Out of the Darkness and Into the Light?
Development Effects of Rural Electrification

Fiona Burlig Louis Preonas

Supplementary Appendix: For online publication

A Additional results

A.1 First stage

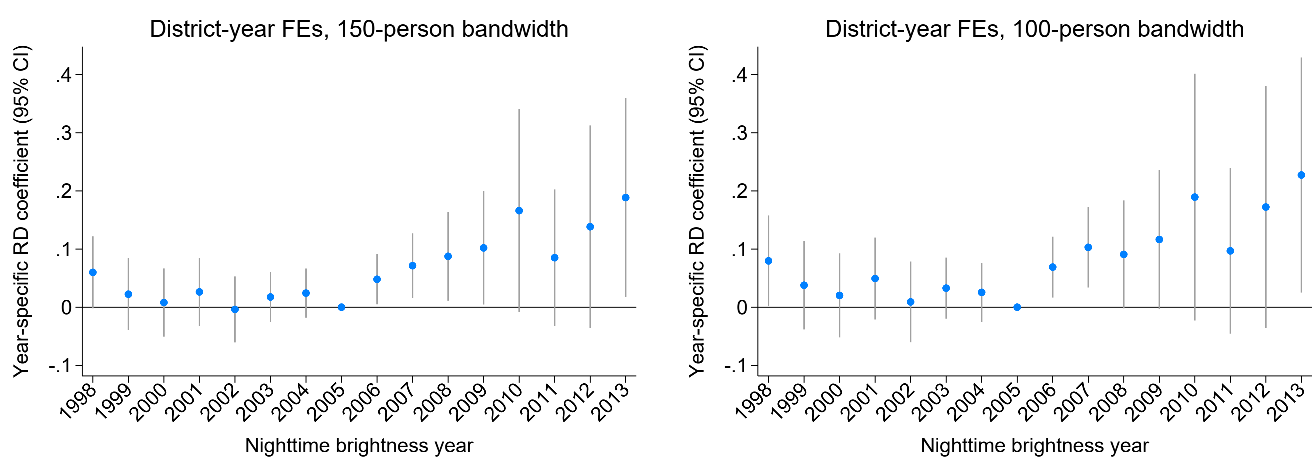
Our main `rdrobust` estimates are cross-sectional, controlling for nighttime brightness in years prior to RGGVY (or prior to the dependent variable). Comparing across annual cross-sectional RDs in Figure 5, we see null effects prior to 2005 and positive, statistically significant RD effects after 2008. We also estimate a "difference-in-discontinuities" specification to formally test for differential RD effects after vs. before RGGVY:

$$\begin{aligned} Y_{vt} \;&=\; \alpha_{1t}\, Z_v \;+\; \alpha_{2t}\, (P_v - 300) \;+\; \alpha_{3t}\, (P_v - 300) \cdot Z_v \;+\; \zeta_v \;+\; \eta_{dt} \;+\; \varepsilon_{vt} \\ &\text{for} \;\; 300 - h \leq P_v \leq 300 + h \end{aligned} \tag{A1}$$

We modify our cross-sectional RD specification to estimate separate RD coefficients for each year (α_{1t}, α_{2t}, α_{3t}), adding village fixed effects (ζ_v) and district-by-year fixed effects (η_{dt}). We impose a common RD bandwidth of $h = \{100, 150\}$ across all years, and weight villages using a triangular kernel in distance from the cutoff (which aligns with our preferred `rdrobust` specification). Table A1 reports results for all-sector power access, where the 2011 RD coefficients are differenced relative to 2001 (the omitted year).[1] Figure A1 reports results for nighttime brightness, where the omitted year is 2005. These differenced RD results are quite similar to our `rdrobust` first-stage estimates in Table 2 and Figure 5.

Tables A2–A4 apply our DD strategy to Census power access dummies and nighttime brightness. Table A5 reports DD estimates for household appliance ownership using Census data, which are analogous the estimates in Table 3.

Figure A1: Difference in discontinuities – village-level nighttime brightness



Note. — Village-level RD coefficients by year, estimated jointly via Equation (A1). Yearly coefficients plot differential RD LATEs relative to the RD LATE of 2005 (the omitted year). Each regression controls for separate linear polynomials for each year (both above and below the cutoff), village fixed effects, and district-by-year fixed effects. We weight observations by distance from the cutoff using a triangular kernel, which aligns with our preferred `rdrobust` specification. The left panel imposes a common RD bandwidth of 150 above/below the 300-person cutoff for all years; the right panel narrows this bandwidth to 100. Our estimation sample includes within-bandwidth single-habitation villages in RGGVY 10th-Plan districts. Whiskers display 95% confidence intervals, with standard errors clustered by Census block.

1. We cannot estimate Equation (A1) using commercial access or hours of power supply, since these variables are missing from the 2001 Census.

Table A1: Difference in discontinuities – village-level electricity access

	Dummy for power access, all 3 sectors	
	(1)	(2)
$\mathbf{1}[2001 \text{ pop} \geq 300] \times \mathbf{1}[2011]$	0.039***	0.036**
	(0.015)	(0.018)
Mean of dep var	0.452	0.461
RD bandwidth	150	100
Village-year observations	33,474	22,444

Note. — Difference-in-discontinuities estimates applying Equation (A1) to a two-period Census panel. We report the differential RD LATE for 2011, relative to the RD LATE in 2001 (the omitted year). The outcome is a dummy variable for electricity access in all sectors of the village economy, the only statistically significant outcome from Table 2 that we observe in the 2001 Census. Each regression imposes a common RD bandwidth across both 2001 and 2011 Census years; we weight observations by distance from the 300-person cutoff using a triangular kernel, which aligns with our preferred `rdrobust` specification. Regressions control for separate linear polynomials in the running variable for each year (both above and below the cutoff), village fixed effects, and district-by-year fixed effects. Our estimation sample includes within-bandwidth single-habitation villages in RGGVY 10th-Plan districts. Standard errors are clustered by Census block. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A2: Village-level DD of electricity access dummies

	Outcome: Dummy for any power access in village, by sector			
	Any 1 sector	Domestic	Agricultural	All sectors
	(1)	(2)	(3)	(4)
$\mathbf{1}[\text{10th-Plan district}] \times \mathbf{1}[2011]$	0.077***	0.081***	0.052***	0.037*
	(0.017)	(0.018)	(0.017)	(0.020)
Mean of dep var	0.831	0.823	0.650	0.530
Village-year observations	1.16M	1.15M	1.03M	0.97M

Note. — Village-level DD using 2001 and 2011 Census dummy variables for electricity access. Unfortunately, the 2001 Census did not collect information on electricity access for commercial use, meaning we cannot estimate that DD model. Missing data from the 2001 Census prevent us from constructing the "all sector" dummy variable for roughly 22% of villages. All regressions include: district fixed effects; year fixed effects; state-specific linear trends; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita (trends in NSS expenditure quantiles align with our NSS DD specification from Table 3). Estimation samples include villages in 10th-Plan districts, 11th-Plan districts, and non-RGGVY districts, without restricting village size. Standard errors are clustered at the district level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A3: District-level DD of electricity access dummies

	Outcome: Share of villages with any power access, by sector			
	Any 1 sector	Domestic	Agricultural	All sectors
	(1)	(2)	(3)	(4)
$\mathbf{1}[\text{10th-Plan district}] \times \mathbf{1}[2011]$	0.055***	0.059***	0.045***	0.038**
	(0.011)	(0.012)	(0.016)	(0.019)
Mean of dep var	0.897	0.890	0.671	0.571
District-year observations	1126	1126	1126	1126

Note. — This table replicates Table A2 at the district level, rather than the village level. We collapse the 2-wave Census panel of villages weighting by the number of rural households in each district (which aligns with NSS sampling weights). All regressions include: district fixed effects; year fixed effects; state-specific linear trends; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Estimation samples include villages in 10th-Plan districts, 11th-Plan districts, and non-RGGVY districts, without restricting village size. Standard errors clustered at the district level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A4: Village-level DD of nighttime brightness

	Outcome: Max village brightness		
	1st wave × 2005	1st wave × funds	Both waves × funds
	(1)	(2)	(3)
1[District is RGGVY-eligible]	0.408***	0.351***	0.218**
	(0.108)	(0.110)	(0.099)
Mean of dep var	5.75	5.75	5.75
Village-year observations	6.28M	6.28M	6.28M

Note. — Village-level DD using annual nighttime brightness from 1998 to 2013. The outcome variable is maximum (raw) nighttime brightness in each year, across each village polygon. Column (1) defines treatment at the district-level, for 10th-Plan districts only, and turns on treatment in 2005. Column (2) staggers the timing of treatment based on when each 10th-Plan district received RGGVY project funds. Column (3) expands the definition of treatment to include 11th-Plan villages, with staggered treatment turning on the year each district received RGGVY project funds. All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Estimation samples include villages in 10th-Plan districts, 11th-Plan districts, and non-RGGVY districts, without restricting village size, and in states with reliable village shapefiles. Standard errors are clustered at the district level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A5: District-level DD of household appliance ownership, Census variables

	Outcome: Share of rural households with	
	Electric lighting	TV
	(1)	(2)
1[10th-Plan district] × **1**[2011]	0.034***	0.004
	(0.011)	(0.005)
Mean of dep var	0.528	0.274
District-year observations	1126	1126

Note. — This table replicates Columns (4) and (6) of Table 3 using analogous variables from the 2001 and 2011 Census. We collapse 2011 village-level outcomes to the district level, weighting by the number of rural households (which aligns with NSS sampling weights). 2001 electric lighting and TV ownership rates are only available at the Census block level, and we use the same method to collapse them to the district level. Estimation samples include villages in 10th-Plan districts, 11th-Plan districts, and non-RGGVY districts, without restricting village size. All regressions include: district fixed effects; year fixed effects; state-specific linear trends; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A6: District-level DD, NSS variables, without state-specific linear trends

	HH electricity use (kWh/month)		
	1$[Q > 0]$	Levels	Logs
	(1)	(2)	(3)
1[10th-Plan district] × **1**[2010]	0.048***	2.47*	0.306***
	(0.014)	(1.32)	(0.068)
Mean of dep var	0.590	31.45	3.038
Clusters	552	552	550
Observations	1670	1670	1661

Note. — Regressions are identical to Columns (1)–(3) of Table 3, except that they remove state-specific linear trends. All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A7: District-level DD split by 2001 village population

	$\mathbf{1}$[kWh > 0]		kWh/month		log (kWh/month)	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{1}$[10th-Plan district] × $\mathbf{1}$[2010]	0.072***	0.043**	3.69**	6.42***	0.303***	0.325***
	(0.019)	(0.021)	(1.80)	(2.27)	(0.086)	(0.091)
2001 village population split	≤ 2000	> 2000	≤ 2000	> 2000	≤ 2000	> 2000
Mean of dep var	0.616	0.683	32.49	38.77	3.158	3.287
Clusters	535	501	535	501	516	495
Observations	1088	1014	1088	1014	1048	1000

Note. — District-level DD with two NSS years (2005, 2010), splitting on 2001 village population before collapsing to the district level using NSS sampling weights. Regressions are otherwise identical to those in Table A6. We do not observe village populations in the 2000 NSS wave, meaning that in order to estimate a 3-period panel that removes extremely large villages, we can only split on the distribution of village weights. All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A8: District-level first stage DD – villages in weight quintile 1

	$\mathbf{1}$[kWh > 0]		kWh/month		log (kWh/month)	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{1}$[10th-Plan district] × $\mathbf{1}$[2010]	0.202***	0.173***	6.46	9.32**	0.263*	0.175
	(0.049)	(0.059)	(3.92)	(4.71)	(0.151)	(0.177)
State-specific linear trends		Yes		Yes		Yes
Mean of dep var	0.632	0.632	38.28	38.28	3.385	3.385
Clusters	162	162	162	162	136	136
Observations	418	418	418	418	355	355

Note. — Regressions are identical to those in Tables 3 and A6, except that we restrict the sample to villages in the first quintile of NSS village weights before collapsing to the district level using sampling weights. We do not observe village populations for the 2000 NSS wave, meaning that we cannot estimate a 3-period panel splitting directly on village size. However, isolating the first quintile of NSS sampling weights shifts the distribution of 2001 village populations (as observed in 2005 and 2010 waves) towards smaller villages. The 50th (90th) percentile village in weight quintile 1 had a 2001 population of 1043 (4875). All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A9: District-level first stage DD – villages in weight quintiles 2–5

	$\mathbf{1}$[kWh > 0]		kWh/month		log (kWh/month)	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{1}$[10th-Plan district] × $\mathbf{1}$[2010]	0.043***	0.044***	2.54*	3.20	0.324***	0.139*
	(0.015)	(0.015)	(1.48)	(2.13)	(0.074)	(0.084)
State-specific linear trends		Yes		Yes		Yes
Mean of dep var	0.576	0.576	31.35	31.35	3.011	3.011
Clusters	494	494	494	494	493	493
Observations	1488	1488	1488	1488	1483	1483

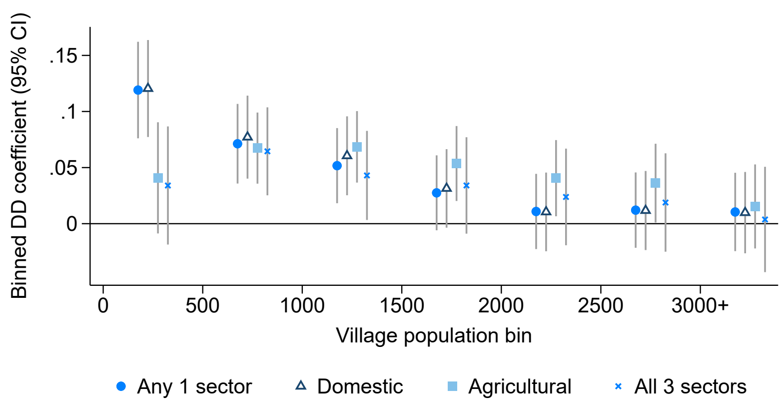
Note. — Regressions are identical to Table A8, except using the opposite set of NSS villages in sampling weight quintiles 2–5. The 50th (90th) percentile village in weight quintiles 2–5 had a 2001 population of 2076 (7291). See notes under Table A8 for details on fixed effects, linear trends, and standard errors. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A6 replicates Columns (1)–(3) of Table 3 removing state-specific linear trends. Table A7 reports NSS first-stage DD estimates split by village population size (above vs. below 2000 people), using the 2005 and 2010 NSS waves. These are similar to our results if we add the 2000 NSS wave and split on the Q1 vs. Q25 subsamples of sampling weights, which we report in Tables A8–A9. Column (1) of Tables A8–A9 corresponds to the first stage of our two-stage least squares estimates in Table 7 (Columns (2)–(3) and (5)–(6)). Figures A2–A3 report analogous DD population splits for Census power access dummies and nighttime brightness.

Table A10 and Figure A5 report numerical and graphical results to accompany Figure 5 of the main text. (We report results for 2002–2006 brightness RDs below in Table B5 and Figure B6). Figure A4 presents RD plots for electricity access in the domestic and agricultural sectors, to accompany Figure 4 from the main text.

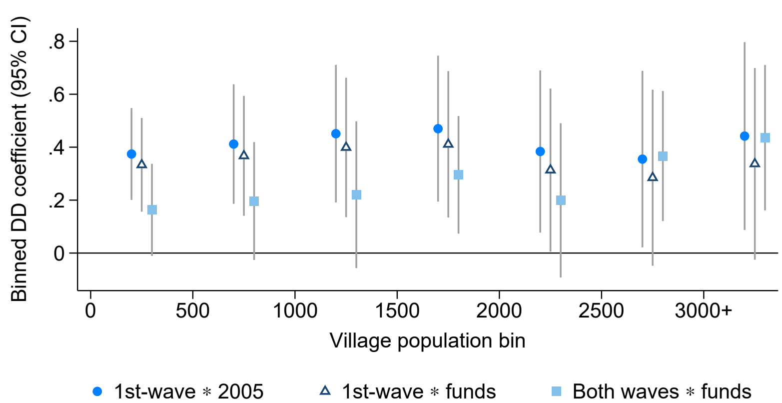
Finally, we might worry that RGGVY-induced increases in electricity purchases are crowding out self-generation of electricity. However, Table A11 finds no statistical evidence that households reduced spending on fuel for self-generation. For our preferred estimates (with state-specific trends), we can reject crowding out greater than 12% of our first-stage kWh/month point estimates.

Figure A2: Village-level DD of electricity access dummies, by village population bin



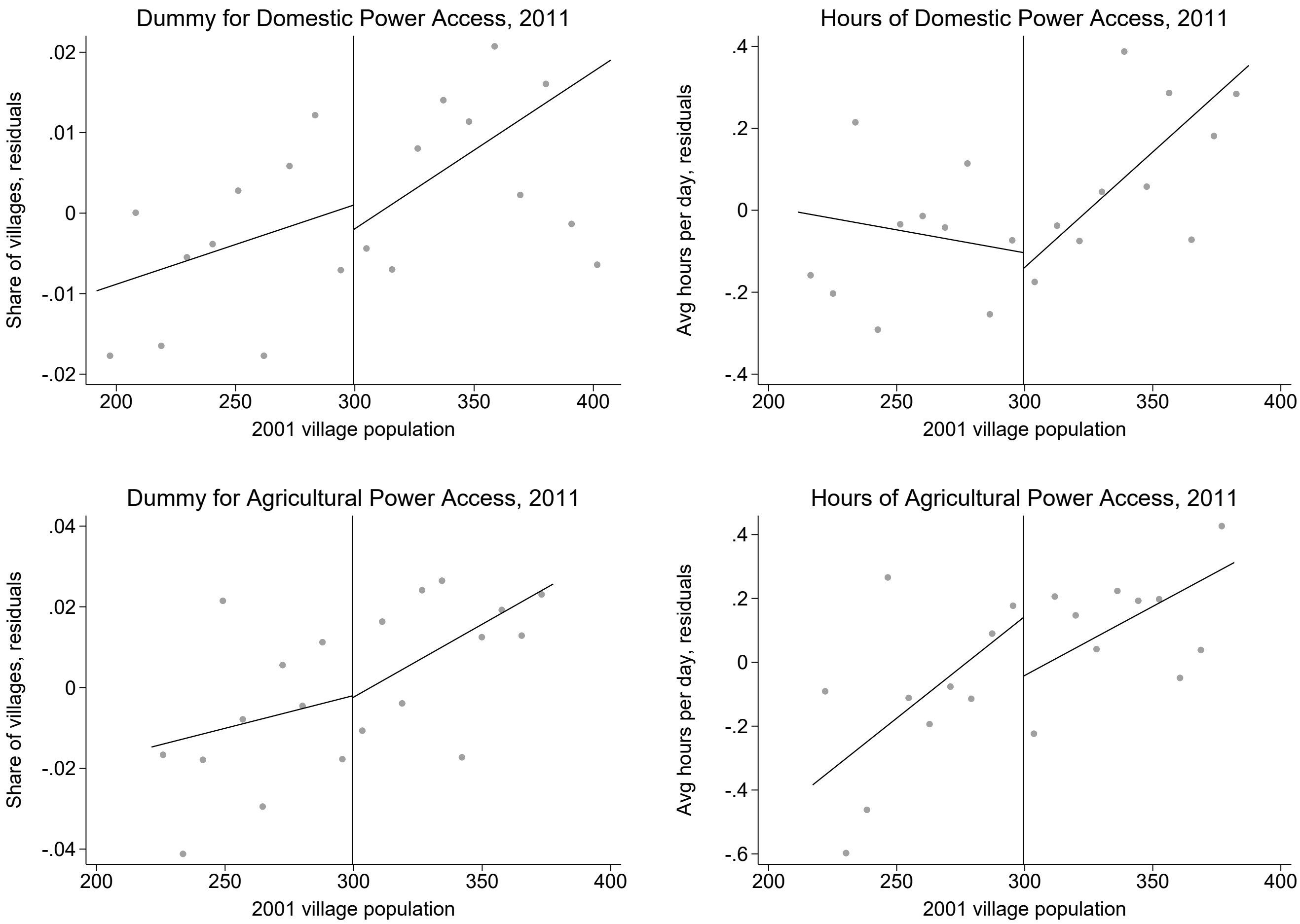
Note. — This figure plots binned DD coefficients for power access by sector We estimate four DD regressions analogous to those in Table A2. Each regression interacts the DD coefficient with 500-person bins in 2001 village population (≤500, 501–1000, . . . , 2501–3000, >3000). We interact year fixed effects with population bins, and also include village fixed effects, both sets of 2005 expenditure trends, and state-specific trends. Whiskers display 95% confidence intervals, with standard errors clustered by district.

Figure A3: Village-level DD of nighttime brightness, by village population bin



Note. — This figure plots binned DD coefficients for maximum nighttime brightness. We estimate three DD regressions analogous to those in Table A4. Each regression interacts the DD coefficient with 500-person bins (as in Figure A2). We interact year fixed effects with population bins, and also include village fixed effects and both sets of 2005 expenditure trends. Whiskers display 95% confidence intervals, with standard errors clustered by district.

Figure A4: Village-level RDs in 2011 electricity access, domestic and agricultural sectors



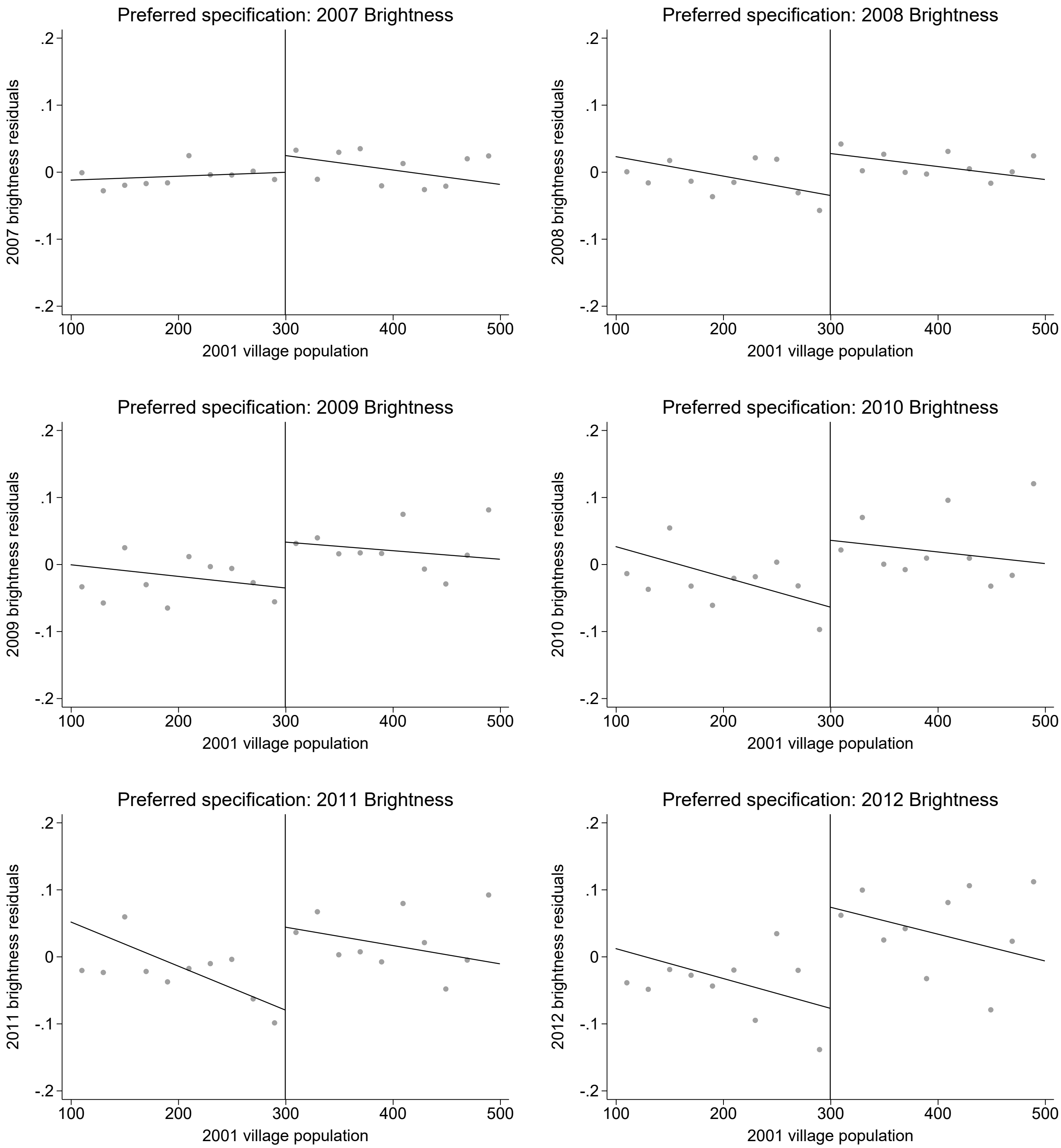
Note. — This figure complements Figure 4, with RD plots corresponding to Columns (1)–(2) of Table 2.

Table A10: Village-level RD in nighttime brightness

	Outcome: Maximum village brightness by year					
	2007	2008	2009	2010	2011	2012
	(1)	(2)	(3)	(4)	(5)	(6)
1[2001 pop $\geq$ 300]	0.040	0.143***	0.097**	0.192**	0.210***	0.253***
	(0.032)	(0.047)	(0.049)	(0.080)	(0.074)	(0.097)
Mean brightness (< 300)	4.327	5.088	5.076	7.501	6.314	7.395
Optimal bandwidth	121	69	118	87	92	98
Village observations	15,094	8,667	14,738	10,933	11,526	12,287

Note. — `rdrobust` specifications are identical to those in Table 2, and regressions are analogous to those in Figure 5. Results are robust to alternative controls, kernels, bandwidth algorithms, and standard errors. See notes under Figure 5 for details. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A5: Village-level RDs in nighttime brightness, by year



Note. — RD plots for nighttime brightness by year, which correspond to Figure 5 and Table A10. We standardize the RD bandwidths and vertical axes of these brightness plots to facilitate visual comparisons across years. See notes under Figure 5 for details. Figure B6 displays analogous RD plots for 2002–06 nighttime brightness.

Table A11: District-level DD of fuel purchases for self-generation

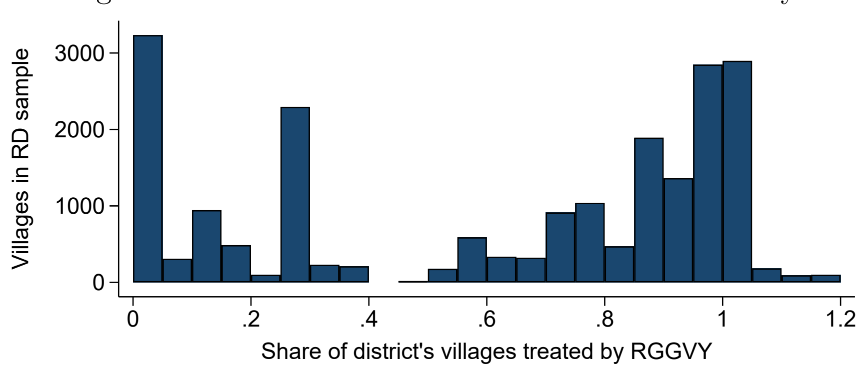
	Diesel/petrol purchases per household (Rs/month)					
	Pooled		Quintile 1		Quintiles 2–5	
	(1)	(2)	(3)	(4)	(5)	(6)
1[10th-Plan] × **1**[2010]	−0.84	−0.15	1.14	−2.75	−0.65	0.26
	(1.50)	(1.57)	(4.75)	(4.90)	(1.69)	(1.76)
Crowding out we can reject (kWh/month)	−0.33	−0.28	−0.72	−1.09	−0.35	−0.28
State-specific linear trends		Yes		Yes		Yes
Mean of dep var (Rs/month)	4.48	4.48	5.18	5.18	4.86	4.86
Clusters	552	552	162	162	494	494
Observations	1670	1670	418	418	1488	1488

Note. — Regressions are identical to Tables 3 and A6, except that Columns (3)–(6) split by quintile of NSS sampling weights before collapsing to a district-year panel. The outcome variable is the total value of diesel and petrol purchased for self-generation of electricity, at the household-month level. The second row converts the lower bound of each 95% confidence interval from Rs/month to kWh/month, assuming a 2010 diesel price of 38.1 Rs/L and a diesel generator that uses 0.3 L to produce 1 kWh. All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A.2 Heterogeneous RGGVY implementation

Here, we present results from our heterogeneity analysis from Section 6.B of the main text. Figure A6 plots RGGVY 10th-Plan implementation intensity at the district level, weighted by the number of districts in our main RD sample. We split on an implementation intensity of 60%, omitting the five 10th-Plan districts where RGGVY claims to have treated an implausibly large number of villages (relative to Census village counts). The "high-intensity" subsample of 90 districts includes 57% of villages in our main RD sample.

Figure A6: RGGVY district-wide treatment intensity



Note. — This histogram summarizes RGGVY district-level treatment intensity for the 130 10th-Plan districts in our main RD sample. As a proxy for treatment intensity, we divide the count of villages treated by RGGVY each district (per program administrative data) by the total number of villages in that district. Below, we isolate the 90 districts on the right end of this distribution, for which RGGVY treated at least 60% of their constituent villages. Our split sample analysis omits 5 districts for which RGGVY administrative data report an implausibly large number of treated villages.

