more variation and allows for a more detailed investigation of non-linearities.

Species begin exhibiting a sharp and monotonically increasing decline in districts below the fourth quintile of forest cover. The critical habitat threshold corresponds to 19% forest cover. These shifts are statistically significant and sizeable: a 1 km^2 intrusion into the third forest cover quintile reduces species diversity by 14% of the average checklist. The equivalent decline is 20% in districts with even higher fragmentation. This result is important for understanding spatial heterogeneities in the development-biodiversity tradeoff as well as for designing targeted conservation policy.

5.4 Dynamics of Species Changes

This section considers more flexible specifications to understand the dynamics of biodiversity changes. Documenting dynamics is important for two reasons. First, a delay between project approval and forest diversion suggests that my main effects underestimate the full development-biodiversity tradeoff. Second, even if deforestation begins immediately, a lag may indicate an "extinction debt", an ecology term describing a delay between habitat disturbance and species responses as the affected ecosystem processes accumulate (Tilman et al., 1994). Under this scenario, forest policy should involve habitat restoration, otherwise species declines may ensue from past fragmentation even if the forest is currently protected.

Event Study Results. I begin by estimating an event study regression, which has the dual advantage of informally revealing the dynamics of species diversity as well as providing evidence of pre-trends, a potential threat to identification. I define an "event" as the date of forest diversion in a district and compare species richness in a window before and after. Because of

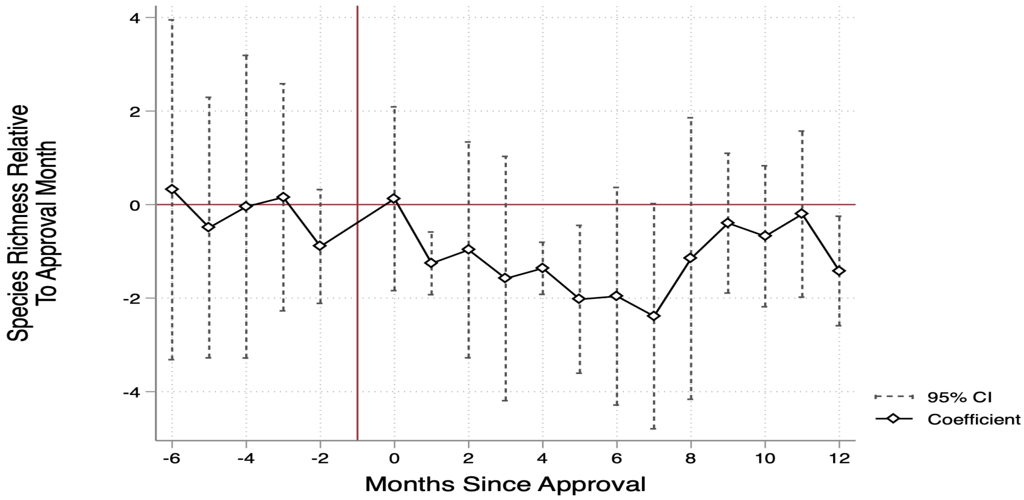

Figure 4: Event Study Results

Note: White diamonds are coefficients from a regression of species richness per user-district-month on the number of days before and after the first project was constructed in a district. Time zero is the approval month and all coefficients are normalized relative to the month before this date. Dotted lines represent 95% confidence intervals. The regression is weighted by number of trips, includes user-year, district, and state-month fixed effects, and controls for non-forest area, percent tree cover, access, temperature, rainfall, and reporting type. Standard errors are clustered by biome.

aggregation, the event could be a single project or multiple ones approved in the same month. I focus on the first event in order to clearly designate before and after groups without overlap from subsequent projects. I estimate the following specification:

$$SR_{idt} = \sum_{k=a}^{k=b} \beta_{1k} \times \mathbb{1}[t - e_d = k] + \beta_2 [X]_{idt} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idt} \tag{4}$$

where e_d is the event date in district d and $\mathbb{1}[t - e_d = k]$ is a dummy that switches on k periods before or after the event. The dummy for $k = -1$ is normalized to zero and omitted for the model to be identified. All fixed effects and covariates are the same as equation (1). Each β_{1k} captures mean species richness k months relative to the month before the event date[14]. Non-existence of pre-trends is indicated by $\beta_{1k} = 0 \; \forall \, k < 0$.

There are two important takeaways from the results in Figure 4. First, species declines are detected almost immediately after approval and persist for about seven months. The declines are strongest one and four months after approval. After five months, the declines dissipate, although districts where no subsequent projects were sited for 8-12 months are likely selective. Overall, the evidence of lagged effects from the event study are weak at best.

[14] I use the window [-6, 12] because a wider one creates variable sample composition across estimates. Since I study the first event only, the pre-post period differs depending on subsequent event timing. For example, with a 24-month post-period, $\beta_{1,24}$ from equation (4) would be noisily estimated off the few districts where no subsequent projects were approved for two years after the first.

Second, Figure 4 provides visual evidence that the parallel trends assumption holds. This corroborates my assumption that projects are not selectively allocated based on pre-existing biodiversity dynamics. Overall, this bolsters evidence that my main estimates can be interpreted causally and do not merely reflect the continuation of a pre-existing decline in species diversity.

Cumulative Lag Results. I estimate a cumulative dynamic lag specification to study treatment effect dynamics more formally. This reveals the persistence of species diversity changes in the medium run. Sharp declines in cumulative effects are evidence of lagged construction or the extinction debt being paid. I simply estimate equation (1) with three lags of $Infrastructure_{dt}$ and report the sum of the baseline effect and the first three lags, which is the net impact of habitat loss three months later. In Figure A3, the cumulative, medium-run impact is nearly equivalent to the baseline impact, with stable point estimates under a narrower (1 month lag) and wider (5 month lag) window. Stated differently, the baseline effect persists in the medium-run. This makes sense since initial forest intrusions trigger species exit, and the habitat remains degraded as the project is built. Although firms fund afforestation, tree-planting typically occurs one to two years later and the offsetting effect is not captured by my estimates (see section 6).

