

Local collectivism-individualism cultural cleavage, institutions and prosocial behaviour: Evidence from household electricity theft in India

Supplementary Appendix 1.

This section briefly presents the recursive bivariate probity model. In order to correct for this remaining concern about the likely endogeneity problem, we make use of a recursive bivariate probit model to obtain the consistent full information maximum likelihood parameter estimates. We may employ the following recursive bivariate probit (RBP) model (Greene, 1998; 2012, pp. 745–749; Panda 2015; 2022; Waters 1999).

$$electheftdum_i = I(\phi collectivistic_i + \beta controls_i + \theta_i + \varepsilon_i > 0) \quad (2)$$

$$collectivistic_i = I(\lambda controls_i + \psi instrument + v_i > 0) \quad (3)$$

where $I(\cdot)$ is the indicator function taking the value one if the statement in the brackets are true and equals zero otherwise and λ , β , ϕ and ψ are the parameters to be estimated. The same framework is also used in case of explaining the likelihood of stealing power in the equation 2. We impose an exclusion restriction on the model in order to improve identification, with *instrument* being the instrumental variable is a dummy that =1 if there is some or a lot of conflict amongst communities or jatis in village or neighborhood, =0 if there's not much conflict. The two error terms ε_i and v_i , respectively for the power theft dummy equation and the collectivism equation are assumed to be distributed bivariate normal (Monfardini and Radice, 2008) which indicates that $E(\varepsilon_i) = E(v_i) = 0$, $var(\varepsilon_i) = var(v_i) = 1$, and $cov(\varepsilon_i, v_i) = \rho = corr(\varepsilon_i, v_i)$. In an event of the unobserved determinants of both *electheftdum_i* and *collectivistic_i* (that are, the probabilities of the

household obtaining a BPL ration card and being politically connected) could be correlated. Wilde (2000) argues that identification in such an RBP model can be achieved if both equations contain at least one varying exogenous regressor, even if no exclusion restriction is introduced. It would certainly be preferable to include such an instrument in the model, especially since in the absence of such variable, the reliability of the exogeneity test (i.e., the test that $\rho = 0$) depends on the assumption that the residuals are distributed bivariate normal (Monfardini and Radice, 2008). It is common practice to impose additional exclusion restrictions to improve identification. Monfardini and Radice (2008) show that the availability of an instrument allows to relax the assumption of bivariate normality of the two error terms, that is, inference on exogeneity relies heavily on a valid exclusion restriction because that preserves validity of the Likelihood Ratio test of exogeneity even when the distributional assumption of bivariate normality fails or is relaxed. They explain that, unless there is a very large sample, the likelihood ratio or Wald test of the cross-equation correlation coefficient is the best test for endogeneity of the collectivistic dummy variable. If the correlation coefficient between the two error terms is significantly different from zero, then the collectivistic dummy is indeed endogenous.

The correlation coefficient, ρ , captures the correlation between the two error terms i.e. unobservables that impact the binaries of receiving a ration card and being politically connected. If the two error terms are positively correlated, then $\rho > 0$ and the coefficient on the collectivistic variable is overestimated. However, if the correlation goes in opposite directions (i.e. $\rho < 0$), the collectivistic coefficient is underestimated. When ρ is zero, the model for *electheftdum_i* becomes the standard probit model (Greene, 2012; Panda 2015).

Aside from the estimate of ϕ , estimates of λ , β and ψ provide information on who is

from a collectivistic locality. We derive the consistent and asymptotically efficient parameter estimates of the collectivism effect for the unconstrained correlation coefficient (ρ) between the two sets of error terms by maximum likelihood estimation of the RBP model.

