Online Appendix

Infrastructure, Institutions, and the Conservation of Biodiversity in India

Raahil Madhok, University of Minnesota

Contents

S1 Appendix Tables

Table S1: eBird Summary Statistics (2015-2020)

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

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

Table S2: Correlation Between Infrastructure Permits and Nightlights

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

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

Table S3: Variation in Species Richness Under Various Fixed Effects

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

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

Table S4: Impact of Forest Infrastructure on Species Diversity

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

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

Table S5: Robustness—Spatial Spillovers

	(1)	(2)	(3)	(4)	(5)
Infrastructure (Standard Deviations)	-0.391**	-0.396**	-0.402**	-0.408**	-0.410**
	(0.176)	(0.173)	(0.178)	(0.177)	(0.182)
Infrastructure (district j $\neq$ d) (Standard Deviations)	-0.137	0.025	0.081	0.399	0.441
	(0.255)	(0.473)	(0.258)	(0.647)	(0.421)
Distance Cutoff	Neighbors	100km	200km	500km	None
User $\times$ Year FEs	✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓
State $\times$ Month FEs	✓	✓	✓	✓	✓
Observations	161896	161896	161896	161896	161896
R^2	0.690	0.690	0.690	0.690	0.690

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome is mean species richness across user's trips in a district-month. All infrastructure variables are standardized for comparability. In all columns, Infrastructure (row 1) is cumulative area of forest occupied by infrastructure in district d during a year-month. In column 1, infrastructure in other districts j refers to those adjacent to district d. In column 2, it refers to cumulative encroachment area in other districts within 100km of the focal district d. In each time period, $Infrastructure_{dsym}$ is multiplied by a $N \times N$ (where N is the number of districts in India) dimensional weight matrix W with elements $w_{dj} = 1/distance_{dj}$ for districts j within 100km of d and zero otherwise. Columns 3 and 4 extend the distance cutoff to 200km and 500km, respectively. Column 5 applies the inverse distance weight to all districts. Section 5.3.2 elaborates the procedure. All regressions control for: temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log experience, log group size, and log spatial coverage. Standard errors clustered by biome.

Table S6: Tests of Endogenous Sorting

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

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

Table S7: Impact of Forest Infrastructure on Species Diversity by Category

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

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

Table S8: Robustness—Other

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

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

Table S9: Robustness: Close Election Design

	(1)	(2)	(3)	(4)	(5)	(6)
Infrastructure (km^2)	-1.850**	-2.695**	-1.308	-2.303**	-2.044**	-2.232*
	(0.713)	(1.171)	(0.779)	(1.005)	(0.801)	(1.144)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
First Stage F-Statistic	8.842	9.043	8.981	6.414	8.442	6.579
Bandwidth	2	3	5	2	2	2
Polynomial Order	1	1	1	2	3	1
User FEs	✓	✓	✓	✓	✓	
User x Year FEs						✓
District FEs	✓	✓	✓	✓	✓	✓
State $\times$ Month FEs	✓	✓	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓	✓	
Observations	134448	134448	134448	134448	134448	129704

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the user-district-month level as in the main specifications. The outcome is mean species richness across user's trips in a district-month. Coefficients are 2SLS estimates as specified in Equation 8. Infrastructure is instrumented with the fraction of constituencies in a district where the incumbent won in a close race during the most recent state election. All regressions control for the same user- and district-level covariates as Equation 2 as well as the fraction of close-election district constituencies where the incumbent ran, election year, and the interaction of victory margin with an indicator for whether any incumbent ran in the district. Column 1 defines close election as a win margin of 2 percent. Columns 2-3 expand the win margin to 3 and 5 percent, respectively. Columns 4 and 5 uses a second- and third -rder polynomial in the win margin. Column 6 uses user-by-year fixed effects.

Table S10: Robustness: Treatment Heterogeneity by Institutional Type

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

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

Table S11: Mechanisms: Institutions and Sustainable Infrastructure

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

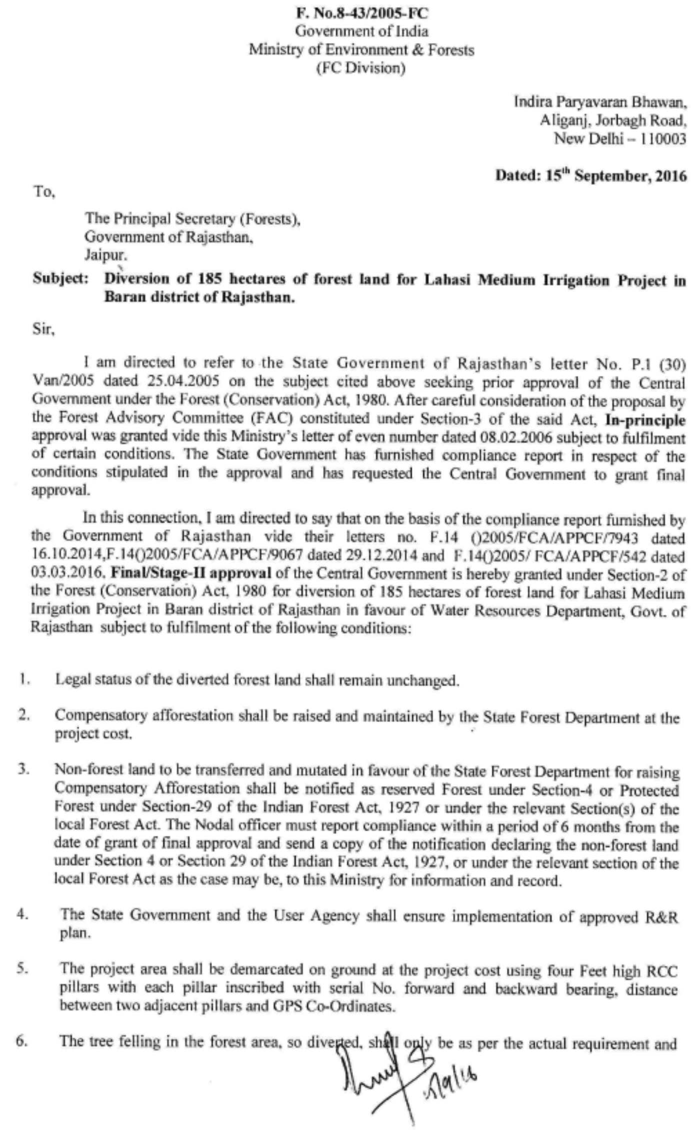
Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the project level for the digital subsample. Inclusive (=1) means the district has historically inclusive institutions. Sample restricted to 163 districts in Banerjee and Iyer (2005) under British Rule. Informed consent indicates whether the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