Overall, the event study and medium-run estimates provide evidence that species declines are triggered soon after habitat disturbance and persist in the medium run. Although I find little evidence of an extinction debt, I show that forest regeneration is nevertheless a critical component of forest policy in section 6.

5.5 Sensitivity Checks

In what follows I present a range of sensitivity checks of the estimated development-biodiversity tradeoff. I first consider the influence of spatial spillovers, unauthorized construction, different specifications of seasonality, and extreme values. Next, I consider alternative measures of species diversity and several sample restrictions before turning to alternative clustering techniques. All results are shown in Tables A6 and A7.

Spatial Spillovers. I test whether my estimates are partially driven by development in other districts. Table A6 shows that spatial spillovers pose a minimal threat, as argued in section 4. In column 1, “other districts” are defined as contiguous to focal district d[15]. Conditional on their development, the local tradeoff remains within the confidence interval of the baseline estimate (Figure 3A). This confirms that the few projects (4%) spanning two districts induce minimal bias. In columns 2-5, “other districts” are all others in India, weighted by inverse distance, and with various distance cutoffs since the area of encroachments in one district should be unrelated to those hundreds of kilometres away. The main coefficient remains stable across all specifica-

[15]For each time period t, I multiply $Infrastructure_{dt}$ in equation 1 by a $N \times N$ dimensional (where N is the number of districts in India) spatial weight matrix W with $w_{dj} = 1$ if district j is adjacent to d. W is row normalized so that the weighting operation reflects an average of neighbouring forest infrastructure.

tions. The spillover coefficient is uniformly positive (and significant in three of five columns), implying that species emigrate to other districts when facing disturbance in their focal district.

Premature Construction. My estimates may be underestimated if enough firms begin construction before receiving stage-II approval. In this case, my data records zero *approved* infrastructure even though deforestation work may have begun. I develop a straightforward test to address this concern. Given that stage-II projects negatively impact species diversity, the same should be true for stage-I projects that begin construction early. I add the area of pending stage-I projects to equation (1) and check whether it affects species diversity. Table A7 column 1 shows that, conditional stage-II approved projects, an additional km^2 of stage-I projects has no effect on species diversity. Although unauthorized logging does occur in practice—confirmed by media reports—my results suggest this is not widespread, at least to the extent that it can be picked up by reductions in local biodiversity.

User-specific Seasonality. The main estimating equation accounts for state-specific cycles of eBird activity. However, the knowledge needed to identify migratory species may vary seasonally such that winter observations are disproportionally reported by expert users (Johnston et al., 2018). If seasonality differs across individuals, then my estimates may be biased by not accounting for changing distributions of user types across months. I address this concern with user-by-month fixed effects, an especially demanding specification that relies on comparisons across districts and years within a user-month. To control for learning, I include a user-specific linear trend as well as the birding rate, measured as the number of months per year of birdwatching (this is swept away by the user-year fixed effect in equation (1)). This results in a decline of 0.73 species per km^2 of development (column 2), similar to the baseline estimate.

Because of the restrictiveness of user-month fixed effects, I also estimate a middle-ground specification that fits a cubic user-time trend to the baseline specification. The coefficient is nearly equivalent to the baseline estimate (column 3). Overall, user-specific seasonality is unlikely to bias my estimates.

Observer Effort. Despite the saturated baseline model, the specification may still suffer from selection on observer effort. In addition to trip duration and reporting type, I add distance (in km) and number of birdwatchers in the group as controls in column 4. These enter as user-district-month means to be consistent with equation (1). The decline in species diversity remains statistically significant and well within the confidence interval of the baseline specification, suggesting that my main estimates are not biased by unobserved effort.

Outliers and Functional Form. My estimation sample excludes three mega-projects. I transform the project size distribution in two alternate ways to check robustness to other ways of removing outliers. First, I first add back the three outlier projects and then winsorize size distribution at the 99th percentile before aggregating. This has the dual advantage of limiting extreme values without reducing sample size. The coefficient remains negative and significant (column 5), but drops nearly three times in magnitude relative to the baseline estimate. This is because winsorizing eliminates 27 large projects, some of which add useful variation and are not necessarily outliers, leaving the coefficient to be estimated off of relatively smaller projects.

Second, rather than winsorize project size, I instead apply the inverse hyperbolic sine (arcsinh) transformation[16]. Small changes in $archsinh(x)$ approximately reflect proportional changes in x and can be interpreted similar to the standard logarithmic transformation (Bellemare and Wichman, 2020). There are two relevant advantages of arcsinh in my context. First, it is defined at zero, which is a common data value in districts with no forest or no project approvals (see Figure 1B). Second, since it mimics the natural log, arcsinh-transformed variables also reduce the influence of outliers. The coefficient on forest infrastructure again implies statistically significant species diversity declines from development (column 6).

Alternative Diversity Measures. Species richness has been criticized for conveying scant information on abundance or rarity. A location with one pigeon and 99 crows, and another with fifty of each, both have a species richness of two despite the latter being more "even". I compute two of the main diversity measures that account for evenness. The Shannon Index is measured as $SH_j = -\sum_{s=1}^{S} p_{sj} ln(p_{sj})$, where p_{sj} is the proportion of all observations on trip j belonging to species s, and is increasing in diversity. The Simpson Index is measured as $SI_j = 1/\sum_{s=1}^{S} p_{sj}^2$ and reflects the probability that two randomly drawn individuals belong to the same species (Magurran, 2013). For comparability, I use $1 - SI_j$ so that the index increases in diversity. The disadvantage of these measures in my context is that eBird abundance data is notoriously imprecise, given the difficulties of recording accurate counts of quickly moving flocks.

Columns 7 and 8 show that these alternative measures also indicate species loss from development. However, the coefficients are imprecisely estimated as expected. In terms of magnitude, the effects on Shannon and Simpson diversity are 0.7% and 0.6% of their respective means.