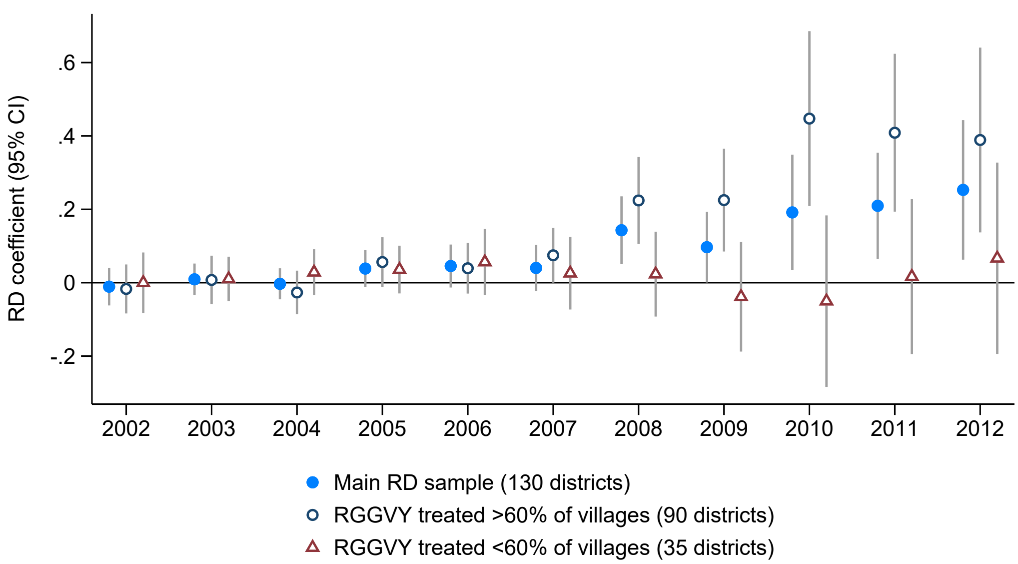
We present heterogeneous first-stage results in Tables A12–A14 and Figure A7, including electric lighting RD estimates in Table A13. Table A15 reports reduced-form RD results for the high-intensity subsample, while Table A16 reports the corresponding fuzzy RD results for expenditure per capita. Table A17 reports NSS DD estimates split by high- vs. low-intensity districts; we lack the statistical power to produce corresponding two-stage least squares estimates.

Table A12: Village-level RD in 2011 electricity access, high-intensity districts

	Outcome: Village-level electricity access			
	Domestic (1)	Agricultural (2)	Commercial (3)	All 3 sectors (4)
A. Dummy for any power access				
1[2001 pop $\geq$ 300]	0.017	0.020	0.056**	0.056**
	(0.013)	(0.019)	(0.024)	(0.024)
Mean of dep var (< 300)	0.900	0.821	0.622	0.619
Optimal bandwidth	106	92	96	100
Village observations	7,528	6,530	6,803	7,098
B. Hours/day of power supply				
1[2001 pop $\geq$ 300]	0.502*	0.182	0.973***	0.636**
	(0.274)	(0.236)	(0.334)	(0.264)
Mean of dep var (< 300)	10.062	5.927	5.297	6.572
Optimal bandwidth	82	95	77	117
Village observations	4,230	4,800	3,895	8,092

Note. — Regressions are identical to those in Table 2, except that we restrict the sample to the subset of RGGVY 10th-Plan districts where the program treated at least 60% of all villages in the district (per RGGVY district-level implementation data). This reduces the main RD sample from 130 districts in 12 states to 90 districts in 11 states. See notes under Table 2 for further details. Results are robust to alternative kernels and bandwidth algorithms. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A7: Village-level RD in nighttime brightness, by RGGVY implementation intensity



Note. — Solid blue markers reproduce our main nighttime brightness RD results from Figure 5. Hollow markers report RD results from identical regressions on two split samples: (i) the 90 districts where implementation data report that RGGVY treated over 60 percent of all villages in the district (navy circles); and (ii) 35 districts where RGGVY treated less than 60% of district villages (maroon triangles). Split samples omit 5 10th-Plan districts with implausibly high counts of treated villages in RGGVY administrative data. See notes under Figure 5 for further details. Results are robust to alternative kernels and bandwidth algorithms. Whiskers display 95% confidence intervals.

Table A13: Village-level RD – households' main source of lighting

	Share of HHs whose main source of lighting is ...			
	Electricity		Kerosene	
	(1)	(2)	(3)	(4)
1[2001 pop $\geq$ 300]	0.007	0.037***	−0.005	−0.040***
	(0.012)	(0.014)	(0.013)	(0.014)
High-intensity districts only		Yes		Yes
Mean of dep var (< 300)	0.549	0.650	0.437	0.342
Optimal bandwidth	105	110	95	105
Village observations	13,142	7,788	11,887	7,459
Number of 10th-Plan districts	129	89	129	89

Note. — Regressions are identical to those in Table 2, except Columns (2) and (4) restrict the sample to the subset of RGGVY 10th-Plan districts where the program treated at least 60% of all villages in the district (per RGGVY district-level implementation data). This reduces the main RD sample from 129 districts in 12 states to 89 districts in 11 states. Columns (1) and (3) use the main RD sample. The outcome variables are the share of households in the village using electricity (or kerosene) as primary lighting source. See notes under Table 2 for further details. Results are robust to alternative kernels and bandwidth algorithms. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A14: District-level DD – first stage, high- vs. low-intensity districts

	1[kWh > 0]			kWh/month		
	(1)	(2)	(3)	(4)	(5)	(6)
1[10th-Plan district] × **1**[2010] ×						
> 60% of villages treated	0.039**	0.116**	0.044**	3.75**	1.56	4.52**
	(0.016)	(0.056)	(0.017)	(1.88)	(5.02)	(2.06)
< 60% of villages treated	0.024	−0.008	0.022	0.30	5.43	0.26
	(0.019)	(0.116)	(0.020)	(1.46)	(8.82)	(1.60)
p-value on test of equality	0.501	0.297	0.321	0.081	0.666	0.045
Village weight quintiles	Pooled	1	2–5	Pooled	1	2–5
Mean of dep var, > 60% treated	0.632	0.683	0.620	28.23	28.94	28.07
Mean of dep var, < 60% treated	0.578	0.629	0.562	31.94	40.61	31.76
Clusters	530	142	478	530	142	478
Observartions	1605	367	1440	1605	367	1440

Note. — We interact the DD treatment variable with indicators for 10th-Plan districts where RGGVY treated more/less than 60% of villages. Regressions split on village weight quintiles, as in Table 7. The third row reports p-values for a test of equality of the two interacted DD coefficients. All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A15: Village-level RD – reduced-form outcomes, high-intensity districts

	RD estimate	Std error	95% CI	Mean Y_v
A. Consumption and income (2011)				
Expenditure per capita (Rs/month)	2.411	(22.307)	$[-41.310, 46.131]$	1434.858
Expenditure per capita (logged)	-0.005	(0.016)	$[-0.035, 0.026]$	9.713
Share HH with poverty indicator	-0.033^{*}	(0.018)	$[-0.067, 0.002]$	0.507
Share HH rely on cultivation income	0.008	(0.016)	$[-0.023, 0.038]$	0.458
Share HH earning > Rs 5k/mth	0.015	(0.013)	$[-0.011, 0.040]$	0.087
Share HH with salaried job	0.005	(0.004)	$[-0.003, 0.012]$	0.012
B. Village demographics (2011)				
Population	5.633	(4.598)	$[-3.378, 14.644]$	293.539
Share population age 0–6	-0.001	(0.002)	$[-0.005, 0.003]$	0.141
Average household size	0.017	(0.035)	$[-0.051, 0.085]$	4.913
C. Workers as share of population (2011)				
Ag workers, total	-0.010	(0.008)	$[-0.027, 0.006]$	0.428
Ag workers, male	-0.011	(0.007)	$[-0.024, 0.002]$	0.477
Ag workers, female	-0.009	(0.012)	$[-0.033, 0.014]$	0.377
Non-ag workers, total	0.009^{*}	(0.005)	$[-0.000, 0.017]$	0.071
Non-ag workers, male	0.007	(0.005)	$[-0.003, 0.018]$	0.091
Non-ag workers, female	0.010^{**}	(0.005)	$[0.000, 0.020]$	0.050
D. Firm outcomes (2013)				
Number of firms	0.382	(0.865)	$[-1.314, 2.078]$	6.870
Number of firm employees	-3.600	(7.140)	$[-17.594, 10.394]$	14.124
E. School outcomes (2014–15 school year)				
# students enrolled, grades 1–5	2.744	(3.956)	$[-5.008, 10.497]$	42.150
# students enrolled, grades 6–8	-0.867	(2.561)	$[-5.887, 4.153]$	9.022
# students passed, grades 4–5	-0.081	(0.573)	$[-1.205, 1.042]$	4.988
# students passed, grades 7–8	-0.386	(0.462)	$[-1.290, 0.519]$	1.401

Note. — RD estimates in this table are identical to those in Table 4, except that here we restrict our RD sample to the subset of RGGVY 10th-Plan districts where the program treated at least 60% of all villages in the district (per RGGVY district-level implementation data). Optimal bandwidths in the table range from 85 to 147 above/below 300 people. See notes under Tables 4 and A12 for further details. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A16: Fuzzy RD in expenditure per capita, high-intensity districts

	Expenditure per capita (Rs/month)			
	Levels	Logs	Levels	Logs
	(1)	(2)	(3)	(4)
Hours/day of commercial power	−21.934	−0.021		
	(32.388)	(0.023)		
LATE for a 10-hour increase	−219.336	−0.212		
95% CI for a 10-hour increase	$[-854.1, 415.5]$	$[-0.657, 0.232]$		
Units of nighttime brightness			17.973	−0.008
			(64.673)	(0.044)
LATE for a 2.6-unit increase			46.730	−0.020
95% CI for a 2.6-unit increase			$[-282.8, 376.3]$	$[-0.246, 0.206]$
Mean of dep var (< 300)	1427.8	9.710	1426.3	9.712
Optimal bandwidth	75	82	83	85
Village observations	3,606	3,935	5,577	5,703

Note. — Regressions are identical to those in Table 6, except that we restrict the sample to the subset of RGGVY 10th-Plan districts where the program treated at least 60% of all villages in the district (per RGGVY district-level implementation data). This reduces the main RD sample from 130 districts in 12 states to 90 districts in 11 states. See notes under Table 6 for further details. Results are robust to alternative kernels and bandwidth algorithms. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A17: District-level DD – reduced form, high- vs. low-intensity districts

	Expenditure per capita (Rs/month)					
	Levels			Logs		
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{1}$[10th-Plan district] $\times$ $\mathbf{1}$[2010] $\times$						
$>$ 60% of villages treated	50.27	−150.26	82.44**	0.045*	−0.044	0.063**
	(38.14)	(133.87)	(37.97)	(0.027)	(0.088)	(0.028)
$<$ 60% of villages treated	−37.58	−105.60	−20.88	−0.027	−0.053	−0.023
	(24.40)	(136.18)	(27.68)	(0.024)	(0.150)	(0.026)
p-value on test of equality	0.024	0.794	0.013	0.021	0.957	0.009
Village weight quintiles	Pooled	1	2–5	Pooled	1	2–5
Mean of dep var, $>$ 60% treated	918.76	1065.55	898.81	6.784	6.928	6.763
Mean of dep var, $<$ 60% treated	987.97	1157.61	957.64	6.840	6.979	6.811
Clusters	530	142	478	530	142	478
Observations	1605	367	1440	1605	367	1440

Note. — Regressions are identical to those in Table A14, except that they use monthly expenditure per capita as the outcome variable. The third row reports p-values for a test of equality of the two interacted DD coefficients. All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

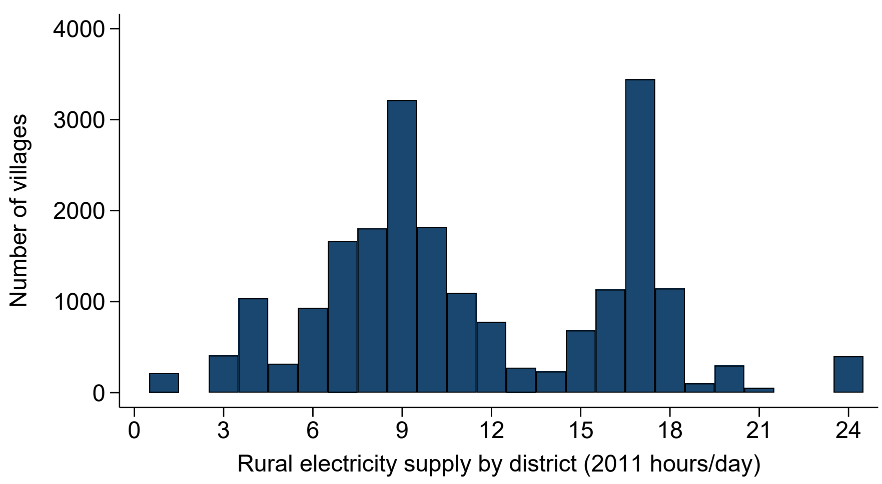
A.3 Heterogeneous power supply

Poor power quality is common on the frontier of rural electrification, including in our context. Figure A8 reports the distribution of 2011 daily power supply for electrified villages in our main RD sample (i.e. conditioning on positive hours per day of supply). This reveals that nearly half of villages in our RD analysis are located in districts where electrified villages tended to receive less than 10 hours of power per day. The following heterogeneity analysis splits on this measure of rural power quality: districts where the average electrified village receives above vs. below 10 hours of power per day. Since individual villages are small relative to these district-wide averages, we are not concerned about village-level endogeneity in hours of power due to RGGVY treatment.[2]

Table A18 estimates a heterogeneous difference-in-discontinuities specification (modifying Equation (A1)) using the all-sector power access dummy. In districts with relatively less power *supply*, RGGVY have a greater impact on the extensive margin of village power *access*: 4.8 pp vs. 2.5 pp, using an RD bandwidth of 100 people. While these effects are not statistically distinguishable from one another, they do suggest that districts with relatively worse power provision experienced greater gains on the connection margin. Figure A9 presents the analogous nighttime brightness results. We do not find evidence that more consistent power supply generated meaningful differences in RGGVY's impact on brightness—even though the more continuous brightness measure should pick up intensive-margin differences in power availability. It seems plausible that a weaker extensive-margin effect (per Table A18) is offsetting a potentially stronger brightness-intensity effect for RGGVY-eligible villages with more hours of supply.

Table A19 repeats our main reduced-form RD analysis using only districts with at least 10 hours per day of power supply. This does not reveal meaningful economic impacts in places where RGGVY connections likely yielded greater increases in electricity consumption. These results are quite similar to our results using the full RD sample (reported in Table 4).

Figure A8: Rural power supply by district (2011, hours per day)



Note. — This histogram summarizes the quality of rural power supply for districts in our main RD sample. We calculate this variable by averaging hours of all-sector power (from the 2011 Census, at the village level) over all electrified villages in each district. Below, we split on districts with at least 10 hours per day of power to the average electrified village. This histogram weights districts by the number of villages in our main RD sample.

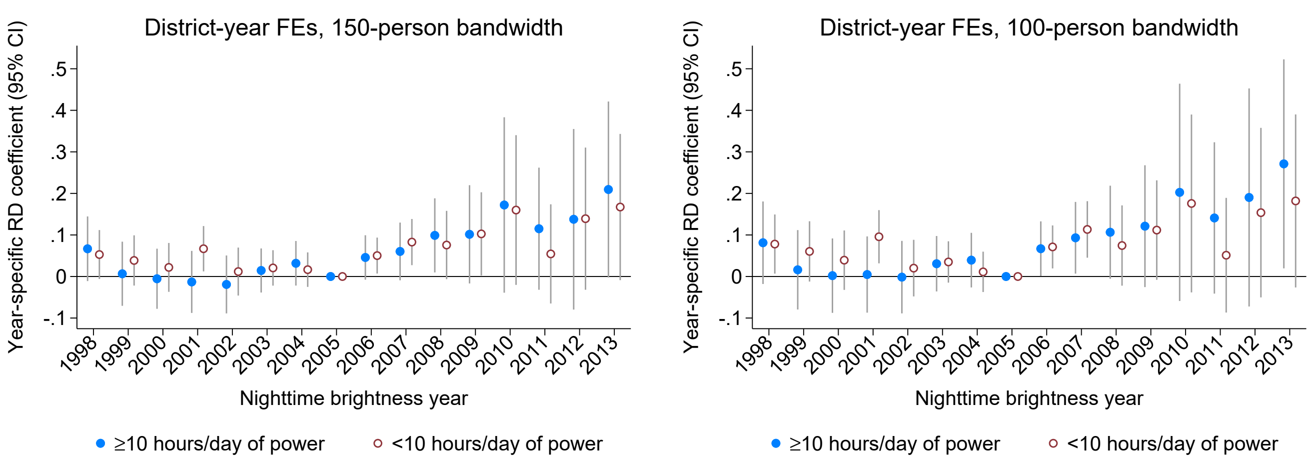
2. A previous version of this heterogeneity analysis split on state-wide power supply surplus/deficit, as reported by the Central Electricity Authority. However, this metric ignored substantial within-state between-district heterogeneity in power quality, and also conflated aggregate (urban and rural) vs. rural power quality.

Table A18: Difference in discontinuities – village-level electricity access × high/low power supply

	Dummy for power access, all 3 sectors	
	(1)	(2)
$\mathbf{1}[2001 \text{ pop} \geq 300] \times \mathbf{1}[2011] \times \mathbf{1}[\geq 10 \text{ hours/day}]$	0.030*	0.025
	(0.017)	(0.020)
$\mathbf{1}[2001 \text{ pop} \geq 300] \times \mathbf{1}[2011] \times \mathbf{1}[< 10 \text{ hours/day}]$	0.049***	0.048**
	(0.017)	(0.021)
p-value on test of equality	0.290	0.255
Mean of dep var	0.452	0.461
RD bandwidth	150	100
Village-year observations	33,474	22,444

Note. — Difference-in-discontinuities estimates are identical to those in Table A1, except that we interact the RD indicator with dummies for districts where rural villages receive over/under 10 hours per day of power supply. We calculate this variable by averaging hours of all-sector power (from the 2011 Census, at the village level) over all electrified villages in each district. We report p-values for a test of equality of the two interacted DD coefficients. Both regressions include village fixed effects and district-by-year fixed effects. Standard errors are clustered by Census block. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A9: Difference in discontinuities – village-level nighttime brightness × high/low power supply



Note. — Difference-in-discontinuity estimates are identical to those in A1, except that we interact the RD indicator with dummies for districts where rural villages receive over/under 10 hours per day of power supply. We calculate this variable by averaging hours of all-sector power (from the 2011 Census, at the village level) over all electrified villages in each district. Whiskers display 95% confidence intervals, with standard errors clustered by Census block.

Table A19: Village-level RD – reduced-form outcomes, districts with $\geq$ 10 hours/day of power

	RD estimate	Std error	95% CI	Mean Y_v
A. Consumption and income (2011)				
Expenditure per capita (Rs/month)	-9.064	(21.035)	$[-50.292, 32.163]$	1382.064
Expenditure per capita (logged)	-0.009	(0.015)	$[-0.038, 0.020]$	9.677
Share HH with poverty indicator	-0.002	(0.016)	$[-0.033, 0.030]$	0.562
Share HH rely on cultivation income	-0.002	(0.015)	$[-0.032, 0.028]$	0.397
Share HH earning > Rs 5k/mth	0.003	(0.008)	$[-0.013, 0.020]$	0.055
Share HH with salaried job	0.002	(0.005)	$[-0.009, 0.013]$	0.016
B. Village demographics (2011)				
Population	3.336	(4.679)	$[-5.835, 12.507]$	263.933
Share population age 0–6	-0.000	(0.002)	$[-0.005, 0.004]$	0.127
Average household size	-0.020	(0.027)	$[-0.072, 0.033]$	4.703
C. Workers as share of population (2011)				
Ag workers, total	-0.010	(0.008)	$[-0.026, 0.006]$	0.386
Ag workers, male	-0.008	(0.007)	$[-0.022, 0.007]$	0.466
Ag workers, female	-0.013	(0.012)	$[-0.036, 0.010]$	0.302
Non-ag workers, total	0.002	(0.005)	$[-0.008, 0.013]$	0.087
Non-ag workers, male	0.004	(0.007)	$[-0.009, 0.017]$	0.110
Non-ag workers, female	-0.001	(0.006)	$[-0.013, 0.011]$	0.064
D. Firm outcomes (2013)				
Number of firms	1.116	(1.049)	$[-0.941, 3.173]$	9.902
Number of firm employees	-3.990	(7.003)	$[-17.716, 9.736]$	19.558
E. School outcomes (2014–15 school year)				
# students enrolled, grades 1–5	-8.080	(4.939)	$[-17.760, 1.601]$	38.174
# students enrolled, grades 6–8	-7.593^{**}	(3.097)	$[-13.664, -1.522]$	8.359
# students passed, grades 4–5	-1.600	(0.977)	$[-3.515, 0.315]$	5.478
# students passed, grades 7–8	-1.262^{*}	(0.730)	$[-2.692, 0.169]$	1.764

Note. — RD estimates in this table are identical to those in Table 4, except that here we restrict our RD sample to the subset of districts where electrified villages received at least 10 hours per day of all-sector power supply in 2011. Optimal bandwidths in the table range from 87 to 188 above/below 300 people. The results for grades 6–8 enrollment and passing are not robust to alternative kernels and bandwidth algorithms. See notes under Tables 4 and A18 for further details. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

We repeat this heterogeneity analysis using NSS data. While there is some concern about estimating a heterogeneous DD model that splits on an endogenous district-level variable, it seems plausible that RGGVY's rollout did not impact the average hours of power supplied *conditional on there being non-zero supply*.[3] Table A20 reveals stronger first-stage DD effects in districts with higher rural power quality, yet these extensive-margin estimates are not statistically distinguishable between high- vs. low-power-quality groups. Table A21 replicates our DD-IV analysis, restricting the NSS sample to districts with at least 10 hours of rural power supply. For the Q1 subsample, we find tighter null effects than in Table 7. This implies that even with relatively reliable power supply, "full electrification" is unlikely to generate large economic benefits in small-to-medium villages. On the other hand, we find positive, weakly significant expenditure impacts for the Q25 subsample. While these point estimates are very large, they suggest that electrification with reliable power supply may spur economic development in larger villages. We use estimates from Columns (3)–(6) of Table A20 in the internal rate of return calculations in the bottom row of Table 9.

Table A20: District-level DD – first stage, heterogeneous effects in districts' power supply

	HH electricity use (kWh/month)					
	$\mathbf{1}[Q > 0]$	Levels	$\mathbf{1}[Q > 0]$	Levels	$\mathbf{1}[Q > 0]$	Levels
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{1}$[10th-Plan district] $\times$ $\mathbf{1}$[2010] $\times$						
$\geq$ 10 hours/day of power	0.068***	7.19***	0.212***	11.00**	0.048**	6.47**
	(0.017)	(2.35)	(0.055)	(4.80)	(0.019)	(3.00)
$<$ 10 hours/day of power	0.042**	0.22	0.045	3.78	0.040*	−0.16
	(0.021)	(1.64)	(0.163)	(10.83)	(0.021)	(1.78)
p-value on test of equality	0.303	0.002	0.336	0.527	0.759	0.011
Village weight quintiles	Pooled	Pooled	1	1	2–5	2–5
Mean of dep var, $\geq$ 10 hrs/day	0.668	38.31	0.654	40.83	0.662	39.57
Mean of dep var, $<$ 10 hrs/day	0.459	19.70	0.540	27.36	0.442	18.45
Clusters	552	552	162	162	494	494
Observations	1670	1670	418	418	1488	494

Note. — We interact the DD treatment variable with indicators for rural households in the district receiving over/under 10 hours per day of power supply. We calculate this variable by averaging hours of domestic power (from the 2011 Census, at the village level) over all electrified villages in each district. The third row reports p-values for a test of equality of the two interacted DD coefficients. Regressions are otherwise identical to Columns (1)–(2) of Table 3, and Columns (2) and (4) of Tables A8–A9. All regressions include: district fixed effects; year fixed effects; state-specific linear trends; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

3. If anything, we might expect RGGVY to have weakly lowered hours of power supplied to the average electrified village, if the total rural supply allocation did not increase to meet the demand of RGGVY-induced expansions in the customer base. This source of endogeneity would bias us against finding differentially larger first-stage impacts in districts with more hours of power. For these NSS regressions, we split hours of *domestic* (rather than all-sector) power supply. This distinction is almost trivial, since the two averages are highly correlated at the district level.

Table A21: District-level DD-IV – districts with power supply $\geq$ 10 hours/day

	Expenditure per capita (Rs/month)					
	Levels			Logs		
	(1)	(2)	(3)	(4)	(5)	(6)
1[HH consumes any elec]	608.7	−757.1	1428.0*	0.501	−0.286	0.942*
	(587.3)	(540.5)	(769.0)	(0.429)	(0.306)	(0.560)
95% confidence	[−546.4, 1763.7]	[−1826.6, 312.4]	[−85.3, 2941.2]	[−0.342, 1.344]	[−0.891, 0.320]	[−0.160, 2.045]
Village weight quintiles	Pooled	1	2–5	Pooled	1	2–5
50th pctile of 2001 pop	1946	1016	2135	1946	1016	2135
90th pctile of 2001 pop	6888	4910	7400	6888	4910	7400
Mean of dep var	1064.1	1179.6	1030.1	6.914	7.002	6.881
Clusters	352	130	306	352	130	306
Observations	1054	339	909	1054	339	909
First-stage estimate	0.071***	0.227***	0.064***	0.071***	0.227***	0.064***
(standard error)	(0.015)	(0.051)	(0.016)	(0.015)	(0.051)	(0.016)
First-stage F-statistic	21.39	19.68	15.80	21.39	19.68	15.80

Note. — Regressions are identical to those in Table 7, except that they isolate the subset of districts where rural households receive (on average) at least 10 hours per day of power supply. We calculate this variable by averaging hours of domestic power (from the 2011 Census, at the village level) over all electrified villages in each district. All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. The bottom three rows report the first-stage point estimates and standard errors, along with Kleibergen-Paap first-stage F-statistics. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A.4 Reduced form

Table A22 reports reduced-form RD results for additional village-level outcomes. Panel G includes all household-level Census variables relating to electricity-using assets. Panel L reports null effects for five aggregated outcomes, providing evidence against false precision due to multiple hypothesis testing across disaggregated outcomes for (respectively): income and wealth (Panel A); household characteristics (Panels G–H); community amenities (Panel I); labor (Panels C and F); and schooling (Panels E, J, and K).

The income/wealth index includes five SECC variables: the share of households (i) relying on cultivation income, (ii) earning at least Rs 500 per month, (iii) with a salaried job, (iv) owning any land, and (v) with**out** a poverty indicator. The household characteristics index includes 16 Census variables: share of households with (i) telephone, (ii) TV, (iii) bicycle, (iv) motorcycle, (v) electric/gas cooking, (vi) in-home kitchens, (vii) banking, (viii) water, (ix) latrine; share of households with**out** (x) thatched roofs, (xi) mud floors; share of households (xii) that are owned, (xiii) that are permanent; share of (xiv) total households, (xv) residences, and (xvi) residence-cum-other classified as "good" condition. The community amenities index includes 27 (desirable) Census variables: indicators for (i) landline phone service, (ii) mobile phone service, (iii) post office, (iv) agricultural credit society, (v) any banking, (vi) commercial banking, (vii) cooperative banking, (viii) bus service, (ix) automobiles, (x) rickshaws, (xi) taxis, (xii) tractors, (xiii) vans, (xiv) all-weather roads, (xv) drinking water; number of (xvi) primary schools, (xvii) middle schools, (xviii) secondary schools, (xix) senior secondary schools, (xx) colleges, (xxi) training schools, (xxii) primary health subcenters, (xxiii) alternative medicine hospitals, (xxiv) family welfare centers, (xxv) maternity and child welfare centers; (xvi) land area irrigated, and (xvii) percent of total land area irrigated.[4] The aggregate employment variable pools worker counts across genders, and across agricultural, non-agricultural, and household categories. The aggregate school enrollment variables sums across both genders and all available grades, and average across four school years.

Table A23 applies our DD strategy to the subset of RD outcome variables from Table 4 with pre- and post-RGGVY observations. We see significant differential increases in non-agricultural employment using Census data (Panel C), but this does not carry over to firm employee counts from the Economic Census (Panel D). Panels E and L show some evidence that RGGVY induced differential increases in school enrollment. However, these increases are not robust to the timing of RGGVY's district-level rollout (i.e. if we use district-specific treatment timing, as in Table A4).

Figures A10–A12 present RD plots for the remaining outcomes in Table 4 not included in the main text (for the sake of brevity). Finally, Table A24 reports NSS reduced-form results for the six regressions in Table 7.