Supplementary Appendix 2

Table A1. Mean differences in characteristics of the groups of household heads

	All household heads	Those who do not steal electricity	Those who steal electricity	Non-stealers (mean)	Stealers (mean)	Difference means	in t-statistic
Collectivistic locality (binary)	23,247	21,213	2,034	0.7279 (0.003)	0.7645 (0.009)	-0.037*** (0.010)	-3.5527
Log fuel kerosene expenditure	23,247	21,213	2,034	8.013 (0.007)	8.066 (0.021)	-0.053** (0.024)	-2.198
Log total HH electricity expenditure	23,247	21,213	2,034	6.6411 (0.018)	1.364 (0.064)	5.276*** (0.062)	84.731
Consumption inequality (Gini)	23,247	21,213	2,034	0.4023 (0.0001)	0.4018 (0.0003)	0.0005 (0.0003)	1.369
Outstanding debt with local shopkeeper (binary)	23,148	21,115	2,033	0.221 (0.003)	0.348 (0.011)	-0.127*** (0.010)	-12.960
BPL card holder (binary)	23,247	21,213	2,034	0.4604 (0.003)	0.4115 (0.011)	0.049*** (0.012)	4.2289
Education years	23,247	21,213	2,034	5.442 (0.032)	3.950 (0.097)	1.492*** (0.109)	13.728
Log household income	23,247	21,213	2,034	11.182 (0.007)	10.770 (0.020)	0.412*** (0.023)	18.226
Political connections (binary)	23,247	21,213	2,034	0.399 (0.003)	0.384 (0.011)	0.0144 (0.011)	1.275
Casual job (binary)	23,247	21,213	2,034	0.456 (0.003)	0.579 (0.011)	-0.124*** (0.012)	-10.703
Self-employed (binary)	23,247	21,213	2,034	0.708 (0.003)	0.802 (0.009)	-0.094*** (0.010)	-8.9721
Connection to govt. employee	23,210	21,178	2,032	0.325	0.202	0.123***	11.441

Connection to big officer	23,215	21,181	2,034	(0.003) 0.182	(0.009) 0.086	(0.011) 0.096***	10.961
Newspaper reading in HH	23,247	21,213	2,034	(0.003) 0.5052	(0.006) 0.362	(0.009) 0.143	12.398
TV watching in family	23,247	21,213	2,034	(0.003) 0.842	(0.011) 0.671	(0.012) 0.171***	19.611
Years lived in the place	23,247	21,213	2,034	(0.003) 79.193	(0.010) 85.117	(0.009) -5.924***	-10.977
Local crime dummy	23,247	21,213	2,034	(0.163) 0.052	(0.366) 0.066	(0.540) -0.014***	-2.756
Civic meeting participation dummy	23,247	21,213	2,034	(0.002) 0.343	(0.006) 0.301	(0.006) 0.042***	3.824
Consumption expenditure per capita	23,246	21,212	2,034	(0.003) 24370.51	(0.010) 17068.10	(0.011) 7302.41***	13.707
				(162.522)	(295.362)	(532.76)	

Notes: statistical significance is shown as ***p < 0.01. Standard errors are in parenthesis. Except age, education and log household income, all others are binary variables.

Table A2. Mean differences in characteristics of the groups of household heads

	All household heads	Individualistic	Collectivistic	Individualistic (mean)	Collectivistic (mean)	Difference in means	t-statistic
Electricity theft (binary)	23,247	6,250	16,997	0.077 (0.003)	0.091 (0.002)	-0.015*** (0.004)	-3.553
Log fuel kerosene expenditure	23,247	6,250	16,997	7.948 (0.014)	8.043 (0.007)	-0.095*** (0.015)	-6.155
Log total hh electricity expenditure	23,247	6,250	16,997	6.190 (0.038)	6.176 (0.024)	0.014 (0.045)	0.306
Consumption inequality (Gini)	23,247	6,250	16,997	0.402 (0.0002)	0.402 (0.0001)	-0.0002 (0.0002)	-0.725
Outstanding debt (binary)	23,148	6,204	16,944	0.246 (0.005)	0.228 (0.003)	0.019 (0.006)	3.002
BPL card holder (binary)	23,247	6,250	16,997	0.448 (0.006)	0.459 (0.004)	-0.011 (0.007)	-1.534
Education years	23,247	6,250	16,997	5.332 (0.060)	5.303 (0.036)	0.029 (0.070)	0.418
Log household income	23,247	6,250	16,997	11.155 (0.012)	11.144 (0.008)	0.012 (0.015)	0.811
Political connections (binary)	23,247	6,250	16,997	0.373 (0.006)	0.407 (0.004)	-0.033*** (0.007)	-4.612
Casual job (binary)	23,247	6,250	16,997	0.447 (0.006)	0.474 (0.004)	-0.027*** (0.007)	-3.637
Self-employed (binary)	23,247	6,250	16,997	0.704 (0.006)	0.721 (0.003)	-0.017** (0.007)	-2.533
Connection to govt. employee	23,210	6,241	16,969	0.278	0.328	-0.049***	-7.179