Table S12: Balance Table: Project Category Distribution by Institutional Type

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

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

S2 Appendix Figures



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

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

Dated: 15th September, 2016

To,

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

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

Sir,

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

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

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

Figure S1: Example Approval Letter

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

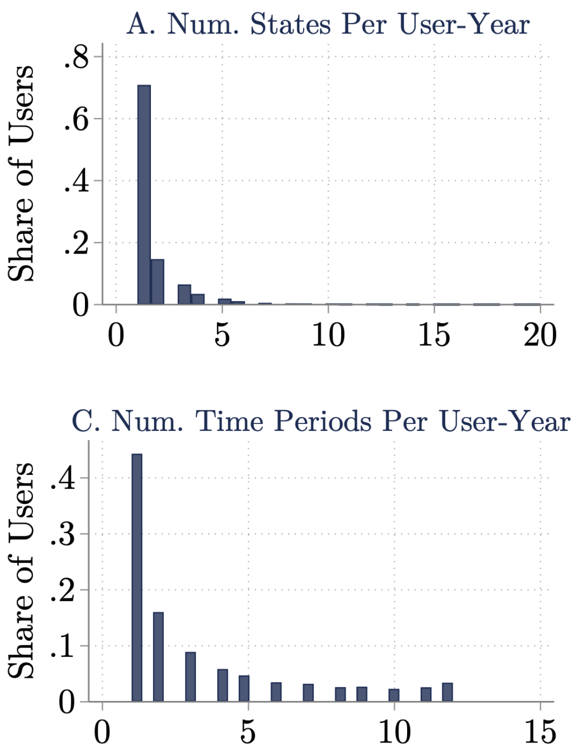


Figure S2: Within-User Distribution of Spatiotemporal Activity

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

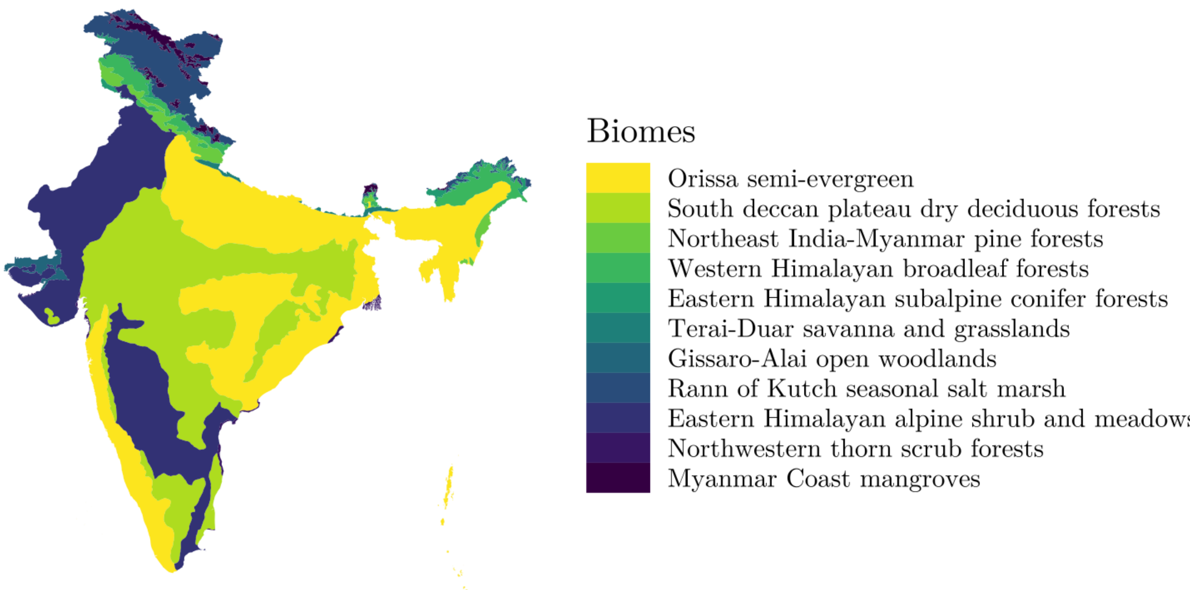


Figure S3: Biomes of India

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

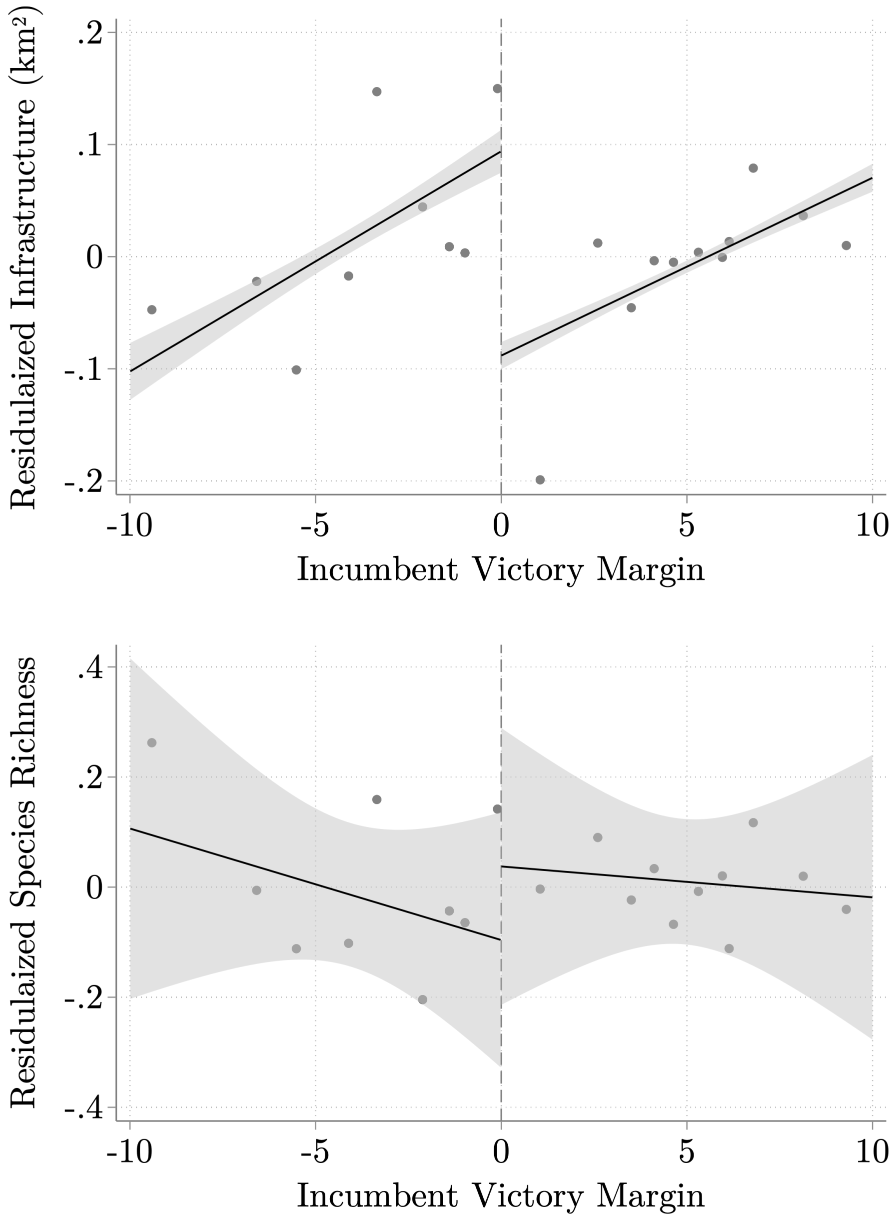


Figure S4: Margin of Victory, Infrastructure Project Approvals, and Species Richness

Note: Panel A (top) shows the first stage (Equation 7) and Panel B (bottom) shows the reduced form. Each figure plots a binscatter of mean incumbent win margins across close elections (x-axis) in the district against the outcome (y-axis) residualized on user, district, state-month and year fixed effects as well as the same set of covariates as the main TWFE specification. Points to the left of zero denote districts where incumbents lost in close elections. A linear fit is generated separately for each side of zero, with 95% confidence intervals displayed.

