

Online Appendix for “Clientelism in Indian Villages”

By SIWAN ANDERSON, PATRICK FRANCOIS AND ASHOK KOTWAL

Appendix A: Summary Statistics

Appendix B: Alternative Estimations

Appendix C: Additional Theoretical Results and Proofs

Appendix D: Independence of MLD and MPROP

ONLINE APPENDIX A: SUMMARY STATISTICS

Table A1 - Summary Statistics - GP and Village Measures

Variable	Mean	Standard Deviation	Observations
Maratha Pradhan	0.41	0.49	300
Reserved	0.58	0.49	319
Population	2271.2	659.9	319
Proportion Maratha	0.41	0.31	310
Proportion SC/ST	0.25	0.17	310
Maratha Land Dominated	0.61	0.49	320
Distance to Water	2.85	2.19	318
Distance to Road	2.60	2.75	318
Distance to Rail	22.6	18.9	318
River/Canal	0.26	0.44	320
Topsoil Nitrogen	2.02	0.89	318
Topsoil Organic Carbon	0.21	0.93	318
Topsoil Ph	0.53	1.21	318
Rainfall	70.98	20.06	318
Longitude	76.21	1.19	320
Latitude	19.46	1.04	320
Elevation	483.0	138.4	320
All Programs	5.36	2.53	304
BPL Programs	1.73	0.88	304
EGS	0.20	0.21	304
Income Programs	4.85	2.33	304
Non-Income Programs	0.51	0.28	304
Revenu (1)	149.8	360.8	229
Revenue (2)	9.7	27.6	318
Expenditure	8.9	25.4	318
BDO Meetings	3.22	6.31	319
MP Meetings	1.72	8.23	318
DC Meetings	1.26	4.59	319

Note: Information on the caste of the Pradhan (the elected leader of the village government) and whether the position is reserved comes from our GP Questionnaire. Village population numbers, caste proportions, and caste land ownership patterns come from our Village Questionnaire. SC/ST refers to the Scheduled Castes and Schedule Tribes, the lowest ranking group in the caste hierarchy. Maratha land dominated is equal to one if a village is dominated by Marathas in terms of land ownership and equal to zero if instead the majority of landholdings in the village are in the hands of a lower caste. Distance to water, road, and rail come from GPS Census data. Whether there is a canal or river in the village comes from the 2001 Village Census. The three variables pertaining to topsoil (30 cm) content come from FAO-UNESCO soil maps. Rainfall information, which is only available at the district level, comes from the India Meteorological Department. Latitude, Longitude, and Elevation measures come from the GPS Census Data. Total Programs refers to the total number of the 15 Government Schemes implemented in the village.^a BPL refers to the number of the possible 8 programs targeted at individuals below the poverty line (these include: Housing Support Scheme; Sanitation Support Scheme; Indira Awas Yojana IAY, a housing construction program; Targeted Public Distribution System (TPDS)). EGS refers to the Employment Guarantee Scheme, which is the precursor to the present NREGA. Income programs is the set of programs that likely directly or indirectly affect household labour decisions and income (such as public good and housing improvement schemes which do rely on villagers' labour in their construction and the targeted public distribution system). Non-Income Programs refers to programs which do not directly affect household labour decisions such as those targeted towards children (child development (ICDS) and mid-day meals) and the elderly (social security pensions and foodgrains (Annapurna)). We obtained information on the availability of programs from our household survey and aggregated this information to the village level. Revenue (1) refers to data collected from the balance sheets of submitted by the GPs, these are computed per capita of the GP population. We obtained the majority of this information using the RTI Act. The information covers the last 24 months. Revenue (2) and Expenditure are annual per capita values from the 2001 Village Census. BDO (Block Development Officer) , MP (Member of Parliament), and DC (District Collector) meetings all refer to the number of times in the last year that the Gram Pradhan has met with officials from higher level governments to seek resources. This information is from the GP Questionnaire.

^aThe complete list of programs: Housing Support Scheme; Sanitation Support Scheme; EGS; SGYR (Sampoom Gram Rojgar Yojana); IAY (Indira Awas Yojana); SGSY (Swamjayanti Grameen Sawa Rogar Yojana); ICDS (Integrated Child Development Scheme); Social Security Pension, Mid-day Meal Program; ARWSP (Accelerated Rural Water Supply Program); PMGSY (Pradhan Mantri Gram Sadak Yojana); TPDS (Targeted Public Distribution System); Annapurna; Watershed Development Programs under DRAP and DDP; Total Sanitation Campaign; Swajaidhara; Business Support Program; FFW (Food for work program); PDS (Public Distribution Scheme).

Table A2 - Summary Statistics - Household Measures

Variable	Mean	Standard Deviation	Observations
Maratha	0.38	0.49	8671
OBC	0.31	0.46	8671
SC/ST	0.29	0.45	8671
Landless	0.34	0.47	9136
Low Land (0-2.5 Acres)	0.21	0.41	9136
Large Land Owners (>5 Acres)	0.29	0.45	9136
Less than Primary School	0.28	0.45	8948
Insured (1)	0.63	0.48	9135
Insured (2)	0.61	0.49	9132
Insured (3)	0.62	0.49	9129
Insured (4)	0.64	0.48	9133
Insured (5)	0.63	0.48	9127
Insured (6)	0.66	0.47	9120
Insurer	0.79	0.41	9128
Voted-Personal	0.24	0.43	8266
Trust	2.57	0.95	9110
No Cheat	0.89	0.31	8772
Repair	0.36	0.48	9084
Donated Cash	0.52	0.50	9140
Donated Labour	0.29	0.45	9140
Agree	0.55	0.50	9133
Share Water	0.70	0.46	5093
Target Village	2.20	0.58	8964
Shared Funds	1.65	1.20	8329
Festivals	0.59	0.49	9132

Note: All information comes from our Household Questionnaire. OBC refers to Other Backward Castes and SC/ST refers to Scheduled Castes and Scheduled Tribes. These caste groups are ranked below Marathas in the caste hierarchy, where the SC/ST category is the lowest ranked. Household education is measured by the highest level of education that any household male has reached. Less than primary refers to the highest category being less than primary school. Insured (1): "Would most people in your village help you with some money in times of crisis?". Insured (2): "Would a higher caste member of your village help you with some money in times of crisis?". Insured (3): "Would most people in your village help a lower caste villager with some money in times of crisis?". Insured (4)-(6) are the same questions with "money" replaced by "grain". Insurer: "Suppose a lower caste man asks to borrow a good sum of money from you because someone in his family has fallen ill. He is from the village and has the ability to repay the amount. Would you lend it to him?". Voted - Personal equals to one if the household voted for a candidate due to a personal connection rather than due to the characteristics of the candidate (honesty, good reputation, qualifications). Samples are conditional on voting. Trust is response to: "Would you say that the large landholders can be trusted? 1=Almost none, 2=Some; 3=Majority; 4=Almost. Cheat refers to answering someone from a higher caste is most likely to cheat you (compared to other castes or wealth levels). Repair is the answer to "If someone from your village noticed something wrong on your farm they would?" repair it themselves (compared to conditional answers, such as "alert you if he is from a lower caste...etc). Donated cash or labour are dummy variables equal to one if the household did donate (cash or labour respectively) in the past year to a development project within the village. Agree refers to answering that most people in the village would agree on the type of development project the village should have (compared to differences of opinions within the village). Share Water is equal to 1 if the household shares a water source with members of the Maratha caste. Samples are conditional on sharing a water source. Target Village refers to GP funds should be targeted to the village as a whole, compared to poor or low caste individuals. Shared funds refers to GP funds are shared across the village (e.g. for development projects; public goods) compared to going directly to the poor or low status; the rich and high status; or to GP members or other government officials directly. Festivals is equal to one if there are village projects to finance festivals.

Table A3 - Summary Statistics - Household and Individual Measures

Variable	Mean	Standard Deviation	Observations
Daily Wage	41.49	15.96	15004
Male	0.55	0.50	15021
Illiterate	0.43	0.50	15014
Age	39.4	15.7	15007
Log Kharif Yields	8.91	1.17	5539
Log Kharif Profit	8.31	1.43	4269
Labour/Total Costs (Kharif)	0.31	0.23	5648
Household Members	5.51	2.63	9132
Jowar	0.45	0.50	5874
Rainfed	0.68	0.47	6105
Black Soil	0.53	0.50	6128
Clay Soil	0.64	0.48	6122
Salinity	0.18	0.39	6125
Percolation	0.29	0.45	6126
Drainage	0.29	0.46	6127
Maratha Trader	0.42	0.49	6341
Outside Maratha Trader	0.26	0.44	5945
Maratha Lender	0.43	0.49	901
Terms of Payment (Inputs)	1.02	0.47	21496
Interest Rate on Loan	20.0	22.1	920

Note: The sample of labourers are all those who work for a daily wage in agriculture. The gender, literacy rate, and age of these workers are reported above. Yields, profits, and proportion of labour costs are all measured per acre of land. Kharif yields are the total value of output per acre of land for a given crop, summed over all of the kharif crops for each household. Kharif profit is yields net of input costs (seeds, fertilizer, irrigation, electricity, pesticides, and labour). Workers include parttime and fulltime. Jowar is a dummy variable equal to one if the household grows this main staple crop. Rainfed refers to the percentage of land which is rainfed as opposed to irrigated. Black soil, clay soil, salinity, percolation, and drainage are all measures of the soil quality of the household land. Maratha Trader is equal to one if the household has traded with a Maratha for any tradeable good (which includes agricultural inputs and outputs, farm enterprise and non-farm enterprise goods) conditional on trading goods. Outside Maratha Trader refers to the trader residing outside of the village conditional on trading goods. Maratha Lender refers to borrowing money from a Maratha. Terms of payments is an index variable equal to 0 if the trader requires advanced payments; 1 if full payment is required at the time of sale; and 2 if instead payment in installments is acceptable. Terms of payments and interest rate are reported per individual loan.