Connection to big officer	23,215	6,245	16,970	(0.006) 0.155	(0.004) 0.180	(0.007) -0.025***	-4.399
Newspaper reading in HH	23,247	6,250	16,997	(0.005) 0.476	(0.003) 0.499	(0.006) -0.023***	-3.091
TV watching in family	23,247	6,250	16,997	(0.006) 0.821	(0.004) 0.829	(0.007) -0.008	-1.463
Years lived in the place	23,247	6,250	16,997	(0.005) 77.788	(0.003) 80.418	(0.006) -2.631***	-7.639
Local crime dummy	23,247	6,250	16,997	(0.321) 0.069	(0.172) 0.047	(0.344) 0.022***	6.543
Civic meeting participation dummy	23,247	6,250	16,997	(0.003) 0.297	(0.002) 0.355	(0.003) -0.058***	-8.266
Consumption expenditure percapita	23,246	6,249	16,997	(0.006) 24943.68	(0.004) 23285.90	(0.007) 1657.76***	4.865
				(308.41)	(172.72)	(340.75)	

Notes: Statistical significance is shown as ***p < 0.01. Standard errors are in parenthesis. Except age, education and log household income, all others are binary variables.

Supplementary Appendix 3

Table- A3. Electricity theft and collectivism-individualism cleavages: Recursive bivariate probit model results

Dependent variables:	National		Rural		Urban	
	Collectivistic dummy	Electricity theft dummy	Collectivistic dummy	Electricity theft dummy	Collectivistic dummy	Electricity theft dummy
Collectivistic Culture		-0.464*** [-0.046] (0.177)		-0.368* [-0.042] (0.204)		-0.960*** [-0.058] (0.330)
Outstanding debt with local shop dummy	-0.0405 (0.0457)	0.177*** [0.016] (0.0481)	-0.00506 (0.053)	0.172*** [0.019] (0.05)	-0.156*** (0.0560)	0.188** [0.010] (0.0843)
Log fuel expenditure		0.0724*** [0.007] (0.0200)		0.0751*** [0.009] (0.0222)		0.0828* [0.005] (0.0436)
Log electricity expenditure		-0.229*** [-0.023] (0.0124)		-0.235*** [-0.027] (0.0131)		-0.204*** [-0.012] (0.0150)
Consumption inequality Gini	-2.474*** (0.264)	-8.986*** [-0.942] (0.326)	-2.643*** (0.349)	-9.197*** [-1.128] (0.371)	-11.91*** (1.626)	-27.65*** [-1.778] (7.478)
Household subjective wellbeing: better off	0.000796 (0.0184)	-0.0850* [-0.008] (0.0438)	0.007 (0.027)	-0.0999* [-0.011] (0.0551)	-0.00668 (0.0564)	0.0571 [0.003] (0.0650)

Household subjective wellbeing: worse off	-0.108*** (0.0361)	-0.0566 [-0.008] (0.0765)	-0.141*** (0.0388)	-0.0941 [-0.015] (0.0903)	-0.00876 (0.0732)	0.120 [0.007] (0.111)
Female	-0.0163 (0.0254)	-0.0417 [-0.004] (0.0593)	-0.0263 (0.0247)	-0.0644 [-0.008] (0.0596)	-0.00564 (0.0638)	0.114 [0.007] (0.112)
Age	-0.00249*** (0.0007)	-0.00710*** [-0.001] (0.00128)	-0.0025*** (0.000912)	-0.006*** [-0.001] (0.002)	-0.00291** (0.00119)	-0.0138*** [-0.001] (0.00366)
BPL card holder	0.0613 (0.0382)	-0.170*** [-0.015] (0.0454)	0.0839* (0.0502)	-0.169*** [-0.017] (0.0513)	-0.0162 (0.0484)	-0.185** [-0.011] (0.0791)
Education years	-0.00344 (0.00517)	-0.0216*** [-0.002] (0.00452)	0.001 (0.006)	-0.019*** [-0.002] (0.005)	-0.0136** (0.006)	-0.0310*** [-0.002] (0.011)
Social group: Forward caste	-0.150 (0.109)	-0.299*** [-0.033] (0.0961)	-0.102 (0.152)	-0.272*** [-0.034] 0.094	-0.202** (0.0824)	-0.477** [-0.031] (0.235)
Social group: Other Backward Castes	-0.0879 (0.117)	-0.320*** [-0.034] (0.0803)	-0.0126 (0.155)	-0.304*** [-0.035] (0.0898)	-0.198** (0.0807)	-0.421*** [-0.027] (0.136)
Social group: Scheduled Castes	-0.0256 (0.136)	-0.185 [-0.019] (0.120)	0.0212 (0.179)	-0.194* [-0.021] (0.107)	-0.103 (0.123)	-0.152 [-0.010] (0.204)
Social group: Scheduled Tribes	0.0742 (0.156)	-0.348*** [-0.032] (0.0777)	0.151 (0.208)	-0.403*** [-0.041] (0.0696)	-0.174 (0.120)	0.309 [0.017] (0.224)
Social group: Muslim	-0.115	-0.0327 [-0.006]	-0.116	-0.008 [-0.004]	-0.142	-0.166 [-0.011]