4. These three indices are not perfectly centered at zero within each RD regression. This is because we standardize across villages with 0–1000 people, let `rdrobust` select a smaller bandwidth, and report means below the RD cutoff.

Table A22: Village-level RD – additional reduced-form outcomes

	RD estimate	Std error	95% CI	Mean Y_v
F. Workers as share of population (2011)				
Household workers, male	−0.001	(0.001)	[−0.003, 0.001]	0.009
Household workers, female	−0.001	(0.003)	[−0.006, 0.004]	0.014
Workers, $\geq$ 6 months of year	−0.001	(0.007)	[−0.015, 0.012]	0.313
Workers, $<$ 6 months of year	−0.003	(0.007)	[−0.016, 0.011]	0.172
G. Household asset ownership (2011)				
Share HH with telephone	−0.042***	(0.014)	[−0.069, −0.015]	0.480
Share HH with TV	0.003	(0.007)	[−0.011, 0.016]	0.225
Share HH with bicycle	0.000	(0.008)	[−0.016, 0.017]	0.473
Share HH with motorcycle	0.002	(0.004)	[−0.007, 0.010]	0.121
Share HH with radio	0.010	(0.007)	[−0.005, 0.024]	0.162
Share HH with computer, no internet	−0.006	(0.007)	[−0.019, 0.007]	0.037
Share HH with computer internet	0.000	(0.001)	[−0.001, 0.001]	0.003
Share HH with no assets	0.008	(0.008)	[−0.008, 0.024]	0.265
H. Housing characteristics (2011)				
Share HH with elec/gas cooking	−0.003	(0.004)	[−0.010, 0.004]	0.033
Share HH with kerosene lighting	−0.005	(0.013)	[−0.030, 0.020]	0.437
Share HH with mud floors	0.014*	(0.008)	[−0.003, 0.030]	0.719
Share HH with thatched roofs	−0.003	(0.010)	[−0.023, 0.017]	0.235
Share HH that are dilapidated	−0.006	(0.005)	[−0.016, 0.004]	0.074
I. Community-wide outcomes (2011)				
Share of village land planted	0.004	(0.012)	[−0.019, 0.028]	0.511
Share of village land irrigated	−0.006	(0.009)	[−0.024, 0.013]	0.270
1/0 village has tubewell	0.007	(0.022)	[−0.037, 0.050]	0.466
1/0 village has ag credit society	0.013	(0.009)	[−0.004, 0.030]	0.028
1/0 village has cell service	0.018	(0.016)	[−0.013, 0.049]	0.773
J. School outcomes (2014–15 school year)				
# students enrolled, grades 1–8	2.528	(5.683)	[−8.610, 13.666]	56.539
# boys enrolled, grades 1–8	0.540	(2.995)	[−5.330, 6.410]	28.661
# girls enrolled, grades 1–8	1.678	(2.748)	[−3.709, 7.064]	27.584
K. School outcomes (2011–12 school year)				
# students enrolled, grades 1–5	4.667	(3.883)	[−2.944, 12.278]	53.250
# students enrolled, grades 6–8	−2.350	(1.926)	[−6.124, 1.424]	8.889
# students enrolled, grades 1–8	2.346	(5.232)	[−7.909, 12.601]	61.486
# boys enrolled, grades 1–8	0.845	(2.788)	[−4.619, 6.310]	31.278
# girls enrolled, grades 1–8	1.008	(2.499)	[−3.889, 5.906]	30.053
L. Indexed and aggregated outcomes (2011)				
Index of 5 income/wealth variables	0.021	(0.017)	[−0.013, 0.054]	−0.010
Index of 16 household characteristics	0.000	(0.015)	[−0.029, 0.029]	−0.029
Index of 27 community outcomes	−0.006	(0.011)	[−0.029, 0.016]	−0.088
Total workers as share of population	−0.002	(0.006)	[−0.014, 0.009]	0.485
School enrollment, grades 1–8, 2011–15 avg.	1.424	(4.551)	[−7.495, 10.344]	53.136

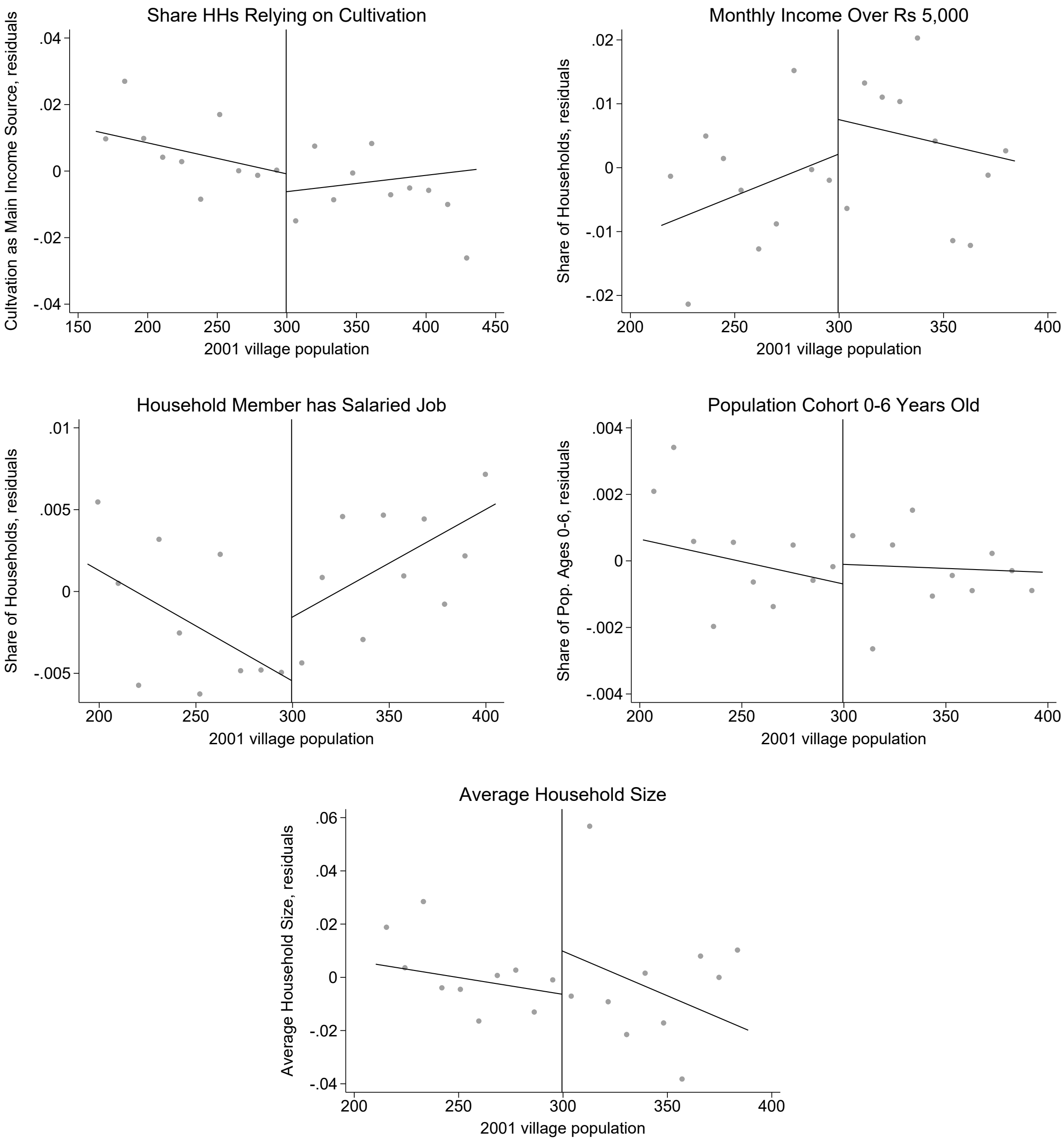
Note. — Each row reports results from a separate RD regression, for outcomes not included in Table 4. In Panels G–I and Panel L (rows 2–4), we control for the 2001 level of the outcome variable (except for computers, where 2001 data are unavailable). In Panels J–K and Panel L (row 5), we control for the 2005–06 level of the outcome variable. Panel L outcomes are pooled indexes and averages that reflect outcomes in: Panel A (income/wealth); Panels G–H (household); Panel I (community); Panels C and F (workers); and Panels E and J–K (schooling). RD robust models are otherwise identical to those in Table 2. Optimal bandwidths in the table range from 51 to 138 above/below 300 people. Results are broadly robust to alternative controls, kernels, bandwidth algorithms, and standard errors. Estimates for telephones and mud floors are not robust to alternative weighting kernels and bandwidth algorithms. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A23: Village-level and district-level DD – reduced-form outcomes

	DD estimate	Std error	95% CI	Mean Y_v
B. Village demographics (2001 & 2011)				
Share population age 0–6	0.001	(0.001)	$[-0.001, 0.003]$	0.160
Average household size †	-1.203	(1.516)	$[-4.181, 1.774]$	35.707
C. Workers as share of population (2001 & 2011)				
Ag workers, total	-0.005^{*}	(0.003)	$[-0.012, 0.001]$	0.351
Ag workers, male	-0.004	(0.003)	$[-0.009, 0.001]$	0.406
Ag workers, female	-0.006	(0.005)	$[-0.015, 0.002]$	0.292
Non-ag workers, total	0.003^{**}	(0.002)	$[0.000, 0.006]$	0.081
Non-ag workers, male	0.002	(0.002)	$[-0.002, 0.005]$	0.116
Non-ag workers, female	0.004^{**}	(0.002)	$[0.001, 0.008]$	0.044
D. Firm outcomes (1990, 1998, 2005, 2013)				
Number of firms	-1.203	(1.516)	$[-4.181, 1.774]$	35.707
Number of firm employees	-1.644	(3.910)	$[-9.324, 6.036]$	78.326
E. School outcomes (2005–15)				
# students enrolled, grades 1–5	5.577^{**}	(2.423)	$[0.818, 10.335]$	187.609
# students enrolled, grades 6–8	0.566	(1.207)	$[-1.804, 2.937]$	74.819
# students passed, grades 4–5	-0.770	(0.597)	$[-1.943, 0.403]$	6.264
# students passed, grades 7–8	-0.290	(0.385)	$[-1.046, 0.467]$	3.763
L. Indexed and aggregated outcomes (2001 & 2011)				
Index of 19 community outcomes	0.017	(0.016)	$[-0.014, 0.048]$	0.285
Total workers as share of population	-0.002	(0.003)	$[-0.009, 0.005]$	0.445
School enrollment, grades 1–8, 2005–15	6.152^{*}	(3.273)	$[-0.277, 12.582]$	262.428

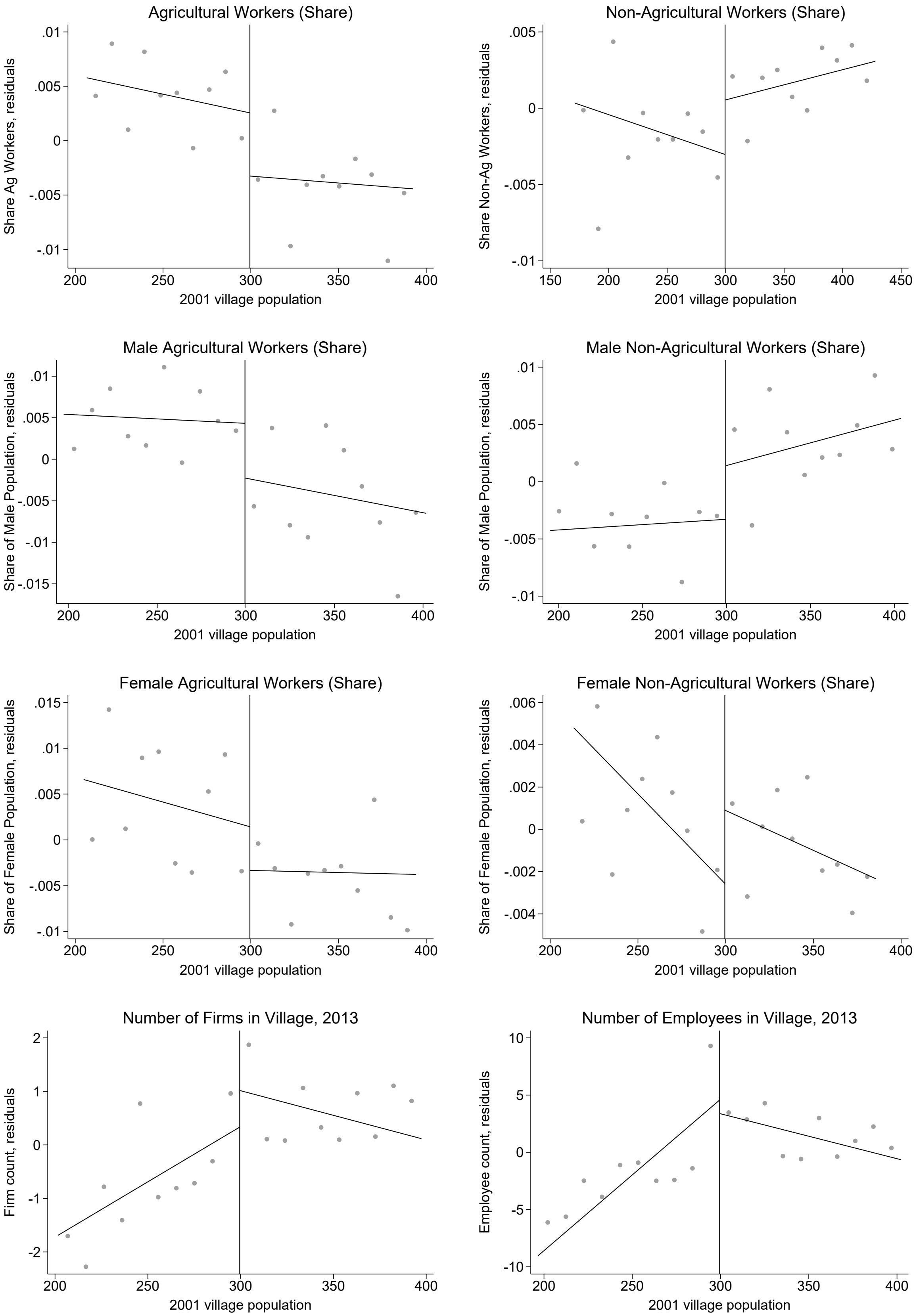
Note. — Each row reports results from a separate DD regression pooling villages of all sizes, with panel years indicated in subheaders (as dictated by each respective data source). A dagger (†) indicates collapsed district-level DD regressions (as in Table A3), for variables that are unavailable at the village level in the 2001 Census. All other regressions use village-level data and village fixed effects (as in Table A2). For yearly school outcomes, we estimate a staggered rollout of 10th-Plan funds (as in Column (2) of Table A4). We lack pre-RGGVY observations for outcomes in Panel A of Table 4, which precludes applying our DD strategy. All regressions include year fixed effects, both sets of 2005 expenditure trends, and state-specific trends. Standard errors are clustered by district. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A10: Village-level RDs in income and demographic indicators



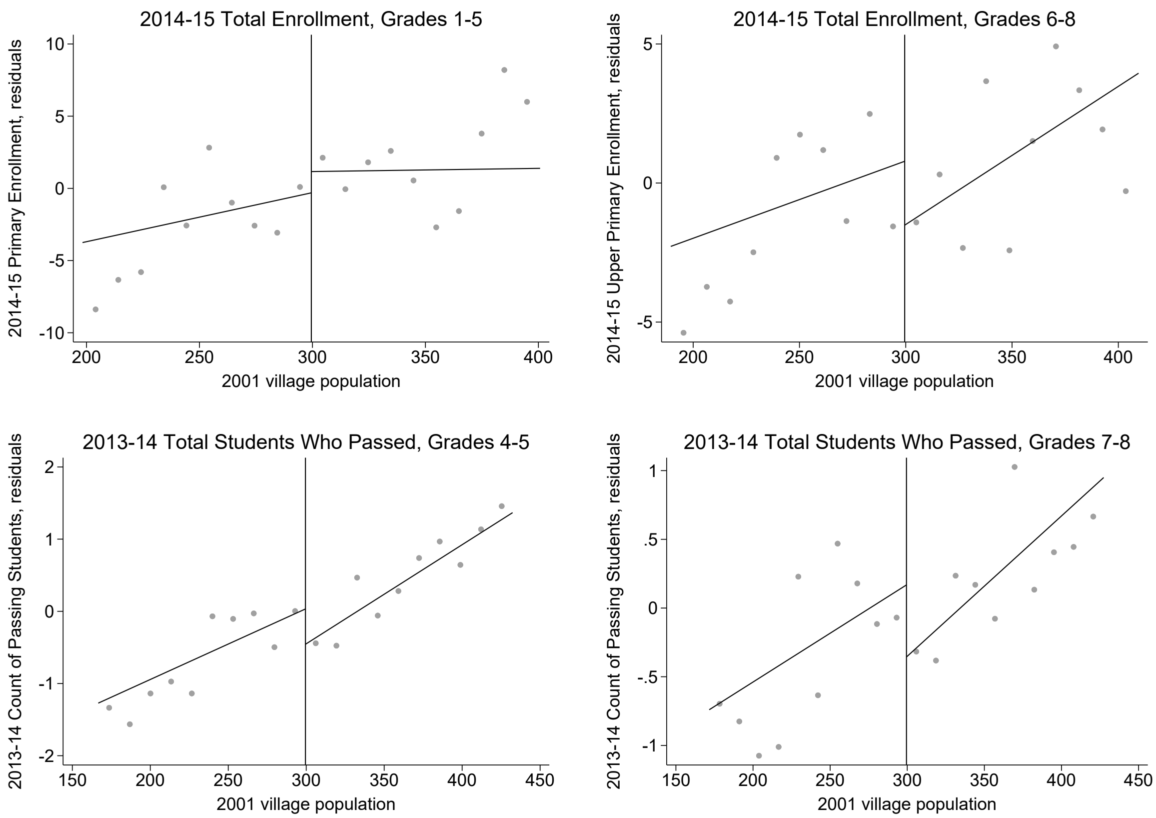
Note. — These RD plots correspond to the regressions in Panels A–B of Table 4. See table notes for details.

Figure A11: Village-level RDs in labor and firm outcomes



Note. — These RD plots correspond to the regressions in Panels C–D of Table 4. See table notes for details.

Figure A12: Village-level RDs in school outcomes



Note. — These RD plots correspond to the regressions in Panel E of Table 4. Pass rate data reported for the 2014–2015 school year reflect student outcomes from the previous academic year. See table notes for details. The statistically imprecise discontinuity estimate for grade 1–5 enrollment changes signs when we increase the RD bandwidth to 200 (from the optimal bandwidth of 101 in row 1 of Panel E).

Table A24: District-level DD – reduced form by village weight quintiles

	Expenditure per capita (Rs/month)					
	Levels			Logs		
	(1)	(2)	(3)	(4)	(5)	(6)
1[10th-Plan district] × **1**[2010]	15.15	−137.56	38.57	0.017	−0.046	0.027
	(25.42)	(93.65)	(25.80)	(0.020)	(0.062)	(0.021)
Village weight quintiles	Pooled	1	2–5	Pooled	1	2–5
Mean of dep var	978.15	1128.21	948.16	6.833	6.957	6.804
Clusters	552	162	494	552	162	494
Observations	1670	418	1488	1670	418	1488

Note. — These regressions are the reduced form analogs of the regressions in Table 7. Expenditures per capita are denominated in 2010 rupees, and net out per capita spending on electricity. All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A.5 Impacts of electrification

A.5.1 Benchmarking "full electrification"

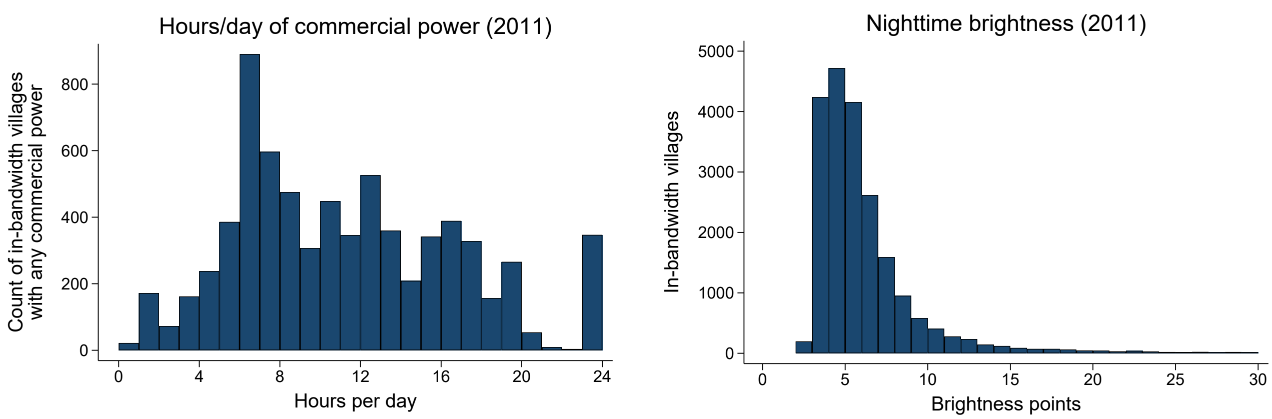
Section 6.A of the main text describes how we benchmark "full electrification" using two fuzzy RD "treatment" variables. The left panel of Figure A13 presents a histogram of 2011 hours of commercial power within our main RD sample, conditional on non-zero commercial power supply. Following these data, we scale our fuzzy RD estimates to a 10-hour-per-day increase in hours commercial power, which corresponds with moving commercial power supply from 0 hours to the median.

The right panel of Figure A13 reports the distribution of 2011 nighttime brightness for villages in our main RD sample, which is highly right-skewed. We scale up our fuzzy RD estimates to a 2.6-unit increase in brightness, corresponding to shifting a village from the 25th percentile to the 75th percentile of brightness within our RD sample.[5] Figure A14 plots median nighttime brightness by bins of village-level electric lighting penetration (i.e. electric lighting in 0–10% of households, 10–20% of households, etc.) and bins of hours of commercial power access (i.e. 0–4 hours per day, 5–8 hours per day, etc.). We see strong positive correlations between nighttime lights and both Census variables in the cross-section, which supports our use of nighttime brightness as a proxy for electrification in this setting.

A.5.2 Robustness of expenditure-based ROI simulations

Figure A15 plots the sampling distributions for expenditure per capita, which we use to conduct our benefit-cost analysis in Table 8. To convert these into ROI, for the fuzzy RD, we apply the above scaling factors (10 hours per day, 2.6 brightness units) to our estimates from Columns (1) and (3) of Table 6; convert from 2011 to 2010 rupees; and finally multiply by 12 to convert from monthly to annual expenditure. For the two DD-IV distributions from Columns (2)–(3) of Table 7, we simply multiply our monthly estimates (already in 2010 rupees) by 12.

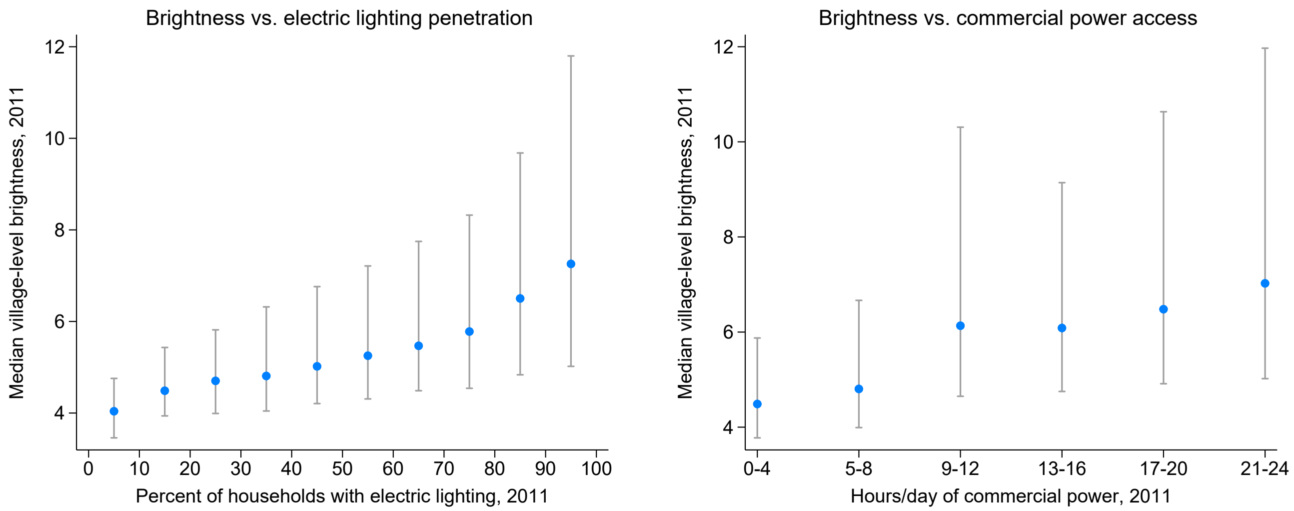
Figure A13: Histograms of fuzzy RD "treatment" variables



Note. — This figure reports histograms of the two endogenous "treatment" variables that we use to estimate fuzzy RDs. The right histogram includes all 20,880 villages in our main RD sample with 2001 populations between 150–450, for which the distribution of 2011 nighttime brightness has a mean of 6.2, a standard deviation of 3.4, and an interquartile range of 2.6 (which we use to scale our fuzzy RD confidence intervals by 2.6). The left histogram uses the same sample, while also conditioning on villages having non-zero commercial power access in 2011. This leaves a sample of 7,107 villages. This subsample has a mean of 10.9, a standard deviation of 5.5, a median of 10, and an interquartile range of 9 hours per day of commercial power supply (we rescale our fuzzy RD confidence intervals by 10).

5. Raw brightness data are integer-valued. We use linear projection to remove measurement error across brightness years, which introduces fractional (more continuous) brightness units. See Appendix C.3 below for further details.

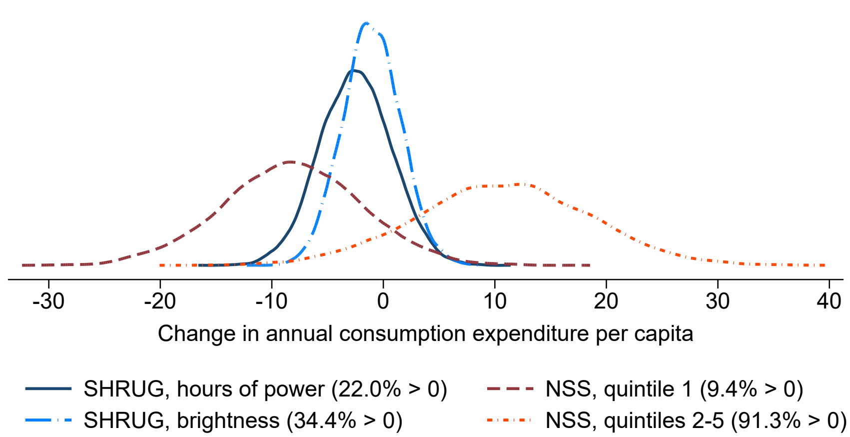
Figure A14: Comparing nighttime brightness to village-level Census variables



Note. — Both panels plot median 2011 nighttime brightness at the village level, with whiskers showing the interquartile range. The left panel splits villages by the percent of households with electric lighting (i.e., between 0–10%, 10–20%, etc.). A 3.2-unit increase in brightness (roughly) corresponds to a 100-percentage-point increase in the penetration of residential electric lighting for the median village. The right panel splits villages by hours per day of commercial power access (i.e., between 0–4 hours, 5–8 hours, etc.) A 2.5-unit increase in brightness (roughly) corresponds to a 24-hour increase in commercial power access for the median village.

Banerjee et al. (2014) report two cost benchmarks for electrifying villages under RGGVY: Rs 1.8 million per village, and Rs 1.3 million per village. We use the higher cost estimate in the main text (reported in Tables 8–9); it is still below the costs of electrification reported in other studies (e.g. Lee, Miguel, and Wolfram (2020)). However, if we use the lower fixed cost estimate of Rs 1.3 million per village, we find very similar distributions of return on investment (see Panel A of Table A25) and internal rates of return (see Table A26). Panel B of Table A25 also shows that our ROI simulations are robust to ignoring population growth.

Figure A15: Sampling distributions used for expenditure-based ROI simulations



Note. — This figure plots the distributions of annual consumption expenditures used to construct our ROI simulations in Table 8. For SHRUG estimates, we rescale the sampling distributions of our fuzzy RD point estimates in Column (1) of Table 6 by 10 (for hours of commercial power) and 2.6 (for nighttime brightness), while also deflating from 2011 rupees to 2010 rupees. For NSS estimates, the sampling distributions of our IV point estimates in Columns (2)–(3) of Table 7 are already in 2010 rupees. We also rescale all distributions by 12, to convert from monthly to annual expenditures.