Table A4 - Characteristics by Caste

Variable	Marathas	OBCs	SCs
Cultivator (Overall)	0.83 (0.38)	0.65 (0.48)	0.33 (0.47)
Cultivator (Maratha Land Dominated)	0.84 (0.36)	0.66 (0.47)	0.35 (0.48)
Agricultural Labourer (Overall)	0.10 (0.30)	0.19 (0.40)	0.53 (0.50)
Agricultural Labourer (Maratha Land Dominated)	0.09 (0.29)	0.18 (0.38)	0.52 (0.50)
Landless (Overall)	0.13 (0.34)	0.31 (0.46)	0.62 (0.48)
Landless (Maratha Land Dominated)	0.12 (0.32)	0.29 (0.45)	0.60 (0.49)
Average Land Owned (Overall)	6.74 (7.38)	6.00 (6.22)	3.67 (3.28)
Average Land Owned (Maratha Land Dominated)	6.86 (7.53)	5.70 (5.96)	3.47 (2.99)
> 5 Acres (Overall)	0.39 (0.49)	0.35 (0.48)	0.15 (0.36)
> 5 Acres (Maratha Land Dominated)	0.40 (0.49)	0.33 (0.47)	0.14 (0.35)
> 10 Acres (Overall)	0.16 (0.37)	0.13 (0.33)	0.03 (0.18)
> 10 Acres (Maratha Land Dominated)	0.16 (0.37)	0.12 (0.33)	0.03 (0.18)
Voted	0.89 (0.31)	0.93 (0.26)	0.90 (0.30)
Supposed to Vote	0.08 (0.28)	0.09 (0.29)	0.10 (0.29)
Forced Vote	0.002 (0.04)	0.0004 (0.02)	0.001 (0.03)
Unopposed Election	0.09 (0.29)	0.04 (0.21)	0.07 (0.26)
Raise concerns to Pradhan	0.96 (0.19)	0.96 (0.18)	0.96 (0.18)
Met Pradhan	0.97 (0.16)	0.95 (0.22)	0.96 (0.20)
Observations	3259	2659	2019

Note: Standard deviations are in parentheses. OBC refers to Other Backward Castes and SC/ST refers to Scheduled Castes and Scheduled Tribes. These caste groups are ranked below Marathas in the caste hierarchy, where the SC/ST category is the lowest ranked. Occupation categories (Cultivator and Agricultural Labourer) refer to main source of livelihood for household. Total land owned is in acres and are reported conditional on owning land. Voted refers to voted in the last GP election. Supposed to vote refers to "supposed to vote - does not mean anything". Forced vote refers to forced to vote for a given candidate by friends, family, or villagers. Unopposed election - refers to single candidate election (this was the main reason for not voting). The fourth variable is the response to "Do you feel you can raise concerns (bring oral requests) directly to the Gram Pradhan?"

Table A5- Control of Panchayat Measures

Variable	Overall	<i>MLD</i>	<i>MLD</i>	<i>MLD</i>
			Maratha Majority	Non-Maratha Majority
Population Proportion of Marathas	0.41 (0.31)	0.54 (0.26)	0.71 (0.13)	0.28 (0.19)
Maratha Pradhan	0.41 (0.49)	0.57 (0.50)	0.64 (0.48)	0.45 (0.50)
Maratha Pradhan - Unreserved	0.63 (0.48)	0.82 (0.38)	0.94 (0.24)	0.64 (0.49)
Maratha Pradhan - Reserved for Women	0.62 (0.49)	0.89 (0.31)	0.95 (0.23)	0.78 (0.44)
Reserved Pradhan	0.58 (0.49)	0.57 (0.50)		
Reserved Pradhan - Women	0.27 (0.45)	0.26 (0.44)		
Reserved Pradhan - SC/ST	0.24 (0.43)	0.24 (0.43)		
Reserved Pradhan - OBC	0.49 (0.50)	0.50 (0.50)		
Proportion Reserved on GP	0.59 (0.19)	0.56 (0.16)		
Observations	315	193	120	73

Note: MLD denotes Maratha Land Dominated . Data on proportion of Marathas comes from the village survey. Data on characteristics of the Pradhan (elected leader of village government) come from the GP Survey. OBC refers to Other Backward Castes and SC/ST refers to Scheduled Castes and Scheduled Tribes. These caste groups are ranked below Marathas in the caste hierarchy, where the SC/ST category is the lowest ranked.

ONLINE APPENDIX B: ALTERNATIVE ESTIMATIONS

B1. Estimations without Village Controls

As a robustness check, we ran analogous estimations to those reported in Tables 2 to 5, which exclude all of the village-level controls. These estimation results are reported in Tables B1 and B2 below. We see that the main results on the significance of the coefficients on *MLD* and *MLD · MPROP*, discussed in the paper, all go through. We have also run estimations which additionally exclude all of the household level controls in the household level regressions. Though not reported here, the main results are also robust to excluding these control variables as well.

Table B1 - Estimations of GP Measures without Village Controls

Dependent Variable	Coefficient (β_1) <i>MLD</i>	Coefficient (β_2) <i>MPROP</i>	Coefficient (β_3) <i>MLD · MPROP</i>	Coefficient $\beta_1 + \beta_3$	Observations
Maratha Pradhan	0.47 (0.13)***	0.93 (0.20)***	-0.50 (0.26)**	-0.03 (0.18)	292
All Programs (2)	-1.48 (0.53)***	-1.55 (0.86)*	1.82 (1.08)*	0.34 (0.76)	293
BPL Programs (2)	-0.58 (0.18)***	-0.51 (0.29)*	0.83 (0.37)**	0.25 (0.26)	293
EGS (2)	-0.10 (0.04)**	0.03 (0.10)	0.19 (0.11)*	0.09 (0.09)	293
Income Programs (2)	-1.42 (0.50)***	-1.49 (0.80)*	1.88 (1.01)*	0.45 (0.70)	293
Non-Income Programs (2)	-0.05 (0.05)	-0.06 (0.08)	-0.06 (0.10)	-0.12 (0.08)	293
Revenue (1)	-125.7 (61.8)**	-131.2 (198.2)	82.4 (208.0)	-43.3 (216.2)	222
Revenue (2)	-13.5 (4.9)***	-18.9 (6.5)***	27.9 (9.6)***	14.4 (5.1)	308
Expenditure	-12.4 (4.8)***	-17.8 (6.2)***	26.0 (9.4)***	13.6 (4.9)***	308
BDO Meetings	-2.35 (1.20)**	0.21 (3.86)	-0.09 (3.9)	-2.44 (3.32)	290
MP Meetings	-2.92 (1.46)**	-2.71 (1.76)	2.74 (1.75)	-0.18 (1.19)	290
DC Meetings	-2.19 (1.00)**	-3.91 (1.49)***	2.96 (1.42)**	0.77 (0.64)	290
Meetings (AES)	-2.5 (1.2)**	-2.15 (2.0)	1.9 (2.4)	-0.6 (1.7)	289

Note: All estimations include regional fixed effects. A single asterix denotes significance at the 10% level, double for 5%, and triple for 1%. Robust standard errors are in parentheses. Acronyms used are: Maratha land dominated (MLD); Maratha population proportion (MPROP); Below Poverty Line (BPL); Employment Guarantee Scheme (EGS); Block Development Officer (BDO); District Collector (DC); Member of Parliament (MP); and Average Effect Size (AES). Programs (2); BPL Programs (2); EGS (2), Income Programs (2), and Non-Income Programs (2) are village level variables defined as in the notes to Table 2. Revenue (1) refers to data collected from the balance sheets (covers last 24 months) submitted by the GPs (obtained using RTI Act). Revenue (2) and Expenditure are annual per capita values from the 2001 Village Census. Meetings (AES) is the estimated average effect size of the three variables: BDO Meetings, DC Meetings, and MP Meetings.