	(0.122)	(0.103)	(0.152)	(0.114)	(0.118)	(0.128)
Social group:		-0.603**	0.146	-0.423	-0.212	-6.378***
Christian/Sikh/Jain	-0.0294	[-0.060]		[-0.044]		[-0.390]
	(0.197)	(0.288)	(0.322)	(0.275)	(0.188)	(0.304)
Log household income	-0.0140	-0.00121	-0.0234	-0.0004	0.0248	-0.030
	(0.0231)	[-0.005]		[-0.001]		[-0.002]
		(0.0270)	(0.0291)	(0.029)	(0.0239)	(0.0506)
Main Source of Income-1	0.0285	-0.0788 [-0.007]	0.0537	-0.057 [-0.005]	-0.0768*	-0.245*
	(0.0386)					[-0.016]
		(0.0492)	(0.041)	(0.05)	(0.0461)	(0.140)
Main Source of Income-2	0.0481	-0.0301	0.0885*	-0.0437	-0.0875	-0.0957
	(0.0405)	[-0.002]		[-0.002]		[-0.007]
		(0.0799)	(0.0462)	(0.126)	(0.0587)	(0.118)
Main Source of Income-3	-0.00756	-0.0390	0.0202	-0.024	-0.138***	-0.148
	(0.0572)	[-0.004]		[-0.002]		[-0.011]
		(0.0480)	(0.077)	(0.0477)	(0.0468)	(0.101)
Local political connection	0.103**	-0.0280 [-0.0003]	0.113*	-0.021 [0.001]	0.0483	-0.00692
	(0.0438)					[-0.00004]
		(0.0368)	(0.0616)	(0.0403)	(0.0521)	(0.115)
Casual job dummy	0.0159	0.0113 [0.001]	0.0323	0.0276[0.004]	-0.0517	-0.0517
	(0.0152)					[-0.004]
		(0.0434)	(0.0239)	(0.0493)	(0.0498)	(0.105)
Newspaper reading in household	0.0917**	0.0709* [0.009]	0.0916**	0.0884**	0.0779	-0.0217
	(0.0366)			[0.013]		[-0.001]
		(0.0396)	(0.0368)	(0.0393)	(0.0683)	(0.122)
TV watching in household	0.006	-0.078* [-0.008]	-0.002	-0.08* [-0.009]	0.0342	-0.103
	(0.043)					[-0.006]
		(0.04)	(0.043)	(0.046)	(0.104)	(0.0745)

Years lived in the place	0.003*** (0.001)	0.004*** [0.0005] (0.002)	0.005*** (0.002)	0.005** [0.001] (0.003)	0.00221*** (0.0005)	0.00267* [0.0002] (0.001)
Local crime incidence dummy	-0.216*** (0.045)	-0.144* [-0.019] (0.078)	-0.229*** (0.0703)	-0.178** [-0.027] (0.089)	-0.157 (0.103)	0.0486 [0.002] (0.0889)
Civic body meeting participation	0.13*** (0.046)	-0.049 [-0.002] (0.035)	0.115** (0.0485)	-0.0647* [-0.004] (0.036)	0.139* (0.0708)	0.0267 [0.003] (0.130)
Confidence in state govt: great deal	-0.162*** (0.04)	-0.008 [-0.005] (0.054)	-0.175*** (0.0513)	-0.007 [-0.006] (0.0532)	-0.133** (0.0552)	0.0337 [0.001] (0.110)
Confidence in state govt: hardly any	-0.0518 (0.0399)	0.0101 [-0.0002] (0.0282)	-0.0397 (0.0372)	0.012 [0.0002] (0.0402)	-0.0889 (0.0644)	0.035 [0.001] (0.105)
Confidence in local body: great deal	-0.0111 (0.0214)	0.0385 [0.004] (0.0657)	-0.0419 (0.0417)	0.030 [0.002] (0.0699)	0.0862 (0.0533)	0.106 [0.007] (0.105)
Confidence in local body: hardly any	-0.196*** (0.0465)	0.0215 [-0.003] (0.0533)	-0.215*** (0.0589)	-0.009 [-0.008] (0.0617)	-0.124*** (0.0350)	0.204** [0.012] (0.104)
Confidence in police: great deal		-0.160*** [-0.015] (0.0448)		-0.157*** [-0.017] (0.0519)		-0.232** [-0.013] (0.105)
Confidence in police: hardly any		0.108 [0.011] (0.0789)		0.134* [0.016] (0.079)		-0.0377 [-0.002] (0.104)
Relative of police		0.0768 [0.008] (0.0672)		0.0901 [0.010] (0.0754)		-0.184 [-0.011] (0.274)