Sample Restrictions. I check the sensitivity of my results to the chosen sample composition. First, my estimates may still be biased by non-representative checklists. This bias is likely to be minor since 92% of checklists in my sample are complete (report all observed species). Nevertheless, I estimate the baseline model on a restricted sample of only complete checklists (column 9). The coefficient on species diversity is virtually unchanged from the main result, and precision improves, suggesting that my main results are not driven by non-representative checklists.

Second, I drop districts that never had an approved project during the study period. In the baseline specification, these districts form part of the counterfactual and may bias the results if biodiversity dynamics differ between districts with and without projects. The impact of forest infrastructure remains negative and precision improves (column 10). The point estimate is

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

equivalent to a 3.8% decline, nearly equivalent to the magnitude of baseline result.

Third, I restrict the sample by dropping districts in the first decile of eBird activity, measured as the total number of trips in a district. This allows me to focus on areas with high data collection activity. The resulting coefficient is remarkably similar to the baseline effect (column 11), suggesting that my estimate is not driven by peculiarities in districts with sparse eBird usage.

Clustering. Whereas the unobservable determinants of species diversity are assumed to be arbitrarily dependent within biomes with similar biophysical conditions, the treatment nevertheless varies at the district level. When standard errors are clustered by district, precision slightly declines but remains weakly significant at the 10% level (column 12).

6 Policy Analysis and Discussion

Can slowing the fast-track policy neutralize the development-biodiversity tradeoff? This question is important given the clash between anthropogenic pressures on species diversity on one hand, and the accelerating pace of infrastructure approvals on the other. Following the 2014 fast-track initiative, annual forest area earmarked for approved and pending projects was 60% higher than during the previous 40 years.

In the absence of pre-2014 data to conduct a formal policy evaluation, I instead answer this with a simple simulation. First, I estimate species diversity in differentially fragmented districts. Second, I calculate marginal encroachments under alternative project approval rates and identify the corresponding species responses from the first step. This reveals whether slowing approvals can improve species resiliency. Lastly, I discuss the importance of complementing policy changes with forest regeneration.

6.1 Slowing the Fast-Track Policy

I estimate a binned version of equation (1), which allows the impacts of forest infrastructure to materialize non-linearly. The separate coefficient for each fragmentation bin enables a mapping between hypothetical habitat loss and corresponding species diversity impacts. I estimate:

$$SR_{idt} = \alpha + \sum_{j=1}^{6} \beta_{1j} \mathbb{1}\{Area_{jdt}\} + \sum_{k=1}^{6} \beta_{2k} [Share]_{kdt} + \beta_3 [X]_{idt} + \phi_{iy} + \gamma_d + \theta_{sm} + \epsilon_{idt} \tag{5}$$

where $Area_{jdt} = 1 \; \forall \, Infrastructure_{dt} \in [j, j+1)$. Infrastructure area is binned into groups of length $2km^2$ to balance sample size and curvature: (0-2], (2-4], ..., 8+. Districts with no forest infrastructure are in bin $j = 0$ and omitted. The second summation term is the share of approved projects in category k, which disentangles area effects from the category effect per se. Remaining controls and fixed effects are defined as in equation (1).

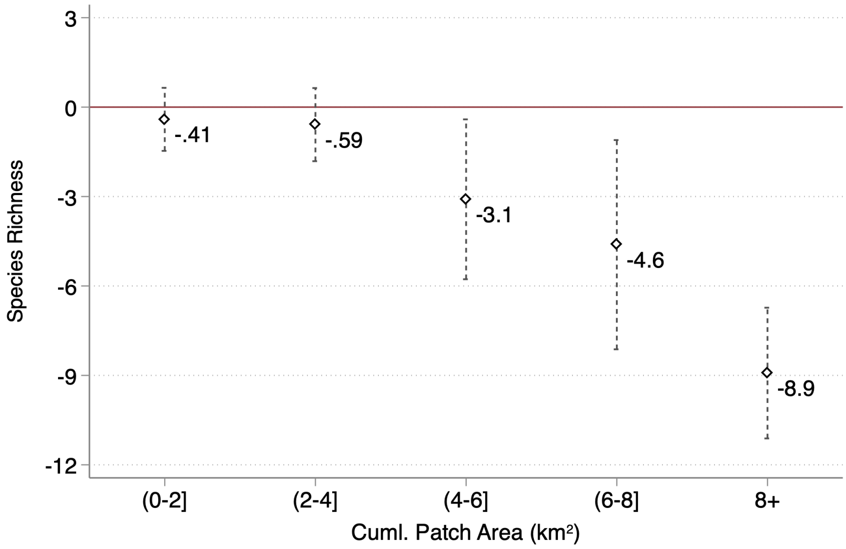

Figure 5: Species Richness Over the Patchiness Distribution

Note: The graph shows species richness over the district fragmentation distribution. Each coefficient describes mean species richness in a patch area bin relative to districts with no infrastructure projects. For example, the coefficient for the (4-6] bin shows average observed species richness by eBird users in districts with 4-6km^2 of encroachment area relative to pristine districts. 95% confidence intervals are shown by the dashed bars. All regressions are weighted by the number of user-trips in a district-time period and include controls for: non-forest area, percent tree cover, access, temperature, rainfall, and reporting type. Standard errors clustered by biome.

Figure 5 shows that species are resilient in relatively intact forests, mirroring the results in Figure 3C. After a 2km^2 threshold, species richness declines monotonically with increasing fragmentation. Districts with 8+ km^2 of projects experience particularly dramatic species declines (61% of mean species richness in the bin) and mostly comprise Madhya Pradesh, home to 12% of the national forest area and 25 wildlife sanctuaries.

To illustrate the effect of slowing the fast-track policy, consider a back-of-the-envelope exercise. At the end of 2018 there were 26,569 pending applications to clearcut 4000km^2 of India's forests. If these were all approved tomorrow in the 135 districts having ever received forest infrastructure, resulting marginal fragmentation would be 30km^2. This is equivalent to siting 1.5 times the world's largest coal mine in each district and is politically unthinkable. If rulings were instead spread over two years, and assuming three decision periods[17], then the per-period marginal fragmentation of 5km^2 (4000/135/6) is perhaps more realistic but triggers significant species loss (see bin (4-6]). Over a three year ruling window, marginal intrusions are 3.3km^2, a level small enough to be politically feasible and at which species are tolerant (bin (2-4]).