Table A25: Sensitivity of expenditure-based return on investment simulations

	Pr (20-year ROI > 0), by village population			
	300	300	1000	2000
A. Low fixed costs				
$r = 0.05$	0.186	0.285	0.091	0.910
$r = 0.10$	0.174	0.264	0.090	0.909
$r = 0.15$	0.160	0.241	0.089	0.909
B. No population growth				
$r = 0.05$	0.171	0.260	0.090	0.909
$r = 0.10$	0.152	0.231	0.088	0.909
$r = 0.15$	0.135	0.203	0.085	0.907
Expenditure/capita	SHRUG	SHRUG	NSS	NSS
Endog. variable	Hours of power	Brightness	$\mathbf{1}$[HH elec > 0]	$\mathbf{1}$[HH elec > 0]
Instrument	300-person RD	300-person RD	1st-wave district	1st-wave district
Estimation sample	RD bandwidth	RD bandwidth	Quintile 1	Quintiles 2–5

Note. — This table presents two alternate versions of Table 8. Our preferred simulations in Table 8 assume fixed costs of Rs 1.8 million per village, applying the "high" fixed cost norm from Banerjee et al. (2014, p. 51). Panel A applies the same authors' "low" fixed cost norm of Rs 1.3 million per village. All ROI calculations apply variable costs of Rs 2,200 per household (Banerjee et al. 2014, p. 51). We inflate all costs from 2008 to 2010 rupees. Table 8 also applies the annual population growth rates from Panel G of Table A28. Panel B assumes zero population growth, which slightly reduces the ROIs relative to Table 8—since the undiscounted sum of per-capita benefits within each village is now constant over time (rather than increasing).

A.5.3 Constructing IRRs using consumer surplus

Table A27 reports first-stage regression results splitting the NSS sample by quartiles of household expenditure per capita rather than village size.[6] We split on expenditure in levels to see whether there are meaningful differences in the effects of RGGVY on electricity use along this dimension. This allows us to adjust consumer surplus by expenditure quartile. We estimate these regressions on an uncollapsed panel of NSS households with nationally representative regression weights.

Table A28 presents the components of our IRR calculations in Table 9. All calculations start with an assumed monthly kWh per newly electrified household, which we derive from the ratio of two first-stage econometric estimates: DD coefficients for the NSS outcome variables of kWh per household-month (as in Column (4) of Table A8) and an indicator for positive consumption (as in Column (2) of Table A8). Dividing the former by the latter lets us approximate monthly consumption per newly electrified household, which we report in the first rows of Panels A–C.[7] We follow Lee, Miguel, and Wolfram (2020) in assuming linear demand, which lets us construct consumer surplus triangles from (i) a quantity of kWh consumed, (ii) a retail electricity price (2.64 Rs/kWh, from the 2010 NSS), and (iii) a demand elasticity (0.62, from Burgess et al. (2020)). For Panel C, we report the inner product of expenditure-quartile-specific kWh per household estimates and village-size-specific quartile shares (reported in Panel D).[8] We characterize expenditure growth by

6. We do not have the statistical power to split on both dimensions.

7. Since our RD strategy cannot provide direct estimates of household electricity consumption, we use our first-stage estimates for the Q1 subsample to derive consumer surplus in a 300-person village.

8. The point estimates in Table A27 imply $\langle 45.7, 46.4, 57.4, 78.2\rangle$ kWh/month per newly electrified household in expenditure quartiles $\langle 1, 2, 3, 4\rangle$. While Columns (1)–(4) of Table A27 are statistically imprecise, we arrive at similar estimates using quartile-specific means rather than point estimates: $\langle 40.9, 47.6, 56.7, 71.3\rangle$ kWh/month.

shifting the composition of households across the four expenditure quartiles using the annualized percentage point changes reported in Panels E–F (corresponding to rows 5–6 of Table 9). Since higher-quartile households consume more electricity, this has the effect of increasing the average household's consumer surplus. These increases in consumption come on top of the 3% annual growth in electricity consumption growth that we assume in our preferred scenario.

Table A26: Internal rate of return from electrification, assuming low fixed costs

		IRRs by village population		
First-stage NSS estimates	Scenario	300	1000	2000
Q1 vs. Q25 splits	No population or kWh growth	–	13%	36%
Q1 vs. Q25 splits	No kWh growth	0%	14%	38%
Q1 vs. Q25 splits	**Preferred**	**3%**	**18%**	**43%**
Expenditure quartile splits	Preferred	4%	19%	33%
Expenditure quartile splits	NSS expenditure growth	4%	21%	35%
Expenditure quartile splits	3% expenditure growth	5%	20%	35%
Q1 vs. Q25 splits, districts with high power quality	Preferred	3%	17%	100%

Note. — Calculations in this table are identical to those in Table 9, except that they assume lower fixed costs per electrified village. Our preferred estimates in Table 9 assume fixed costs of Rs 1.8 million per village, applying the "high" fixed cost norm from Banerjee et al. (2014, p. 51). Here, we apply the same authors' "low" fixed cost norm of Rs 1.3 million per village. Both sets of IRR calculations apply variable costs of Rs 2,200 per household (Banerjee et al. 2014, p. 51). We inflate all costs from 2008 to 2010 rupees. In row 1, electrification is welfare-decreasing for 300-person villages even with a 0% annual discount rate. In row 7, electrification is benefit-cost positive for 2000-person villages after 1 year. See notes under Tables 9 and A28 for further detail.

Table A27: First-stage DD estimates split by household expenditure quartiles

	kWh/month				$\mathbf{1}$[kWh > 0]			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbf{1}$[10th-Plan]	3.39	2.74*	3.07	2.98	0.074***	0.059***	0.053**	0.038**
× $\mathbf{1}$[2010]	(2.18)	(1.53)	(1.96)	(3.12)	(0.025)	(0.021)	(0.022)	(0.018)
Expenditure quartile	1	2	3	4	1	2	3	4
Max Rs/month	574	763	1,052	405,503	574	763	1,052	405,503
Mean of dep var	15.73	24.49	35.97	56.27	0.385	0.515	0.634	0.789
Clusters	428	437	438	438	428	437	438	438
Observations	41,505	47,820	56,556	73,933	41,505	47,820	56,556	73,933

Note. — District-level DD-IV with three NSS years (2000, 2005, 2010), splitting the uncollapsed panel of NSS households by quartiles of household monthly expenditure. While all other NSS regressions collapse to a district-year panel using NSS sampling weights, these regressions use a household-year panel with nationally representative regression weights: NSS sampling weights (representative at the district level) multiplied by each district's rural population. The outcome variables are the household's monthly electricity consumption, and an indicator for whether the household consumed any electricity. Expenditure quartiles are net of spending on electricity. We divide the point estimates in Columns (1)–(4) by the point estimates in Columns (5)–(8) to construct quartile-specific estimates of kWh/month per newly electrified household, which serve as an input to rows 4–6 of Table 9. All regressions include: district fixed effects; year fixed effects; state-specific linear trends; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A28: Components used to calculate internal rates of return from electrification

	Village population		
	300	1000	2000
A. Year 1 consumption and surplus, Q1 vs. Q25 splits			
kWh/month per newly electrified HH	53.9	53.9	73.4
CS per newly electrified HH (Rs/year)	1,327	1,327	1,807
CS summed across all HH in village (Rs/year)	79,601	265,337	722,706
B. Year 1 consumption and surplus, Q1 vs. Q25 high power quality splits			
kWh/month per newly electrified HH	51.9	51.9	136.0
CS per newly electrified HH (Rs/year)	1,279	1,279	3,347
CS summed across all HH in village (Rs/year)	76,721	255,738	1,338,833
C. Year 1 consumption and surplus, expenditure quartile splits			
kWh/month per newly electrified HH	55.4	58.6	59.2
CS per newly electrified HH (Rs/year)	1,364	1,442	1,458
CS summed across all HH in village (Rs/year)	81,814	288,464	583,214
D. Share of households in each expenditure quartile (2010 NSS)			
Share of HH in quartile 1, $[0\ ,\ 574]$ Rs/month	0.289	0.224	0.234
Share of HH in quartile 2, $(574\ ,\ 763]$ Rs/month	0.272	0.238	0.219
Share of HH in quartile 3, $(763\ ,\ 1,052]$ Rs/month	0.228	0.227	0.212
Share of HH in quartile 4, $(1,052\ ,\ 405,503]$ Rs/month	0.211	0.311	0.336
E. Observed changes in expenditure shares (2005 & 2010 NSS)			
Annualized change in share of HH in quartile 1	−0.015	−0.019	−0.021
Annualized change in share of HH in quartile 2	+0.008	−0.005	−0.001
Annualized change in share of HH in quartile 3	+0.002	+0.001	+0.001
Annualized change in share of HH in quartile 4	+0.006	+0.023	+0.020
F. Changes in expenditure shares, under 3% annual expenditure growth			
Annualized change in share of HH in quartile 1	−0.022	−0.007	−0.023
Annualized change in share of HH in quartile 2	−0.003	−0.013	−0.003
Annualized change in share of HH in quartile 3	+0.011	+0.010	+0.007
Annualized change in share of HH in quartile 4	+0.014	+0.011	+0.018
G. Additional assumptions and summary statistics			
Annualized population growth rate (2001 & 2011 Census)	1.38%	1.51%	1.57%
Annualized growth rate in kWh per electrified HH (2005 & 2010 NSS)	2.15%	-0.69%	1.06%
Total variable costs: Rs 2,732 per household	163,913	546,377	1,092,753

Note. — Panels A–C report kWh per electrified household used in Table 9, for the initial year post-electrification. We calculate consumer surplus assuming linear electricity demand, a demand elasticity of 0.62 (Burgess et al. (2020)), a retail electricity price of Rs 2.64 per kWh (the median price in the 2010 NSS), and 5 people per household. For rows 1–3 of Table 9, Panel A divides Column (4) by Column (2) in Tables A8–A9 (using Q1 estimates for 300- and 1000-person villages, and Q25 estimates for 2000-person villages). For row 7 of Table 9, Panel B uses the analogous first-stage estimates from Table A20. For rows 4–6 of Table 9, Panel C takes the inner product of expenditure-quartile-specific kWh per household (ratios of coefficients from Table A27) and village-size-specific expenditure quartile shares (reported in Panel D). For row 5 of Table 9, we use the observed annualized shifts in expenditure quartile shares (reported in Panel E) translate year-on-year growth in expenditures into increased electricity consumption, by recalculating the inner product with updated expenditure quartile shares for each year post-electrification. For row 6 of Table 9, we do the same using the annualized shifts in expenditure quartile shares implied by 3% annual growth in expenditure (following Lee, Miguel, and Wolfram (2020); reported in Panel F). Rows 2–7 of Table 9 impose annual population growth rates for each village size, as implied by the 2001 and 2011 Censuses (reported in Panel G). Rows 3–7 of Table 9 impose 3% annual growth in kWh, which exceeds the annualized growth rates implied by the 2005 and 2010 NSS (reported in Panel G; in rows 5–6 of Table 9, this 3% kWh growth in on top of expenditure-driven increases in kWh). All IRR calculations include variable costs of 2,732 per household (reported in Panel G), inflating Rs 2,200 per household (Banerjee et al. 2014, p. 51) from 2008 to 2010 rupees. Expenditures and consumer surplus are also in 2010 rupees. NSS summary statistics impose nationally representative weights, inflating NSS weights by each district's rural population. The average NSS household with positive electricity consumption consumed 53.5 kWh per month. Village-size-specific summary statistics use villages with populations of [250, 350], [950, 1050], and [1950, 2050], respectively.

A.5.4 Additional fuzzy RD and DD-IV results

Our preferred fuzzy RD estimates in Table 7 remove (most) larger villages by splitting on NSS sampling weights. Isolating the first weight quintile shifts the distribution of village populations towards smaller villages and allows us to include the 2000 NSS wave (where village populations are not reported). Table A29 instead reports DD-IV results selecting directly on villages with populations of 2000 or less. These estimates have slightly smaller first-stage F-statistics than our preferred Q1 estimates, since filtering by village population forces us to drop 2000 from our NSS panel. However, the resulting point estimates and confidence intervals are quite similar to Columns (2) and (5) of Table 7. This assuages concerns that there may be unexpected correlations between NSS sampling weights and unobserved (population-invariant) village characteristics, which might otherwise undermine our use of these weights as a population proxy.

Table A29: District-level DD-IV of consumption expenditures – sub-2000-person villages

	Expenditure per capita (Rs/month)					
	Levels			Logs		
	(1)	(2)	(3)	(4)	(5)	(6)
1[HH consumes any elec]	−486.7	−613.2	−392.2	−0.183	−0.289	−0.262
	(467.4)	(462.7)	(446.0)	(0.349)	(0.335)	(0.359)
95% confidence	[−1404.9, 431.5]	[−1522.2, 295.8]	[−1268.3, 483.9]	[−0.867, 0.502]	[−0.946, 0.369]	[−0.967, 0.444]
Village population (2001)	≤ 2000	≤ 2000	≤ 2000	≤ 2000	≤ 2000	≤ 2000
50th pctile of 2001 pop	991	991	991	991	991	991
90th pctile of 2001 pop	1781	1781	1781	1781	1781	1781
Linear trends by:						
State exp. quartiles		Yes	Yes		Yes	Yes
Nat'l exp. deciles			Yes			Yes
Mean of dep var	948.9	948.9	948.9	6.787	6.787	6.787
Clusters	535	535	535	535	535	535
Observations	1088	1088	1088	1088	1088	1088
First-stage estimate	0.076***	0.077***	0.072***	0.076***	0.077***	0.072***
(standard error)	(0.017)	(0.018)	(0.019)	(0.017)	(0.018)	(0.019)
First-stage F-statistic	18.89	18.67	14.49	18.89	18.67	14.50

Note. — District-level DD-IV with two NSS years (2005, 2010), splitting on 2001 village population before collapsing to the district level using NSS sampling weights. We do not observe village populations in the 2000 NSS wave, meaning that in order to estimate a 3-period panel that removes extremely large villages, we can only split on the distribution of village weights. These regressions include NSS villages with 2000 or fewer people in 2001, with results that compare favorably to Columns (2) and (5) of Table 7 (over similar population supports). Regressions are otherwise identical to those in Table 7, except for fewer linear trends in in Columns (1)–(2) and (4)–(5). Including both sets of linear trends in expenditure causes our F-statistics in the 2-period panel to drop below the Stock-Yogo critical value of 16.38. Since these trends help our identification, we report the 3-period panels as our main results. All regressions include district fixed effects and year fixed effects. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. The bottom three rows report the first-stage point estimates and standard errors, along with Kleibergen-Paap first-stage F-statistics. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Tables A30–A31 report fuzzy RD results for all outcomes in Table 4, except for SHRUG expenditure per capita (already reported in Table 6). We omit these results from the main text for the sake of brevity.

Table A30: Village-level fuzzy RD outcomes – hours of commercial power

	RD estimate	Std error	95% CI × 10	Mean Y_v
A. Consumption and income (2011)				
Share HH with poverty indicator	0.011	(0.020)	$[-0.289, 0.507]$	0.546
Share HH rely on cultivation income	−0.036	(0.028)	$[-0.914, 0.186]$	0.419
Share HH earning > Rs 5k/mth	0.003	(0.012)	$[-0.209, 0.261]$	0.070
Share HH with salaried job	0.004	(0.006)	$[-0.066, 0.156]$	0.012
B. Village demographics (2011)				
Population	12.055	(9.130)	$[-58.389, 299.493]$	293.897
Share population age 0–6	−0.000	(0.003)	$[-0.060, 0.059]$	0.141
Average household size	0.021	(0.049)	$[-0.736, 1.165]$	4.897
C. Workers as share of population (2011)				
Ag workers, total	−0.008	(0.013)	$[-0.331, 0.162]$	0.398
Ag workers, male	−0.011	(0.011)	$[-0.338, 0.110]$	0.465
Ag workers, female	−0.004	(0.017)	$[-0.382, 0.295]$	0.328
Non-ag workers, total	0.005	(0.007)	$[-0.096, 0.191]$	0.075
Non-ag workers, male	0.005	(0.009)	$[-0.122, 0.225]$	0.095
Non-ag workers, female	0.005	(0.008)	$[-0.109, 0.207]$	0.053
D. Firm outcomes (2013)				
Number of firms	1.043	(1.189)	$[-12.870, 33.740]$	7.910
Number of firm employees	−8.508	(9.458)	$[-270.459, 100.298]$	15.916
E. School outcomes (2014–15 school year)				
# students enrolled, grades 1–5	15.882	(22.183)	$[-275.947, 593.592]$	46.509
# students enrolled, grades 6–8	−6.021	(13.282)	$[-320.537, 200.125]$	10.047
# students passed, grades 4–5	−2.822	(2.729)	$[-81.713, 25.270]$	5.202
# students passed, grades 7–8	−4.123	(3.974)	$[-119.114, 36.661]$	1.450

Note. — This table reports fuzzy RD result for the outcomes in Table 4, using 2011 hours of commercial power access as the endogenous "treatment" variable. We omit expenditure outcomes, reported in Columns (1)–(2) of Table 6. We apply a scaling factor of 10 to the 95% confidence intervals. See notes under Table 6 for further details. Optimal bandwidths in the table range from 106 to 164 above/below 300 people. Results are broadly robust to alternative controls, kernels, bandwidth algorithms, and standard errors. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A31: Village-level fuzzy RD outcomes – nighttime brightness

	RD estimate	Std error	95% CI $\times 2.6$	Mean Y_v
A. Consumption and income (2011)				
Share HH with poverty indicator	-0.026	(0.076)	$[-0.452, 0.318]$	0.550
Share HH rely on cultivation income	-0.058	(0.085)	$[-0.583, 0.283]$	0.422
Share HH earning > Rs 5k/mth	0.063	(0.063)	$[-0.156, 0.486]$	0.071
Share HH with salaried job	0.021	(0.026)	$[-0.076, 0.188]$	0.013
B. Village demographics (2011)				
Population	32.401	(25.117)	$[-43.750, 212.233]$	296.447
Share population age 0–6	0.004	(0.009)	$[-0.035, 0.054]$	0.141
Average household size	0.099	(0.132)	$[-0.413, 0.928]$	4.899
C. Workers as share of population (2011)				
Ag workers, total	-0.026	(0.038)	$[-0.262, 0.126]$	0.398
Ag workers, male	-0.034	(0.036)	$[-0.270, 0.096]$	0.466
Ag workers, female	-0.021	(0.051)	$[-0.315, 0.207]$	0.328
Non-ag workers, total	0.024	(0.023)	$[-0.056, 0.179]$	0.075
Non-ag workers, male	0.030	(0.031)	$[-0.082, 0.236]$	0.095
Non-ag workers, female	0.019	(0.023)	$[-0.066, 0.167]$	0.054
D. Firm outcomes (2013)				
Number of firms	4.802	(5.133)	$[-13.674, 38.643]$	7.955
Number of firm employees	-12.225	(31.404)	$[-191.819, 128.249]$	15.824
E. School outcomes (2014–15 school year)				
# students enrolled, grades 1–5	2.799	(12.087)	$[-54.319, 68.875]$	46.472
# students enrolled, grades 6–8	-4.745	(7.564)	$[-50.884, 26.210]$	10.111
# students passed, grades 4–5	-1.750	(2.088)	$[-15.187, 6.089]$	5.184
# students passed, grades 7–8	-1.114	(1.598)	$[-11.041, 5.246]$	1.501

Note. — This table reports fuzzy RD result for the outcomes in Table 4, using 2011 nighttime brightness as the endogenous "treatment" variable. We omit expenditure outcomes, reported in Columns (3)–(4) of Table 6. We apply a scaling factor of 2.6 to the 95% confidence intervals. See notes under Table 6 for further details. Optimal bandwidths in the table range from 97 to 198 above/below 300 people. Results are broadly robust to alternative controls, kernels, bandwidth algorithms, and standard errors. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B Identification checks and sensitivity analysis

B.1 First-stage RD

Figures B1–B2 summarize an array of sensitivity analyses for our first-stage RD estimates. First, we compare the (`rdrobust` default) triangular kernel to the Epanechnikov and uniform kernels, for weighting observations by distance from the cutoff. Next, we compare the (`rdrobust` default) MSE-optimal bandwidth algorithm (`mserd`) to two alternative bandwidth selectors: the MSE-sum algorithm, which optimizes the sum (rather than the difference) of mean squared errors (`msesum`); and the common coverage error rate algorithm (`cerrd`).[9] Our preferred estimates apply `rdrobust`'s nearest-neighbor variance estimator, which performs best in finite samples (Calonico, Cattaneo, and Titiunik (2014)). The resulting standard errors consider "neighbors" in terms of the RD running variable rather than physical space, however our first-stage results are robust to allowing for spatially correlated errors within Census blocks and districts.[10] While our preferred RD estimates include state fixed effects, we show our results are broadly robust to removing these fixed effects; including district fixed effects slightly attenuates our RD point estimates, causing our electricity access RDs to lose significance. While adding a second-order polynomial reduces precision using the Census electricity variables, our nighttime brightness RDs are robust to this alternative functional form.

We test for robustness to two RD sample criteria. First, our main RD sample excludes the 11% of villages classified as "single-habitation", but whose official village Census populations differ from their matched habitation populations by over 20% (see discussion in Appendix C.5). Second, we also exclude 0.2% of villages with extremely implausible 10-year swings in nighttime brightness (see discussion at the end of Appendix C.3). Including villages with population mismatches slightly attenuates our RD estimates for Census power access dummies (losing significance), likely due to right-hand-side measurement error from villages that are not truly single-habitation. Including brightness outlier villages substantially attenuates our brightness RD estimates, but does not impact our Census electricity access RD estimates.

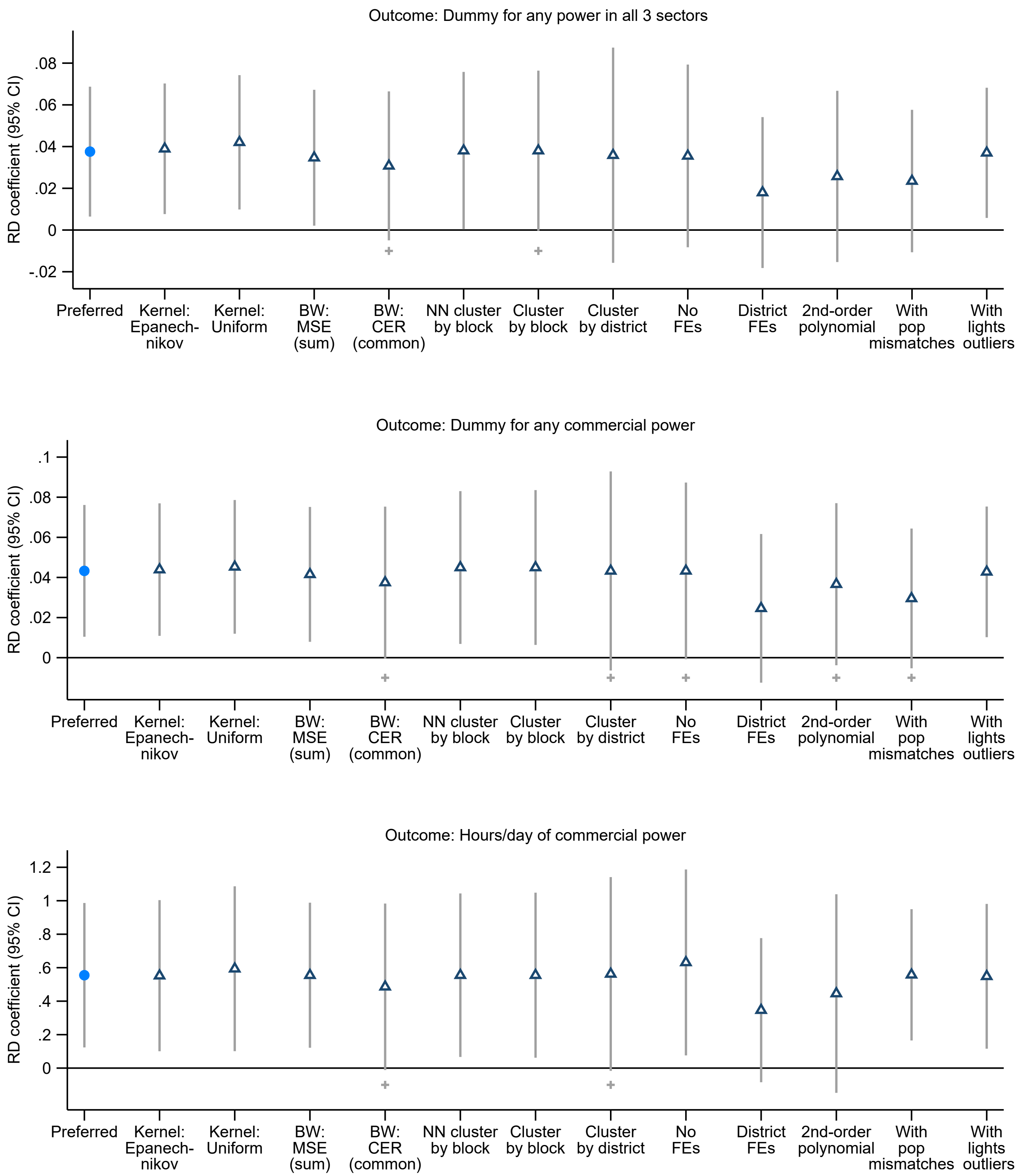
The bottom-right panel of Figure B2 also reports results for alternative brightness variables. Our preferred estimates remove satellite-induced measurement error via linear projection, and assign the maximum brightness across all village pixels. Our results are similar if we use mean projected brightness, maximum brightness averaged over 3 or 5 years, or two alternative satellite-derived data products. Raw (un-projected, un-averaged) brightness data yield statistically insignificant results. Appendix C.3 provides more detail on nighttime brightness data.

Next, we perform a falsification exercise based on the implementation details of the RGGVY program. Our RD sample includes only villages in RGGVY 10th-Plan districts, for which the relevant eligibility cutoff was 300 people. It also includes only villages with exactly one habitation, for which 2001 village population is the appropriate running variable (since habitation populations determined RGGVY eligibility). Figures B3–B4 show that our first-stage RD estimates are all statistically indistinguishable from zero if we use villages in 11th-Plan districts (where 300 people is

9. A previous version of this paper imposed a common 150-person bandwidth across all RD regressions. We now use `rdrobust` bandwidth algorithms that almost always select bandwidths narrower than 150 people on either side of the 300-person cutoff. Our results are still robust to imposing ad hoc bandwidths between 50 and 150 (results available upon request).

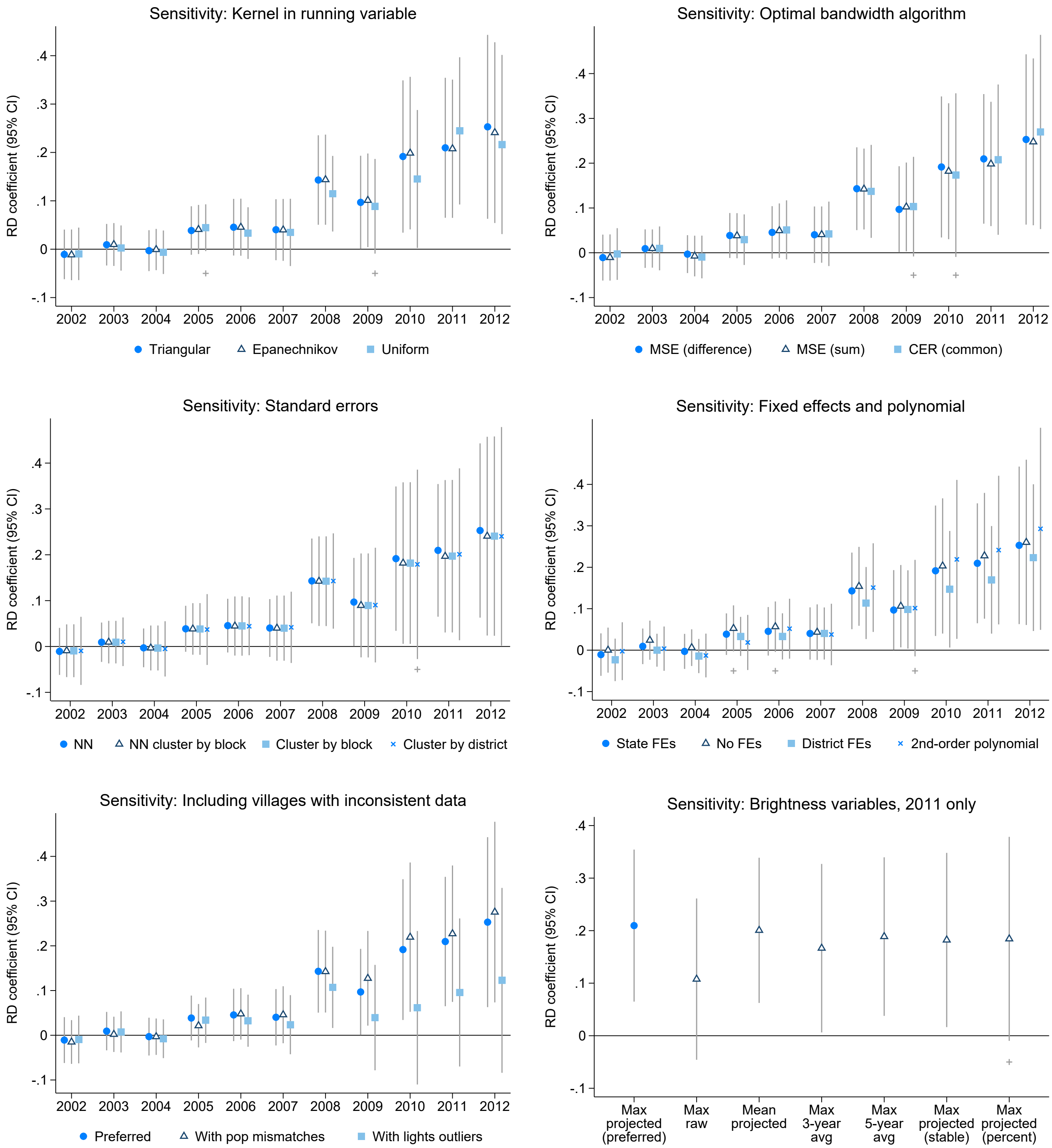
10. Census block is the administrative unit that is smaller than district but larger than village. While RGGVY projects were approved based on district-specific implementation plans (or DPRs), the level of (quasi-)randomization in our village-level RD design is the village.