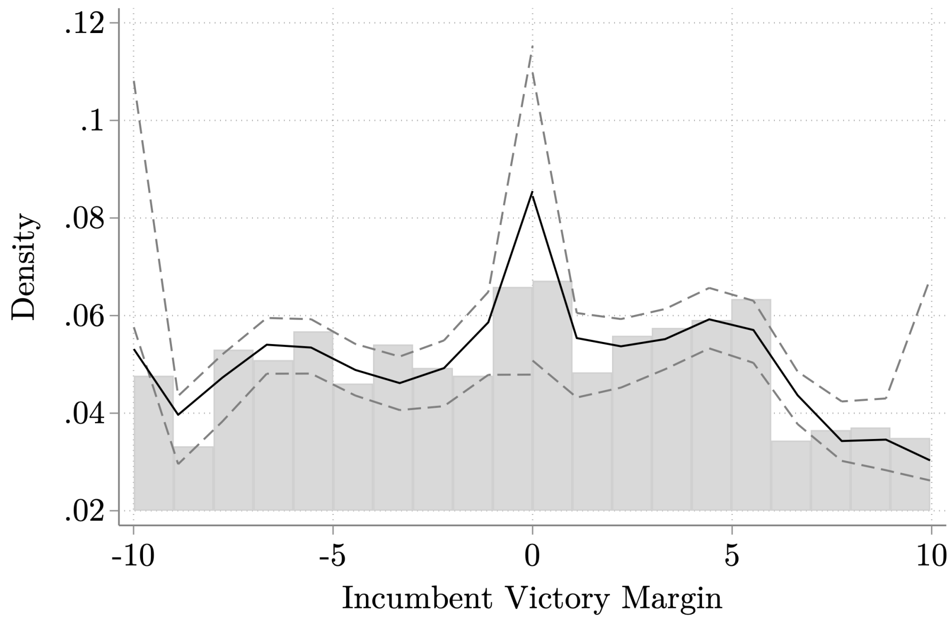


Figure S5: Density Discontinuity Test for Manipulation

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

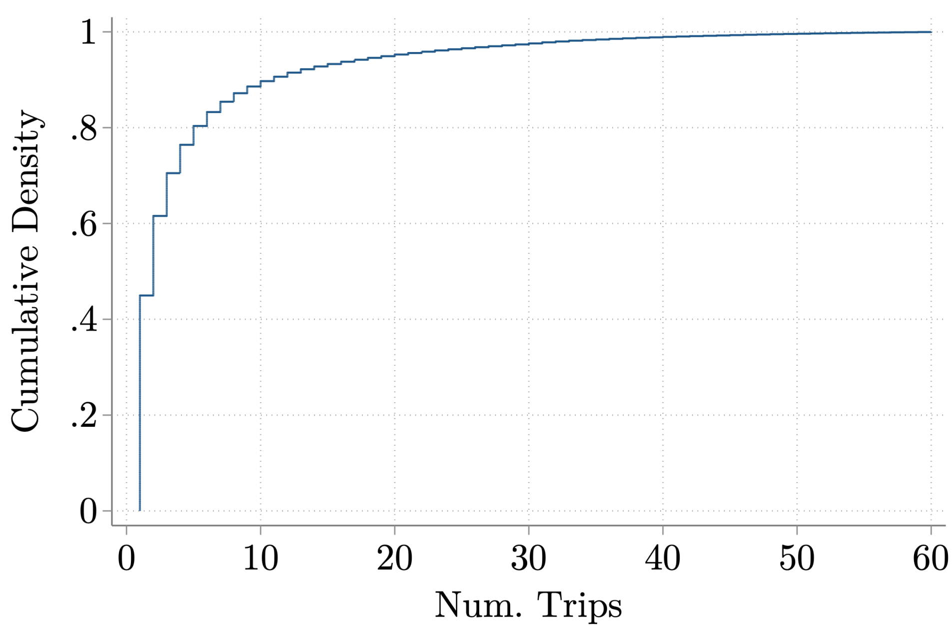
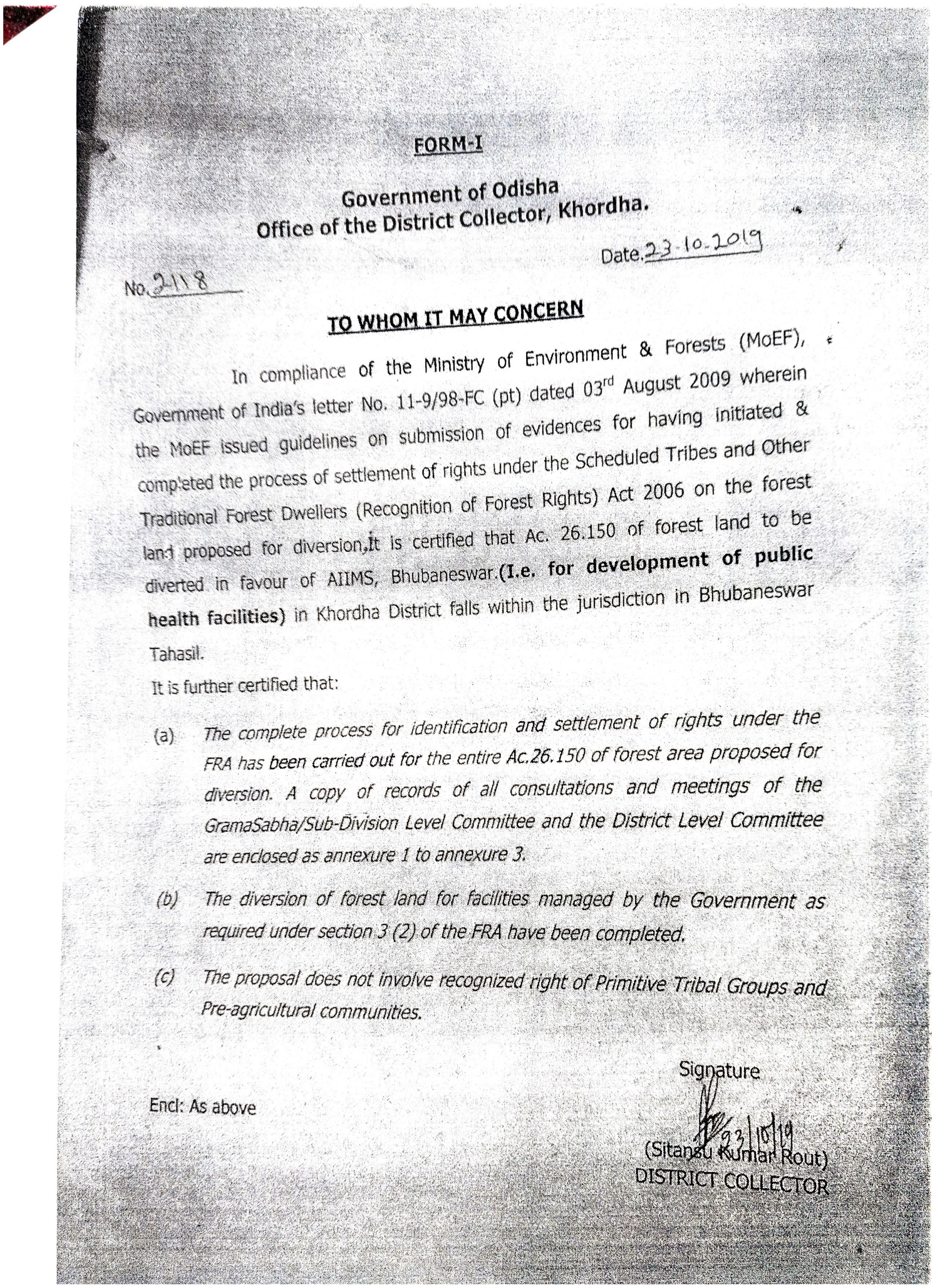


Figure S6: Cumulative Distribution of Weighting Variable

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



FORM-I

Government of Odisha
Office of the District Collector, Khordha.

No. 2118

Date. 23.10.2019

TO WHOM IT MAY CONCERN

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

It is further certified that:

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

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

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

Encl: As above

Signature
23/10/19
(Sitansu Kumar Rout)
DISTRICT COLLECTOR

Figure S7: Example Letter of Informed Consent

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

S3 Infrastructure Data

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

S3.1 Forest Clearance Process

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

To receive stage-II approval, the firm pays into a tree-planting fund and is checked for compliance with the Forest Rights Act. The infrastructure data sample consists of projects with stage-II approval.

S3.2 Infrastructure Sample Construction

Table S13: Procedure for Project Categorization

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

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

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

The manual subsample was digitized as follows. First, a PDF of each approval letter was downloaded. The district of each project was extracted from the subject header (see Figure S1) and cross-checked online. If only the village was given, the district was identified on Google. Second, project size (hectares approved for forest diversion) was also extracted from the letter. For projects that span multiple districts (e.g. roads), a separate document called "Form A" (also available in the portal) was downloaded to identify hectares per district. 26 multi-district projects did not specify a district-wise breakdown, in which case total project size was divided equally across districts. Lastly, project category was extracted from the letter. In some cases it was taken from the Form A document which includes a detailed project description[1] .