Table B2 - Estimations of Household Measures without Village Controls

Dependent Variable	Sub-Sample	Coeff. (β_1) <i>MLD</i>	Coeff. (β_2) <i>MPROP</i>	Coeff (β_3) <i>MLD · MPROP</i>	Coefficient $\beta_1 + \beta_3$	Obs
Insured (AES)	Landless	0.11 (0.03)***	0.02 (0.06)	-0.18 (0.07)**	-0.06 (0.06)	2583
	Large Land Owners	0.08 (0.02)***	0.06 (0.05)	-0.19 (0.06)***	-0.10 (0.05)**	2519
Daily Wage	All Labourers	-2.06 (0.99)**	0.69 (2.0)	3.79 (2.36)*	1.73 (1.93)	13581
Daily Wage	Males	-2.45 (1.18)**	-1.48 (2.31)	4.99 (2.87)*	2.54 (2.38)	7502
Log Kharif Yields	Large Land Owners	0.29 (0.12)***	0.23 (0.17)	-0.50 (0.23)**	-0.21 (0.16)	2334
Log Kharif Profit	Large Land Owners	0.31 (0.15)**	0.24 (0.28)	-0.82 (0.32)***	-0.50 (0.24)**	1849
Maratha Trader	Low Castes	0.13 (0.04)***	0.22 (0.09)	-0.05 (0.11)	0.08 (0.09)	3025
Outside Maratha Trader	Low Castes	0.11 (0.03)***	0.19 (0.06)***	-0.19 (0.08)***	-0.08 (0.06)	2800
Maratha Lender	Low Castes	0.29 (0.07)***	0.47 (0.15)***	-0.23 (0.19)	0.06 (0.14)	454
Terms of Payment (Inputs)	Low Castes	0.09 (0.05)*	0.27 (0.10)***	-0.27 (0.11)**	-0.18 (0.09)*	10044
Interest Rate on Loan	Low Castes	-8.60 (4.0)**	7.85 (8.70)	1.78 (11.54)	-6.81 (9.89)	252
Voted-Personal	Low Castes	0.09 (0.03)***	0.15 (0.08)	-0.22 (0.10)**	-0.12 (0.08)	2121
Social Capital (AES)	Low Castes	0.07 (0.01)***	0.03 (0.02)	-0.12 (0.03)***	-0.06 (0.02)**	4176
Share Water	Low Castes	0.31 (0.06)***	0.55 (0.14)***	-0.34 (0.16)**	-0.02 (0.12)	2947
Target Village	Low Castes	1.09 (0.31)***	0.70 (0.59)	-1.73 (0.72)***	-0.64 (0.52)	4888
Shared Funds	Low Castes	0.95 (0.30)***	1.43 (0.65)**	-1.87 (0.75)***	-0.92 (0.62)	4608

Note: All estimations include household-level controls (education, land ownership, and caste identity), and regional fixed effects. Regression disturbance terms are clustered at the village level. A single asterix denotes significance at the 10% level, double for 5%, and triple for 1%. Acronyms used are: Maratha land dominated (MLD); Maratha population proportion (MPROP); and Average Effect Size (AES). Insured (AES) is the estimated average effect size of the six variables, Insured (1) to (6), defined in the notes of Table 3. Insurer is a dummy variable equal to one if respondents answer yes to: "Suppose a lower caste man asks to borrow a good sum of money from you because someone in his family has fallen ill. He is from the village and has the ability to repay the amount. Would you lend it to him?". Large land owners have ≥ 5 acres. The sample of labourers are all those who work for a daily wage in agriculture. Additional individual controls (gender, age, education) are included in the wage estimations. Regression disturbance terms are clustered at the household and village level for these estimations using the approach of Cameron, Gelbach and Miller (2011). The sample for the yields, profits, proportion of labour costs regressions is all large cultivators (>5 acres of land). All measures are per acre of land. Kharif yields are the total value of output per acre of land for a given crop, summed over all of the kharif crops for each household. Kharif profit is yields net of input costs (seeds, fertilizer, irrigation, electricity, pesticides, and labour). Workers include partime and fulltime, same results held if restricted ourselves to fulltime workers. Additional crop controls are included in the yields and profits estimations. Maratha Trader is equal to one if the household has traded with a Maratha for any tradeable good (which includes agricultural inputs and outputs, farm enterprise and non-farm enterprise goods) conditional on trading goods. Outside Maratha Trader refers to the trader residing outside of the village conditional on trading goods. Maratha Lender refers to borrowing money from a Maratha. These estimations on Maratha trading relationships are probit estimations, where the coefficients reported are the partial derivatives of the predicted probability. Terms of payments is an index variable equal to 0 if the trader requires advanced payments; 1 if full payment is required at the time of sale; and 2 if instead payment in installments is acceptable. These are ordered probit estimations. Voted - Personal equals to one if the household voted for a candidate due to a personal connection rather than due to the characteristics of the candidate (honesty, good reputation, qualifications). Samples are conditional on voting. The sample of low castes in the voting regressions is SC/STs. Social Capital (AES) is the estimated average effect size of the six variables: Trust, No Cheat, Repair, Donated Cash, Donated Labour, and Agree as defined in the notes of Table 5. Target Village refers to GP funds should be targeted to the village as a whole, compared to poor or low caste individuals. Shared funds refers to GP funds are shared across the village (e.g. for development projects; public goods) compared to going directly to the poor or low status; the rich and high status; or to GP members or other government officials directly. These two estimations are estimated as multinomial logit models.

B2. *Estimations with Village Controls Interacted with MPROP*

As a robustness check, we ran analogous estimations to those reported in Tables 2 to 5 in the paper, which include additional interaction terms. In these estimations the key village-level variables, discussed in Table 1 (measuring presence of river/canal, distance to natural water sources, distance to railways and national roads, soil quality measures, proportion of the population that is SC/ST as well as total village population) are interacted with *MPROP*. These estimation results are reported in Tables B3 and B4 below. We see that the main results on the significance of the coefficients on *MLD* and *MLD · MPROP*, discussed in the paper, again all continue to hold.

Table B3 - *Estimations of GP Measures with Village Controls Interacted with MPROP*

Dependent Variable	Coefficient (β_1) <i>MLD</i>	Coefficient (β_2) <i>MPROP</i>	Coefficient (β_3) <i>MLD · MPROP</i>	Coefficient $\beta_1 + \beta_3$	Observations
Maratha Pradhan	0.44 (0.14)***	0.83 (0.53)	-0.59 (0.28)**	-0.14 (0.19)	290
All Programs (2)	-1.27 (0.59)**	-2.23 (2.01)	2.02 (1.12)*	0.75 (0.75)	291
BPL Programs (2)	-0.43 (0.21)**	-0.35 (0.70)	0.77 (0.39)**	0.34 (0.26)	291
EGS (2)	-0.08 (0.04)*	0.22 (0.21)	0.21 (0.11)**	0.14 (0.08)	291
Income Programs (2)	-1.25 (0.55)**	-2.17 (1.91)	2.12 (1.05)**	0.86 (0.70)	291
Non-Income Programs (2)	-0.02 (0.05)	-0.05 (0.18)	-0.10 (0.11)	-0.11 (0.07)	291
Revenue (1)	-188.1 (112.7)*	165.0 (328.3)	146.8 (276.4)	-41.3 (233.7)	220
Revenue (2)	-13.1 (6.7)***	-20.6 (31.1)	30.2 (12.8)***	17.1 (7.0)***	307
Expenditure	-12.6 (6.6)**	-17.6 (31.1)	29.0 (12.6)**	16.4 (6.8)**	307
Meetings (AES)	-1.2 (1.1)	5.9 (4.2)	-0.4 (2.1)	-1.6 (1.5)	289

Note: All estimations include village-level controls (presence of river/canal, distance to natural water sources, distance to railways and national roads, soil quality measures, proportion of the population that is SC/ST, and total village population) and regional fixed effects. The estimations also include interaction terms between each of these village-level controls and *MPROP*. A single asterix denotes significance at the 10% level, double for 5%, and triple for 1%. Robust standard errors are in parentheses. Acronyms used are: Maratha land dominated (*MLD*); Maratha population proportion (*MPROP*); Below Poverty Line (*BPL*); Employment Guarantee Scheme (*EGS*); and Average Effect Size (*AES*). *Programs (2)*; *BPL Programs (2)*; *EGS (2)*, *Income Programs (2)*, and *Non-Income Programs (2)* are village level variables defined as in the notes to Table 2. *Revenue (1)* refers to data collected from the balance sheets (covers last 24 months) submitted by the GPs (obtained using RTI Act). *Revenue (2)* and *Expenditure* are annual per capita values from the 2001 Village Census. *Meetings (AES)* is the estimated average effect size of the three variables: BDO Meetings, DC Meetings, and MP Meetings as defined in the notes to Table 2.