Figure B1: Sensitivities for village-level RD in 2011 electricity access



Note. — Each panel reports `rdrobust` point estimates and 95% confidence intervals, for sensitivity analysis on one of three outcomes from Table 2: all-sector power access dummy (top panel); commercial power access dummy (middle panel); and hours per day of commercial power supply (bottom panel). In each panel, the left-most point plots our preferred RD estimates from Table 2. First, we compare Epanechnikov and uniform weighting kernels to our preferred triangular kernel. Next, we compare our preferred MSE-optimal bandwidth selection algorithm (`mserd`) with the MSE-optimal bandwidth based on the sum (`msesum`) rather than the difference, and with the coverage error rate optimal bandwidth algorithm (`cerrd`). Next, we test alternative standard errors: nearest-neighbor clustering by Census block, clustering by Census block, and clustering by district. Next, we remove state fixed effects, and then add district fixed effects. Next, we use a quadratic polynomial in the running variable. Finally, we add in (i) the 11% of villages with population discrepancies in merged habitation data, and (ii) the 0.2% of villages with extremely implausible 10-year changes in nighttime brightness. Pluses indicate significance at the 10% level, but not at the 5% level.

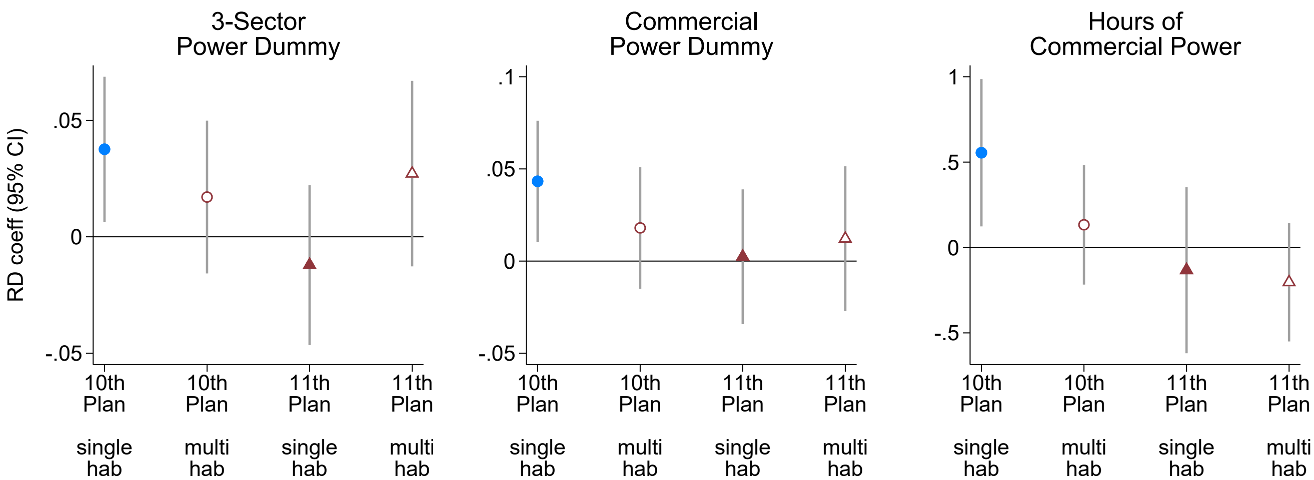
Figure B2: Sensitivities for village-level RD in nighttime brightness



Note. — Each panel reports `rdrobust` point estimates and 95% confidence intervals, for a separate set of sensitivity analyses. In all 6 panels, the solid blue circles report our preferred results from Figure 5. The top-left panel compares our preferred triangular weighting kernel to the Epanechnikov and uniform kernels. The top-right panel compares our preferred MSE-optimal bandwidth selection algorithm (`mserd`) with the MSE-optimal bandwidth based on the sum (`msesum`) rather than the difference, and with the coverage error rate optimal bandwidth algorithm (`cerrd`). The middle-left panel applies alternative standard errors: nearest-neighbor clustering by Census block, clustering by Census block, and clustering by district. The middle-right panel removes state fixed effects, adds district fixed effects, and uses a 2nd-order polynomial in the running variable (with state fixed effects). The bottom-left panel alternately adds in (i) the 11% of villages with population discrepancies in merged habitation data, and (ii) the 0.2% of villages with extremely implausible 10-year changes in nighttime brightness. The bottom-right panel tests our preferred outcome variable (spatial maximum of NOAA's "average" nighttime brightness, linearly projected across adjacent years to remove measurement error) for sensitivity to using: raw brightness, spatial mean brightness, unweighted 3- and 5-year average brightness, and two alternative NOAA lights products (more processed "stable" lights, and "percent" lights that weight by observed frequency). Pluses indicate significance at the 10% level, but not at the 5% level.

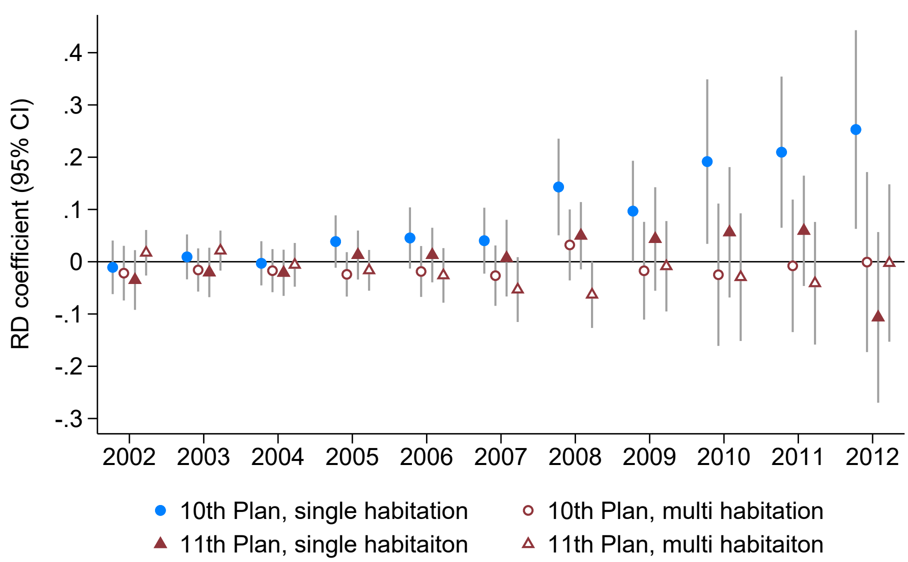
the wrong cutoff) or multi-habitation villages (where 2001 village population is the wrong running variable). This provides further evidence that the RGGVY program is the true mechanism behind our RD results.

Figure B3: Falsification test for village-level RD in 2011 electricity access



Note. — This figure compares our main RD estimates to analogous RD estimates using three separate "falsification" samples of villages. All RDs apply the same specification as in Table 2, using 2001 Census populations as the running variables with a 300-person cutoff. Solid blue circles report our results in Table 2: Column (4), Panel A (left panel); Column (3), Panel A (middle panel); and Column (3), Panel B (right panel). These estimates use the subset of villages in RGGVY 10th-Plan districts (i.e. eligible for first-wave funding), with a single habitation (i.e. for which 2001 Census populations determined RGGVY eligibility). For multi-habitation villages, 2001 Census populations do not match the habitation-based 300-person cutoff—meaning that these RDs are using the wrong running variable. For villages in RGGVY 11th-Plan districts, the population cutoff was reduced to 100 people—meaning that these RDs are using the wrong cutoff. Whiskers display 95% confidence intervals.

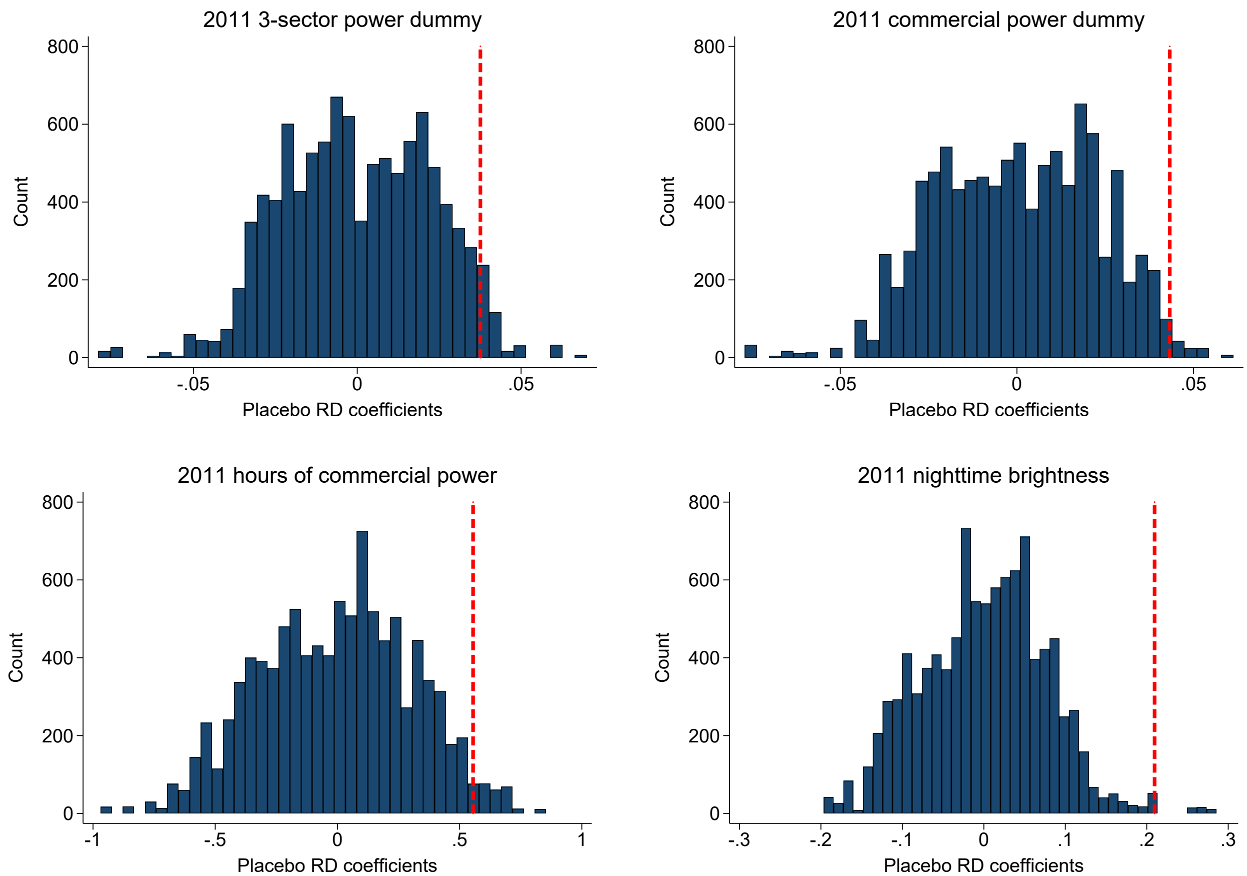
Figure B4: Falsification test for village-level RD in nighttime brightness



Note. — This figure conducts three falsification tests on our main nighttime brightness RD estimates. The "falsification" samples are identical to those in Figure B3; see notes under Figure B3 for further detail. Whiskers display 95% confidence intervals.

Finally, we draw 10,000 placebo thresholds from a uniform distribution $U \sim [151, 275] \cup [325, 1000]$.[11] For each placebo threshold, we estimate our `rdrobust` specification using four first-stage outcome variables main. Figure B5 reveals that our true first-stage RD point estimates (red dashed lines) are each above the 96th percentile of their respective sampling distributions. This provides further evidence that our RD results are not spurious.

Figure B5: Placebo tests for first-stage village-level RD estimates



Note. — Each panel reports the density of 10,000 placebo RD estimates, which are otherwise identical to our preferred specification from Table 2. We randomly generate 10,000 placebo cutoffs between 151 and 1000, excluding cutoffs between 275 and 325 (close to the true 300-person cutoff). The true RDs estimates (red dashed lines) fall above the 96th, 98th, 97th, and 99th percentiles of the placebo distributions for the 3-sector dummy, the commercial dummy, commercial hours of power, and 2011 nighttime brightness, respectively.

11. We omit (275,325) placebo thresholds to avoid possible contamination with the true 300-person RD effect.

B.2 Reduced-form RD

Tables B1–B3 repeat the above `rdrobust` kernel and bandwidth sensitivity analyses using our reduced-form RD estimates from Table 4. Table B4 adds 10 states with unreliable or missing village shapefiles to our RD sample: Arunachal Pradesh, Assam, Himachal Pradesh, Jammu and Kashmir, Meghalaya, Mizoram, Nagaland, Sikkim, Uttarakhand, and Uttar Pradesh (see Table C3). Our results are largely unchanged, with smaller upper confidence intervals on expenditure per capita and population; while sectoral labor shifts gain statistical significance, their confidence intervals change only slightly.

Table B1: Village-level RD – reduced-form outcomes, Epanechnikov kernel

	RD estimate	Std error	95% CI	Mean Y_v
A. Consumption and income (2011)				
Expenditure per capita (Rs/month)	−4.374	(18.174)	[−39.995, 31.246]	1366.071
Expenditure per capita (logged)	−0.009	(0.013)	[−0.035, 0.016]	9.669
Share HH with poverty indicator	−0.005	(0.013)	[−0.031, 0.022]	0.546
Share HH rely on cultivation income	−0.008	(0.012)	[−0.032, 0.015]	0.420
Share HH earning > Rs 5k/mth	0.004	(0.010)	[−0.015, 0.023]	0.069
Share HH with salaried job	0.003	(0.004)	[−0.003, 0.010]	0.011
B. Village demographics (2011)				
Population	6.162	(3.911)	[−1.503, 13.826]	301.939
Share population age 0–6	0.001	(0.002)	[−0.003, 0.004]	0.141
Average household size	0.027	(0.024)	[−0.021, 0.074]	4.911
C. Workers as share of population (2011)				
Ag workers, total	−0.007	(0.007)	[−0.020, 0.006]	0.399
Ag workers, male	−0.008	(0.006)	[−0.019, 0.004]	0.466
Ag workers, female	−0.005	(0.009)	[−0.024, 0.013]	0.329
Non-ag workers, total	0.004	(0.003)	[−0.003, 0.010]	0.075
Non-ag workers, male	0.004	(0.004)	[−0.005, 0.013]	0.095
Non-ag workers, female	0.005	(0.004)	[−0.004, 0.013]	0.052
D. Firm outcomes (2013)				
Number of firms	0.810	(0.720)	[−0.600, 2.220]	8.187
Number of firm employees	−1.827	(4.651)	[−10.942, 7.289]	16.459
E. School outcomes (2014–15 school year)				
# students enrolled, grades 1–5	5.327	(3.905)	[−2.328, 12.981]	47.490
# students enrolled, grades 6–8	−1.898	(2.535)	[−6.867, 3.070]	10.266
# students passed, grades 4–5	−0.391	(0.532)	[−1.434, 0.651]	5.226
# students passed, grades 7–8	−0.566	(0.448)	[−1.444, 0.313]	1.501

Note. — RD estimates in this table are identical to those in Table 4, except that here we apply RD robust with an Epanechnikov kernel instead of our preferred triangular kernel. Optimal bandwidths in the table range from 65 to 120 above/below 300 people. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B2: Village-level RD – reduced-form outcomes, uniform kernel

	RD estimate	Std error	95% CI	Mean Y_v
A. Consumption and income (2011)				
Expenditure per capita (Rs/month)	−2.116	(19.099)	[−39.549, 35.317]	1367.701
Expenditure per capita (logged)	−0.011	(0.014)	[−0.038, 0.016]	9.671
Share HH with poverty indicator	−0.008	(0.014)	[−0.037, 0.020]	0.544
Share HH rely on cultivation income	−0.016	(0.015)	[−0.044, 0.013]	0.422
Share HH earning > Rs 5k/mth	0.011	(0.009)	[−0.007, 0.029]	0.071
Share HH with salaried job	0.004	(0.003)	[−0.002, 0.011]	0.011
B. Village demographics (2011)				
Population	3.795	(3.844)	[−3.740, 11.330]	304.748
Share population age 0–6	0.001	(0.002)	[−0.002, 0.004]	0.141
Average household size	0.015	(0.023)	[−0.029, 0.060]	4.906
C. Workers as share of population (2011)				
Ag workers, total	−0.009	(0.007)	[−0.023, 0.005]	0.398
Ag workers, male	−0.010	(0.006)	[−0.022, 0.003]	0.465
Ag workers, female	−0.009	(0.009)	[−0.027, 0.009]	0.328
Non-ag workers, total	0.003	(0.004)	[−0.003, 0.010]	0.074
Non-ag workers, male	0.005	(0.005)	[−0.005, 0.015]	0.095
Non-ag workers, female	0.004	(0.004)	[−0.005, 0.012]	0.052
D. Firm outcomes (2013)				
Number of firms	0.593	(0.732)	[−0.842, 2.028]	8.440
Number of firm employees	−1.511	(4.951)	[−11.215, 8.192]	17.794
E. School outcomes (2014–15 school year)				
# students enrolled, grades 1–5	6.686	(4.951)	[−3.018, 16.390]	48.552
# students enrolled, grades 6–8	−2.623	(2.651)	[−7.820, 2.574]	11.144
# students passed, grades 4–5	−0.284	(0.561)	[−1.384, 0.815]	5.323
# students passed, grades 7–8	−0.619	(0.500)	[−1.599, 0.361]	1.614

Note. — RD estimates in this table are identical to those in Table 4, except that here we apply RD robust with a uniform kernel instead of our preferred triangular kernel. Optimal bandwidths in the table range from 49 to 95 above/below 300 people. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B3: Village-level RD – reduced-form outcomes, CER bandwidth

	RD estimate	Std error	95% CI	Mean Y_v
A. Consumption and income (2011)				
Expenditure per capita (Rs/month)	−0.408	(20.266)	$[-40.128, 39.313]$	1369.979
Expenditure per capita (logged)	−0.005	(0.015)	$[-0.034, 0.023]$	9.673
Share HH with poverty indicator	−0.004	(0.015)	$[-0.033, 0.025]$	0.545
Share HH rely on cultivation income	−0.014	(0.013)	$[-0.040, 0.012]$	0.419
Share HH earning > Rs 5k/mth	−0.009	(0.010)	$[-0.030, 0.011]$	0.070
Share HH with salaried job	0.004	(0.003)	$[-0.003, 0.010]$	0.010
B. Village demographics (2011)				
Population	7.559*	(4.523)	$[-1.305, 16.423]$	323.320
Share population age 0–6	−0.000	(0.002)	$[-0.004, 0.004]$	0.141
Average household size	0.014	(0.029)	$[-0.042, 0.070]$	4.906
C. Workers as share of population (2011)				
Ag workers, total	−0.006	(0.007)	$[-0.020, 0.009]$	0.398
Ag workers, male	−0.007	(0.006)	$[-0.019, 0.006]$	0.465
Ag workers, female	−0.006	(0.011)	$[-0.026, 0.015]$	0.326
Non-ag workers, total	0.005	(0.004)	$[-0.003, 0.013]$	0.074
Non-ag workers, male	0.004	(0.005)	$[-0.006, 0.014]$	0.095
Non-ag workers, female	0.006	(0.005)	$[-0.004, 0.016]$	0.052
D. Firm outcomes (2013)				
Number of firms	0.929	(0.854)	$[-0.745, 2.603]$	8.732
Number of firm employees	−3.907	(5.659)	$[-14.998, 7.184]$	17.850
E. School outcomes (2014–15 school year)				
# students enrolled, grades 1–5	5.196	(4.255)	$[-3.143, 13.536]$	48.117
# students enrolled, grades 6–8	−0.569	(2.629)	$[-5.722, 4.583]$	11.165
# students passed, grades 4–5	−0.183	(0.533)	$[-1.228, 0.862]$	5.438
# students passed, grades 7–8	−0.354	(0.425)	$[-1.186, 0.478]$	1.660

Note. — RD estimates in this table are identical to those in Table 4, except that here we apply RD robust using the CER-optimal bandwidth selection algorithm (`cerrd`) instead of our preferred MSE-optimal algorithm (`mserd`). Optimal bandwidths in the table range from 41 to 78 above/below 300 people. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B4: Village-level RD – reduced-form outcomes, including unreliable/missing shapefiles

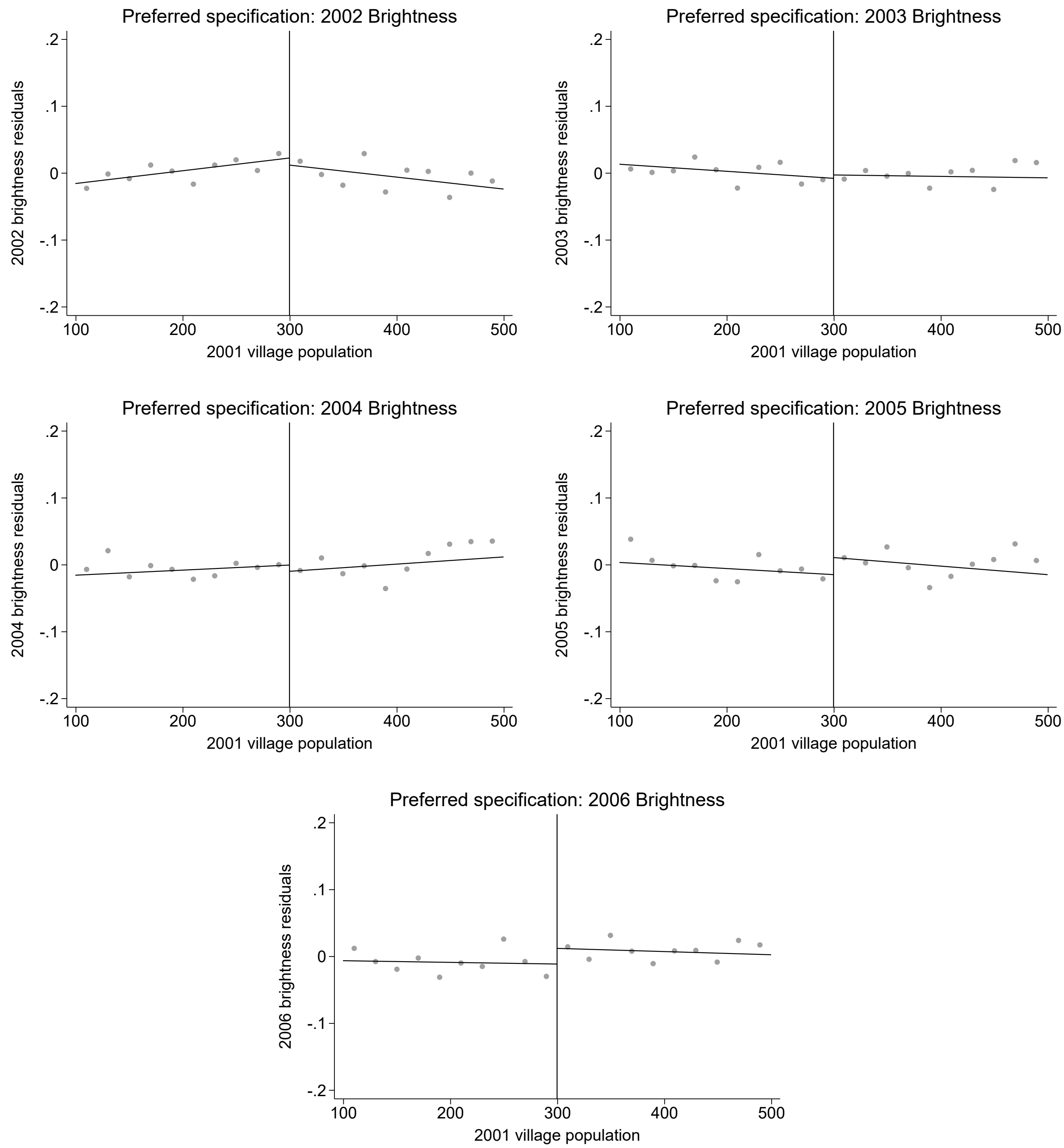
	RD estimate	Std error	95% CI	Mean Y_v
A. Consumption and income (2011)				
Expenditure per capita (Rs/month)	−3.933	(13.352)	[−30.103, 22.236]	1403.729
Expenditure per capita (logged)	−0.004	(0.009)	[−0.022, 0.013]	9.694
Share HH with poverty indicator	−0.003	(0.011)	[−0.025, 0.019]	0.479
Share HH rely on cultivation income	−0.015	(0.013)	[−0.039, 0.010]	0.442
Share HH earning > Rs 5k/mth	0.002	(0.008)	[−0.013, 0.018]	0.080
Share HH with salaried job	0.003	(0.004)	[−0.004, 0.010]	0.021
B. Village demographics (2011)				
Population	3.382	(3.928)	[−4.316, 11.081]	309.793
Share population age 0–6	0.000	(0.001)	[−0.002, 0.003]	0.145
Average household size	0.019	(0.021)	[−0.022, 0.060]	5.148
C. Workers as share of population (2011)				
Ag workers, total	−0.010**	(0.005)	[−0.020, −0.001]	0.358
Ag workers, male	−0.011**	(0.005)	[−0.020, −0.002]	0.427
Ag workers, female	−0.009	(0.007)	[−0.022, 0.004]	0.284
Non-ag workers, total	0.005*	(0.003)	[−0.000, 0.011]	0.074
Non-ag workers, male	0.005	(0.004)	[−0.002, 0.012]	0.099
Non-ag workers, female	0.007*	(0.004)	[−0.000, 0.014]	0.048
D. Firm outcomes (2013)				
Number of firms	0.454	(0.538)	[−0.600, 1.508]	7.677
Number of firm employees	−1.400	(3.865)	[−8.976, 6.176]	15.701
E. School outcomes (2014–15 school year)				
# students enrolled, grades 1–5	1.480	(2.647)	[−3.708, 6.669]	37.410
# students enrolled, grades 6–8	−0.639	(1.854)	[−4.272, 2.995]	9.620
# students passed, grades 4–5	−0.354	(0.380)	[−1.099, 0.391]	4.208
# students passed, grades 7–8	−0.398	(0.331)	[−1.047, 0.252]	1.406

Note. — RD estimates in this table are identical to those in Table 4, except that here, we (i) expand our estimation sample to include 10 states with unreliable (or missing) village shapefiles, and (ii) do not control for pre-RGGVY village-level brightnness. Optimal bandwidths in the table range from 65 to 134 above/below 300 people. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

B.3 Smoothness of predetermined covariates

Tables B5–B6 and Figures B6–B7 provide evidence that pre-RGGVY village covariates were smooth at the 300-person eligibility cutoff. We find no statistically significant discontinuities in nighttime brightness prior to RGGVY's 2005 announcement. We also recover null RD estimates for all four electricity power access dummies in the 2001 Census (see RD plots in Figure B7). While we find RD estimates that are statistically distinguishable from zero at the 10% level for four 2001 covariates, these estimates are not robust to alternative kernels and fixed effects.

Figure B6: Village-level nighttime brightness RDs by year, 2002–2006



Note. — RD plots for nighttime brightness by year, which correspond to Figure 5 and Table B5. As in Figure A5, we standardize the RD bandwidths and vertical axes of these brightness plots to facilitate visual comparisons across years. See notes under Figure 5 for details.