Project Categorization Verbatim project categories often refer to the same type of project. To simplify the analysis, I re-categorize projects according to Table S13[2]. Because of inconsistencies in raw data entry by firms, I manually examine and reassign projects mistakenly categorized as "other". "Other" project descriptions with the word "power", "substation", and "kv" are placed in the Electricity category. "Other" projects with the word "resettle", "relocate", and "pattayam"[3] are placed in the Resettlement category.

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

[1]The category of each project in the manual subsample was available digitally, and scraped, but the majority were listed as "Other". I manually categorized them based on the subject header (see Figure S1)

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

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

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

S3.3 Summary Statistics by Project Ownership and Shape

Table S14: Summary Statistics of Projects by Ownership and Shape

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

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

Table S15: Percent of Projects in Each Category by Ownership and Shape

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

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

Table S14 shows summary statistics by ownership and shape. Over 70% of projects are publicly owned (Panel A). These are about five times larger than non-public projects. Grouped by shape, only 10% of projects are nonlinear, but these are five times larger than linear projects (Panel B). Table S15 shows the project distribution by ownership, shape, and category. Public projects are largest because resettlement and irrigation—the second and third largest category—are almost all publicly owned (column 1). Mining and “other” feature a more even public-private split than any other category. Nonlinear projects are largest because nearly all mining and resettlement projects are non-linear (column 5).