[17] A decision period refers to a Forest Approval Committee (FAC) meeting where final-stage proposals are approved. The Forest Conservation Rules (2003) suggest the FAC meet once a month, however, this is not a hard rule (MoEFCC, 2003). Moreover, meetings often result in proposal revisions instead of granting approval. Assuming three decision periods realistically allows flexibility in the FAC process—either meetings with more than one month in between or, regular meetings with approvals granted in a fraction of them.

Overall, slowing the approval process results in less and less per-period fragmentation until species are tolerant. In contrast, fewer species are observed when large forest swathes are cleared at once. This exercise abstracts away from several complexities. First, I assume 100% approval rates. Only 1% of projects were rejected during my study period, making it safe to assume that all pending projects will eventually be approved. Second, the *cumulation* of marginal intrusions may still negatively impact species diversity in the medium run, unless accompanying regeneration efforts are deployed. I explore this further in the next section.

6.2 Implications for Conservation Policy

The UN declared the 2020s the "Decade on Ecosystem Restoration". Impressively, half of the developing world has drafted regeneration plans, making it clear that, conditional on forest loss, biodiversity restoration is possible (FAO and UNEP, 2020). Yet, broader conservation policy governing where and how to grant forest rights in the first place receives less attention.

Optimal forest policy features: 1) large total forest area and 2) small, interspersed pockets of economic activity. These pockets may fuse together as economies grow, making careful deployment of regeneration crucial for muting biodiversity decline. The first feature is supported by the axiomatic species-area theory (Arrhenius, 1921) and the second by the fact that, as more tracts are diverted, surviving forest becomes smaller, more isolated, and closer to human activity. This restricts species recolonization after local extinction (Haddad et al., 2015).

The previous section showed that slowing the fast-track policy yields less and less fragmentation until the development-biodiversity tradeoff is no longer a tradeoff. In the exercise, however, total deforestation is the same whether cleared tomorrow or gradually over three years. In the latter case, despite species resilience to small intrusions, each subsequent encroachment is made into a patchier district. My results in Figure 3C suggest that the gradual cumulation of small intrusions will have a net negative impact on species diversity in the medium-run. Effective regeneration is therefore paramount to offset erstwhile encroachments and produce a connected landscape mosaic preserving original biodiversity.

India's Compensatory Afforestation (CA) policy forms the blueprint for an effective regeneration regime. Firms must pay into a CA fund before receiving stage-II approval and afforestation work begins within one to two years on adjacent forestland[18]. My results are not offset by this since data are monthly. There are at least two limitations of India's CA policy (see Saxena (2019) for more). First, over half of plantations are monocultures with substantially fewer tree species than natural forests. Second, existing CA efforts harbour lower overall biodiversity than assisted natural regeneration (ANR) and native plantations (Osuri et al., 2020). Nevertheless, CA is vital to "compensate" forest loss. As my simulation suggests, species diversity can gradually recover if CA policy is revamped for maximum biodiversity protection (e.g. using ANR) and deployed alongside the optimal landscape patchwork.

[18]If adjacent land is unavailable, it can be found elsewhere in the district. If still unavailable, it can be found elsewhere in the state and, as last resort, in an adjacent state with the approval of the State Secretary (MoEFCC, 2003)

Slow bureaucracies are generally considered costly (Olken and Pande, 2012). Therefore, a policy that slows the rate of approvals appears counterintuitive when viewed under a political economy lens. In the context of environmental reviews, however, slowing the approval rate would be accompanied by more stringent biodiversity protections rather than constitute bureaucratic delay ceteris peribus. These protections could be in the form of: more detailed site inspections, special attention to surrounding fragmentation or project category (as shown in this paper), and involving affected communities in the ruling decision[19].

7 Conclusion

Economic development in the tropics raises concern about harmful impacts that may percolate through our planet's fragile ecosystems. This paper formally connects these phenomena and provides rigorous evidence on the development-biodiversity tradeoff in a biodiverse developing nation. I find that, between 2015-2018, development in India's forests accounted for 60% of its decline in bird diversity, an important proxy of overall biodiversity.

My results are especially relevant as emerging economies prioritize the types of infrastructure studied here. China's Belt and Road Initiative is the textbook example, envisioning enhanced regional connectivity across Asia and Africa. Surprisingly, studies from these regions find limited ecological costs of large-scale development projects (Asher et al., 2020; Garg and Shenoy, 2021; Baehr et al., 2021). In the absence of biodiversity data, however, these studies use tree cover as a broad ecosystem health measure, whereas I leverage several million actual species sightings. This novel data yields robust evidence of anthropogenic species collapse and can be used to inform infrastructure planning as economies expand throughout the 21st century.

Village resettlements are the main driver of species declines. Several thousand of communities await relocation from India's 600 PAs and other areas, primarily due to human-wildlife conflict and forced evictions (Lasgorceix and Kothari, 2009). For example, Kerala's proposed Silver Line railway project is set to divert backwater wetlands and forests and displace enroute communities. My results imply that species loss from construction of such projects can be compounded if displaced villages are resettled by large-scale fragmentation of forests elsewhere. Relocation to non-forested areas or less fragmented districts can achieve a better net outcome.

As humans modify natural landscapes, biodiversity in Earth's surviving forests depends on optimally allocating conservation efforts. My results guide conservation policy in at least two ways. First, protecting relatively fragmented habitats avoids species collapse. Second, the optimal landscape matrix features low fragmentation, high connectivity, and cutting-edge restorative techniques. This can be achieved by slowing the pace of project approvals, resulting in smaller, infrequent intrusions which, when offset by tree-planting, results in species resilience. Rejecting more applications or circumventing forests altogether would yield a similar outcome.