Table B5: Village-level RD in nighttime brightness, pre-RGGVY

	Outcome: Maximum village brightness by year				
	2002	2003	2004	2005	2006
	(1)	(2)	(3)	(4)	(5)
$\mathbf{1}$[2001 pop $\geq$ 300]	−0.011	0.009	−0.003	0.039	0.045
	(0.026)	(0.022)	(0.022)	(0.026)	(0.030)
Mean brightness (< 300)	4.443	3.609	3.919	3.512	3.860
Optimal bandwidth	143	129	136	99	88
Village observations	17,754	16,047	16,907	12,409	11,050

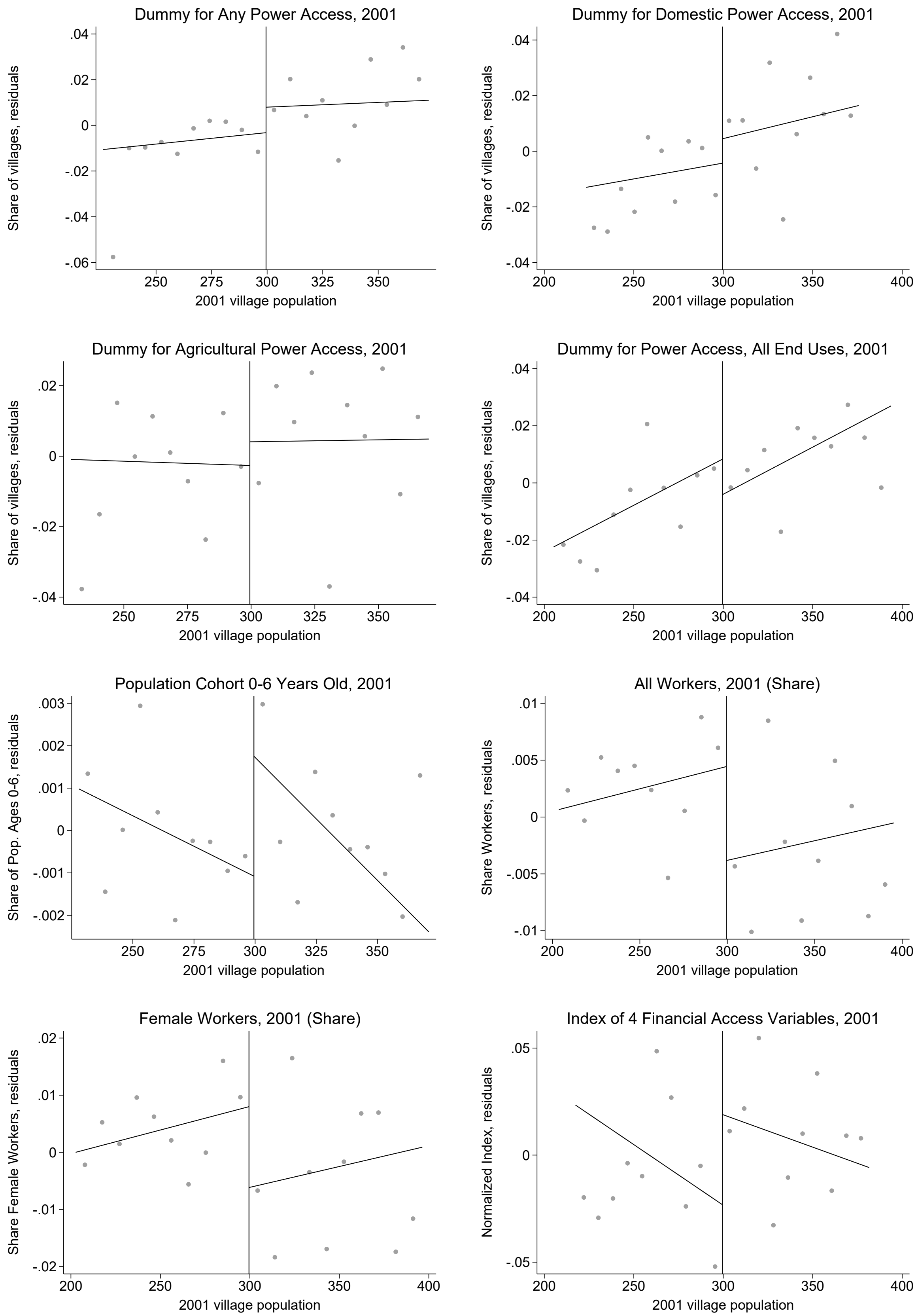
Note. — `rdrobust` specifications are identical to those in Table 2. Results are robust to alternative controls, kernels, bandwidth algorithms, and standard errors. See notes under Figure 5 for details. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B6: Village-level RD – smoothness in pre-determined covariates

	RD estimate	Std error	95% CI	Mean Y_v
A. Electricity access in village (2001)				
1/0 anywhere in village	0.017	(0.019)	[−0.020, 0.055]	0.657
1/0 domestic sector	0.014	(0.019)	[−0.022, 0.051]	0.647
1/0 agricultural sector	0.011	(0.017)	[−0.022, 0.045]	0.524
1/0 all sectors	−0.013	(0.014)	[−0.041, 0.015]	0.409
B. Village demographics (2001)				
Share population age 0–6	0.004*	(0.002)	[−0.001, 0.008]	0.171
Literacy rate	−0.007	(0.008)	[−0.022, 0.009]	0.447
Share Scheduled Caste/Scheduled Tribe	0.017	(0.019)	[−0.021, 0.055]	0.389
C. Workers as share of population (2001)				
All workers	−0.010*	(0.005)	[−0.020, 0.001]	0.479
Male workers	−0.002	(0.004)	[−0.010, 0.005]	0.555
Female workers	−0.017*	(0.009)	[−0.034, 0.001]	0.398
Ag workers, total	−0.011	(0.007)	[−0.024, 0.002]	0.403
Non-ag workers, total	0.006	(0.004)	[−0.002, 0.013]	0.063
D. Village characteristics (2001)				
1/0 village has drinking water	−0.001	(0.004)	[−0.009, 0.008]	0.995
1/0 village has tubewell	−0.008	(0.021)	[−0.048, 0.033]	0.588
Share of village land irrigated	−0.002	(0.012)	[−0.026, 0.021]	0.283
Distance to nearest town (km)	−0.903	(0.887)	[−2.642, 0.836]	23.980
E. Indexes of village amenities (2001)				
Schooling access (7 variables)	0.001	(0.016)	[−0.030, 0.032]	−0.074
Health services access (8 variables)	0.009	(0.016)	[−0.022, 0.041]	−0.034
Financial services access (4 variables)	0.053*	(0.029)	[−0.005, 0.110]	−0.051
Communication access (4 variables)	0.025	(0.023)	[−0.021, 0.070]	−0.105
Transportation access (4 variables)	0.028	(0.030)	[−0.031, 0.087]	−0.177

Note. — This table is analogous to Tables 4 and A22, using 2001 Census covariates as dependent variables. RD robust regressions control only for state fixed effects, but are otherwise identical to our preferred specification. Of the four covariates with weak statistical significance, three are not robust to alternative kernels and fixed effects (share population age 0–6, all workers, and financial services access). Optimal bandwidths in the table range from 64 to 135 above/below 300 people. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure B7: Village-level RDs in pre-determined covariates

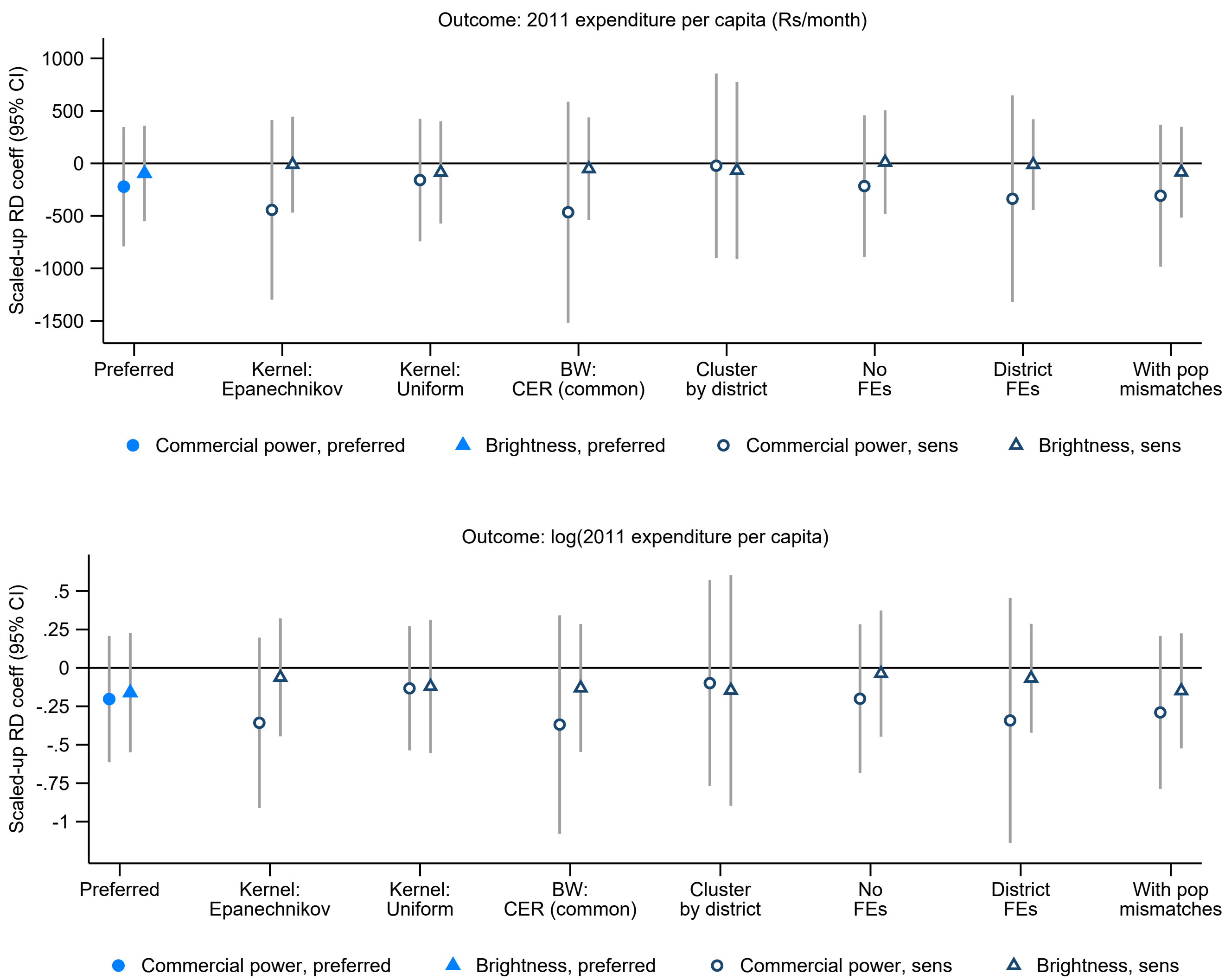


Note. — These RD plots correspond to eight regressions in Table B6: the four electricity access dummies in the 2001 Census, and the four pre-determined covariates with weakly statistical significant RD estimates. See table notes for details.

B.4 Fuzzy RD

Figure B8 presents sensitivity analyses for our fuzzy RD expenditure per capita estimates, applying the scaling factors of 10 hours of power and 2.6 brightness units in order to facilitate comparisons across both "treatment" variables. We can reject large increases in expenditure per capita across most sensitivities, except when clustering by district (which is likely overly conservative). We describe each sensitivity analysis above in Appendix B.2.

Figure B8: Sensitivities for village-level fuzzy RD in expenditure per capita



Note. — Each panel reports fuzzy RD sensitivity analysis for expenditure outcomes in Table 6. Circles plot fuzzy RD point estimates using hours of commercial power as the endogenous "treatment" variable, scaled up by a factor of 10. Triangles plot fuzzy RD point estimates using nighttime brightness as the endogenous "treatment" variable, scaled up by a factor of 2.6. Whiskers denote (scaled-up) 95% confidence intervals. Solid markers reproduce our preferred estimates from Table 6, while hollow markers report sensitivities. First, we compare Epanechnikov and uniform weighting kernels to our preferred triangular kernel. Next, we compare our preferred MSE-optimal bandwidth selection algorithm (`mserd`) with the coverage error rate optimal bandwidth algorithm (`cerrd`). Next, we test alternative standard errors, clustering by district. Next, we remove state fixed effects, and then add district fixed effects. Finally, we add in the 11% of villages with population discrepancies in the merged habitation data.

B.5 Difference-in-differences

In the main text, we report NSS differential pre-trend estimates in Panel B of Table 1. We present these full pre-trend regression results here, in Tables B7–B8. We find no evidence of differential pre-trends in household electricity consumption, expenditure per capita, or appliance ownership (except TVs). If anything, 10th-Plan districts exhibit negative pre-trends for these NSS electricity indicators, which would bias against finding first-stage DD effects.

We use randomization inference to assess the likelihood of recovering our first-stage NSS results if RGGVY 10th-Plan districts were assigned at random, by reestimating Equation (2) 10,000 times with scrambled DD "treatment" vectors. Figure B9 reports the resulting distributions of randomized DD coefficients. The true first-stage estimates (from Columns (1)–(3) of Table 3) all fall above the 99.8th percentile of their respective randomized DD distributions, as indicated by the red dashed lines. In fact, the true first-stage DD estimate for the extensive margin of household electricity consumption is greater than the *maximum* of its randomized distribution.

Table B7: Pre-trends for district-level DD in household electricity access and usage

	HH elec use (kWh/month)			$\mathbf{1}$[HH owns electric appliance]				
	$\mathbf{1}[Q > 0]$	Levels	Logs	Lighting	Fan	TV	Fridge	AC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbf{1}$[10th-Plan]	−0.015	−2.06	−0.171	−0.018	−0.019	−0.025**	−0.004	−0.006
× $\mathbf{1}$[2005]	(0.016)	(1.37)	(0.111)	(0.016)	(0.017)	(0.012)	(0.004)	(0.005)
Mean of dep var	0.529	27.00	2.871	0.545	0.309	0.222	0.039	0.028
Clusters	539	539	535	539	539	539	539	539
Observations	1096	1096	1086	1096	1096	1096	1096	1096

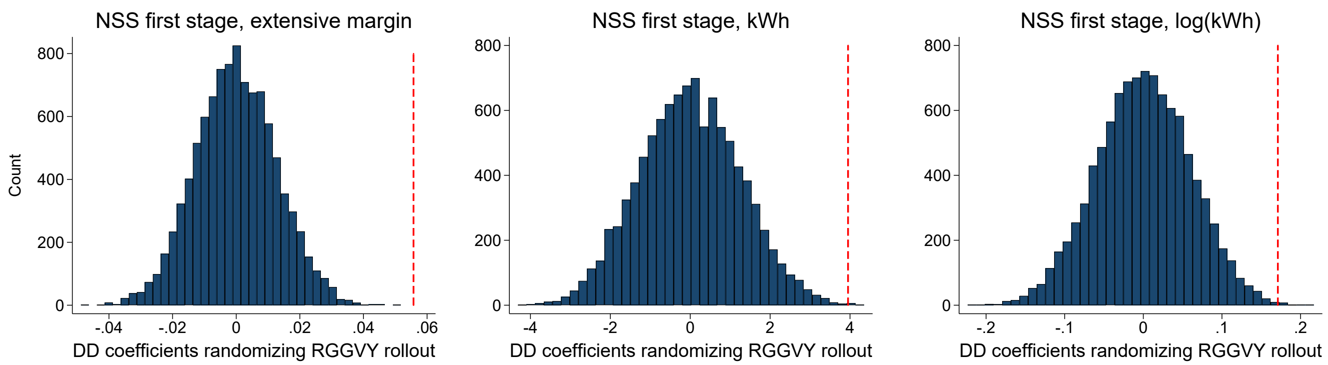
Note. — District-level DD with two NSS years (2000, 2005), using 2005 (rather than 2010) as a "post" period to test for differential pre-treatment trends. Regressions are otherwise identical to those in Table 3. All regressions include: district fixed effects; year fixed effects; state-specific linear trends; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B8: Pre-trends for district-level DD of household consumption expenditures

	Expenditure per capita (Rs/month)	
	Levels	Logs
	(1)	(2)
$\mathbf{1}$[10th-Plan] × $\mathbf{1}$[2005]	16.19	0.019
	(20.44)	(0.020)
Mean of dep var	952.51	6.813
Clusters	539	539
Observations	1096	1096

Note. — District-level DD with two NSS years (2000, 2005), using 2005 (rather than 2010) as a "post" period to test for differential pre-treatment trends. Regressions are otherwise identical to those in Table 5. Expenditures per capita are denominated in 2010 rupees, and net out per capita spending on electricity. All regressions include: district fixed effects; year fixed effects; state-specific linear trends; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure B9: Randomization tests for household electricity consumption



Note. — Each panel reports the density of 10,000 randomized DD estimates, where each iteration re-estimates our main DD specification after randomly permuting the set of 10th- vs. 11th-Plan RGGVY districts. From left to right, panels correspond to Columns (1)–(3) of Table 3. The true DD estimates (red dashed lines) each fall above the 99.8th percentile of their respective randomized distributions.

Table B9: District-level DD – RGGVY single-Plan districts only

	HH elec use (kWh/month)			Exp. per capita (Rs/month)	
	$\mathbf{1}[Q > 0]$	Levels	Logs	Levels	Logs
	(1)	(2)	(3)	(4)	(5)
$\mathbf{1}$[10th-Plan district] $\times$ $\mathbf{1}$[2010]	0.062***	5.00***	0.216**	33.29	0.033
	(0.015)	(1.92)	(0.084)	(26.48)	(0.024)
Mean of dep var	0.589	31.57	3.039	972.71	6.830
Clusters	488	488	486	488	488
Observations	1470	1470	1461	1470	1470

Note. — These regressions remove 30 districts from the "treated" group that had RGGVY projects in both the 10th and 11th Plans. They also remove 37 districts from the "control" group that had no RGGVY projects under either Plan. This leaves two "pure" groups of 199 10th-Plan-only "treated" districts and 297 11th-Plan-only "control" districts. Regressions are otherwise identical to Columns (1)–(3) of Table 3, and Columns (1)–(2) of Table 5. All regressions include: district fixed effects; year fixed effects; state-specific linear trends; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B9 re-estimates our NSS regressions using only the subset of districts exclusively assign to either RGGVY 10th-Plan funding or RGGVY 11th-Plan funding. This removes 30 (potentially contaminated) "treated" districts that received funding under both Plans, and 37 (potentially poor) "control" districts that never received RGGVY funding. The resulting first-stage estimates increase slightly in magnitude, while the reduced-form estimates remain statistical zeros.

Our preferred DD-IV estimates in Table 7 do not include state-specific time trends, since doing so lowers the first-stage F-statistic for the Q1 and Q25 subsamples. Table B10 replicates Table 7 adding state-specific time trends. The resulting first-stage F-statistics are larger for the pooled samples (Columns (1) and (4)), and the upper confidence intervals similar to those in Table 7. While the F-statistics are much lower for the Q1 subsample when we include state-specific trends (Columns (2) and (5)), the resulting point estimates are similar (albeit less precise).

Table B11 presents the un-instrumented OLS versions of the DD-IV regressions in Table 7. In the presence of weak instrument bias, we should expect our two-stage least squares estimates to be biased in the direction of these (endogenous) OLS coefficients. Table B11 suggests that if anything, our DD-IV results are biased towards finding positive effects of electrification on expenditure, given the positive district-level correlations between household electricity access and expenditure per capita. Because our first stages are weakest for the Q25 subsample, this may help explain why these estimate are so large.

Table B10: District-level DD-IV adding state-specific linear trends

	Expenditure per capita (Rs/month)					
	Levels			Logs		
	(1)	(2)	(3)	(4)	(5)	(6)
1[HH consumes any elec]	492.8	−657.0	1339.4	0.528	−0.103	0.971
	(455.1)	(634.2)	(825.4)	(0.413)	(0.411)	(0.670)
95% confidence	[−401.1, 1386.7]	[−1909.5, 595.4]	[−282.3, 2961.2]	[−0.283, 1.338]	[−0.914, 0.707]	[−0.346, 2.288]
Village weight quintiles	Pooled	1	2–5	Pooled	1	2–5
50th pctile of 2001 pop	1913	1043	2076	1913	1043	2076
90th pctile of 2001 pop	6854	4875	7291	6854	4875	7291
Mean of dep var	978.2	1128.2	948.2	6.833	6.957	6.804
Clusters	552	162	494	552	162	494
Observations	1670	418	1488	1670	418	1488
First-stage estimate	0.056***	0.173***	0.044***	0.056***	0.173***	0.044***
(standard error)	(0.014)	(0.059)	(0.015)	(0.014)	(0.059)	(0.015)
First-stage F-statistic	15.71	8.70	8.17	15.70	8.70	8.17

Note. — This table is identical to Table 7, except that it adds state-specific linear time trends, which reduce our first-stage F-statistics for the Q1 and Q25 subsamples. All regressions include: district fixed effects; year fixed effects; state-specific linear trends; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. The bottom three rows report the first-stage point estimates and standard errors, along with Kleibergen-Paap first-stage F-statistics. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B11: District-level DD household consumption expenditures, without instrumenting

	Expenditure per capita (Rs/month)					
	Levels			Logs		
	(1)	(2)	(3)	(4)	(5)	(6)
1[HH consumes any elec]	131.0**	32.1	147.3**	0.129**	0.113	0.135**
	(65.4)	(205.7)	(64.9)	(0.051)	(0.136)	(0.054)
95% confidence	[2.5, 259.6]	[−374.1, 438.4]	[19.9, 274.8]	[0.028, 0.229]	[−0.155, 0.382]	[0.028, 0.242]
Village weight quintiles	Pooled	1	2–5	Pooled	1	2–5
Mean of dep var	978.2	1128.2	948.2	6.833	6.957	6.804
Clusters	552	162	494	552	162	494
Observations	1670	418	1488	1670	418	1488

Note. — Regressions are identical to those in Table 7, except that estimate (endogenous) OLS models rather than instrumenting with the interaction $\mathbf{1}$[10th-Plan district] $\times$ $\mathbf{1}$[2010]. All regressions include: district fixed effects; year fixed effects; linear trends in state quartiles of 2005 household expenditures per capita; and linear trends in national deciles of 2005 household expenditures per capita. Standard errors are clustered at the district level, collapsing to a single cluster for (i) districts that split in the 2001 Census, but which the 2000 NSS sampled based on 1991 district definitions; and (ii) the few cases where an RGGVY DPR included multiple districts. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

C Data

C.1 RGGVY program data

The Rural Electrification Corporation published an online database of RGGVY implementation status, with separate portals for "Villages Covered" and "Villages Completed" under the 10th and 11th Plans.[12] While these portals reported village-specific details on RGGVY's proposed and actual implementation, neither portal reported the dates on which RGGVY project work was sanctioned, begun, or completed. These data are also rife with missing information, and internal inconsistencies that conflict with village-level Census data. As we show in Table C1, RGGVY program outcomes are missing for 65% of villages that were eligible under the 10th Plan. RGGVY data also report a greater number of covered habitations than exist for 32% of villages.[13]

Table C1: RGGVY microdata irregularities

Type of Irregularity	Percent of Villages	Percent of 10th-Plan Villages
RGGVY outcomes disagree across Covered and Completed datasets	26.8	32.7
Outcomes missing from either Covered or Completed dataset	77.9	65.3
Outcomes missing from both Covered and Completed datasets	74.4	59.9
All outcomes missing from both Covered and Completed datasets	33.4	22.3
Completed dataset reports status not energised	24.4	14.1
RGGVY covers more habitations than exist in village	32.2	31.7

Note. — This table shows data irregularities across the RGGVY Covered and Completed village datasets, which we do not use in our analysis. We report the percent of all villages and of 10th-Plan villages that satisfy each irregularity criterion, where the denominator excludes missing and unmatched villages. Program outcomes considered in the first four rows include the count of household connections, aggregate transformer capacity installed, and aggregate transmission capacity installed. (The first three rows count villages where *any* outcome disagrees or is missing; the fourth row counts only villages for which *all* of these outcomes are missing.)

Even if these RGGVY microdata precisely identified which villages were treated, at which point in time, we would still need to construct a control group from the subset of villages *not* included in RGGVY's ledgers. Any imperfect merge or missing RGGVY microdata would cause us to misclassify a village. We might also worry about manipulation of RGGVY village-level outcomes, if implementing agencies had an incentive to overstate realized program outcomes.[14]For these reasons, we exclude the RGGVY village-level data from our analysis entirely.

Instead, we rely on RGGVY's district-level Detailed Project Reports (DPRs) to determine (i) the Five-Year Plan under which each electrification project was sanctioned, (ii) the implementing agency responsible for implementation, and (iii) the approximate timing of electrification.[15] As this is a matter of public record and involves large transfers of public funds from the Rural Electrification Corporation to decentralized implementing agencies, we are much more confident in the accuracy of this aggregate information.

12. We downloaded these data in 2014 from http://rggvy.gov.in, which has since been deactivated.

13. We use total habitation counts from our merge of the Habitation Census to the village-level Census panel, considering only RGGVY villages that match to a *matched* village from the Habitation-to-Census merge (see below).

14. Asher and Novosad (2020) document striking irregularities in administrative population data from India's flagship road construction program.

15. Most districts were covered by exactly one DPR. For districts with multiple DPRs, we have aggregated DPRs up to the district-level, to create a one-to-one mapping between districts and DPRs. Of the 530 RGGVY districts, only 30 districts had DPRs under both the 10th and 11th Plans.

Table C2: Summary statistics – RGGVY implementation and scope

Type of Implementing Agency	States	Districts	Award Dates	Unelectrified Villages	Electrified Villages	BPL Connections
A. 10th Plan						
Public Sector Undertakings	11	57	2005–07	32,637	19,740	2,363,435
State Departments of Power	5	10	2007–10	519	972	31,044
State Electricity Boards	3	6	2006–07	4,216	1,785	358,766
Distribution Companies	14	156	2005–08	26,399	76,044	4,587,207
Total	25	229		63,771	98,541	7,340,452
B. 11th Plan						
Public Sector Undertakings	9	78	2008–11	29,974	59,833	6,331,673
State Departments of Power	5	33	2008–10	2,355	2,896	106,673
State Electricity Boards	3	34	2008–11	76	16,319	143,057
Distribution Companies	15	183	2008–10	12,482	128,246	6,567,015
Rural Electricity Coops	2	5	2008–11	0	736	75,368
Total	25	331		44,887	208,030	13,223,786

Note. — This table summarizes RGGVY program outcomes at the DPR (district) level. Public sector undertakings include government-owned generating companies, such as Power Grid Corporation of India and National Hydroelectric Power Corporation. The right three columns show the number of previously unelectrified and previously electrified villages treated by the program, as well as the the number of below poverty line households that received electric connections. Villages classified as electrified had basic electricity infrastructure with at least 10% of households electrified prior to RGGVY implementation. 23 (of 27) states contain both 10th and 11th Plan districts, while 30 (of 530) individual districts were targeted under both Plans. For a few districts, we correct financial award dates reported to have occurred before their respective project sanction dates or before the official announcement of the program.

Table C2 summarizes electrification achieved under RGGVY, at the DPR level.[16] Under both the 10th and 11th Plans, most DPRs were implemented by local electricity distribution companies. However, there were also many districts whose implementing agencies included large public sector undertakings, state departments of power, and state electricity boards. Most villages treated under both Plans were categorized as "electrified" villages, meaning that power access existed prior to RGGVY. RGGVY's 10th-Plan projects provided grid connections to 7.3 million below-poverty-line (BPL) households—nearly 10% of the *total* number of rural households in 10th-Plan districts, and nearly 1 in 5 previously unelectrified households in these districts.[17]

Table C2 highlights the staggered nature of RGGVY's rollout, which is key to our DD identification strategy. Nearly all 10th-Plan funds were awarded prior to the earliest 11th-Plan funds in January 2008. Our nighttime brightness RDs (Figure 5) show positive effects starting in 2008, roughly two years after most 10th-Plan funds were awarded. Similarly, our brightness event studies (Figure 6) detect effects starting two years after receipt of funds. Together, these patterns suggest that we should not expect 11th-Plan DPRs to cause meaningful increases in electrification prior to 2011, which assuages concerns about contamination of 11th-Plan "control" districts in 2010 NSS data.

16. The state of Goa was excluded from RGGVY along with all 7 union territories, because 100% of their villages were electrified prior to 2005 (Ministry of Power (2012)). We treat Telangana as part of Andhra Pradesh, since its 2014 split from Andhra Pradesh occurred after our period of analysis.

17. This aligns with statistics reported by Sreekumar and Dixit (2011) on the extent of RGGVY's household electrification outcomes, and ignores newly connected above-poverty-line households.

C.2 Geospatial data

Our main source of geospatial information is ML Infomap's VillageMap. These data include village boundary shapefiles for nearly every village in India, as defined by the 2001 Census.[18] We take 2001 boundaries as fixed, which conforms with our decision to use the 2001 village as our unit of analysis.[19] Every square meter in India belongs to a village or city/town (urban area); the only "blank spaces" in the village maps are forests, bodies of water, etc. Village boundaries are set by the Census Organization of India (Census of India (2011)).

We were unable to acquire village shapefiles for Arunachal Pradesh, Meghalaya, Mizoram, Nagaland, and Sikkim. This forces us to drop these 5 small states from our geospatial dataset. To maintain an internally consistent sample, we exclude these states from our entire RD analysis. Together, they represent less than 1% of India's rural population, and only 1% of villages covered by RGGVY 10th-Plan DPRs.

We also exclude Assam, Himachal Pradesh, Jammu and Kashmir, Uttarakhand, and Uttar Pradesh, due to the extremely poor quality of our their village shapefiles. These 5 states have correlations less than 0.35 between the village areas derived from shapefile polygons and village areas reported in the 2001 Census (the official body that defines village boundaries). Table C3 reports these correlations by state, demonstrating a clear gap between 12 "correlated" vs. 5 "uncorrelated" states. "Uncorrelated" states represent 39% of villages in RGGVY 10th-Plan districts, most of which are in Uttar Pradesh. Pre-RGGVY nighttime brightness is an important cross-sectional control for our RD sample, as it adds significant statistical power. We therefore restrict our RD analysis to the 12 states in Table C3 with village area correlations above 0.35. Our DD analysis, which instead uses village or district fixed effects, does not impose this sample restriction.