S4 Additional Results

This appendix section provides additional results.

S4.1 Timing of Permitting Decisions

This Appendix provides additional evidence for the quasi-randomness of permitting decisions. I conduct a test based on Deshpande and Li (2019) that predicts the likelihood and timing of project approval based on observable characteristics. I find that neither bureaucratic nor geographic characteristics consistently predict approval or the timing of approval, helping make the case that the timing of permits is plausibly random.

First, I predict whether observable characteristics predict the likelihood of project approval in each year between 2016-2019 with the following equation:

$$Approved_i = \alpha + \beta_1 ProcessingTime_i + \beta_2 Apps_i + \Gamma X'_i + \theta_s + \epsilon_i \tag{1}$$

where $Approved_i$ is a dummy indicating whether project i was approved in a given year. $ProcessingTime_i$ is the mean processing time (number of days between proposal submission and approval) in the district of project i in the previous year. $Apps_i$ is the number of projects approved in i's district in the previous year. X'_i is a vector of geographic characteristics in the district of project i including: distance to coast, slope, elevation, water area within 50km, and area of nearest coal deposit[4]. θ_s is a state fixed effect which account for state-specific factors influencing project approval.

Next, I investigate whether the same local characteristics predict the timing of project approvals between 2016-2019 for projects under review in that year and which will be approved in the future. I estimate the following equation:

$$ApprovalYear_i = \alpha + \beta_1 ProcessingTime_i + \beta_2 Apps_i + \Gamma X'_i + \theta_s + \epsilon_i \tag{2}$$

where $ApprovalYear_i$ is the year in which project i was approved.

Table S16 presents the estimates. Columns 1-4 present estimates of Equation 1 of how local characteristics affects the likelihood of project approval for projects under consideration in the column year. Columns 5-8 present estimates of equation 2 of how these characteristics predict the timing of project approval conditional on the project being eventually approved. The sample consists of projects under consideration in the column year and will be approved by 2020. Overall, there are no observable characteristics that consistently

[4]Coal deposits are obtained from the USGS (Trippi and Tewalt, 2011), Water body shapefiles are from Natural Earth Data, and gridded elevation is from the NOAA.

predict project approval in each year. There are also no characteristics that consistently predicts the timing of approval. These ancillary results help build confidence that the timing of permit awards is effectively random.

Table S16: Determinants of Permit Award Timing

	Project Approved				Timing of Approval			
	(1) 2016	(2) 2017	(3) 2018	(4) 2019	(5) 2016	(6) 2017	(7) 2018	(8) 2019
Processing Time (previous year)	-0.0015	-0.0018**	0.0008	-0.0012	0.0125*	0.0032	-0.0084**	0.0038
	(0.0009)	(0.0009)	(0.0012)	(0.0015)	(0.0072)	(0.0059)	(0.0041)	(0.0032)
Applications (previous year)	0.1904	-0.0848	0.1292	0.0695	-0.7973	-0.6073	-0.9736**	-0.6083***
	(0.1257)	(0.1034)	(0.0899)	(0.1643)	(0.6632)	(0.5299)	(0.3765)	(0.2316)
Distance to Coast (km)	-0.0115	-0.0133*	0.0096	0.0166**	0.0810*	0.0182	-0.0218	-0.0109
	(0.0073)	(0.0070)	(0.0081)	(0.0075)	(0.0458)	(0.0347)	(0.0252)	(0.0145)
Slope (degrees)	-1.2681	1.6114*	-0.0725	-0.8172	-2.0044	-7.6311**	-1.8914	-0.2557
	(0.8519)	(0.8329)	(1.1301)	(0.7877)	(2.9915)	(3.6630)	(3.3840)	(1.6727)
Elevation (m)	0.0061	-0.0066	-0.0002	0.0049	0.0023	0.0287	0.0037	-0.0069
	(0.0044)	(0.0044)	(0.0070)	(0.0044)	(0.0165)	(0.0206)	(0.0194)	(0.0087)
Water Area (km2)	-0.0000	0.0000	0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Coal Area (km2)	-0.0006	0.0021*	-0.0011	0.0004	0.0005	-0.0043	0.0006	-0.0004
	(0.0012)	(0.0012)	(0.0009)	(0.0014)	(0.0047)	(0.0036)	(0.0026)	(0.0029)
State FEs	✓	✓	✓	✓	✓	✓	✓	✓
Observations	4929	5684	5951	6222	4156	4054	3392	2537
R^2	0.029	0.025	0.019	0.036	0.091	0.078	0.082	0.101

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Data are a balanced project year panel of projects under review in the year indicated in the column heading and will be approved by 2020. Columns 1-4 presents estimates from Equation 1 of how observable characteristics predict the likelihood of project approval for projects under consideration in the year indicated in the column heading. Columns 5-8 present estimates from Equation 2 of how observable characteristics predict the timing of project approval, conditional on approval. *ProcessingTime* is the mean number of days between proposal submission and approval in the district of project i. *Applications* is the number of proposals approved in the district of project i. Geography variables are at the district-level. Standard errors clustered at the district level.

S4.2 Ruling Out Reverse Causality

To rule out reverse causality and improve causal interpretation of the main estimates, I test whether lagged bird diversity predicts project approval. To do this, I first partial out the effort covariates from individual-level species richness data and aggregate residuals to the district level. This represents "natural" variation in local species diversity, separate from variation based on user characteristics, and aligns with what district authorities might consider in project approval decisions. Second, I estimate the following 12-month lagged equation:

$$Infrastructure_{dst} = \gamma_d + \theta_{sy} + \delta_m + SR_{dst} + \sum_{k=1}^{12} \beta_k \cdot SR_{ds(t-k)} + \epsilon_{dst} \tag{3}$$

where $Infrastructure_{dst}$ is cumulative infrastructure area approved in district d of state s in year-month t. SR_{dst} is species richness in district d, residualized on individual bird-watching effort. γ_d are district fixed effects, which account for time-invariant local determinants of project placement. θ_{sy} are state-year fixed effects, which account for regional demand shocks. δ_m are month fixed effects, which account for seasonality. $\beta_k^{\prime}s$ capture the impact of species diversity k months ago on present-day infrastructure approvals.