Table B4 - Estimations of Household Measures with Village Controls Interacted with MPROP

Dependent Variable	Sub-Sample	Coeff. (β_1) <i>MLD</i>	Coeff. (β_2) <i>MPROP</i>	Coeff. (β_3) <i>MLD · MPROP</i>	Coefficient $\beta_1 + \beta_3$	Obs
Insured (AES)	Landless	0.09 (0.04)***	-0.04 (0.14)	-0.18 (0.08)**	-0.08 (0.06)	2565
	Large Land Owners	0.05 (0.02)*	0.07 (0.12)	-0.14 (0.06)**	-0.09 (0.05)*	2501
Daily Wage	All Labourers	-2.03 (1.0)**	-7.03 (4.3)	4.19 (2.14)**	2.16 (1.50)	13546
Daily Wage	Males	-2.22 (1.26)*	-7.02 (5.58)	5.22 (2.72)**	3.0 (1/95)	7480
Log Kharif Yields	Large Land Owners	0.24 (0.14)*	-0.24 (0.51)	-0.43 (0.25)*	-0.18 (0.15)	2320
Log Kharif Profit	Large Land Owners	0.29 (0.18)*	-0.55 (0.61)	-0.77 (0.37)**	-0.47 (0.24)**	1838
Maratha Trader	Low Castes	0.12 (0.04)***	0.09 (0.21)	-0.06 (0.11)	0.06 (0.08)	3012
Outside Maratha Trader	Low Castes	0.11 (0.03)***	0.09 (0.14)	-0.19 (0.08)**	-0.08 (0.06)	2793
Maratha Lender	Low Castes	0.20 (0.08)***	-0.50 (0.37)	-0.21 (0.20)	-0.02 (0.15)	452
Terms of Payment (Inputs)	Low Castes	0.10 (0.05)**	0.05 (0.15)	-0.25 (0.11)**	-0.15 (0.09)*	10034
Interest Rate on Loan	Low Castes	-8.48 (4.0)**	2.77 (17.9)	-5.67 (12.95)	-14.15 (11.48)	250
Voted-Personal	Low Castes	0.09 (0.04)**	0.12 (0.19)	-0.15 (0.11)	-0.06 (0.09)	2108
Social Capital (AES)	Low Castes	0.05 (0.01)***	-0.002 (0.07)	-0.12 (0.03)***	-0.07 (0.03)***	4693
Share Water	Low Castes	0.35 (0.06)***	0.84 (0.25)***	-0.46 (0.14)***	-0.11 (0.10)	2929
Target Village	Low Castes	0.85 (0.37)**	-0.27 (1.41)	-1.48 (0.84)*	-0.63 (0.60)	4865
Shared Funds	Low Castes	1.24 (0.35)***	2.11 (1.58)	-2.78 (0.83)***	-1.55 (0.66)	4584

Note: All estimations include village-level controls (presence of river/canal, distance to natural water sources, distance to railways and national roads, soil quality measures, proportion of the population that is SC/ST, and total village population), household-level controls (education, land ownership, and caste identity), and regional fixed effects. The estimations also include interaction terms between each of these village-level controls and MPROPA single asterix denotes significance at the 10% level, double for 5%, and triple for 1%. Acronyms used are: Maratha land dominated (MLD); Maratha population proportion (MPROP); and Average Effect Size (AES). Insured (AES) is the estimated average effect size of the six variables, Insured (1) to (6), defined in the notes of Table 3. Insurer is a dummy variable equal to one if respondents answer yes to: "Suppose a lower caste man asks to borrow a good sum of money from you because someone in his family has fallen ill. He is from the village and has the ability to repay the amount. Would you lend it to him?". Large land owners have greater than 5 acres. The sample of labourers are all those who work for a daily wage in agriculture. Additional individual controls (gender, age, education) are included in the wage estimations. Regression disturbance terms are clustered at the household and village level for these estimations using the approach of Cameron, Gelbach and Miller (2011). The sample for the yields, profits, proportion of labour costs regressions is all large cultivators (greater than 5 acres of land). All measures are per acre of land. Kharif yields are the total value of output per acre of land for a given crop, summed over all of the kharif crops for each household. Kharif profit is yields net of input costs (seeds, fertilizer, irrigation, electricity, pesticides, and labour). Workers include parttime and fulltime, same results held if restricted ourselves to fulltime workers. Additional crop controls are included in the yields and profits estimations. Maratha Trader is equal to one if the household has traded with a Maratha for any tradeable good (which includes agricultural inputs and outputs, farm enterprise and non-farm enterprise goods) conditional on trading goods. Outside Maratha Trader refers to the trader residing outside of the village conditional on trading goods. Maratha Lender refers to borrowing money from a Maratha. These estimations on Maratha trading relationships are probit estimations, where the coefficients reported are the partial derivatives of the predicted probability. Terms of payments is an index variable equal to 0 if the trader requires advanced payments; 1 if full payment is required at the time of sale; and 2 if instead payment in installments is acceptable. These are ordered probit estimations. Voted - Personal equals to one if the household voted for a candidate due to a personal connection rather than due to the characteristics of the candidate (honesty, good reputation, qualifications). Samples are conditional on voting. The sample of low castes in the voting regressions is SC/STs. Social Capital (AES) is the estimated average effect size of the six variables: Trust, No Cheat, Repair, Donated Cash, Donated Labour, and Agree as defined in the notes of Table 5. Target Village refers to GP funds should be targeted to the village as a whole, compared to poor or low caste individuals. Shared funds refers to GP funds are shared across the village (e.g. for development projects; public goods) compared to going directly to the poor or low status; the rich and high status; or to GP members or other government officials directly. These two estimations are estimated as multinomial logit models.

B3. Estimations with Maratha Land Holdings

We now report the results from analagous regressions to those estimated in the paper. Here, instead of using a binary variable, MLD which equals 1 if Marathas are the land dominant group, and 0 otherwise, as we did in the paper, we use an alternative source of information on Maratha land holdings from our household surveys. From these 30 households per village we obtain an estimate of the overall proportion of village lands held by Marathas $\widetilde{MLD} \in [0, 1]$ for each village. Refer to Section 10.1 in Online Appendix D for more details on this variable.

These estimation results are reported in Tables B5 and B6 below. Importantly, the main results discussed in the paper all go through as well in these alternative estimations.

Table B5 - Estimations of GP Measures with Maratha Land Holdings

Dependent Variable	Coefficient (β_1) \widetilde{MLD}	Coefficient (β_2) $MPROP$	Coefficient (β_3) $\widetilde{MLD} \cdot MPROP$	Coefficient $\beta_1 + \beta_3$	Observations
Maratha Pradhan	0.61 (0.26)**	1.16 (0.22)***	-0.71 (0.40)*	-0.10 (0.23)	275
Revenue (1)	-117.6 (89.6)	-119.4 (83.4)	107.7 (133.2)	-9.87 (84.1)	193
Revenue (2)	-20.7 (9.1)**	-20.5 (8.3)***	49.4 (19.1)***	28.7 (11.2)	291
Expenditure	-19.1 (9.1)**	-19.5 (8.2)***	46.6 (19.0)***	27.5 (11.1)	291
Programs (1)	-1.42 (0.65)**	-1.84 (0.58)***	3.63 (1.09)***	2.21 (0.65)***	7752
BPL Programs (1)	-0.60 (0.25)**	-0.56 (0.21)***	1.46 (0.42)***	0.86 (0.25)***	7752
EGS (1)	-0.16 (0.08)**	-0.05 (0.07)	0.36 (0.12)***	0.20 (0.08)	7725
Income Programs (1)	-1.40 (0.61)	-1.73 (0.54)***	3.53 (1.02)***	2.13 (0.61)***	7752
Programs (2)	-1.70 (0.72)***	-2.00 (0.61)***	3.80 (0.83)***	2.10 (0.70)***	275
BPL Programs (2)	-0.70 (0.28)***	-0.62 (0.21)***	1.50 (0.45)***	0.80 (0.26)***	275
EGS (2)	-0.19 (0.08)***	-0.04 (0.07)	0.39 (0.14)***	0.20 (0.09)	275
Income Programs (2)	-1.67 (0.68)	-1.90 (0.56)***	3.73 (1.10)***	2.07 (0.65)***	275
BDO Meetings	-4.70 (1.64)***	0.64 (2.58)	4.75 (3.61)	0.05 (2.58)	275
MP Meetings	-4.92 (2.86)*	-1.81 (2.30)	5.87 (3.04)**	0.96 (1.72)	275
DC Meetings	-3.94 (1.28)***	-3.51 (1.52)**	5.83 (1.98)***	1.90 (1.10)*	275

Note: MLD refers to the proportion of village land that is owned by Marathas constructed from the household level data. All estimations include village-level controls (latitude, longitude, elevation, presence of river/canal, distance to natural water sources, distance to railways and national roads, soil quality measures, rainfall levels, proportion of the population that is SC/ST, total village population, and whether the GP is reserved) and regional fixed effects. A single asterix denotes significance at the 10% level, double for 5%, and triple for 1%. Robust standard errors are in parentheses. Acronyms used are: Maratha population proportion (MPROP); Below Poverty Line (BPL); Employment Guarantee Scheme (EGS); Block Development Officer (BDO); District Collector (DC); and Member of Parliament (MP). Revenue (1) refers to data collected from the balance sheets (covers last 24 months) submitted by the GPs (obtained using RTI Act). Revenue (2) and Expenditure are annual per capita values from the 2001 Village Census. Information on programs (Programs (1); BPL Programs (1); EGS (1); Income Programs (1); Non-Income Programs) are reported from household level data, and regression disturbance terms are clustered at the village level for these estimations and household level controls are also included (education, land ownership, and caste identity). Programs (2); BPL Programs (2); EGS (2), Income Programs (2), and Non-Income Programs (2) are variables which aggregate this household level information up to the village level. Estimations are OLS except EGS (1), which is a probit estimation, where the reported coefficients are the partial derivatives of the predicted probability.