Urban	-0.0206 (0.065)	-0.380*** [-0.038] (0.050)				
Caste conflict	-0.304*** (0.0213)	[-0.007]	-0.233*** (0.0318)	[-0.007]	-0.497*** (0.0366)	
Constant	1.143*** (0.247)	2.607*** (0.442)	1.016*** (0.367)	2.419*** (0.584)	6.003*** (0.745)	12.75*** (3.205)
Rho $\rho(\epsilon, u)$ (unconstrained)	0.409	0.409	0.360	0.360	0.723	0.723
Wald test $\rho=0$ chi-squared (p-value)	13.382*** (0.0003)	13.382*** (0.0003)	6.816*** (0.009)	6.816*** (0.009)	9.552*** (0.002)	9.552*** (0.002)
Observations	23,148	23,148	16,201	16,201	6,947	6,947
Mean dependent variable	0.734	0.102	0.745	0.126	0.703	0.037

Note: Statistical significance is shown by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Huber-White robust standard errors clustered at the village/neighbourhood level are in parentheses. All the regressions used the village/neighbourhood fixed effects and sampling weights. For the social group dummies, the benchmark category is the Brahmin households. The bracket [] contains the mean marginal effects. Main income source dummies are defined such that it is =0 if cultivation or allied agriculture or others, =1 if labour, agriculture & non-agriculture, =2 petty trade or business, =3 salaried or profession. Cultivation or agriculture or others is the benchmark category.

Table- A4. Electricity theft and collectivism-individualism cleavages: Recursive bivariate probit model results

Dependent variable	Male household heads		Female household heads	
	Collectivistic dummy	Electricity theft dummy	Collectivistic dummy	Electricity theft dummy
Collectivistic Culture		-0.0536 [-0.0046] (0.282)		-0.130 [-0.013] (0.714)
Outstanding debt with local shop dummy	-0.0465 (0.0449)	0.180*** [0.014] (0.0565)	0.00433 (0.102)	0.245** [0.024] (0.101)
Log fuel expenditure		0.0746*** [0.006] (0.0252)		0.109** [0.011] (0.0535)
Log electricity expenditure		-0.243*** [-0.021] (0.0115)		-0.209*** [-0.020] (0.0201)
Consumption inequality Gini	-2.827*** (0.229)	-4.742*** [-4.742] (0.311)	-2.113* (1.102)	5.986*** [5.986] (1.321)
Household subjective wellbeing: better off	0.0137 (0.0254)	-0.0815 [-0.007] (0.0559)	-0.0862 (0.105)	-0.0904 [-0.011] (0.120)
Household subjective wellbeing: worse off	-0.0513 (0.047)	-0.0291 [-0.004] (0.093)	-0.298*** (0.0745)	-0.109 [-0.020] (0.162)
Age	-0.00268** (0.001)	-0.00718*** [-0.001] (0.00231)	-0.0005 (0.00163)	-0.00649 [-0.001] (0.00431)
BPL card holder	0.0590 (0.038)	-0.162*** [-0.012] (0.0512)	0.0922 (0.0750)	-0.299*** [-0.027] (0.0868)
Education years	-0.00259	-0.0203*** [-0.002]	-0.00372	-0.0396***