Table S17 presents the estimates. Project placement does not appear to be influenced by local bird diversity. Lagged coefficients are near-zero and statistically insignificant across three- (column 1), six- (column 2), and 12-month (column 3) windows. These findings improve confidence that reverse causality is not driving my main results, further supporting the assumption of quasi-random project approvals.

Table S17: Impact of Lagged Species Diversity on Infrastructure Placement

outcome: infrastructure (km^2)	(1)	(2)	(3)
Species Richness (Lag 0)	-0.002	-0.007	0.023
	(0.027)	(0.033)	(0.033)
Species Richness (Lag 1)	0.005	0.005	0.031
	(0.027)	(0.032)	(0.035)
Species Richness (Lag 2)	-0.007	-0.005	-0.006
	(0.008)	(0.009)	(0.008)
Species Richness (Lag 3)	-0.008	-0.011	-0.018
	(0.009)	(0.010)	(0.013)
Species Richness (Lag 4)		-0.010	-0.008
		(0.009)	(0.014)
Species Richness (Lag 5)		-0.006	-0.020
		(0.007)	(0.013)
Species Richness (Lag 6)		-0.007	-0.021
		(0.009)	(0.012)
Species Richness (Lag 7)			-0.012
			(0.012)
Species Richness (Lag 8)			-0.014
			(0.012)
Species Richness (Lag 9)			-0.006
			(0.009)
Species Richness (Lag 10)			-0.013
			(0.010)
Species Richness (Lag 11)			-0.010
			(0.009)
Species Richness (Lag 12)			-0.005
			(0.012)
District FEs	✓	✓	✓
State × Year FEs	✓	✓	✓
Month FEs	✓	✓	✓
Observations	19924	18247	15250
R^2	0.903	0.916	0.935

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Data are at the district-yearmonth level. The outcome in all specifications is cumulative area of infrastructure encroachment. Explanatory variables are monthly lags of species richness. Species richness is the mean over eBird users' birdwatching trips in a district, residualized on effort covariates. Standard errors clustered by biome.

S4.3 Impacts by Project Ownership and Shape

Table S18 groups estimates by project ownership and shape. Public projects are the main threat to biodiversity (Panel A), mainly because resettlement, transport, and irrigation projects—the three most harmful categories—are almost all publicly owned (Table S15). In contrast, the effect of private projects is noisy, likely because "other" projects are largely privately owned (Figure 4B of main text). Similar logic helps interpret the estimates by project shape (Panel B). Nonlinear projects have a robust negative impact on species diversity. The magnitude is a combination of the small mining and large resettlement coefficients in Figure 4B of the main text, which are almost all nonlinear (Table S15). The tradeoff for linear projects is weaker but the magnitude is twice as large. The large magnitude is a combination of the transportation, electricity, and irrigation coefficients, which are mostly linear (Table S15). Wide standard errors likely come from "other" projects, which are also mostly linear.

Table S18: Impact of Infrastructure on Species Diversity by Project Ownership and Shape

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

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

S4.4 Impacts by Species Type

To estimate impacts by species type, I manually code each species in the eBird dataset (N=1,640 species) with their migratory status, habitat specialization, endemicity, and range size. These characteristics are obtained via the State of India's Birds database, which contains a searchable repository of Indian birds along with their detailed physiology. Migration categories include: non-migrant, long migrant (winter or summer migrant), and short migrant (within India migrant). Habitat specialization categories include: forest, grassland, wetland, and generalist (observed in all habitats). Endemicity is either endemic or non-endemic.

In the same way that species abundance by IUCN status is calculated (Section 3.2 of main text), I count the number of times that each user recorded species in each category during a district and year-month. I then estimate Equation 2 in the main text with these user-level abundance measures as outcomes. I report Poisson estimates since the outcome is a count variable. Table S19 presents the estimates. Results are discussed in Section 6.3 of the main text.

Table S19: Estimates by Habitat Type and Endemicity

	Habitat Specialization				Endemicity		Migration Type		
	(1) Forest	(2) Wetland	(3) Grass	(4) General	(5) Endem.	(6) Non-endem.	(7) Non-Mig.	(8) Short	(9) Long
Infrastructure (km^2)	-0.015***	-0.007	-0.005**	-0.010***	-0.013***	-0.012***	-0.012***	-0.009***	-0.006***
	(0.003)	(0.006)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
User x Year FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
State × Month FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	158662	159502	158883	161723	160323	161871	161769	160972	159507

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

S4.5 Impacts by Habitat Conservation Importance

To measure how important a districts' habitat is for bird conservation, I use the Important Bird Area (IBA) designation created by BirdLife International. IBAs are spatial units of conservation priority and delineated based on a standardized, globally agreed upon set of criteria determining species vulnerability. Regional IBAs are also created on a country-by-country basis taking into account local threats and thresholds. Importantly, IBAs are delineated such that they represent a policy-relevant and manageable unit, such as a protected area (Donald et al., 2019).

I obtained global IBA shapefiles by contacting BirdLife International directly. I then compute two key measures of bird conservation importance at the district level to study treatment heterogeneity. First, I measure the number of pre-period IBAs in a district. Second, I compute the proportion of district area covered by IBAs created during the pre-period. For both of these variables, I use an above-median indicator as a measure of whether district habitat is of high conservation importance for bird populations or not.

Lastly, I estimate Equation 4 in the main text. Results are in Table S20 and discussed in Section 6.4 of the main text.

Table S20: Treatment Effects by Conservation Importance

	(1)	(2)
Infrastructure (km^2)	-0.232*** (0.056)	-0.125*** (0.034)
Infrastructure (km^2) $\times$ 1[Num. IBAs $>$ Median]	0.178** (0.069)	
Infrastructure (km^2) $\times$ 1[IBA Coverage $>$ Median]		0.068 (0.094)
Infrastructure $\times$ Category Shares	Yes	Yes
Controls	Yes	Yes
User x Year FEs	✓	✓
District FEs	✓	✓
State $\times$ Month FEs	✓	✓
Observations	161896	161896
R^2	0.690	0.690

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcome is mean species richness across users' trips in a district-month. *Infrastructure* is cumulative area of infrastructure in a district-month. Num. IBAs is the number of pre-period IBAs in a district. IBA Coverage is the proportion of district area designated as an IBA. All specifications include user-year, district, and state-month fixed effects as well as controls for temperature, rainfall, traveling trips, log nightlights, log duration, log distance, log group size, and log spatial coverage. Columns 2 and 4 additionally include six interactions terms of infrastructure with the baseline district share of projects in each category. Standard errors clustered by biome.