C.3 Nighttime brightness

We use the National Oceanic and Atmospheric Administration (NOAA)'s Defense Meteorological Satellite Program-Operational Line Scan (DMSP-OLS) Nighttime Lights data.[20] Descriptions of these data can be found in Elvidge et al. (1997) and Doll (2008).[21] NOAA compiles DMSP-OLS images from nightly satellite photographs taken between 8:30 and 10:00 PM local time; they are extremely high resolution, with pixels that are 30 arc-second squares.[22] Each pixel is assigned a digital number (DN) indicating brightness, ranging from 0 to 63. This DN is approximately proportional to average luminosity (Chen and Nordhaus (2011)). While the images often top-code very bright locations such as urban centers (Henderson, Storeygard, and Weil (2012)), our analysis focuses on rural areas with very low risk of top-coding.

18. We link village polygons to other village-level datasets using 2001 Census identifiers. Min (2011) and Asher and Novosad (2017), among others, also use this data product to map villages in India.

19. In the 2% of cases where a 2001 village matches to multiple 2011 villages in the 2001/2011 Census concordance, we aggregate 2011 data up to the 2001 village definition.

20. The data are available for download here: http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

21. Numerous researchers have also these nighttime brightness data as proxies for economic activity, including: Noor et al. (2008); Bleakley and Lin (2012); Henderson, Storeygard, and Weil (2012); Li, Ge, and Chen (2013); Michalopoulos and Papaioannou (2013); and Michalopoulos and Papaioannou (2014). Given that we study electrification directly, we refrain from using nighttime brightness as a proxy for GDP.

22. These pixels are squares with approximately 1,000-meter sides near the equator. The Indian subcontinent alone contains 417,876 pixels.

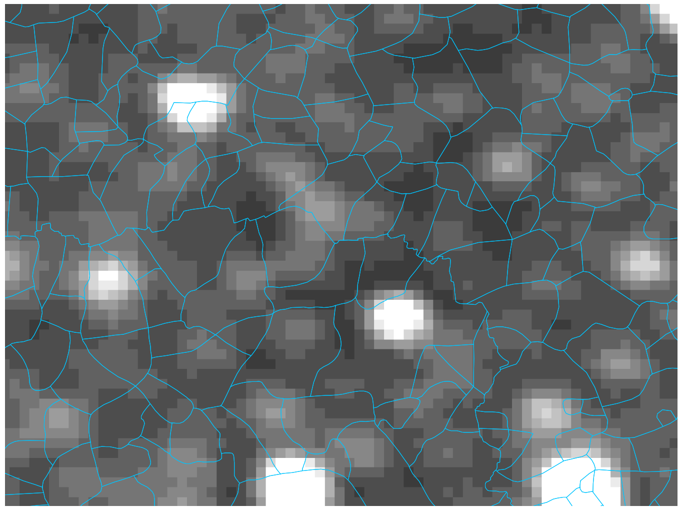
Table C3: Correlation of shapefiles with village areas

State	Area Correlation	Percent of Total Villages	Percent of 10th-Plan Villages
Jharkhand	0.978	5.0	5.9
West Bengal	0.954	6.4	10.8
Bihar	0.932	6.6	9.8
Gujarat	0.896	3.1	0.8
Haryana	0.873	1.1	0.4
Karnataka	0.806	4.6	6.5
Maharashtra	0.781	7.0	1.5
Andhra Pradesh	0.772	4.5	7.2
Rajasthan	0.714	6.8	9.9
Odisha	0.680	8.1	2.5
Madhya Pradesh	0.638	8.9	3.4
Chhattisgarh	0.605	3.3	1.2
Uttarakhand	0.326	2.7	5.3
Uttar Pradesh	0.281	16.6	31.8
Himachal Pradesh	0.138	3.0	0.4
Assam	0.106	4.1	1.0
Jammu and Kashmir	0.002	1.1	0.5
Arunachal Pradesh	missing	0.6	0.3
Meghalaya	missing	0.9	0.3
Mizoram	missing	0.1	0.1
Nagaland	missing	0.2	0.1
Sikkim	missing	0.1	0.1
States with correlation > 0.35		68.3	59.9
States with correlation < 0.35		28.8	39.0
States with missing shapefiles		2.9	1.0

Note. — This table reports the correlation between polygon areas (calculated from village shapefiles) and village areas reported in the Census's 2001 Village Directory. Our RD analysis includes the 12 states for which this correlation is at least 0.35. We omit the 5 states with shapefile areas that are uncorrelated with reported village areas, a sign of low quality shapefiles. The middle column reports the percent of Indian villages contained in each state; the right column use total villages in RGGVY 10th-Plan districts as a denominator. This table omits 3 states without any RGGVY 10th-Plan districts (Goa, Punjab, and Tamil Nadu). It also omits 2 states which were eligible under RGGVY's 10th Plan, but contain no single-habitation 10th-Plan villages in our RD bandwidth (Kerala and Tripura).

A substantial body of evidence demonstrates that DMSP-OLS instruments are sensitive enough to detect the subtle changes in brightness associated with rural electrification. Elvidge et al. (1997) and Elvidge et al. (2001) use DMSP-OLS data to estimate electrification rates around the world at the national level. Several studies have mapped nighttime brightness to electrification rates at the sub-national level, including Ebener et al. (2005); Doll, Muller, and Morley (2006); Chand et al. (2009); and Townsend and Bruce (2010). Four papers (at least) use DMSP-OLS data to study electrification in rural villages. Min et al. (2013) show that electrified villages in Senegal and Mali are significantly brighter than their unelectrified counterparts. Min and Gaba (2014) find a strong correlation between nighttime brightness and ground-based measures of both streetlights and household electrification in Vietnam. Machemedze et al. (2017) correlate household electrification with brightness in South Africa. Finally, Dugoua, Kennedy, and Urpelainen (2018) find strong

Figure C1: Example of nighttime brightness with village boundaries



Note. — This image shows a close-up of an image overlaid with village boundaries, for an area in Rajasthan. The $\approx 1\text{km}^2$ pixels in this image range in brightness values from 3 to 38. We construct our nighttime brightness variable L_v^t by assigning each village v the value of its brightest pixel from the composite satellite image in each year t.

correlations between brightness and the number of electrified households in rural Indian villages. Based on these findings, we are confident that DMSP-OLS data are capable of detecting rural electrification in our context. Any activity in rural India bright enough to be visible from space is likely to require electricity.[23] Any bottom coding would lead us to underestimate the effects of RGGVY on nighttime brightness.[24]

NOAA releases three different DMSP-OLS lights products: "average visible lights"; "stable lights"; and "average visible × percent lights". The "average visible lights" dataset contains the annual average DN for each pixel. The "stable lights" dataset is a more heavily processed version of the average visible lights, removing intermittent brightness from fires and other sporadic events. Finally, the "average visible × percent lights" dataset multiplies the average visible DN in a pixel by its observed frequency.[25] All three datasets remove any sunlit hours, glare, cloud cover, forest fires, the aurora phenomena, and other irregularities. We use the "average visible lights" data in our analysis, which are best equipped to detect the low levels of lighting associated with electrifying small villages (Min et al. (2013)). The bottom-right panel of Figure B2 shows that our RD results are similar (though slightly less precise) using the other two DMSP-OLS data products.

We construct village-level brightness by overlaying 2001 village boundaries on top of annual "average visible lights" images (see Figure C1). We assign each village the brightness of its brightest pixel—either the maximum DN value over all pixels with centroids inside the village polygon, or the value of the pixel at the village centroid (for villages too small to contain a pixel centroid). This reflects the typical organizational structure of South Asian villages, with concentrated inhabited regions surrounded by fields. Maximum brightness best captures RGGVY program outcomes (i.e.

23. Two obvious exceptions are agricultural fires and car headlamps. Ephemeral fires do not appear in DMSP-OLS annual composite brightness images (see details in NOAA's DMSP-OLS data description). Roads are sparse in rural India, and high traffic density is extremely unlikely. Traffic-induced brightness would also likely be too erratic to show up in annual average brightness images (Min and Gaba (2014)).

24. Suppose, for example, that the satellite can only detect brightnesses greater than λ, and that at baseline, villages A and B both have brightness of $\lambda - 5$ (bottom-coded as λ). If village B is electrified under RGGVY and now has brightness of $\lambda + 5$, we only observe a difference of 5 units of brightness, rather than the true difference of 10 units.

25. Alam (2013) uses this lights product to study power quality in India.

electrification of households and public places), while avoiding averaging over unlit agricultural land. Our RD results are nearly identical using the spatial mean of brightness across each village polygon (see bottom-right panel of Figure B2).[26] Between 1999 and 2007, NOAA had two satellites operating DMSP-OLS equipment; for these years, we calculate village brightness separately using each annual image, and then take an unweighted average across satellites.[27]

DMSP-OLS sensors tend to degrade over time, and the satellites do not contain on-board calibration equipment. NOAA's Earth Observation Group, which manages DMSP-OLS data, has conducted ex-post calibration in order to align DN values across satellite-years. This process is imperfect and not fully transparent.[28] Most economists who use DMSP-OLS data include satellite or year fixed effects to control for inconsistencies in sensors over time (e.g. Henderson, Storeygard, and Weil (2012)). Our empirical applications either include year fixed effects (in brightness DD regressions) or rely on within-satellite-year cross-sectional variation (in brightness RD regressions).

We undertake an additional procedure to remove measurement error from DMSP-OLS data. After constructing a village-year brightness panel, we linearly project each village's brightness value on its brightness values in adjacent years. We estimate the following OLS regressions for each brightness year, weighting by village area:[29]

$$L_v^t \;=\; \kappa_0 \;+\; \kappa_1 L_v^{t-2} \;+\; \kappa_2 L_v^{t-1} \;+\; \kappa_3 L_v^{t+1} \;+\; \kappa_4 L_v^{t+2} \;+\; \upsilon_v \tag{C1}$$

L_v^t is the GIS-assigned maximum (or mean) brightness of village v in year t. We use the fitted values $\widehat{L}_v^t$ as the outcomes of our RD regressions, which removes year-specific noise and isolates more stable year-on-year brightness changes associated with electrification. The bottom-right panel of Figure B2 compares our preferred "projected" brightness (i.e. $\widehat{L}_v^t$) to both "raw" brightness (i.e. L_v^t) and unweighted averages of brightness across adjacent years. We use "raw" (unprojected) brightness in DD regressions, since year fixed effects also remove measurement error common to all brightness observations in a given year (and are collinear with the $\widehat{\kappa}$ coefficients). Finally, our main RD sample removes the 0.2% of villages with extremely implausible changes in brightness. We drop villages with greater than 20-point brightness swings between 2001 and 2011, equivalent to shifting from the 1st percentile to the 98th percentile of $\widehat{L}_v^t$ in our RD bandwidth. Such dramatic shifts in brightness almost certainly reflect measurement error, and are inconsistent with the scale of RGGVY-induced electrification. Including these extreme outliers substantially attenuates our brightness RD estimates (see bottom-left panel of Table B2).

C.4 Census data

We construct a village-level panel dataset using three datasets from the Census of India's 2001 and 2011 decennial Censuses. These datasets are all available from the Census of India's website.[30]

26. We use ArcGIS's `Zonal Statistics as Table` operation to calculate both maximum and mean brightness.

27. Active satellites were: F12 (1994–1999), F14 (1997–2003), F15 (2000–2007), F16 (2004–2009), F18 (2010–2013).

28. For example, it assumes that brightness in one region (e.g. the island of Sicily) remains fixed over time, and calibrating sensors to that region's DN values.

29. Unweighted regressions produce nearly identical results.

30. http://censusindia.gov.in. We downloaded these data between September 2014 and August 2015. We treat Telangana as part of Andhra Pradesh, since the 2104 split occurred after our study period.

Primary Census Abstract The Primary Census Abstract (PCA) reports village population and employment for all geographic units across Indian states and union territories. These data include the number of households in each village, along with village population broken down by gender, the 0–6 age cohort, scheduled caste, and scheduled tribe.[31] They also report employment counts by gender, and by job category. We combine "agricultural laborers" and "cultivators" into a single agricultural labor category; the former work for wages, while the latter operate their own land. "Household industry workers" represent a small subset of total workers. Our non-agricultural labor category includes all "other" workers, covering non-farm, non-household employment (e.g. government workers, teachers, microenterprises, factory workers, etc.). Importantly, PCA employment counts cover individuals who *live in the village*, regardless of the location of their work. These counts also separate workers into "main" vs. "marginal" groups, depending on whether they work at least 6 months of the year; we consider each group separately in Panel I of Tale A22, but ignore this distinction elsewhere. Finally, the PCA reports literacy rates by village. The 2001 PCA dataset comprises 593,643 villages, while the 2011 dataset comprises 597,483 villages.

Houselisting Primary Census Abstract The Houselisting Primary Census Abstract (HPCA) reports on a variety of household-related variables. These include physical housing stock characteristics such as type of floor/wall/roof and number of rooms; drinking water source; type of latrine; primary cooking fuel; and main source of in-home lighting (e.g., electricity vs. kerosene). HPCA data also report the number of household members; the number of married couples per household; and whether houses are owned or rented. Finally, they report household asset ownership, including telephones, televisions, motorcycles, radios, and other durable goods. The 2011 HPCA dataset reports the above variables as village-level household shares (i.e., share of households in the village meeting each criterion), for 597,502 total villages. The 2001 HPCA is only available at higher levels of aggregation (either Census block or district), and reports average village-level household shares for (most of) the same variables.

Village Directory The Village Directory (VD) provides detailed village-level amenity data, analogous the community survey that often accompanies household survey data.[32] These community-level characteristics are not specific to individual households. They include the number of primary/middle/secondary schools and other educational facilities; the number of hospitals, community health centers, and other health facilities; community drinking water sources; phone, post office, and other communication services; bus service, rail service, and road quality; and the presence of banking facilities and credit societies. Importantly for our purposes, VD data include 1/0 indicators for the availability of electric power services anywhere in the village, broken out by domestic, agricultural, and commercial end-use sectors; the 2011 VD also reports average hours per day of electric power received by each sector.[33] We use these electric power variables in our first-stage analysis (in Table 2). Finally, the VD contains several geographic indicators, including village area,

31. Scheduled castes (SC) and scheduled tribes (ST) are official designations for historically disadvantaged castes and ethnic groups.

32. The 2001 VD was a standalone product, while the 2011 VD was part of the District Census Handbook (DCHB).

33. Power access indicators are coded as 1 even if there are technical supply disruptions or temporary bans on new connections. The "domestic" category includes "residential houses; bungalows; clubs; hostels and hospitals run on non-commercial basis; charitable, educational, and religious institutions." Agricultural power includes "all electricity given to farmers for conducting agricultural activities including irrigation." Commercial power includes "electricity connections given for workshops, industries etc. or for any commercial purposes." While the 2001 VD reports all-

area of cropland irrigated (by water source), distance to the nearest road and navigable waterway, and distance to the nearest town. After removing villages with populations that are either zero or missing, the 2001 and 2011 VD datasets contain 593,643 and 596,615 villages, respectively.

Census panel dataset We link villages across the above datasets using their official census codes. Within each Census year (2001 and 2011), state, district, and village codes are coded consistently across PCA, HPCA, and VD datasets.[34] We use the Census's 2001–2011 concordance to link villages across Census years. We treat the 2001 PCA village as our master cross-sectional unit, re-aggregating any 2001 villages that split into multiple villages by 2011.[35] Our final panel includes only villages that match to all 5 village level datasets: 2001/2011 PCA, 2011 HPCA, and 2001/2011 VD. To merge in the more aggregated 2001 HPCA, we match villages to their parent blocks and/or districts. Where necessary and possible, we harmonize variable definitions that change across Census years, in order to construct a two-wave village-level Census panel of 580,643 villages.

C.5 Habitation merge

RGGVY determined eligibility using habitation populations: only villages with constituent habitations of at least 300 people were eligible for electrification under the 10th Plan.[36] 2001 village populations (as reported in the PCA) are an imperfect indicator of eligibility status, since not all villages with populations above 300 have constituent habitations large enough to meet this criterion. To accurately assign eligibility status in a sharp RD design, we supplement our village-level Census dataset with habitation-level data.

The National Rural Drinking Water Programme conducted a census of habitations in 2003 and 2009 with the purpose of assessing drinking water availability for all rural habitations in India.[37] To the best our knowledge, this is the only comprehensive nationwide habitation-level dataset. The two habitation census waves list habitation names and populations for 483,510 and 567,406 villages, respectively, and over 1.3 and 1.6 million individual habitations, respectively. Together, they cover over 95% of India's villages. Unlike the Census products described above, habitation census datasets do not include Census village codes. This means that we much link them to Census villages by village name, which is not straightforward.

We apply a four-step string matching algorithm to merge the 2003 and 2009 habitation census data into our Census panel. First, for each Census village code, we construct a list of all observed village name transliterations.[38] Second, we search for exact string matches between villages

sector power access (i.e., agricultural, domestic, and commercial), a separate commercial power indicator is missing and cannot be inferred from the information reported.

34. Subdistrict, tehsil, and block codes are not consistently coded across datasets, which reflects different administrative conventions across states. Because village code is *virtually* unique within each district, we can link only on state, district, and village codes, ignoring subdistrict, block, and tehsil codes.

35. Many 2001 villages match to more than one 2011 village, which we interpret as administrative splits. We drop the *very* few cases (i.e. $< 0.01\%$) where multiple 2001 villages match to a single 2011 village.

36. Habitations (a.k.a. hamlets) are sub-village administrative units, similar to neighborhoods. They are frequently used in economic and social policies in rural India, despite not being official Census administrative units.

37. We downloaded these data at: http://indiawater.gov.in/imisreports/nrdwpmain.aspx; they are now available at https://ejalshakti.gov.in. RGGVY program documentation lists this habitation census as a reference to be used by implementing agencies (Ministry of Power (2014)).

38. We include spellings from the 2001 and 2011 PCA, the 2001 VD, the Census 2001–2011 village concordance, and the RGGVY microdata. The latter simply provide another set of variant spellings, in order to increase the chances of an exact string match.

Table C4: Summary of habitation census merge results

Habitation census match	2003 and 2009	2003 only	2009 only	Unmatched
A. Match rates (all villages)				
Exact matches	0.486	0.065	0.132	0.317
+ `reclink`	0.639	0.050	0.126	0.185
+ `Masala merge`	0.651	0.085	0.125	0.139
B. Match rates (150–450 population)				
Exact matches	0.471	0.046	0.140	0.343
+ `reclink`	0.627	0.048	0.139	0.186
+ `Masala merge`	0.641	0.069	0.137	0.153
C. Summary statistics (all villages)				
Average habitations per village	2.672	2.397	3.352	
Share single-habitation villages	0.571	0.402	0.517	
Share with population mismatch $>$ 20%	0.087	0.478	0.095	
D. Summary statistics (150–450 population)				
Average habitations per village	1.875	1.776	2.057	
Share single-habitation villages	0.654	0.517	0.607	
Share with population mismatch $>$ 20%	0.114	0.545	0.106	

Note. — This table shows results from the habitation merge algorithm described above. Panels A and B report the share of villages that have merged after each step of the algorithm. Panels C and D calculate summary statistics on the subset of Census panel villages that successfully merge to the habitation dataset. Panels A and C report match counts and summary statistics for all 580,643 villages, while Panels B and D report only the 129,453 villages with 2001 populations between 150 and 450 (slightly wider than our optimal `rdrobust` bandwidths). Population mismatches occur when the sum of a village's constituent habitation populations deviates from both its 2001 and 2011 Census population by at least 20%.

with Census codes and village names in the habitation census. We repeat this procedure for each transliteration of the village name to increase the likelihood of an exact string match. Third, for the remaining unmatched villages, we repeat this process using Stata's `reclink` fuzzy string matching function. Fourth, for the remaining unmatched villages, we apply the `Masala merge` fuzzy match routine developed by Asher and Novosad (2020). This algorithm computes the Levensthein distance between strings, while accounting for letter substitutions and interpolations common to Hindi transliterations.[39] After completing this matching procedure separately for the 2003 and 2009 habitation datasets, we combine the results into a single village-level dataset to merge with the Census panel. This dataset includes indicators for villages having matched to the 2003 and/or 2009 habitation census; the 2003 and 2009 counts of the number of habitations per village; and the 2003 and 2009 populations summed over all constituent habitations. Overall, 86.1% of villages match to either the 2003 or 2009 habitation census, and 50.6% of matched villages have exactly 1 habitation.[40]

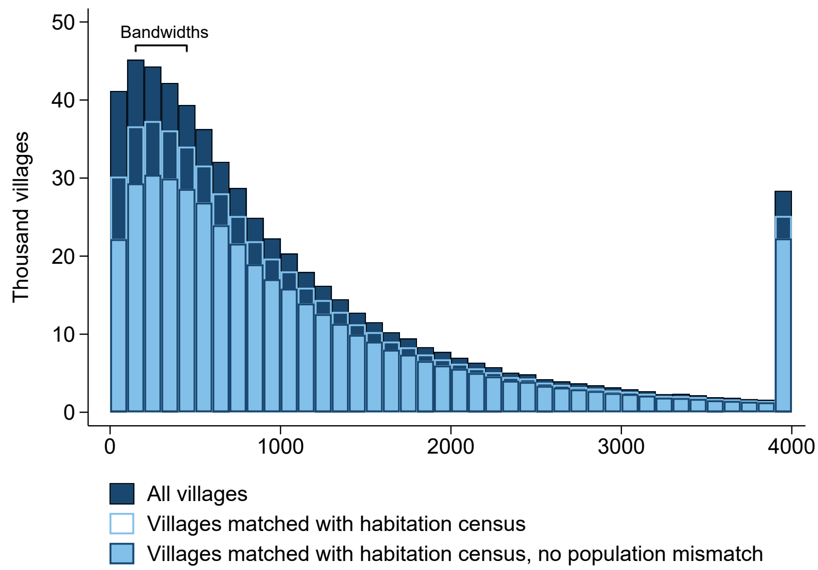
Table C4 reports match rates after each step in this fuzzy merge process. For 91.3% of matches (and 88.6% of matches with 2001 village populations between 150 and 450), the village population implied by the habitation census (i.e. the sum of all habitation populations in a village) deviates

39. `Masala merge` is more accurate and flexible than standard fuzzy merging routines, such as `reclink`. However we use `reclink` to remove close matches before applying the more computationally intensive `Masala merge` algorithm. We thank the authors for sharing this code, which is available here: https://paulnovosad.com/code.html.

40. These match rates are very close to those achieved by Asher and Novosad (2020). For villages that match to both the 2003 and 2009 habitation census, there is a correlation of 0.98 between 2003 and 2009 habitation counts.

from Census's village population by less than 20%.[41] Since large population disparities indicate erroneous matches, our RD analysis excludes the 11.4% villages with population mismatches greater than 20%. Including these population mismatches attenuates our RD estimates for village power indicators (likely due to measurement error at the RD threshold), but does not meaningfully change our other RD results (see Figures B1, B2, and B8). Figure C2 summarizes habitation census match rates across the support of 2001 village populations, with lower match rates for smaller villages closer to our RD bandwidth.

Figure C2: Habitation merge results, by 2001 village population



Note. — This figure shows a histogram of Indian villages by 2001 population (solid navy), and the subset of villages that we successfully match with the habitation census (hollow blue). Solid light blue bars show the subset of matched villages with population disparities of less than 20%, which we include in our RD analysis. We match 84.7% of villages with 2001 populations between 150 and 450. Excluding villages with population mismatches leaves us with 69.6% of villages with 2001 populations between 150 and 450.

C.6 SHRUG dataset

The Socioeconomic High-resolution Rural-Urban Geographic Dataset for India (SHRUG) is an open-source repository of Indian village-level datasets (Asher et al. (2021)).[42] SHRUG data include a small-area measure of consumption expenditure per capita, which we use as our primary village-level economic outcome. To construct this measure, the authors combine household-level asset data enumerated by the Socioeconomic and Caste Census (SECC, described next in Section C.7) with consumption data from the 2011–12 Indian Human Development Survey (IHDS-II), following the method of Elbers, Lanjouw, and Lanjouw (2003). They regress total household consumption on a set of continuous and dummy variables containing all SECC asset and earnings information. Then, they use these regression estimates to predict household-level consumption, aggregate up to the village level to construct annual expenditure per capita, and cross-validate using both IHDS and NSS expenditure data. We follow Asher and Novosad (2020) in using this SHRUG consumption

41. Even with 100% match accuracy, we should not expect village populations from 2003/2009 habitation censuses to correspond exactly with village populations from the 2001/2011 Census, due to timing differences.

42. SHRUG data access are available at http://www.devdatalab.org/shrug. Further details on variable construction are available in the SHRUG codebook: http://paulnovosad.com/pdf/shrug-codebook.pdf.

expenditure measure as a village-level welfare indicator. We also use SHRUG's data on the share of households where cultivation is the main source of income, which Asher et al. (2021) extract from restricted-access SECC data. Merging these outcome variables into our main Census data is trivial, since SHRUG data include Census village codes.

C.7 Socio-Economic Caste Census dataset

We incorporate additional microdata from the 2011 Socioeconomic and Caste Census (SECC) to examine household income and poverty rates. The SECC collected individual- and household-level data with the intention of documenting the socioeconomic status of every person in India.[43] We obtained a subset of these data from the Ministry of Petroleum and Natural Gas, whose liquid petroleum gas subsidy program (Pradhan Mantri Ujjwala Yojana) used SECC data to determine eligibility.[44] As a result, we observe SECC data for the universe of rural individuals that were eligible for this fuel subsidy program. This includes "households having one of the Deprivations [in the SECC]", where a "deprivation" is a household poverty indicator. SECC deprivation criteria were assessed after removing affluent and destitute households, such that assignment to poverty status automatically excluded affluent households and automatically included destitute households.

Our sample of fuel-subsidy-eligible households was constructed by first removing all households that satisfied at least one affluence indicator, or "exclusion".[45] These household-level affluence indicators were: motorized 2/3/4 wheelers or fishing boats; mechanized 3–4 wheeler agricultural equipment; a Kisan credit card (issued by the government to assist farmers) with a credit limit over Rs. 50,000; being a government employee; operating a non-agricultural enterprise registered with the government; earning more than Rs. 10,000 per month; paying income tax; paying professional tax; having 3 or more rooms with "pucca" (essentially permanent) walls and roof; a refrigerator; a landline phone; more than 2.5 acres of irrigated land and irrigation equipment; owning 5 or more acres of irrigated land for two or more crop seasons; **or** at least 7.5 acres of land and irrigation equipment. Next, destitute households were "automatically" included if they were without shelter; were destitute or living on alms; earned income from manually scavenging; belonged to a primitive tribal group; or were engaged in legally released bonded labor.

Our SECC dataset includes all remaining (non-affluent, non-destitute) households that satisfied **at least one** poverty (or "deprivation") criterion. These criteria are: having one or fewer rooms; "kuccha" (non-pucca) walls and roof; no adult members between the age of 18 and 59; a female head-of-household with no adult male member between 16 and 59; a "differently-able" member with no other able-bodied member; being scheduled caste or scheduled tribe; no literate adults above age 25; **or** being landless and deriving a majority of income from manual labor. This yields a dataset of 332 million individuals from 81 million households with no affluence indicators (auto-exclusions), no destitution indicators (auto-inclusions), and at least one poverty indicator (deprivation). This represents roughly half of all households in rural India.[46]

43. See http://www.secc.gov.in/aboutusReport for further details. The SECC was a follow-up to the 2002 Below Poverty Line Census. We were unable to gain access to these 2002 data.

44. We downloaded our data in Excel format from http://lpgdedupe.nic.in/secc/secc_data.html.

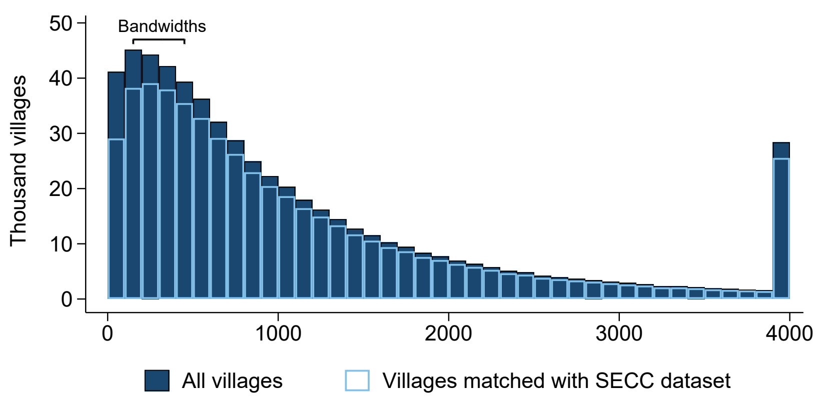
45. See http://www.pmujjwalayojana.in and https://mopng.gov.in/en/marketing/pmuy.