[19] Public hearings with affected communities are technically part of the review process for certain firms. However the consultations suffers from several challenges and in some cases are avoided altogether (Thayyil, 2014).

Whereas I do establish negligible construction before approval, I am unable to test whether deforestation after approval exceeds the requested amount. Thus, my results form an upper bound. They also apply a linearity assumption to the underlying relationship between habitat loss and species diversity. Several studies endorse this, while others advocate for more complexities (Dengler, 2009; Plotkin et al., 2000). Lastly, with a five-year study period I am unable to provide evidence of an Environmental Kuznets Curve for biodiversity (Grossman and Krueger, 1995; Dasgupta et al., 2002), which I leave for future research. Despite these limitations, this study provides powerful insights into the dynamics of biodiversity in human-modified landscapes and is critical for decision-makers tasked with conserving local and global 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.

Aggarwal, Shilpa, "Do rural roads create pathways out of poverty? Evidence from India," *Journal of Development Economics*, 2018, *133*, 375–395.

Arrhenius, Olof, "Species and area," *Journal of Ecology*, 1921, *9* (1), 95–99.

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

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.

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.

Benítez-López, Ana, Rob Alkemade, and Pita A Verweij, "The impacts of roads and other infrastructure on mammal and bird populations: a meta-analysis," *Biological conservation*, 2010, *143* (6), 1307–1316.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How much should we trust differences-in-differences estimates?," *The Quarterly journal of economics*, 2004, *119* (1), 249–275.

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.

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.

Burlig, Fiona and Louis Preonas, "Out of the darkness and into the light? development effects of rural electrification," *Energy Institute at Haas WP*, 2016, *268*, 26.

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.

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.

Dengler, Jürgen, "Which function describes the species–area relationship best? A review and empirical evaluation," *Journal of Biogeography*, 2009, *36* (4), 728–744.

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

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.

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

Gregory, Richard, David Noble, Rob Field, John Marchant, M.J. Raven, and D. Gibbons, "Using birds as indicators of biodiversity," *Ornis Hungarica*, 01 2003, *12*, 11–24.

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

Haddad, Nick M, Lars A Brudvig, Jean Clobert, Kendi F Davies, Andrew Gonzalez, Robert D Holt, Thomas E Lovejoy, Joseph O Sexton, Mike P Austin, Cathy D Collins et al., "Habitat fragmentation and its lasting impact on Earth?s ecosystems," *Science advances*, 2015, *1* (2), e1500052.

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.

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.

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.

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.

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.

Lasgorceix, Antoine and Ashish Kothari, "Displacement and relocation of protected areas: a synthesis and analysis of case studies," *Economic and political weekly*, 2009, pp. 37–47.

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.

Liang, Yuanning, 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.

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.

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.

Murcia, Carolina, "Edge effects in fragmented forests: implications for conservation," *Trends in ecology & evolution*, 1995, *10* (2), 58–62.

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.

Newbold, Tim, Lawrence N Hudson, Andrew P Arnell, Sara Contu, Adriana De Palma, Simon Ferrier, Samantha LL Hill, Andrew J Hoskins, Igor Lysenko, Helen RP Phillips et al., "Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment," *Science*, 2016, *353* (6296), 288–291.

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

Olken, Benjamin A and Rohini Pande, "Corruption in developing countries," *Annual Review of Economics*, 2012, *4* (1), 479–509.

Osuri, Anand M, Abhishek Gopal, TR Shankar Raman, Ruth DeFries, Susan C Cook-Patton, and Shahid Naeem, "Greater stability of carbon capture in species-rich natural forests compared to species-poor plantations," *Environmental Research Letters*, 2020, *15* (3), 034011.

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.

Plotkin, Joshua B, Matthew D Potts, W Yu Douglas, Sarayudh Bunyavejchewin, Richard Condit, Robin Foster, Stephen Hubbell, James LaFrankie, N Manokaran, Lee Hua Seng et al., "Predicting species diversity in tropical forests," *Proceedings of the National Academy of Sciences*, 2000, *97* (20), 10850–10854.

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.

Saxena, KB, "Compensatory Afforestation Fund Act and rules: Deforestation, tribal displacement and an alibi for legalised land grabbing," *Social Change*, 2019, *49* (1), 23–40.

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.

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.

Thayyil, Naveen, "Public participation in environmental clearances in India: Prospects for democratic decision-making," *Journal of the Indian Law Institute*, 2014, pp. 463–492.

Tilman, David, Robert M May, Clarence L Lehman, and Martin A Nowak, "Habitat destruction and the extinction debt," *Nature*, 1994, *371* (6492), 65–66.

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.

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

8 Supplementary Tables and Figures

B1 Tables

Table A1: Trip Protocols in Sample Frame

Protocol	Num. Trips	Pct.
Traveling	445346	61.31
Stationary	222055	30.57
Incidental	47154	6.49
Historical	8719	1.2
Random	1590	0.22
Area	1066	0.15
Nocturnal Flight Call Count	236	0.03
Banding	120	0.02
eBird Pelagic Protocol	111	0.02
CWC Point Count	13	0
International Shorebird Survey (ISS)	3	0
Rusty Blackbird Spring Migration Blitz	6	0
Tricolored Blackbird Winter Survey	1	0

Note: Traveling, stationary, banding, and random trips comprise the analysis sample. The traveling protocol is used if the observer walks or drives during the trip. The stationary protocol is used when the user remains in one place. Many of the remaining protocols describe community initiatives to document specific species. For details of these protocols see manual in eBird Basic Dataset (2019).

Table A2: Summary Statistics of Forest Infrastructure Projects (2015-18)

	Num. Projects	Size (ha.)	SD (ha.)	Total Area (ha.)
Electricity	268	9.79	36.33	2624.60
Irrigation	122	7.64	23.93	932.42
Mining	18	103.23	156.79	1858.06
Other	1144	1.36	17.84	1553.87
Resettlement	9	134.99	108.59	1214.90
Transportation	1209	4.15	18.74	5012.16

Note: Data are arranged at the project level for 2,770 approved projects, prior to aggregating to the district level. Total area (ha.) is the summed deforestation area of all projects in the sample for the particular category.