				[-0.004]
	(0.004)	(0.0052)	(0.0152)	(0.0110)
Social group: Forward caste	-0.181	-0.206 [-0.023]	0.0242	-0.604**
	(0.112)	(0.127)	(0.159)	[-0.059]
			0.071	(0.241)
Social group: Other Backward Castes	-0.113	-0.251*** [-0.025]		-0.564***
	(0.113)	(0.0902)	(0.183)	[-0.053]
				(0.163)
Social group: Scheduled Castes	-0.0298	-0.0935 [-0.009]	-0.00316	-0.570***
	(0.124)	(0.144)	(0.229)	[-0.056]
			0.174	(0.183)
Social group: Scheduled Tribes	0.0602	-0.361*** [-0.029]		-0.296***
	(0.145)	(0.118)	(0.232)	[-0.024]
			0.142	(0.114)
Social group: Muslim	-0.157	0.081 [0.003]		-0.367*
	(0.111)	(0.125)	(0.223)	[-0.032]
Social group: Christian/Sikh/Jain	-0.0233	-0.432 [-0.038]	-0.0482	(0.208)
	(0.212)	(0.282)	(0.214)	-1.579*** [-1.579]
				(0.334)
Log household income	-0.0245	-0.0045 [-0.001]	0.0456	0.0002 [0.001]
	(0.0233)	(0.0243)	(0.0499)	(0.072)
		-0.0853	0.05	-0.0598
Main Source of Income-1	0.0161	[-0.007]		[-0.004]
	(0.0386)	(0.0559)	(0.102)	(0.194)
		-0.0081	0.156	-0.199 [-0.015]
Main Source of Income-2	0.0295	[0.0001]		
	(0.0523)	(0.0972)	(0.0992)	(0.161)
Main Source of Income-3	-0.015	-0.007 [-0.001]	0.00495	-0.134 [-0.013]
	(0.0714)	(0.0429)	(0.0653)	(0.155)
Local political connection	0.111**	-0.070 [-0.003]	0.0489	0.0831 [0.010]

	(0.0455)	(0.0464)	(0.0657)	(0.151)
Casual job dummy	0.018	0.010 [0.001]	0.0396	-0.0003 [0.001]
	(0.0157)	(0.0445)	(0.0879)	(0.0988)
Newspaper reading in household	0.076*	0.0941** [0.010]	0.182***	-0.178 [-0.012]
	(0.0418)	(0.0476)	(0.0643)	(0.118)
TV watching in household	0.0007	-0.119** [-0.010]	-0.00634	0.0226 [0.002]
	(0.0493)	(0.0511)	(0.0666)	(0.0980)
			0.00392**	-0.0003
Years lived in the place	0.00348***	0.00456** [0.0005]		[0.0001]
	(0.00112)	(0.002)	(0.002)	(0.00135)
Local crime incidence dummy	-0.206***	-0.114 [-0.015]	-0.294**	-0.121 [-0.021]
	(0.0576)	(0.0970)	(0.114)	(0.182)
Civic body meeting participation	0.120***	-0.0884*** [-0.004]	0.202***	0.0877 [0.015]
	(0.0452)	(0.0338)	(0.0765)	(0.128)
			-0.140*	-0.155
Confidence in state govt: great deal	-0.170***	0.0401 [-0.001]		[-0.019]
	(0.0443)	(0.0507)	(0.0763)	(0.105)
			-0.0642	-0.0221
Confidence in state govt: hardly any	-0.0515	0.0196 [0.0003]		[-0.004]
	(0.0423)	(0.0373)	(0.100)	(0.0972)
			-0.0437	-0.000159
Confidence in local body: great deal	-0.00561	0.0345 [0.003]		[-0.001]
	(0.0230)	(0.0740)	(0.0583)	(0.0904)
Confidence in local body: hardly any	-0.190***	0.0250 [-0.003]	-0.231*	0.0921 [0.002]
	(0.0400)	(0.0597)	(0.128)	(0.121)
				-0.173
Confidence in police: great deal		-0.174*** [-0.014]		[-0.017]
		(0.0452)		(0.109)
Confidence in police: hardly any	0.018	0.144* [0.013]		-0.0694 [-0.007]

	(0.015)	(0.0761)		(0.135)
Relative of police	-0.023	0.0903 [0.008]		-0.106 [-0.010]
	(0.022)	(0.0823)		(0.282)
Urban	-0.0118	-0.449*** [-0.039]	-0.0665	-0.167 [-0.018]
	(0.0801)	(0.0586)	(0.0571)	(0.104)
Caste conflict dummy	-0.316*** [-0.009***]		-0.240*** [-0.007***]	
	(0.0251)		(0.0568)	
Constant	1.489***	20.15***	-0.128	-28.59***
	(0.276)	(0.473)	(0.872)	(0.976)
Rho $\rho(\epsilon, u)$ (unconstrained)	0.160	0.160	0.249	0.249
Wald test $\rho=0$ chi-squared (p-value)	0.991 (0.319)	0.991 (0.319)	0.294 (0.588)	0.294 (0.588)
Observations	19,924	19,924	3,224	3,224
Mean dependent variable	0.736	0.101	0.718	0.108

Note: Statistical significance is shown by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Huber-White robust standard errors clustered at the village/neighbourhood level are in parentheses. All the regressions used the village/neighbourhood fixed effects and sampling weights. For the social group dummies, the benchmark category is the Brahmin households. The bracket [] contains the mean marginal effects. Main income source dummies are defined such that it is =0 if cultivation or allied agriculture or others, =1 if labour, agriculture & non-agriculture, =2 petty trade or business, =3 salaried or profession. Cultivation or agriculture or others is the benchmark category.