Table B6 - Estimations of Household Measures with Maratha Land Holdings

Dependent Variable	Sub-Sample	Coeff. (β_1)		Coeff. (β_2)		Coeff. (β_3)		Coefficient $\beta_1 + \beta_3$	Obs
		\widetilde{MLD}	$MPROP$	\widetilde{MLD}	$MPROP$	\widetilde{MLD}	$MPROP$		
Insured (1)	Landless	0.19 (0.08)**	0.11 (0.06)*	-0.31 (0.13)**		-0.13 (0.08)	2579		
Insured (2)	Landless	0.19 (0.08)**	0.09 (0.07)	-0.35 (0.13)***		-0.16 (0.08)	2579		
Insured (3)	Landless	0.20 (0.08)***	0.10 (0.07)	-0.33 (0.13)***		-0.13 (0.09)	2579		
Insured (4)	Landless	0.24 (0.08)***	0.12 (0.07)*	-0.35 (0.14)***		-0.11 (0.09)	2579		
Insured (5)	Landless	0.26 (0.08)***	0.07 (0.07)	-0.37 (0.13)***		-0.11 (0.09)	2579		
Insured (6)	Landless	0.24 (0.08)***	0.08 (0.07)	-0.38 (0.13)***		-0.14 (0.09)	2579		
Insurer	Large land owners	0.18 (0.05)***	-0.02 (0.04)	-0.19 (0.07)***		-0.01 (0.04)	2507		
Daily Wage	All Labourers	-1.65 (0.82)**	-1.84 (0.85)**	6.70 (1.46)***		5.05 (0.97)***	13581		
Daily Wage	Low Castes	-2.25 (0.96)**	-3.25 (1.00)***	8.60 (1.85)***		6.35 (1.31)***	9195		
Log Kharif Yields	Large land owners	0.34 (0.18)**	-0.02 (0.21)	-0.47 (0.27)*		-0.13 (0.20)	2323		
Maratha Trader	Low Castes	0.23 (0.08)***	0.17 (0.08)**	-0.11 (0.15)		0.11 (0.10)	3021		
Outside Maratha Trader	Low Castes	0.17 (0.06)***	0.11 (0.06)*	-0.24 (0.11)**		-0.07 (0.07)	2800		
Maratha Lender	Low Castes	0.50 (0.18)***	0.39 (0.17)**	-0.44 (0.32)		0.06 (0.23)	453		
Interest Rate on Loan	Low Castes	-29.1 (10.1)***	2.30 (11.1)	24.7 (19.1)		-4.43 (14.85)	165		
Voted-Personal	Low Castes	0.17 (0.08)**	0.05 (0.07)	-0.20 (0.13)†		-0.04 (0.09)	2116		
Social Capital (AES)	Low Castes	0.10 (0.03)***	0.02 (0.02)	-0.13 (0.05)***		-0.04 (0.03)	4711		
Share Water	Low Castes	0.38 (0.11)***	0.47 (0.11)***	-0.37 (0.18)**		0.01 (0.11)	2942		
Target Village	Low Castes	1.57 (0.57)***	-0.07 (0.62)	-1.93 (1.0)**		-0.36 (0.68)	4883		
Shared Funds	Low Castes	1.17 (0.58)**	0.70 (0.60)	-1.04 (0.99)		0.12 (0.65)	4603		

Note: MLD refers to the proportion of village land that is owned by Marathas, constructed from the household level data. All estimations include village-level controls (latitude, longitude, elevation, presence of river/canal, distance to natural water sources, distance to railways and national roads, soil quality measures, rainfall levels, proportion of the population that is SC/ST, total village population, and whether the GP is reserved), household-level controls (education, land ownership, and caste identity), and regional fixed effects. Regression disturbance terms are clustered at the village level. A single asterisk denotes significance at the 10% level, double for 5%, and triple for 1%. Acronyms used are: Maratha population proportion (MPROP) and Average Effect Size (AES). Insured (1): "Would most people in your village help you with some money in times of crisis?". Insured (2): "Would a higher caste member of your village help you with some money in times of crisis?". Insured (3): "Would most people in your village help a lower caste villager with some money in times of crisis?". Insured (4)-(6) are the same questions with "money" replaced by "grain". Insurer: "Suppose a lower caste man asks to borrow a good sum of money from you because someone in his family has fallen ill. He is from the village and has the ability to repay the amount. Would you lend it to him?". Large land owners have \geq than 5 acres. The sample of labourers are all those who work for a daily wage in agriculture. Additional individual controls (gender, age, education) are included in the wage estimations. Regression disturbance terms are clustered at the household and village level for these estimations using the approach of Cameron, Gelbach and Miller (2011). The sample for the yields and profits regressions is all large cultivators ($>$ 5 acres of land). All measures are per acre of land. Kharif yields are the total value of output per acre of land for a given crop, summed over all of the kharif crops for each household. Kharif profit is yields net of input costs (seeds, fertilizer, irrigation, electricity, pesticides, and labour). Additional crop controls are included in the yields and profits estimations. Maratha Trader is equal to one if the household has traded with a Maratha for any tradeable good (which includes agricultural inputs and outputs, farm enterprise and non-farm enterprise goods) conditional on trading goods. Outside Maratha Trader refers to the trader residing outside of the village conditional on trading goods. Maratha Lender refers to borrowing money from a Maratha. These estimations on Maratha trading relationships are probit estimations, where the coefficients reported are the partial derivatives of the predicted probability. Terms of payments is an index variable equal to 0 if the trader requires advanced payments; 1 if full payment is required at the time of sale; and 2 if instead payment in installments is acceptable. These are ordered probit estimations. Voted - Personal equals to one if the household voted for a candidate due to a personal connection rather than due to the characteristics of the candidate (honesty, good reputation, qualifications). Samples are conditional on voting. The sample of low castes in the voting regressions is SC/STs. Social Capital (AES) is the estimated average effect size of the six variables: Trust, No Cheat, Repair, Donated Cash, Donated Labour, and Agree as defined in the notes of Table 5. Target Village refers to GP funds should be targeted to the village as a whole, compared to poor or low caste individuals. Shared funds refers to GP funds are shared across the village (e.g. for development projects; public goods) compared to going directly to the poor or low status; the rich and high status; or to GP members or other government officials directly. These two estimations are estimated as multinomial logit models.

B4. Additional Estimations

Table B7 - Other Outcomes - Landless Sample

Dependent Variable	Coeff. (β_1) <i>MLD</i>	Coeff. (β_2) <i>MPROP</i>	Coeff (β_3) <i>MLD · MPROP</i>	Observations
Pradhan - Honesty	0.05 (0.12)	-0.28 (0.24)	0.26 (0.28)	2573
Pradhan - Qualifications	0.006 (0.13)	-0.27 (0.22)	0.44 (0.28)	2564
Pradhan - Providing Public Goods	0.08 (0.11)	-0.26 (0.18)	0.45 (0.23)*	2569
Pradhan - Solving Problems	0.001 (0.11)	-0.19 (0.21)	0.43 (0.25)*	2567
Pradhan - Allocating Spending Fairly	0.01 (0.10)	-0.26 (0.24)	0.44 (0.27)	2565
Pradhan - Acquiring Funds	0.16 (0.13)	-0.09 (0.20)	0.29 (0.26)	2566
Pradhan - Not Misusing Funds	-0.14 (0.12)	-0.43 (0.19)**	0.50 (0.26)*	2560
Pradhan - Representing Village to Govt.	-0.06 (0.13)	0.01 (0.17)	0.42 (0.23)*	2566
Share Information	-0.03 (0.03)	-0.07 (0.06)	-0.002 (0.07)	2571
Trust Neighbours	0.01 (0.07)	0.15 (0.13)	-0.22 (0.17)	2582
Trust Own Caste	0.09 (0.08)	0.20 (0.16)	-0.29 (0.18)	2582

Note: All estimations include village-level controls (latitude, longitude, elevation, presence of river/canal, distance to natural water sources, distance to railways and national roads, soil quality measures, rainfall levels, proportion of the population that is SC/ST, total village population, and whether the GP is reserved), household-level controls (education, land ownership, and caste identity), and regional fixed effects. Regression disturbance terms are clustered at the village level. A single asterix denotes significance at the 10% level, double for 5%, and triple for 1%. Acronyms used are: Maratha land dominated (MLD) and Maratha population proportion (MPROP). The sample is landless. The dependent variables for the Pradhan variables are dummy variables equal to one if respondents answered a low rank in terms of their confidence in their Pradhan with regards to the listed characteristics, and zero otherwise. Estimations are probits, where the coefficients reported are the partial derivatives of the predicted probability. "Share Information" refers to a question which asked "Suppose you find employment available at a good wage that others do not know about it, who would you share the information with?", this variable is equal to one if they would share it with everyone in the village as opposed to just their close family and friends. Trust Neighbours is response to: "Would you say that your neighbours can be trusted? 1=Almost none, 2=Some; 3=Majority; 4=Almost. Trust Own Caste is response to: "Would you say that members of your own caste can be trusted? 1=Almost none, 2=Some; 3=Majority; 4=Almost.