Table A3: Summary Statistics for eBird (2015-2018)

	Deforestation Districts			Non-deforestation Districts		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
District Variables						
Num. Users	94.50	141.10	180	55.72	110.30	428
Num. Trips	1114.04	3178.13	180	666.14	2414.56	428
Population Density (per km^2)	506.93	743.65	180	1137.18	3684.78	428
Outcome						
Species Richness	21.08	19.45	39285	18.39	16.25	57685
Covariates						
Tree Cover (%)	18.29	13.39	39285	17.36	15.59	57685
Spatial Coverage (%)	13.85	14.59	39285	22.20	24.67	57626
Rainfall (mm)	2.68	4.40	39285	1.75	3.49	57685
Temperature (° C)	25.22	4.89	36875	25.60	4.16	56650
Duration (min)	132.62	141.38	39285	111.36	117.22	57685
% Full Reporting	0.93	0.24	39285	0.95	0.21	57685

Note: Deforestation districts are districts that had a project constructed at any point during the study period. Non-deforestation districts never experienced forest infrastructure. Species richness, duration, and % Full Reporting are means over users' trips in a district-time period. Remaining variables are at the district-month level. Spatial Coverage (%) captures district accessibility for birdwatching and is measured as the proportion of district area traversed by birdwatchers. % Full reporting is the proportion of user trips with checklists reporting all species observed, as opposed to only a subset (i.e. the highlights). Remaining covariate details are explained in section 3.1.

Table A4: Spatial Span of Projects

Districts spanned	Pct.	Cum.
1	96.32	96.32
2	2.71	99.02
3	0.51	99.53
4	0.29	99.82
5	0.04	99.86
6	0.07	99.93
9	0.04	99.96
10	0.04	100.00

Note: This table summarizes the spatial span of individual projects (N=2,770). The data are a count of the number of districts spanned by each project based on the district names listed in the application. Most projects spanning more than one district are roads, railway lines, or transmission lines.

Table A5: Main Results: Impact of Forest Infrastructure on Species Diversity.

	With Learning Bias	Without Learning Bias
	(1)	(2)
Forest Infrastructure (km^2)	-0.347	-0.622**
	(0.307)	(0.250)
Non-Forest Diversion (km^2)	0.075	0.072
	(0.086)	(0.111)
Experience Index	Yes	No
Outcome Mean	19.728	19.730
User FEs	✓	
User x Year FEs		✓
District FEs	✓	✓
State x Month FEs	✓	✓
State x Year FEs		
Year FEs	✓	
N	89581	87202
R^2	0.596	0.653

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. This table shows full regression results for the main analysis (Figure 3A). See Figure 3A notes for details.

Table A6: Spatial Spillovers

	Neighbours	Inverse Distance			
	(1)	(2)	(3)	(4)	(5)
Infrastructure (km^2)	-0.712**	-0.684**	-0.637**	-0.692**	-0.648**
	(0.238)	(0.239)	(0.263)	(0.255)	(0.245)
Other-district Infrastructure	1.488***	5.480**	5.502	6.972*	2.346
	(0.420)	(2.225)	(9.474)	(3.494)	(2.993)
Distance Cutoff	N/A	None	100km	200km	500km
User x Year FEs	✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓
State x Month FEs	✓	✓	✓	✓	✓
N	87202	87202	87202	87202	87202
R^2	0.653	0.653	0.653	0.653	0.653

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome is mean species richness across a user's trips in a district-month. In column 1, *Other-district infrastructure* describes cumulative infrastructure area in districts adjacent to focal district d. It is computed as follows: in each time period, $Infrastructure_{dt}$ is multiplied by a $N \times N$ (where N is the number of districts in India) dimensional weight matrix W, with elements $w_{dj} = 1$ if district j neighbours d and zero otherwise. W is row normalized such that, if there are D neighbours, then each is given an equal $1/D$ weight. In column 2, the elements of W are $w_{dj} = 1/distance_{dj}$ such that *Other-district infrastructure* is an inverse-distance weighted average of infrastructure encroachments in all other districts. Columns 3-5 also use $w_{dj} = 1/distance_{dj}$ but set $w_{dj} = 0$ if the distance between district d and j exceeds the given cutoff. All regressions are weighted by the number of user-trips in a district-time period and include controls for: non-forest area, percent tree cover, access, temperature, rainfall, and reporting type. Standard errors clustered by biome.

Table A7: Robustness Checks

	(1) Stage-I	(2) User-Month	(3) Cubic	(4) Effort	(5) Winsorize	(6) IHS
Infrastructure	-0.620**	-0.700***	-0.630**	-0.588**	-0.216***	-0.718***
(km^2)	(0.253)	(0.196)	(0.245)	(0.231)	(0.059)	(0.211)
Stage-I Approvals	-0.157					
(km^2)	(0.161)					
User x Month FEs		✓				
User x Year FEs	✓		✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓	✓
State x Year FEs		✓				
State x Month FEs	✓		✓	✓	✓	✓
User Trend	None	Linear	Cubic	None	None	None
Additional Controls	No	No	No	Yes	No	No
N	87202	76614	87202	87199	87202	87202
R^2	0.653	0.603	0.653	0.658	0.653	0.653

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. In columns 1-5, the outcome is mean species richness across a user's trips in a district-month. In column 1, stage-I is the coefficient of interest and captures whether premature construction affects species richness. In column 4, additional effort controls are distance travelled (km) and number of members in the birdwatching group. These enter as means over trips in the district-month. In column 5, project size is winsorized at the 99th percentile prior to aggregation. In column 6, infrastructure area is transformed by the inverse hyperbolic sine, and the coefficient can be interpreted similarly to a level-log specification. All regressions are weighted by the number of user-trips in a district-time period and include controls for: non-forest area, percent tree cover, access, temperature, rainfall, and reporting type. Standard errors clustered by biome.