Table- A5. Electricity theft and collectivism-individualism cleavages: Recursive bivariate probit model results

Dependent variable:	SC/ST/OBC households		Non-SC/ST/OBC households	
	Collectivistic dummy	Electricity theft dummy	Collectivistic dummy	Electricity theft dummy
Collectivistic Culture		-0.439** [-0.044] (0.185)		-0.028 [-0.002] (0.481)
Outstanding debt with local shop	-0.0363 (0.056)	0.173*** [0.016] (0.056)	-0.0511 (0.0621)	0.212*** [0.015] (0.0685)
Log fuel expenditure		0.0896*** (0.0269)		0.0423 [0.003] (0.0266)
Log electricity expenditure		-0.234*** (0.0134)		-0.235*** [-0.019] (0.0199)
Consumption inequality Gini	-1.520*** (0.386)	-5.458*** [-5.459] (0.325)	2.451*** (0.864)	
Household subjective wellbeing: better off	-0.018 (0.0304)	-0.166*** [-0.017] (0.0375)	0.0410 (0.027)	0.037 [0.004] (0.0736)
Household subjective wellbeing: worse off	-0.157*** (0.058)	-0.101 [-0.014] (0.067)	-0.0264 (0.0550)	0.045 [0.003] (0.130)
Female	-0.054 (0.044)	-0.044 [-0.006] (0.0682)	0.0834 (0.0636)	-0.030 [-0.0002] (0.114)
Age	-0.003*** (0.001)	-0.008*** [-0.001] (0.002)	-0.00286 (0.002)	-0.006*** [-0.001] (0.007)
BPL card holder	0.0701**	-0.173*** [-0.016]	0.0325	-0.184*** [-0.014]

	(0.034)	(0.055)	(0.0566)	(0.0526)
Education years	-0.009**	-0.0277*** [-0.003]	0.008	-0.0115 [-0.001]
	(0.004)	(0.009)	(0.008)	(0.0140)
Log household income	-0.019	-0.001 [-0.001]	-0.004	0.00244 [0.0001]
	(0.029)	(0.0367)	(0.019)	(0.0353)
Main Source of Income-1	0.0230	-0.111** [-0.011]	0.0262	-0.0148 [-0.0004]
	(0.0281)	(0.0533)	(0.0846)	(0.0702)
Main Source of Income-2	0.007	-0.103 [-0.010]	0.105	0.118 [0.013]
	(0.0425)	(0.0901)	(0.0712)	(0.0901)
Main Source of Income-3	-0.0197	-0.039 [-0.005]	0.009	-0.040 [-0.003]
	(0.0716)	(0.0665)	(0.0615)	(0.0893)
Local political connection	0.085	-0.017 [0.0004]	0.141**	-0.080 [-0.003]
	(0.057)	(0.049)	(0.0584)	(0.0504)
Casual job dummy	0.022	-0.010 [-0.0004]	0.0164	0.0616 [0.005]
	(0.025)	(0.056)	(0.0295)	(0.0840)
Newspaper reading in household	0.113***	0.120* [0.015]	0.0692	-0.0306 [-0.001]
	(0.0363)	(0.0731)	(0.0689)	(0.0582)
TV watching in household	0.001	-0.0821* [-0.008]	-0.00378	-0.104 [-0.008]
	(0.0448)	(0.0423)	(0.0790)	(0.0990)
Years lived in the place	0.004***	0.005** [0.001]	0.00260***	0.00279 [0.0003]
	(0.001)	(0.002)	(0.001)	(0.003)
Local crime incidence dummy	-0.180***	-0.149** [-0.019]	-0.271***	-0.122 [-0.017]
	(0.0533)	(0.0736)	(0.0543)	(0.126)
Civic body meeting participation	0.170***	-0.056 [-0.001]	0.038	-0.0699 [-0.005]