ONLINE APPENDIX C: ADDITIONAL THEORETICAL RESULTS AND PROOFS

C1. Explicit Treatment of State Contingent Transfers

Instead of modeling the value of insurance in expected value, suppose instead that there are two possible states; a normal state with consumption valued at 1, and a state of need where a worker's marginal valuation of consumption is $\phi > 1$. Suppose the need state arises with probability μ . It is drawn each period independently from a distribution common to all workers. An insurance promise from a landlord, i , to a worker, j , is a commitment by the landlord to a transfer of S_i^j in the worker's need state. Such a state is observable to both landlords and workers but not enforceable by formal/legal mechanisms. Under the assumption that $\mu \cdot \phi = 1$, the analysis exactly yields the equations in the body of the paper. Relaxing this assumption will imply that the parameters ϕ and μ enter into the interpretation of the estimated coefficients. But since we restrict interpretation of coefficients in terms of model parameters to sign implications, the way the model is interpreted is invariant to relaxing this normalization.

C2. Wages and Yields Affected by Programs and EGS

Most landless individuals sell their labor to large landowners. Most large landowners have as their largest input cost labor. The way labor relations work in these villages is that the landless people or small landholders who rely mainly on labor income for their livelihood typically work on the farm of a large landowner in a permanent or semi-permanent capacity. Much of what workers need to do can only be partly or very imperfectly supervised, suggesting that asymmetries of information in production may arise. Such permanent working arrangements are coveted by workers, and though there is a spot market for some labor, it seems that workers prefer the permanent working arrangements greatly. The threat of losing such employment disciplines the use of discretionary effort. For individuals primarily relying on labor income for their livelihood the threat of employment loss, which would send them into poverty, provides great incentive for them to keep contributing un- or partially monitored discretionary effort in their employment on large landholders farms. Large landholders grow various crops and their labor needs, timing of application, and other inputs use are largely fixed through

the crop cycle. However, the quality of crop obtained depends critically on good labor input and diligence through the production process.

These ingredients suggest an efficiency wage model. Workers are required, by the implicit contract of the landlord, to provide e^* units of labor effort and receive a wage w^* . Landlords imperfectly ascertain, ex post, the effort contribution of their worker and decide whether to rehire them in the next period, or dismiss them from their employ. Since production is largely of a fixed factor variety, we can for simplicity simply characterize the optimal incentive compatible contract (e^*, w^*) offered to each worker by the landlord while letting the landlord's landholding and crop choice (which is a function of the conditions) determine the number of workers required.

In this sort of labor market, even though much of the year sees labor only partially employed or unemployed the activities of the panchayat in providing poverty alleviation programs become significant. In the event that workers are not employed by landlords, they will depend on benefits from the state, or on employment from the state for their livelihood. Thus, we can characterize their reservation utility, \bar{u} , as depending positively on the incidence of these programs. For simplicity let this take two values, $\bar{u}(W)$ when W workers control the panchayat and actively seeks out these programs, and $\bar{u}(L)$ when L landlords control it and such programs are shut down. These are taken as given when worker and landlord play the labor/production game.

THE LABOR/PRODUCTION GAME

Given an increasing and concave per worker effort production function, $f(e)$, the landlord chooses the implicit contract parameters (w, e) :

$$\max_{e,w} f(e) - w$$

subject to (w, e) being incentive compatible for the worker. That is any pair w, e chosen must satisfy

$$(C1) \quad \frac{u(w) - c(e)}{1 - r} \geq u(w) + \frac{r}{1 - r} \bar{u}(x), \text{ where } x = W \text{ or } L.$$

The term $u(w)$ is increasing and concave, $c(e)$ is increasing and convex, and $\bar{u}(x)$ reservation employment if dismissed, is increasing in probability of obtaining benefits, probability of obtaining EGS employment, and probability of obtaining another job (which we can set equal to zero for simplicity), so that $\bar{u}(W) > \bar{u}(L)$.

Firstly note that any optimal e, w chosen must ensure that (C1) binds exactly,

$$\frac{u(w) - c(e)}{1 - r} = u(w) + \frac{r}{1 - r} \bar{u}(x)$$

implying.

$$(C2) \quad w = u^{-1} \left(\frac{c(e)}{r} + \bar{u}(x) \right).$$

Substituting this in, the optimization problem becomes:

$$\max_e f(e) - u^{-1} \left(\frac{c(e)}{r} + \bar{u}(x) \right).$$

With a FOC that implies:

$$f'(e) = u^{-1'} \left(\frac{c(e)}{r} + \bar{u}(x) \right) \frac{c'(e)}{r}.$$

This implicitly defines a solution $e^*(\bar{u}(x))$ and from equation (C2) the corresponding w^* .

Proposition The optimal implicit contract (w^*, e^*) has wage strictly increasing and effort strictly decreasing in $\bar{u}(x)$.

Proof: At e^* :

$$f'(e^*(\bar{u})) = u^{-1'} \left(\frac{c(e^*(\bar{u}, r))}{r} + \bar{u} \right) \frac{c'(e^*(\bar{u}, r))}{r}.$$

Differentiating with respect to \bar{u} yields:

$$\begin{aligned} f''(e^*(\bar{u}, r)) \frac{de^*}{d\bar{u}} &= u^{-1''}(\cdot) \left(\frac{c'(e^*(\bar{u}, r))}{r} \frac{de^*}{d\bar{u}} + 1 \right) \frac{c'(e^*(\bar{u}, r))}{r} \\ &\quad + u^{-1'}(\cdot) \left(\frac{c''(e^*(\bar{u}, r))}{r} \frac{de^*}{d\bar{u}} \right), \end{aligned}$$

rearranging:

$$\frac{de^*}{d\bar{u}} = \frac{u^{-1''}(\cdot) \frac{c'(e^*(\bar{u}, r))}{r}}{f''(e^*(\bar{u}, r)) - u^{-1'}(\cdot) \frac{c''(e^*(\bar{u}, r))}{r} - u^{-1''}(\cdot) \left(\frac{c'(e^*(\bar{u}, r))}{r}\right)^2}.$$

Because $u(\cdot)$ is an increasing and concave function, $u^{-1}(\cdot)$ is an increasing and convex function. Then since $c(\cdot)$ is a convex function by assumption it is immediate that the terms on the RHS can be signed as follows:

$$\frac{de^*}{d\bar{u}} = \frac{[+]}{[-] - [+]} < 0.$$

Differentiating equation (C2) with respect to \bar{u} yields:

$$\begin{aligned} \text{sign} \left[\frac{dw}{d\bar{u}} \right] &= \text{sign} \left[\frac{u^{-1''}(\cdot) \left(\frac{c'(e^*(\bar{u}, r))}{r}\right)^2 + f''(e^*(\bar{u}, r)) - u^{-1'}(\cdot) \frac{c''(e^*(\bar{u}, r))}{r}}{-u^{-1''}(\cdot) \left(\frac{c'(e^*(\bar{u}, r))}{r}\right)^2} \right] \\ &= \text{sign} \left[\frac{f''(e^*(\bar{u}, r)) - u^{-1'}(\cdot) \frac{c''(e^*(\bar{u}, r))}{r}}{[-]} \right] \\ &> 0. \end{aligned}$$

Prediction Where GPs are controlled by landlords, wages should be lower and effort should be higher across the village. $w(L) < w(W)$ and $e(L) > e(W)$.

C3. Proofs of propositions

Proof of Proposition 1

Using (3) and (4), vote trading is individually rational for worker j in village k , $U_{jk}(L_i) \geq U_{jk}(W)$, if and only if, $S_i^j \geq \Delta_{wP} + (d_j - d_i^j)T - x_k$. Since landlords transfer no more than necessary to buy a vote, S_i^j is chosen so that this condition binds. The following conditions are thus also sufficient to ensure incentive compatibility and individual rationality:

$$(C3) \quad S_i^j = \Delta_{wP} + (d_j - d_i^j)T - x_k.$$

Substituting (C3) into the landlord's incentive constraint (1) yields:

$$(C4) \quad \Delta_{wP} + (d_j - d_j^i)T - x_k \leq X_i + I_i^j X + x_i^j.$$

There are three specific cases of condition (C4) to consider. For a Maratha worker and landlord pair, $c_i = c_j = M$, we have:

$$(C5) \quad x_i^j + x_k \geq \Delta_{wP} - (X_M + X) \equiv x_{MM}.$$

For a non-Maratha worker, $c_j = N$, and Maratha landlord, $c_i = M$:

$$(C6) \quad x_i^j + x_k \geq \Delta_{wP} - T - X_M \equiv x_{NM}.$$

For Non-Maratha landlords with either type of worker:

$$(C7) \quad x_i^j + x_k \geq \Delta_{wP} - X_N \equiv x_{NN} \equiv x_{MN}.$$

Which correspond to the conditions in the statement of the proposition.