Table A7—Continued: Robustness Checks

	(1) Shannon	(2) Simpson	(3) Complete	(4) Treated	(5) Active	(6) Cluster
Infrastructure	-0.015	-0.005	-0.626**	-0.812***	-0.634**	-0.622*
(km^2)	(0.019)	(0.004)	(0.261)	(0.176)	(0.247)	(0.325)
Outcome Mean	2.001	0.812	19.882	21.263	19.567	19.730
User x Year FEs	✓	✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓	✓
State x Month FEs	✓	✓	✓	✓	✓	✓
Sample Restriction	None	None	Complete	Treated	High-Activity	None
Clustering	Biome	Biome	Biome	Biome	Biome	District
N	87202	85001	83351	32840	78756	87202
R^2	0.618	0.432	0.658	0.693	0.658	0.653

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Columns 1 and 2 replicate the main specification with the Shannon and Simpson diversity indices as the outcome, respectively. In columns 3-6 the outcome is species richness. Column 3 shows results using complete checklists only. Column 4 shows results from a restricted sample dropping districts that never had a project during the study period. Column 5 shows results from a restricted sample dropping districts with sparse data collection activity (< 10th percentile of number of trips ever recorded in the district). All regressions are weighted by the number of user-trips in a district-time period and include controls for: non-forest area, percent tree cover, access, temperature, rainfall, and reporting type.

B2 Figures

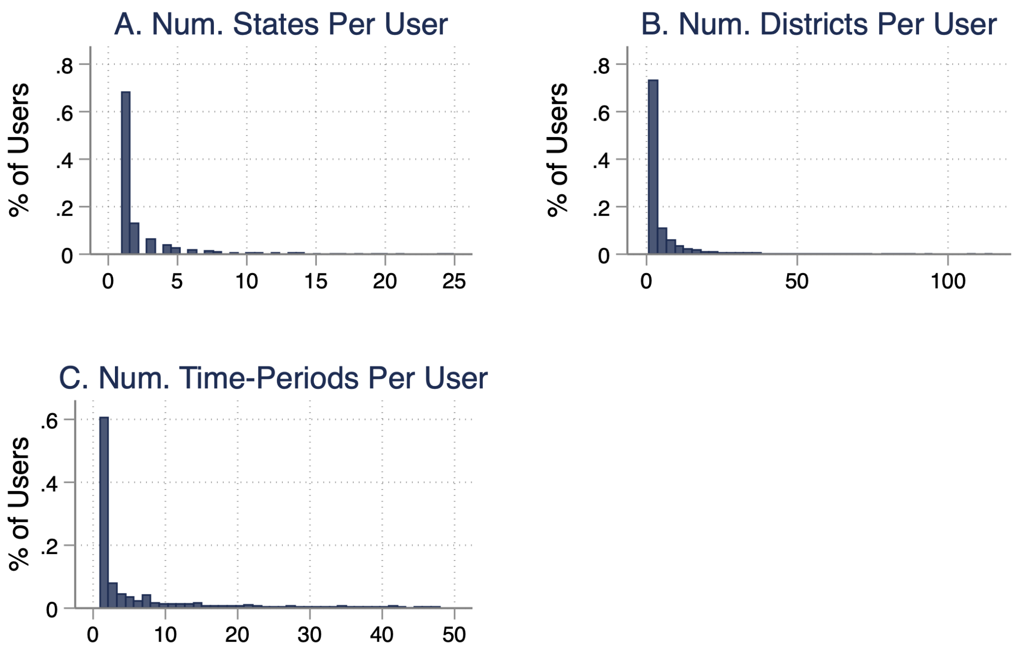

Figure A1: Within-User Distribution of Spatiotemporal Activity

Note: Distributions are based on aggregating eBird data (over all locations and time periods) to the user level (N=10,078 users). Panel A illustrates the distribution of total number of states traversed per user across all their trips during the study period. Panel B shows the same for total number of districts traversed per user. In panel C, a time period is a year-month. There are 48 possible values (12 months*4 years). The distribution describes the number of periods in which a user recorded a trip.

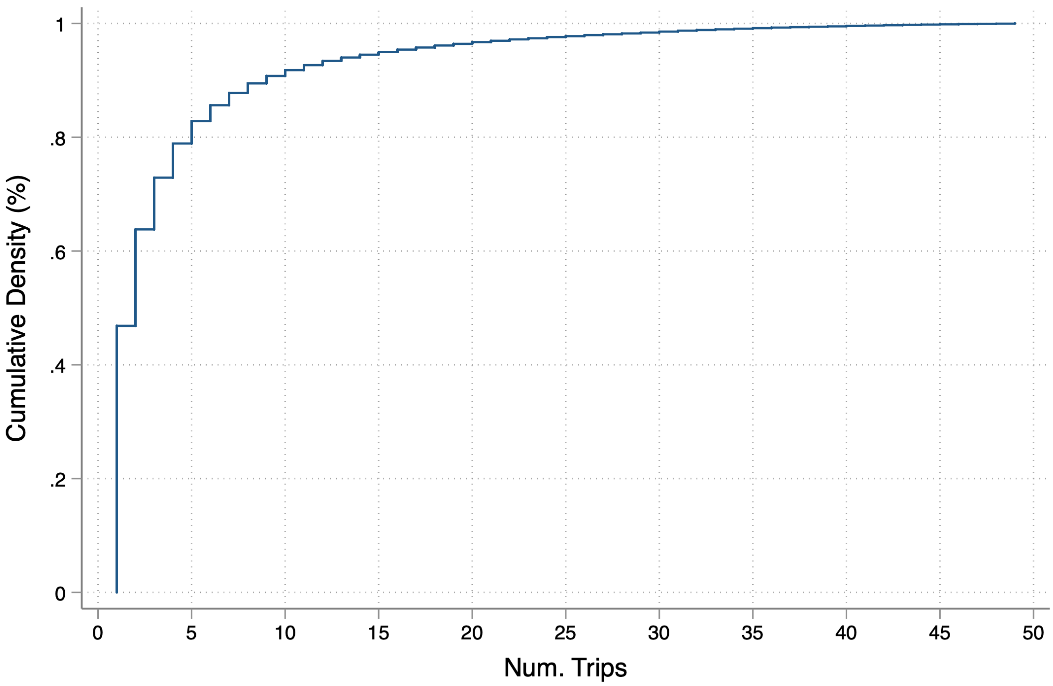

Figure A2: Cumulative Distribution of Weighting Variable

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.