	(0.051)	(0.0458)	(0.0561)	(0.0771)
Confidence in state govt: great deal	-0.137***	-0.0021 [0.004]	-0.211***	0.00155 [-0.005]
	(0.0501)	(0.0777)	(0.0482)	(0.0613)
Confidence in state govt: hardly any	-0.107**	-0.0545 [-0.008]	0.0530	0.160** [0.015]
	(0.0525)	(0.0424)	(0.0507)	(0.076)
Confidence in local body: great deal	-0.036	0.065 [0.006]	0.0455	-0.0249 [-0.001]
	(0.0306)	(0.0783)	(0.0435)	(0.0674)
Confidence in local body: hardly any	-0.206***	0.0605 [0.001]	-0.180***	-0.0438 [-0.008]
	(0.0569)	(0.0705)	(0.0530)	(0.094)
Confidence in police: great deal		-0.237*** [-0.022]		-0.0311 [-0.002]
		(0.0502)		(0.0945)
Confidence in police: hardly any		0.157* [0.017]		0.008 [0.001]
		(0.0878)		(0.105)
Police relative dummy		0.119 [0.012]		0.027 [0.002]
		(0.089)		(0.0605)
Urban	-0.0280	-0.337*** [-0.034]	-0.0342	-0.398*** [-0.032]
	(0.077)	(0.088)	(0.080)	(0.101)
Caste conflict dummy	-0.283*** [-0.007***]		-0.358*** [-0.009***]	
	(0.029)		(0.0528)	
Constant	0.984***	19.78***	-1.893***	-21.84***
	(0.246)	(0.699)	(0.257)	(0.526)
Rho $\rho(\epsilon, \nu)$ (unconstrained)	0.369*** (0.123)	0.369*** (0.123)	0.198 (0.280)	0.198 (0.280)
Wald test $\rho=0$ chi-squared (p-value)	8.974*** (0.002)	8.974*** (0.002)	0.499 (0.480)	0.499 (0.480)

Observations	15,259	15,259	7,889	7,889
Mean dependent variable	0.743	0.107	0.715	0.092

Note: Statistical significance is shown by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Huber-White robust standard errors clustered at the village/neighbourhood level are in parentheses. All the regressions used the village/neighbourhood fixed effects and sampling weights. For the social group dummies, the benchmark category is the Brahmin households. The bracket [] contains the mean marginal effects. Main income source dummies are defined such that it is =0 if cultivation or allied agriculture or others, =1 if labour, agriculture & non-agriculture, =2 petty trade or business, =3 salaried or profession. Cultivation or agriculture or others is the benchmark category.

Concluding remarks

So, we ran a recursive bivariate probit model across specifications and find that in regional samples (in Table A3 in the Supplementary Appendix) imply the collectivistic dummy is endogenous (as the Wald test null hypothesis that $\rho=0$ is rejected) whereas when we subject it to heterogeneity analysis, we find that in case of male, female and non-SC-ST-OBC household head samples, the correlation of error terms is zero meaning we can go with a simple probit model. The evidence of endogeneity of the collectivistic culture dummy is ambiguous. In any case, our making use of the RBP model should allay some concerns of the endogeneity problem. The results for the aforesaid sub-samples (male, female and non-SC-ST-OBC) show it is unnecessary to make use of the RBP model and we can go ahead with using simple probit models as the correlation coefficient between the two error terms is not significantly different from zero and therefore, it implies that the collectivistic dummy is not endogenous. The empirical tests do not provide sufficient evidence of joint endogeneity to justify retaining the recursive bivariate probit model as the preferred specification.

References

- Greene, W.H., (1998). Gender economics courses in liberal arts colleges: Comment. *Journal of Economic Education*, 29(4): 291–300.
- Greene, W. H., (2012). *Econometric Analysis*. Seventh Edition. Prentice Hall: New Jersey.
- Monfardini, C., and R. Radice, 2008. Testing exogeneity in the bivariate probit model: A Monte Carlo study. *Oxford Bulletin of Economics and Statistics*, 70(2): 271–282.
- Panda, S., 2022. Does local elite capture vary by levels of political connections? Evidence from an Indian public housing program. *European Journal of Development Research*, 34(4), 1912–1939.
- Panda, S., 2015. Political connections and elite capture in a poverty alleviation programme in India. *Journal of Development Studies*, 51(1), 50–65.
- Waters, H. R., 1999. Measuring the impact of health insurance with a correction for selection bias – a case study of Ecuador. *Health Economics*, 8, pp. 473–483.

Wilde, J., 2000. Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters*, 69, pp. 309–312.