S4.6 Robustness: Alternative Fixed Effects

Table S21 tests robustness to alternative fixed effects. Columns 1-4 specify alternative forms of seasonality. Column 1 includes user-by-month fixed effects, which relies on comparisons across districts and years within a user-month. This accounts for seasonality exhibiting an individual component i.e., winter migratory species mainly reported by experts (Johnston et al., 2018). Columns 2 and 3 include district-month fixed effects, which accounts for sub-state seasonality[5]. Column 4 uses biome-by-month fixed effects in case seasonality is biome- rather than state-specific. All four estimates are strikingly similar to the baseline estimate, suggesting that biases from individual and regional seasonality are negligible.

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

Column 7 tests robustness to sorting. Recall that user fixed effects account for user heterogeneity, but requires a no-sorting assumption for identification (Section 6.1 of main text). The reverse is to compare species diversity in a fixed location, which obviates this assumption, but then pools checklists by a changing group of users active in a grid cell. I implement this by disaggregating eBird data down to the user-cell-month level and including $10km^2$ cell fixed effect (and no user fixed effect) for identification. To account for user heterogeneity, I add fixed effects for user experience, including cumulative number of trips and months per year of activity. The coefficient remains stable but is less precise. The fact that coefficients are similar, yet the preferred specification includes user fixed effects and shows minimal sorting (Section 6.1 main text), supports the chosen design.

[5]Column 2 of Table S21 uses user-year and district-month fixed effects. This is likely too saturated to yield precise estimates. Column 3 includes user and district-month fixed effects as a compromise.

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

Table S21: Robustness—Alternative Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Infrastructure (km^2)	-0.125**	-0.114	-0.080**	-0.108*	-0.117*	-0.119**	-0.105
	(0.041)	(0.064)	(0.032)	(0.051)	(0.054)	(0.054)	(0.121)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
User x Year FEs		✓		✓	✓	✓	
User × Month FEs	✓						
District FEs	✓			✓	✓	✓	
District × Month FEs		✓	✓				
State × Year FEs	✓		✓				
State × Month FEs					✓	✓	✓
Biome × Month				✓			
Experience FEs					✓		✓
Cell FEs							✓
Year FEs							✓
Time-of-Day FEs						✓	
Observations	143394	161087	166446	161909	161563	161665	282544
R^2	0.702	0.706	0.654	0.688	0.694	0.690	0.581

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

S4.7 Robustness: Alternative Standard Errors

Table S22 shows the baseline estimates adjusted for alternative clustering. Column 1 replicates the baseline, whereby unobservable biophysical determinants of species diversity are assumed to be correlated within biomes, even though treatment varies by district. Columns 2 and 3 show that estimates are quite similar when clustering by district or state, respectively. Clustering by state is a compromise between large biome clusters (Figure S3) and smaller districts. Columns 4-7 investigate spatial correlation more systematically by implementing Conley (1999) standard errors for several choices of the kernel cutoff distance. Reassuringly, precision remains similar, even when allowing for long distance spatial correlation up to 1000km.

Table S22: Robustness—Alternative Standard Errors

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

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

S4.8 Robustness: Category-Wise Robustness Tests

Table S23: Category-Wise Robustness—Alternative Fixed Effects

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

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

Table S24: Category-Wise Robustness—Other

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

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

S4.9 Banerjee and Iyer (2005) Extension

To replicate and extend Banerjee and Iyer (2005), I match district-level data on conflict events and criminal politicians with the raiyatwari/zamindari (inclusive/extractive) district classification provided in the Banerjee and Iyer (2005) replication files. Conflict data are from ACLED (Raleigh et al., 2010), a global georeferenced repository of conflict events. Each event contains a date, GIS coordinates, and a detailed description. To identify environment-related conflicts, I keyword search events with the word ”environment” or ”forests”. To identify conflicts involving minorities, I search for events with the word ”tribe”, ”caste”, or ”scheduled”. Lastly, I build a district-year panel by counting the total number of environmental and minority-involved conflicts in a district and year.

Data on politician characteristics are from sworn affidavits submitted by candidates to the Electoral Commission of India (ECI), including a list of criminal charges at the time of candidacy. These data are digitized by the ECI and publicly distributed through the SHRUG database (Asher et al., 2021). To build a district-year panel, I aggregate the fraction of candidates across constituencies in each district with criminal charges.

To estimate the effect of institutions on conflict and criminality, I replicate the main specification in Banerjee and Iyer (2005). One difference is that my data are a panel whereas theirs is a cross section. I therefore include state-year fixed effects so that cross-sectional comparisons across districts are used for identification. Another difference is that I use Poisson regressions when the outcomes are counts of conflict events.

I estimate the following equation:

$$Y_{dsy} = \alpha + \beta_1 \cdot Inclusive_d + \Gamma X_d' + \theta_{sy} + \epsilon_{dsy} \tag{4}$$

where Y_{dsy} is either the number of conflict events or the share of criminal politicians in district d of state s in year y. $Inclusive_d$ is a dummy for whether district d has inclusive (*raiyatwari*) institutions. X_d' is a set of covariates included in Banerjee and Iyer (2005). θ_{sy} are state-year fixed effects. β_1 is the difference in conflict or criminality in inclusive relative to extractive districts.

Results are in Table S25 and described in Section 7.1 of the main text.

Table S25: Banerjee and Iyer (2005): Institutions, Protests, and Criminal Politicians

	Protests (Number)		Criminal Politicians (Pct.)	
	(1) Environmental	(2) Minorities	(3) Any Crime	(4) Num. Crimes
Inclusive (=1)	-0.906	-0.392**	-0.002	-0.041
	(0.649)	(0.190)	(0.009)	(0.045)
Controls	Yes	Yes	Yes	Yes
Outcome Mean	0.113	0.572	0.200	0.499
State $\times$ Time FEs				
Estimator	Poisson	Poisson	OLS	OLS
N	337	685	207	207
R^2			0.586	0.501

Note: * $p < .1$, ** $p < .05$, *** $p < .01$. Sample is restricted to 163 districts in Banerjee and Iyer (2005). The outcome in column 1 is the number of environmental protests. Column 2 is the number of conflicts involving scheduled castes or tribes. Column 3 is the the share of candidates who ran in the last election having a criminal charge. Column 3 is the mean number of criminal charges across candidates from the previous election. All specifications control for: scheduled caste and tribe share, baseline tree cover, district area, latitude, altitude, and a coastal dummy.