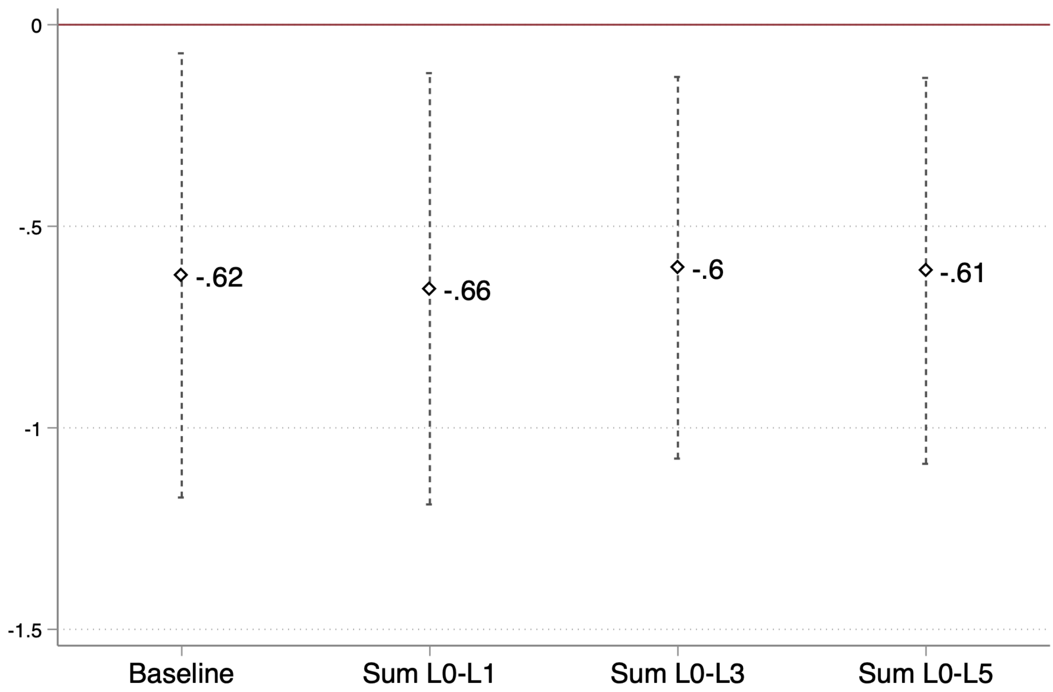

Figure A3: Cumulative Dynamic Lag Results

Note: Results from the cumulative dynamic lag specification. Each coefficient is from a separate regression. "Baseline" repeats the main result. "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" and "Sum L0-L5" sum up to the third and fifth lag, respectively. Dashed lines are 95% confidence intervals. All regressions weighted by number of trips, include user-year, district, and state-month fixed effects, and controls for non-forest area, percent tree cover, access, temperature, rainfall, and reporting type. Standard errors clustered by biome.

B1 Data Appendix

B1 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, approach access, railway	Transportation
canal, irrigation, drinking water	Irrigation
forest village conversion, encroachments, rehabilitation	Resettlement
mining, quarrying	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.

Project Categorization. I web scraped all forest diversion applications (n=19,495 projects) submitted between 2015-18 at all stages of the approval process from the online portal. The digital application has few validation checks for text fields. For example, there are 24 project categories, many of which refer to the same type of project. To simplify the analysis, I re-categorize projects according to Table B1[20].

Because of inconsistencies in raw data entry by firms, I manually examine and reassign projects mistakenly categorized as "other". "Other" project descriptions including the word "pipeline", "water", and "irrigation" together (n=35) are placed in the irrigation category. Projects with the word "approach" (n=413) are placed in the transport category, and those with the word "relocation" in the resettlement category (n=1).

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 97.8% of districts in the application sample. The remaining 2% (10 districts) represent districts with more complex re-

[20] There are 42 projects categorized as "industry" which together make up just 34 ha. This accounts for $< 0.25\%$ of total area cleared during the study period. For this reason, I include industrial projects in the ?other? category.

Table B2: Summary Statistics of Project Size (Non-Truncated Sample)

	N	Mean (ha.)	Std. Dev.	Min.	Max
electricity	269	25.87	266.13	0.001	4334.010
irrigation	124	45.42	314.96	0.002	3168.132
mining	18	103.23	156.79	0.173	517.888
other	1144	1.36	17.84	0.001	564.000
resettlement	9	134.99	108.59	0.201	285.700
transportation	1209	4.15	18.74	0.000	500.600
Total	2773	8.02	109.31	0.000	4334.010

Note: Data are arranged at the project level, prior to aggregating to the district-month panel. Data include three outlier mega-projects that are dropped in the main analysis. The three projects include huge irrigation and electricity projects that skew the distribution (see standard deviation versus mean for these categories.

drawing procedures and are dropped.

Outlier Projects. There are 2,773 projects with stage-II approval. Table B2 shows summary statistics for these projects. The project size distribution is very skewed (mean=8 ha., sd=109.31 ha.) and appears to be driven by the electricity and irrigation category (see mean and sd for these categories). Manual review reveals three outlier "mega-projects" in Telangana state. Two involve the laying of canals, tunnels, and power lines for lifting river water for irrigation. The third is what the Government of India refers to as an Ultra Mega Power Project ($>$ 4000 megawatts). I remove these three projects in the main sample to reduce skewedness and the influence of outliers (see Table A2 for summary stats). As a sensitivity check (see section 5.5), I replicate the analysis with an alternative method of truncating the distribution as well as with the sample including the three mega-projects.

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 and non-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.