46. We were unable to download SECC data from several districts. While most of these districts are urban, six contain a nontrivial number of rural villages. These missing districts are: Chamoli, Uttarakhand (1,246 villages); Jalor, Rajasthan (802 villages); Jalpaiguri, West Bengal (768 villages); Dhanbad, Jharkhand (1,760 villages); Dindigul,

This SECC dataset contains individual-level data on age, gender, employment, caste, and marital status; and household-level data on the housing stock, land ownership, asset ownership, income sources, and the household head. Importantly for our analysis, these data report household income, in the form of an indicator for whether the main income earner in each household receives less than Rs 5,000 per month, or between 5,000 and Rs 10,000 per month.[47] They also report whether any household member has a salaried job, and whether the household owns any land. We average these three household-level outcomes at the village level, across the subset of household with poverty indicators.

We merge these SECC data with our village-level Census dataset using a fuzzy matching algorithm similar to that described in Appendix C.5. While SECC data do not include Census village codes, their reindexed village codes largely preserve the relative order of Census village codes within a block. The first step of our algorithm searches for exact matches on (reindexed) villages codes, allowing for village name discrepancies of up to 2 characters. Second, for the remaining unmatched villages, we search for exact matches on villages name. Third, for the few remaining unmatched villages, we apply the `Masala merge` algorithm discussed above. Ultimately, we match 94.2% of SECC villages to a village in our Census dataset, with very few (i.e., less than 5%) of matches relying on the fuzzy `Masala merge` algorithm.[48] Figure C3 summarizes this SECC-Census merge by village population, where 88.9% of Census villages match to a village in our SECC dataset. This merge enables us to construct an additional SECC outcome variable—by comparing the count of households in the SECC sample to the total number of households in the village (per the 2011 Census), we can infer the share of village households with a poverty indicator.

Figure C3: SECC merge results, by 2001 village population



Note. — This figure shows a histogram of Indian villages by 2001 population (solid navy), and the subset of villages that we successfully match with a village in the SECC dataset (hollow blue). Overall, we match 88.9% of Census villages to the SECC dataset; for 94.1% of these matches, at least 10% of total households are included in our SECC dataset (because they have at least one poverty indicator). We match 88.7% of Census villages with 2001 populations between 150–450; for 91.9% of these matches, our SECC dataset includes at least 10% of total village households.

Tamil Nadu (481 villages); and Thanjavur, Tamil Nadu (516 villages). Together, these districts contain 5,573 villages, representing less than 1% of the total number of villages in our Census dataset.

47. As described above, households whose primary income earner earns above Rs 10,000 per month were automatically excluded from our sample.

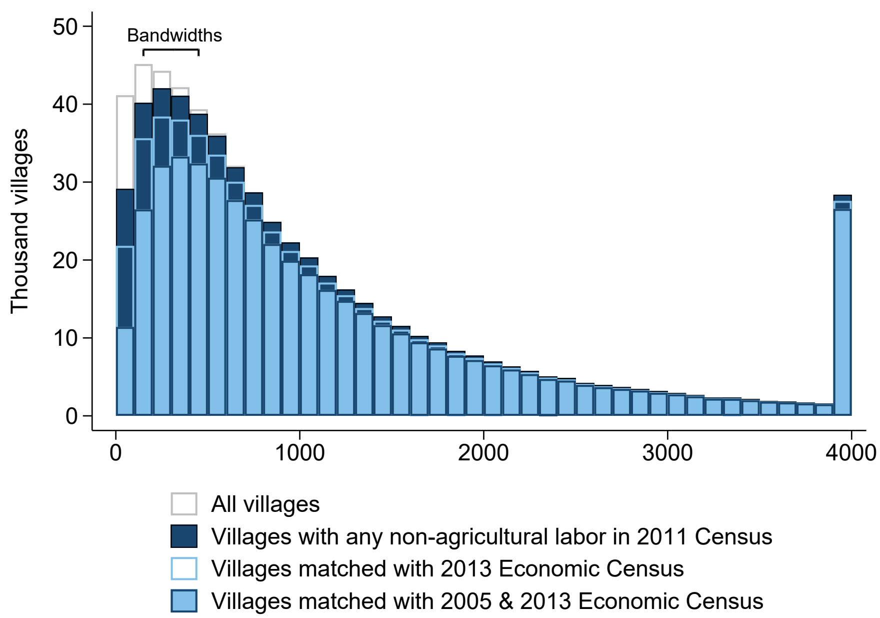
48. This excludes the 2.7% of SECC-Census matches for which the SECC data include either a village population or a village household count over 10% larger than those reported by the 2011 Census. We drop these suspicious matches from our analysis of SECC outcomes.

C.8 Economic Census dataset

The Economic Census is a census of all non-farm establishments, including both formal firms, informal firms, and public sector employers.[49] We use village-level data on the total number of establishments and total number of employees, from the 1990, 1998, 2005, and 2013 Economic Census waves. We follow Asher and Novosad (2020) in dropping any establishments with NIC codes in the defense sector or the agricultural sector. While these data also report firm sectors, sources of finance, and gender of employees, we lack the statistical power to estimate heterogeneous effects (the average village in our RD sample had only 7.6 firms in 2005). Unfortunately for our purposes, the Economic Census stopped collecting information on firm energy source after the 2005 wave—meaning that we cannot use firm-level information on electricity access as a first-stage outcome variable to estimate the impacts of RGGVY.

We use the SHRUG village crosswalk (Asher et al. (2021)) to match villages from the Economic Census to villages in the Census of India. Figure C4 reveals lower match rates for smaller villages in our RD bandwidth, which reflects how small villages are less likely to contain non-farm establishments. If we correct for the share of villages with non-zero non-agricultural labor in the 2011 Census, we match 88% of 150–450-person villages to the 2013 Economic Census, and 75% of 150–450-person villages to both the 2005 and 2013 Economic Census waves.

Figure C4: Economic Census merge results, by 2001 village population



Note. — This figure shows a histogram of Indian villages by 2001 population (hollow white), the subset of villages with any non-agricultural labor in 2011 (solid navy), and the subset of villages that we successfully match to the Economic Census in 2013 (hollow blue) or in 2005 and 2013 (solid light blue). Of villages with 2001 populations between 150 and 450 and non-zero non-agricultural labor, we match 88.1% to the 2013 Economic Census and 75.1 to both 2005 and 2013 years.

C.9 DISE schools dataset

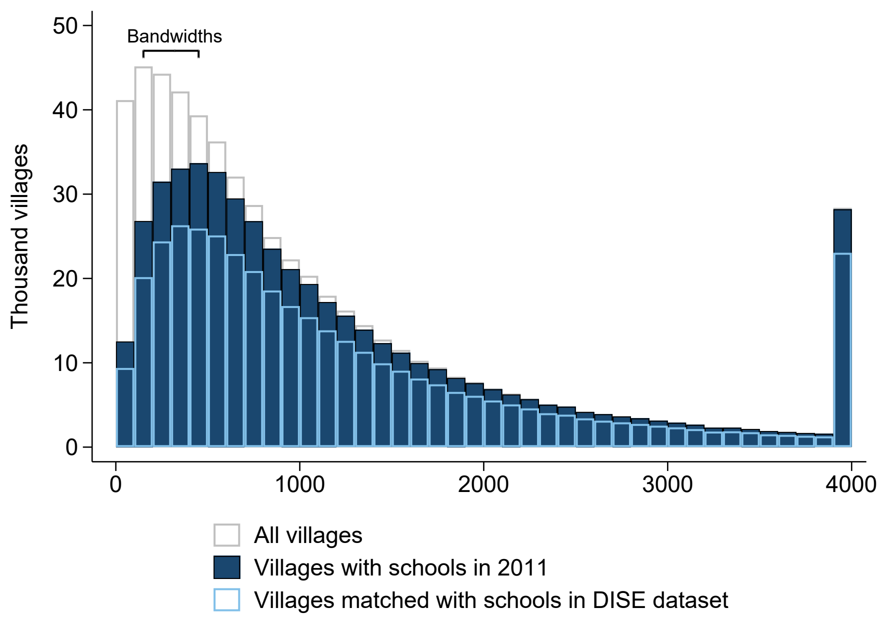
We construct a panel dataset containing the universe of primary (grades 1–5) and upper primary (grades 6–8) schools in India, for the 2005–2006 through 2014–2015 school years. These data are

49. We downloaded these data from http://microdata.gov.in/nada43/index.php/catalog/ECO.

publicly available online through the District Information System for Education (DISE)'s School Report Cards, originally created to inform policymakers about the effectiveness of the District Primary Education Programme (DPEP). DISE data are collected by school headmasters (or head teachers), and submitted to district- and state-level authorities before entering the official national system.[50] In total, these data cover 1.68 million schools across over 600,000 villages. However, the annual panel is quite unbalanced, with the average school appearing in only 7 of 10 years.[51] We observe DISE data for 536,330 villages in 2005–06, with an average of 1.9 schools per village. This increases to 559,442 villages and 2.3 schools per village in 2014–15.

We use two sets of DISE outcomes in our analysis: student enrollment and students passing their examinations. DISE data report the number of students in each school year, broken down by grade and gender. We sum head counts across all schools in a village, and split by primary (grades 1–5) vs. upper primary (grades 6–8). DISE data also report the number of students in grades 4–5 and 7–8 who passed the previous year's end-of-year examination; we divide the number of students who passed with high marks by the previous year's enrollment.[52] DISE data include additional school characteristics that we do not use in our analysis, including whether each school has electricity access. We do not detect first-stage RD effects on school electrification, likely due to the fact that only 40% of villages in our main RD sample had schools.

Figure C5: DISE merge results, by 2001 village population



Note. — This figure shows a histogram of Indian villages by 2001 population (hollow white), the subset of villages with schools 2011 (as reported by the 2011 Census; solid navy), and the subset of villages that we successfully match with a school in the the DISE dataset (hollow blue). We match 79.0% of villages with schools in 2011; for villages with 2001 populations between 150 and 450, we match 78.0% of those with schools in 2011.

50. DISE data can be found here: http://schoolreportcards.in/SRC-New/. Quality control is performed by the cluster resource coordinator, and again at the district level.

51. While this partly reflects new school construction between 2005 and 2014, it also reflects errors we encountered when downloading DISE data. 7% of schools are missing data for at least 1 year between their first and last reporting years. We have no reason to believe that these errors are non-random.

52. Data for passing students are missing for the 2010–11 to 2013–14 school years.

We link schools across years based on unique numerical school codes. DISE also reports state, district, block, and village names, which we use to match schools to Census villages via a fuzzy match procedure similar to our habitation matching algorithm (see Appendix C.5 above). First, we build a concordance between school codes and village names as reported in the DISE dataset. In many cases, one school will be associated with multiple village names in different years, either due to variant spellings or administrative village splits. Next, we search for string matches between village names already linked to Census codes and DISE village names, following the exact string match, `reclink`, and `Masala merge` progression described in Appendix C.5. After establishing a rough Census-DISE village concordance, we use school codes to disambiguate duplicate village matches.[53] Figure C5 summarizes these fuzzy matching results. Smaller villages are both less likely to report having schools in the 2011 Census, and less likely to match into the DISE dataset conditional on having at least one school. Our RD analysis on DISE outcomes ultimately includes 56% of the villages in our main RD sample.

C.10 Village counts by dataset

Each merge between village-level datasets is imperfect. Table C5 reports the number of villages that survive each step in this merging process. Our main RD sample includes single-habitation Census villages in RGGVY 10th-Plan districts, in states with reliable village shapefiles (fourth row, right two columns). The right-most column restricts each sample to villages with 2001 populations between 150 and 450, which is wider than (nearly all of) our `rdrobust` optimal bandwidths. Regressions using SECC, Economic Census, or DISE outcomes include fewer villages, as dictated by the success of their respective matching algorithms.

Table C5: Count of villages by merged dataset

Number of Villages	Total	RGGVY 10th-Plan	RGGVY 10th-Plan Single-Hab.	RGGVY 10th-Plan Single-Hab. 150–450
Raw Census datasets (village-level)	593,000+			
Census panel	580,643	290,067		
Census panel + habitations	499,799	218,841	115,444	29,765
Census panel + habitations + maps	365,431	137,519	70,790	18,686
Census panel + habitations + maps + SHRUG	361,342	137,502	70,784	18,686
Census panel + habitations + maps + SECC	326,821	121,794	62,381	16,240
Census panel + habitations + maps + EC	305,190	116,726	58,685	14,715
Census panel + habitations + maps + DISE	271,587	90,314	44,448	10,529

Note. — All village counts exclude Goa and the 7 Union Territories, which were not covered under RGGVY. Rows 2–8 include villages we can link across 2001 and 2011 Census years. Rows 3–8 include villages we can match to the 2003 or 2009 census of habitations, with populations disparities of less than 20%. Rows 4–8 include villages in states with reliable village shapefiles (i.e. the top 12 rows of Table C4), letting us use nighttime brightness as a pre-RGGVY control in our RD estimation. Row 5 includes villages we can match to the SHRUG dataset (with 2011 expenditure per capita). Row 6 includes villages we can match to our SECC dataset. Row 7 includes villages we can match to both the 2005 and 2013 Economic Census. Row 8 includes villages we can match to the DISE dataset, only 85% of which have nonmissing 2005 school data (a key pre-RGGVY control). We conduct our RD analysis on villages in the right-most column, with optimal RD bandwidths less than 150 above/below the 300-person cutoff.

53. We keep the most frequent match for each school. For example: if school A matches to village 1 in 2005, village 1 in 2006, and village 2 in 2007, we keep the match that links school A with village 1.

C.11 NSS district-level dataset

We use the "thick" (55th, 1999–2000; 61st, 2004–05; and 66th; 2009–10) rounds of India's National Sample Survey (NSS), carried out by the which we refer to as 2000, 2005, and 2010 NSS waves throughout the main text and appendix.[54] These rounds collect data on household consumer expenditures, and surveyed 60,000–80,000 households per round, living all across India.[55] Households are asked about their expenditures on a variety of consumable and durable goods, including foodstuffs, fuel, and other household items such as soap. In each of these rounds, the rural NSS is representative at the district level (Topalova (2010)), though it obscures the identifies of surveyed villages and households. The NSS is collected by the National Sample Survey Office in the Ministry of Statistics and Programme Implementation (MoSPI).[56] The surveys also include household characteristics, district identifiers, and sampling (or "survey") weights, which we use to construct a representative household at the district level. We use 2001 Census district codes to link the 2005 and 2010 NSS waves to our other datasets. For the 2000 NSS wave, which uses district codes from the 1991 Census, we apply a 1991-2001 Census concordance that accounts for administrative district splits.[57]

Sampling The NSS uses a two-stage sampling strategy. MoSPI first selects "first stage units" or Census villages for the rural sample which we use in this paper, and then households.[58] Villages are explicitly allocated to the sample in a manner proportional to population. MoSPI also imposed an explicitly population-based household sampling strategy: enumerators surveyed a proscribed number of households in each sampled village.[59] As a result, in order to reflect the rural population distribution within each district, households in larger villages typically received larger sampling weights, which can be used to construct a representative household at the district level. Because household sampling also takes into account wealth (stratifying on affluence), the correlation between population and sampling weights is strong, but not perfectly mechanical. We take advantage of the relationship between village population and sampling weights to estimate heterogeneous treatment effects on village size.

MoSPI does not typically release information on the population of enumerated villages publicly. However, it keeps internal "multiplier files," which contain all of the information required to build up each household's sampling weights—including village population. We obtained these files for the

54. These "thick" rounds are enumerated quinquennially, and include a substantially larger sample than the intervening "thin" rounds.

55. We use the "Type 1" enumeration for the 66th round, since this uses the same recall period for consumption expenditures as the 55th and 61st rounds (which only had one type); where given a choice, we use 30 day recall periods.

56. The data are available for download from http://microdata.gov.in/nada43/index.php/catalog/CEXP.

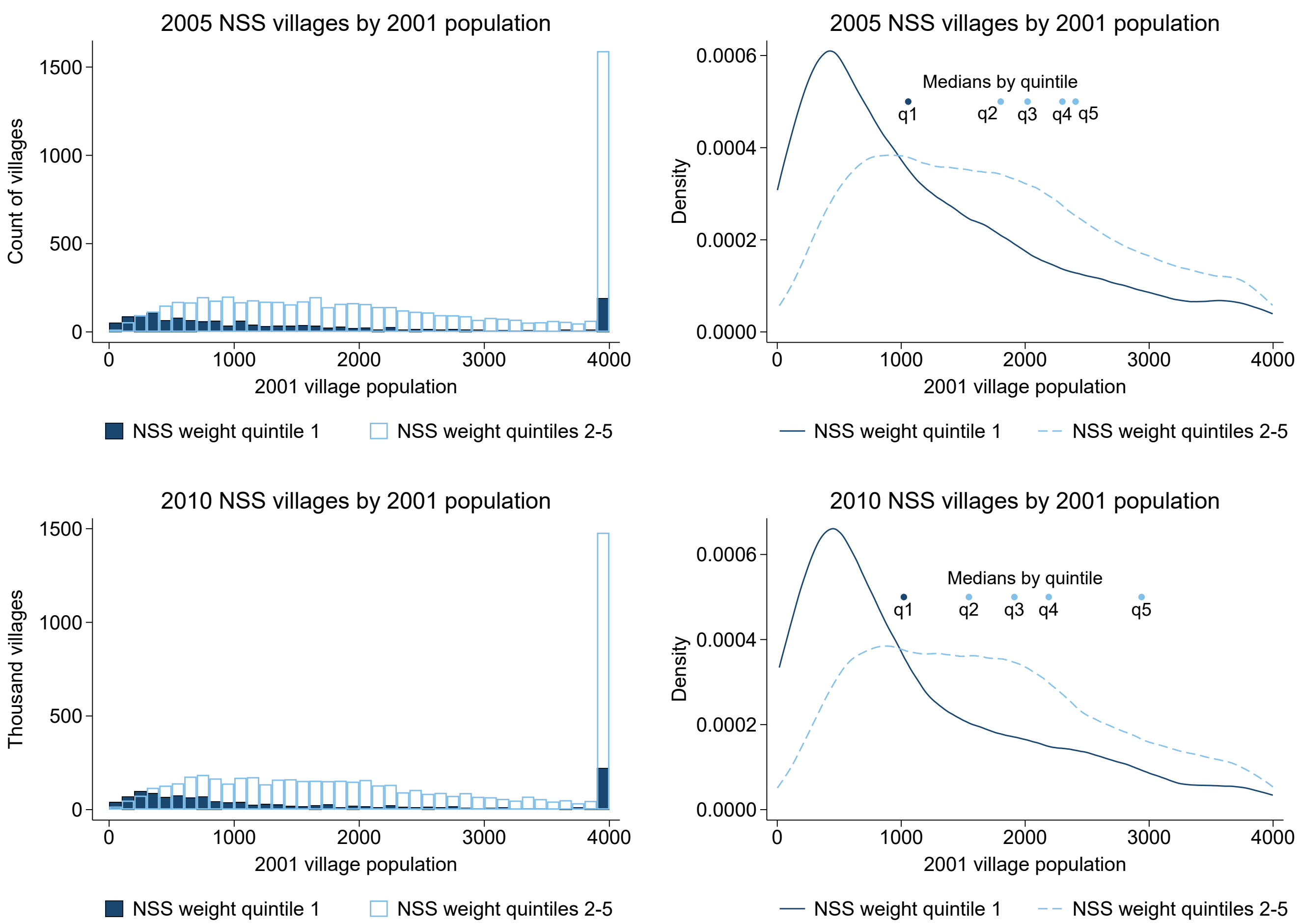
57. One third of 2001 districts were administratively subdivided from more aggregated districts in the 1991 Census. RGGVY implementation applied district definitions from the 2001 Census.

58. For more details on the NSS sampling strategy, see the detailed descriptions available as part of the data archive. As an example, the sampling approach for the 2000 round is described in two documents: one for the first stage (http://www.icssrdataservice.in/datarepository/index.php/catalog/12/download/202) and one for the second stage (http://www.icssrdataservice.in/datarepository/index.php/catalog/12/download/203). The 2005 and 2010 waves use very similar rules. The 2000 wave used the 1991 Census for its base list of villages; the 2005 and 2010 round use the 2001 Census.

59. In the 2010 wave, the NSS surveyed exactly 8 households in 99% of villages sampled. In the 2005 wave, the NSS surveyed exactly 10 households in 99% of villages sampled. In the 2000 wave, the NSS surveyed exactly 12 or 24 households in 93% of villages sampled.

2005 and 2010 waves, but were unable to do so for the 2000 wave. As we discuss in Section 6.A, this leaves us with two options for restricting the NSS sample by village size. First, we can explicitly split on village population in the 2005 and 2010 waves, which necessitates dropping the 2000 wave. Second, we can use sampling weights as proxy for village populations, allowing us to include the 2000 wave, made possible by the sampling strategy described above. We construct weight quintiles within each NSS wave, assigning each village the maximum sampling weight among constituent households. Figure C6 illustrates how isolating the first quintile of NSS sampling weights (i.e. the "Q1" subsample) shifts the distribution towards villages with smaller populations. In the 2005 (2010) wave, the Q1 subsample removes 89% (87%) of villages larger than 4000 people. For all NSS subsamples (splitting either on population or on weight quintiles), we collapse to the district level after isolating the relevant subset of household-level observations, and renormalizing sampling weights to sum to one.

Figure C6: Village population distributions by NSS weight quintile



Note. — The left panels display histograms of 2001 populations for villages in the 2005 and 2010 NSS sampling frames. Solid navy bars report counts of villages in the bottom quintile of NSS sampling weights, while hollow blue bars report counts of villages in the top four quintiles. Histograms are top-coded at 4000. The right panels report the corresponding kernel densities, removing top-coded villages. Dots report medians of the untruncated distributions, for all five quintiles.

Outcome variables We use three sets of NSS outcome variables in our analysis. First, we use the quantity of electricity consumed over the previous 30 days (kWh per household-month), and spending on electricity consumption (rupees per household-month). These two variables let us derive an average price per kWh for each NSS household with non-zero electricity consumption. The NSS electricity consumption variables include all purchased electricity, whether from the grid or from other sources (e.g. a diesel microgrid operator). The NSS also collects each household's spending on diesel and petrol used for self-generation, which we use to test for crowding-out effects (see Table A11).

Second, we use total monthly expenditure per capita for each household, computed using a 30 day recall period and adjusted to 2010 rupees. Because saving is limited in rural India, this is also a close approximation for household income (Atkin et al. (2020)). To make this variable a more accurate proxy for electrification-induced welfare changes, we subtract electricity spending per capita (converting from rupees per household to rupees per person using the household size variable). The average electrified NSS household allocates 3.2% of total expenditures to electricity purchases.

Third, we use dummies for electric appliance ownership at the household level (see outcomes in Columns (4)–(8) of Table 3). We exclude two NSS appliances from our table: electric cookstoves (due to exceedingly low penetration), and sewing machines (since the NSS questionnaire does not specify *electric* sewing machines).

For our pooled DD sample (Tables 3 and 5, and Columns (1) and (4) of Table 7), we average all households into a district-wave panel using NSS sampling weights. This panel uses the 2001 district definitions, which are the same definitions used in RGGVY DPRs. For districts that split between the 1991 and 2001 Censuses, we populate 2000-wave observations using the same 1991-district averages for all constituent 2001-districts. For all regressions using NSS data, we cluster our standard errors by the (more aggregated) 1991 district definitions.

Appendix References

Alam, Muneeza M. 2013. "Coping with Blackouts: Power Outages and Firm Choices." 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." *World Bank Economic Review* 35 (4): 845–871.

Asher, Sam, and Paul Novosad. 2017. "Politics and Local Economic Growth: Evidence from India." *American Economic Journal: Applied Economics* 9 (1): 229–273.

———. 2020. "Rural Roads and Local Economic Development." *American Economic Review* 110 (3): 797–823.

Atkin, David, Benjamin Faber, Thibault Fally, and Marco Gonzalez-Navarro. 2020. "Measuring Welfare and Inequality with Incomplete Price Information." NBER Working Paper 26890.

Banerjee, Sudeshna Ghosh, Douglas Barnes, Bipul Singh, Kristy Mayer, and Hussain Samad. 2014. *Power for All: Electricity Access Challenge in India.* Washington, DC: The World Bank.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and Path Dependence." *Quarterly Journal of Economics* 127 (2): 587–644.

Burgess, Robin, Michael Greenstone, Nicholas Ryan, and Anant Sudarshan. 2020. "Demand for Electricity on the Global Electrification Frontier." Working paper.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Census of India. 2011. "Mapping Activities." http://censusindia.gov.in/2011-common/map.html.

Chand, T. R. Kiran, K. V. S. Badarinath, Christopher D. Elvidge, and B. T. Tuttle. 2009. "Spatial Characterization of Electrical Power Consumption Patterns over India Using Temporal DMSP–OLS Night-Time Satellite Data." *International Journal of Remote Sensing* 30 (3): 647–661.

Chen, Xi, and William D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." *Proceedings of the National Academy of Sciences* 108 (21): 8589–8594.

Doll, Christopher N. H. 2008. *CIESIN Thematic Guide to Night-Time Light Remote Sensing and Its Applications.* Center for Earth Science Information Network (CIESIN).

Doll, Christopher N. H., Jan-Peter Muller, and Jeremy G. Morley. 2006. "Mapping Regional Economic Activity from Night-Time Light Satellite Imagery." *Ecological Economics* 57 (1): 75–92.

Dugoua, Eugenie, Ryan Kennedy, and Johannes Urpelainen. 2018. "Satellite Data for the Social Sciences: Measuring Rural Electrification with Night-Time Lights." *International Journal of Remote Sensing* 39 (9): 2690–2701.

Ebener, Steeve, Christopher Murray, Ajay Tandon, and Christopher D. Elvidge. 2005. "From Wealth to Health: Modelling the Distribution of Income per Capita at the Sub-National Level Using Night-Time Light Imagery." *International Journal of Health Geographics* 4 (1): 5.

Elbers, Chris, Jean Lanjouw, and Peter Lanjouw. 2003. "Micro-Level Estimation of Poverty and Inequality." *Econometrica* 71 (1): 355–364.

Elvidge, Christopher D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis. 1997. "Relation between Satellite Observed Visible-Near Infrared Emissions, Population, Economic Activity and Electric Power Consumption." *International Journal of Remote Sensing* 18 (6): 1373–1379.

Elvidge, Christopher D., Marc L. Imhoff, Kimberly E. Baugh, Vinita Ruth Hobson, Ingrid Nelson, Jeff Safran, John B. Dietz, and Benjamin T. Tuttle. 2001. "Night-Time Lights of the World: 1994–1995." *ISPRS Journal of Photogrammetry and Remote Sensing* 56 (2): 81–99.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2020. "Experimental Evidence on the Economics of Rural Electrification." *Journal of Political Economy* 128 (4): 1523–1565.

Li, Xi, Linlin Ge, and Xiaoling Chen. 2013. "Detecting Zimbabwe's Decadal Economic Decline Using Nighttime Light Imagery." *Remote Sensing* 5 (9): 4551–4570.

Machemedze, Takwanisa, Taryn Dinkelman, Mark Collinson, Wayne Twine, and Martin Wittenberg. 2017. "Throwing Light on Rural Development: Using Nightlight Data to Map Rural Electrification in South Africa." DataFirst Technical Paper 38.

Michalopoulos, Stelios, and Elias Papaioannou. 2013. "Pre-Colonial Ethnic Institutions and Contemporary African Development." *Econometrica* 81 (1): 113–152.

———. 2014. "National Institutions and Subnational Development in Africa." *Quarterly Journal of Economics* 129 (1): 151–213.

Min, Brian. 2011. "Electrifying the Poor: Distributing Power in India." Working paper.

Min, Brian, and Kwawu Mensan Gaba. 2014. "Tracking Electrification in Vietnam Using Nighttime Lights." *Remote Sensing* 6 (10): 9511–9529.

Min, Brian, Kwawu Mensan Gaba, Ousmane Fall Sarr, and Alassane Agalassou. 2013. "Detection of Rural Electrification in Africa Using DMSP-OLS Night Lights Imagery." *International Journal of Remote Sensing* 34 (22): 8118–8141.

Ministry of Power. 2012. *Rajiv Gandhi Grameen Vidyutikaran Yojana.*

———. 2014. *Guidelines for Preparation of DPRs under XII Plan of RGGVY.*

Noor, Abdisalan M., Victor A. Alegana, Peter W. Gething, Andrew J. Tatem, and Robert W. Snow. 2008. "Using Remotely Sensed Night-Time Light as a Proxy for Poverty in Africa." *Population Health Metrics* 6 (1): 5.

Sreekumar, N., and Shantanu Dixit. 2011. "Challenges in Rural Electrification." *Economic & Political Weekly* 46 (43).

Topalova, Petia. 2010. "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India." *American Economic Journal: Applied Economics* 2 (4): 1–41.

Townsend, Alexander C., and David A. Bruce. 2010. "The Use of Night-Time Lights Satellite Imagery as a Measure of Australia's Regional Electricity Consumption and Population Distribution." *International Journal of Remote Sensing* 31 (16): 4459–4480.