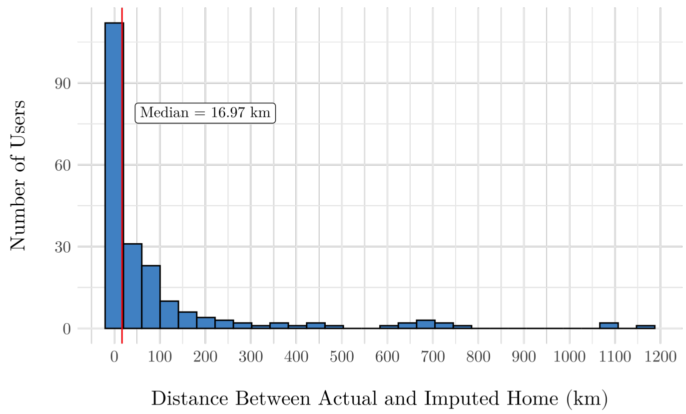


Figure S8: Distance between real and imputed home locations

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

S5 User Demographics

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

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

This method produces a fairly accurate approximation of home location. 210 users volunteered real home locations, which I use for corroboration. I compute the straight-line distance (in km) between their real and imputed home. Figure S8 shows that the median difference is only 17km, which suggests considerable accuracy of the imputation.

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

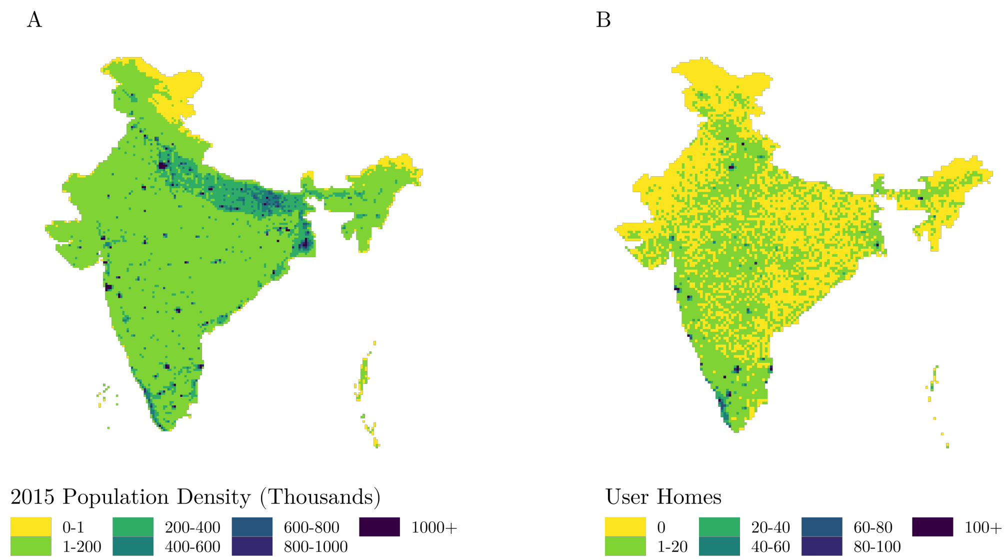


Figure S9: Population density of overall population and eBird users

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

Where Do Users Live? To visualize how representative users are in terms of where they live, I map imputed homes of all 17,634 users against gridded population density data for India. User density is mapped by constructing a 20 $\times$ 20km resolution grid and counting the number of user homes in each cell. Population density for 2015 is obtained from WorldPop[7]. Data are at 1km resolution and aggregated to 20km for consistency.

Dark hotspots in panel A of Figure S9 are India's largest cities. Many of these cities are also home to the highest density of eBird users (Panel B). While it may be unsurprising that eBird users live in big cities, Panel B also shows many remote eBird users (green).

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

Location Profiles from the DHS 2015-16 Survey As a last step to characterize eBird users, I draw on the DHS, a nationally representative household survey of 600,000 households. Households are grouped into georeferenced clusters, usually a village or town.

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

Table S26: T-test for equality of means between matched eBird and DHS samples

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

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

There are 28,395 clusters with available coordinates. My goal is to identify clusters comparable to where eBird users live. One challenge is that cluster centroids are displaced to ensure privacy. Urban and rural clusters are displaced by up to 2 and 10km, respectively.

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

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

Table S26 shows the results. Overall, there are statistically significant differences in wealth between the matched and overall sample, indicating that eBird users live in non-representative locations. Compared to the overall population, eBird users live in places with smaller household sizes and better access to amenities such as a fridge, car, separate kitchen, and flush toilets. These differences persist even within rural and urban subsamples. Put differently, the urban locations where eBird users live are wealthier than the average urban location. Yet it should also be noted that these wealth differences are quantitatively small. Thus, while eBird users live in places that are not nationally representative, these places are not markedly atypical either.

S6 Close Election Design

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

S6.1 Political Context

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

S6.2 Election Data

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

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

S6.3 First Stage Variation

First-stage variation derives from the idea that incumbents target resources toward their supporters (Dixit and Londregan, 1996)[8]. This, however, leads to a bundled estimate of predicted projects: if voters prefer less deforestation, then incumbent winners may *reduce* project approvals as a reward. At the same time, incumbents have been shown to increase public investment overall (Khemani, 2004), which may involve approving *more* projects. I control for nightlights to partially capture the investment channel, which leaves identification to rely on the assumption that, conditional on controls and fixed effects, district-level incumbent strength affects local biodiversity only by sanctioning forest diversion for infrastructure. I acknowledge this is a strong assumption. A second concern is that estimates do not generalize to non-competitive districts. For these reasons, I view this design as a check on coefficient sign rather than another set of main estimates.

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

References

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

Asher, Sam, Tobias Lunt, Ryu Matsuura, and Paul Novosad. 2021. Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in india using the shrug open data platform. *The World Bank Economic Review* 35 (4): 845–871.

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

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

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

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

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

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

Deshpande, Manasi, and Yue Li. 2019. Who is screened out? application costs and the targeting of disability programs. *American Economic Journal: Economic Policy* 11 (4): 213–248.

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

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

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

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

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

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

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