Proof of Proposition 2

Let $P_{VT}(k)$ denote the proportion of workers willing to undertake vote trading in village k , for a given x_k we then have:

$$P_{VT}(k) = \sigma_{MMk} \Pr(x_i^j > x_{MM} - x_k) + \sigma_{MNk} \Pr(x_i^j > x_{MN} - x_k) + \sigma_{NNk} \Pr(x_i^j > x_{NN} - x_k) + \sigma_{NMk} \Pr(x_i^j > x_{NM} - x_k)$$

By Assumption 3

$$I_{VTk} = \begin{cases} 0 & \text{if } P_{VT}(k) < 0.5 \\ 1 & \text{if } P_{VT}(k) \geq 0.5. \end{cases}$$

So that $E[I_{VTk}] = \Pr(P_{VT}(k) < 0.5 | \sigma_{ijk}) \cdot 0 + \Pr(P_{VT}(k) \geq 0.5 | \sigma_{ijk}) \cdot 1 \equiv \Pr(P_{VT}(k) \geq 0.5 | \sigma_{ijk})$. Thus, using (8) we obtain:

$$(C8) \quad E[v_k | \mathbf{Z}_k] = \alpha_v \Pr(P_{VT}(k) \geq 0.5) + \alpha \mathbf{Z}_k + \mu_{vk},$$

We now show that if x_k is drawn from a uniform distribution, then the probability of clientelism occurring in village k is given by:

$$(C9) \quad \Pr(P_{VT}(k) > 0.5 | \sigma_{ijk}) = \sigma_{MMk}(X_M + X) + \sigma_{NMk}(T + X_M) + \sigma_{NNk}(X_N) + \sigma_{MNk}(X_N) + C,$$

where C is a constant.

Assume that x_i^j is randomly drawn from a uniform distribution $F(x_i^j)$ on the interval $[0, \bar{x}]$. We see below that the form of this distribution is irrelevant so this is without loss of generality. Given this, the $Pr(x_i^j > z) \equiv \frac{\bar{x}-z}{\bar{x}}$ for any $z \in [0, \bar{x}]$. Applying the law of large numbers and substituting for x_i^j from equations (5), (6) and (7) yields the proportion of vote traders in village k conditional upon candidate quality x_k :

$$(C10) \quad P_{VT}(k) = \sigma_{MMk} \left(\frac{\bar{x} - \Delta_{wP} + X_M + X + x_k}{\bar{x}} \right) + \sigma_{NMk} \left(\frac{\bar{x} - \Delta_{wP} + T + X_M + x_k}{\bar{x}} \right) \\ + \sigma_{NNk} \left(\frac{\bar{x} - \Delta_{wP} + X_N + x_k}{\bar{x}} \right) + \sigma_{MNk} \left(\frac{\bar{x} - \Delta_{wP} + X_N + x_k}{\bar{x}} \right),$$

which, since the σ_{ij} s sum to one, rearranges to:

$$(C11) \quad P_{VT}(k) = \sigma_{MMk} \left(\frac{X_M + X}{\bar{x}} \right) + \sigma_{NMk} \left(\frac{T + X_M}{\bar{x}} \right) + \sigma_{NNk} \left(\frac{X_N}{\bar{x}} \right) + \sigma_{MNk} \left(\frac{X_N}{\bar{x}} \right) + \frac{\bar{x} - \Delta_{wP} + x_k}{\bar{x}},$$

Since we assume the law of large numbers within the village, the first moment of the distribution determines the number of vote traders arising from within each of the landlord/worker groups fully. Hence the invariance with respect to the distribution of the x_i^j .

For clientelism in village k , necessarily $P_{VT}(k) > 0.5$, which is equivalent to:

$$(C12) \quad x_k > 0.5\bar{x} - \bar{x} + \Delta_{wP} - \sigma_{MMk}(X_M + X) - \sigma_{NMk}(T + X_M) - \sigma_{NNk}(X_N) - \sigma_{MNk}(X_N).$$

Let $G(x)$ denote the uniform distribution of the x_k , with support $[0, 1]$. An additional interval restriction on the variable \bar{x} derives from the assumed support of $G(\cdot)$ and is required for the likelihood of clientelism to be a well defined probability. It is that $2(\Delta_{wP} - T - X_M) \geq \bar{x} \geq 2(\Delta_{wP} - 1 - X_N)$. The probability that $P_{VT}(k) > 0.5$, and hence the probability of clientelism, then lies in the unit interval and is given by $(1 - RHS(C12))$, which rearranges to equation (C9) above with $C = 1 + 0.5\bar{x} - \Delta_{wP}$.

Now substituting using (C9) into (C8) yields:

$$(C13) \quad E[v_k | \sigma_{ijk}, \mathbf{Z}_k] = \alpha_v [C + \sigma_{MMk}(X_M + X) + \sigma_{NMk}(T + X_M) + \sigma_{NNk}(X_N) + \sigma_{MNk}(X_N)] + \alpha \mathbf{Z}_k + \mu_{vk},$$

which is the statement in the proposition.

Proof of Proposition 3

Under random matching: $\sigma_{MMk} = MLD_k \cdot MPROP_k$, $\sigma_{NMk} = MLD_k \cdot (1 - MPROP_k)$, $\sigma_{NNk} = (1 - MLD_k) \cdot (1 - MPROP_k)$ and $\sigma_{MNk} = (1 - MLD_k) \cdot MPROP_k$. Substituting these into equation (9) and rearranging yields the expression in the statement of the proposition.

C4. Generalized Distributions

Here we demonstrate invariance with respect to the distributional assumptions we made for computing the likelihood of vote trading in Section II.B. Recalling that $P_{VT}(k)$ denotes the proportion of workers willing to undertake vote trading in village k , we have:

$$P_{VT}(k) = S_{MMk} \Pr(x_i^j > x_{MM} - x_k) + S_{MNk} \Pr(x_i^j > x_{MN} - x_k) + S_{NNk} \Pr(x_i^j > x_{NN} - x_k) + S_{NMk} \Pr(x_i^j > x_{NM} - x_k)$$

This again applies the law of large numbers within the village. Substituting for the x_i^j from equations (5), (6) and (7), we have:

$$P_{VT}(k) = \sigma_{MMk} \Pr(x_i^j > \Delta_{wP} - (X_M + X) - x_k) + \sigma_{NMk} \Pr(x_i^j > \Delta_{wP} - T - X_M - x_k) + \sigma_{NNk} \Pr(x_i^j > \Delta_{wP} - X_N - x_k) + \sigma_{MNk} \Pr(x_i^j > \Delta_{wP} - X_N - x_k).$$

Substituting for the σ_{ij} s using Assumption 4:

$$\begin{aligned} P_{VT}(k) &= MLD_k \cdot MPROP_k \Pr(x_i^j > \Delta_{wP} - (X_M + X) - x_k) + \\ &MLD_k \cdot (1 - MPROP_k) \Pr(x_i^j > \Delta_{wP} - T - X_M - x_k) + \\ &(1 - MLD_k) \cdot (1 - MPROP_k) \Pr(x_i^j > \Delta_{wP} - X_N - x_k) + \\ &(1 - MLD_k) \cdot MPROP_k \Pr(x_i^j > \Delta_{wP} - X_N - x_k). \end{aligned}$$

Rearranging:

$$\begin{aligned} P_{VT}(k) &= MLD_k \cdot MPROP_k \left[\Pr(x_i^j > \Delta_{wP} - X_M - X - x_k) - \Pr(x_i^j > \Delta_{wP} - T - X_M - x_k) \right] \\ &+ MLD_k \left[\Pr(x_i^j > \Delta_{wP} - T - X_M - x_k) - \Pr(x_i^j > \Delta_{wP} - X_N - x_k) \right] \\ &+ MPROP_k [0] \end{aligned}$$

Without the assumption of uniformity on F and G , but instead persisting with general distributions, we cannot specify the linear form of this equation in Section II.B, but can instead express this equation directly in terms of the respective CDFs. Substituting from the CDF for F .

$$\begin{aligned} P_{VT}(k) &= MLD_k \cdot MPROP_k [1 - F(\Delta_{wP} - X_M - X - x_k) - 1 + F(\Delta_{wP} - T - X_M - x_k)] \\ &\quad + MLD_k [1 - F(\Delta_{wP} - T - X_M - x_k) - 1 + F(\Delta_{wP} - X_N - x_k)] \\ &\quad + MPROP_k [0] \end{aligned}$$

Since, from Assumption 3, we have vote trading in village k if and only if $P_{VT}(k) > 0.5$, then the probability of vote trading is

$$\Pr(P_{VT}(k) > 0.5) \equiv \Pr \left(\begin{array}{l} MLD_k \cdot MPROP_k [F(\Delta_{wP} - T - X_M - x_k) - F(\Delta_{wP} - X_M - X - x_k)] \\ + MLD_k [F(\Delta_{wP} - X_N - x_k) - F(\Delta_{wP} - T - X_M - x_k)] \\ + MPROP_k [0] \end{array} > 0.5 \right) \quad (0.5)$$

The law of large numbers removes the idiosyncratic (pairwise between landlord and worker) uncertainty, the remaining uncertainty in this equation (the outer probability) is driven by the stochastic nature of the candidate quality variable, x_k . This is determined by the distribution G . Using the equation before (9) we have:

$$E[v_k | \mathbf{Z}_k] = \alpha_v \Pr(P_{VT}(k) \geq 0.5) + \alpha \mathbf{Z}_k + \mu_{vk}$$

Now insert this object into the estimating equation to obtain, instead of (9):

$$\begin{aligned} E[v_k | \mathbf{Z}_k] &= \alpha_v \Pr \left(\begin{array}{l} MLD_k \cdot MPROP_k [F(\Delta_{wP} - T - X_M - x_k) - F(\Delta_{wP} - X_M - X - x_k)] \\ + MLD_k [F(\Delta_{wP} - X_N - x_k) - F(\Delta_{wP} - T - X_M - x_k)] \\ + MPROP_k [0] \end{array} > 0.5 \right) + \\ &\quad \alpha \mathbf{Z}_k + \mu_{vk} \end{aligned}$$

The outer probability arises from the random draw on x_k . The sign restrictions on coefficients and interpretations obtained from a linear regression as specified in equation (12) or (13) are the same as in the body of the paper. The fact that the distribution of G is no longer uniform does not affect the interpretations, as the precise form of this distribution simply governs the draws of x_k . The effect of the model parameters T , X , X_M and X_N is to simply shift the required cutoff level of x_k (for a given x_i^j) beyond which vote trading is preferred for a landlord/worker

pair. The shifts vary by the type of pairing, so that the relative magnitudes of these parameters is identified by observing differences in the outcome variables (the vs) as a function of the distribution of landlord/worker pairings (the σ_{ij} that we infer from *MPROP* and *MLD*) in a village.

To see that the sign interpretations are the same when using the equation immediately above, consider each coefficient in turn: The coefficient on $MPROP_k$ is again predicted to be zero. Since this variable is multiplied by zero, its magnitude in the village will not affect the likelihood that the proportion of vote traders is above a half. For variable v such that $\alpha_v > 0 (< 0)$, the coefficient on MLD_k is again predicted positive (negative) since $T > 0$ and $X_M - X_N > 0$. Intuitively, vote trading is more likely if there are Maratha landlords since they possess two benefits in clientelism relative to non-Maratha landlords – access to trading networks and more powerful social cohesion. Consequently, the larger proportion of Maratha landlords, the larger the number of worker/landlord pairs tipped into vote trading by the existence of these clientelist benefits. The coefficient on the interaction $MLD_k \cdot MPROP_k$ again allows us to identify the relative power of trading network access and social cohesion. It is again positive iff $X_M + X > X_M + T$ or $X > T$. So that, for $\alpha_v > 0$, $MLD_k \cdot MPROP_k > (<)$ 0 implies $X > (<) T$. The sign of $[coeff]MLD_k \cdot MPROP_k + [coeff]MLD_k > 0$ iff $F(\Delta_{wP} - X_N - x_k) > F(\Delta_{wP} - X_M - X - x_k)$, i.e., iff. $X + X_M > X_N$. So the sign of this expression again yields an upper bound for the benefits of Maratha social cohesion relative to social cohesion in non-Maratha groups $X_M - X_N$.

C5. Workers Less able to Punish Landlords in MLD

Assume now a violation of Assumption 1 where workers in MLD villages are less able to punish a promise violating landlord than in non-MLD villages. In non-MLD villages then continue to assume that x_i^j is drawn from distribution $F(\cdot)$ which continues to be defined over support $[0, \bar{x}]$, but now suppose that x_i^j in MLD villages is drawn from $\widehat{F}(\cdot)$ defined over $[0 - \zeta, \bar{x} - \zeta]$ with $\zeta > 0$. $\widehat{F}(\cdot)$ is

first-order stochastically dominated by F . Equation (C10) now becomes:

$$P_{VT}(k) = \sigma_{MMk} \left(\frac{\bar{x} - \zeta - \Delta_{wP} + X_M + X + x_k}{\bar{x}} \right) + \sigma_{NMk} \left(\frac{\bar{x} - \zeta - \Delta_{wP} + T + X_M + x_k}{\bar{x}} \right) \\ + \sigma_{NNk} \left(\frac{\bar{x} - \Delta_{wP} + X_N + x_k}{\bar{x}} \right) + \sigma_{MNk} \left(\frac{\bar{x} - \Delta_{wP} + X_N + x_k}{\bar{x}} \right),$$

and equation (C11) now rearranges to:

$$(C14) \\ P_{VT}(k) = \sigma_{MMk} \left(\frac{X_M + X - \zeta}{\bar{x}} \right) + \sigma_{NMk} \left(\frac{T + X_M - \zeta}{\bar{x}} \right) + \sigma_{NNk} \left(\frac{X_N}{\bar{x}} \right) + \sigma_{MNk} \left(\frac{X_N}{\bar{x}} \right) + \frac{\bar{x} - \Delta_{wP} + x_k}{\bar{x}},$$

Using (C14) substituting for the σ s and proceeding identically, expression (10) of Proposition 3 now rearranges to:

$$E[v_k | MLD_k, MPROP_k, \mathbf{Z}_k] = \alpha_v C + MLD_k \cdot \alpha_v [T + X_M - X_N - \zeta] + MPROP_k \cdot \alpha_v [0] \\ + MLD_k \cdot MPROP_k \cdot \alpha_v [X - T] + \alpha \mathbf{Z}_k + \mu_{vk},$$

where C is a constant. This expression thus yields identical implications to those previously obtained under expression (10) except that now the predicted coefficient on MLD in the baseline model is biased towards zero by the ζ term. This implies that the interpretation provided by the model for the sign of the coefficient on MLD, given that it is significant and in the direction predicted by the baseline model, is unchanged. The predictions on all other coefficients are also unchanged. There is now, however, a slight difference in interpretation of the magnitude of the coefficient on MLD under this extension of the model – it is reduced by the ζ which represents less worker cohesion in MLD villages. However, since we have already argued that the normalizations imposed in the model render interpretation of coefficients directly in terms of model parameters meaningless, this has no effect.

C6. Landlord Candidates Less Appealing in MLD

Assumption 2 of the baseline model asserts that whether landlords are Maratha or not does not affect the (dis)utility that workers obtain from lending political support to the landlord's candidate in Panchayat elections. Consider a violation of this assumption where, in MLD villages, villagers resent landlords more than in non-MLD villages, perhaps due to their always having been the dominant

caste, then the disutility workers experience by supporting an otherwise identical landlord's candidate may be worse in the MLD case. A simple modification of the model captures this. Continue to assume that the variable x_k , capturing the cost to workers from voting for a landlord's candidate is drawn from distribution $G(\cdot)$ in all villages. Now, however the realization of the term x_k has subtracted from it the amount (in utility metric) ι in MLD villages. This implies that the same quality draw of candidate from $G(\cdot)$ generates ι less utility for voters if this candidate is drawn in a MLD villages rather than in a non-MLD village. Equation (C10) now becomes:

$$P_{VT}(k) = \sigma_{MMk} \left(\frac{\bar{x} - \Delta_{wP} + X_M + X + x_k - \iota}{\bar{x}} \right) + \sigma_{NMk} \left(\frac{\bar{x} - \Delta_{wP} + T + X_M + x_k - \iota}{\bar{x}} \right) \\ + \sigma_{NNk} \left(\frac{\bar{x} - \Delta_{wP} + X_N + x_k}{\bar{x}} \right) + \sigma_{MNk} \left(\frac{\bar{x} - \Delta_{wP} + X_N + x_k}{\bar{x}} \right).$$

But notice that this modification of the model is equivalent to the modification introduced in Section C.C5 above and yields an equation that is identical to equation (C14) above. This violation of Assumption 2 is thus isomorphic in its effect to the violation of Assumption 1 explored there. The interpretation of coefficients under the model is identical.

C7. Positive Assortative Matching

Assumption 4 of the baseline model asserts random matching, so that caste wise matching simply reflects population frequencies. Now assume that Maratha landlords are matched with Maratha workers at rate μ times their relative frequency. Hence with the proportion of Maratha workers in the village denoted $MPROP$ the probability of a single Maratha landlord being matched with a Maratha worker is $MPROP \cdot \mu$. This implies that the frequency of Maratha landlord/Maratha worker pairs will be given by $MLD \cdot MPROP \cdot \mu = \sigma_{MM}$ as opposed to $MLD \cdot MPROP$ from the baseline model. Similarly for the other frequencies: $\sigma_{NM} = MLD \cdot (1 - MPROP \cdot \mu)$, $\sigma_{NN} = (1 - MLD) \cdot (1 - MPROP) \cdot \mu$ and $\sigma_{MN} = (1 - MLD) \cdot [1 - (1 - MPROP) \cdot \mu]$.

Substitute for these values of σ_{ij} into equation (9) to obtain:

$$E[v_k | MLD_k, MPROP_k, \mathbf{Z}_k] = \alpha_v \left(\begin{array}{l} C + MLD_k \cdot MPROP_k \cdot \mu(X_M + X) + \\ MLD_k \cdot (1 - MPROP_k \cdot \mu)(T + X_M) \\ + (1 - MLD_k) \cdot (1 - MPROP_k) \cdot \mu(X_N) + \\ (1 - MLD_k) \cdot (1 - [1 - MPROP_k] \cdot \mu)(X_N) \end{array} \right) + \alpha \mathbf{Z}_k + \mu_{vk}$$

Rearranging:

$$(C15) \quad E[v_k | MLD_k, MPROP_k, \mathbf{Z}_k] = \alpha_v C + MLD_k \cdot \alpha_v [T + X_M - X_N] + MPROP_k \cdot \alpha_v [0] + MLD_k \cdot MPROP_k \cdot \mu \cdot \alpha_v [X - T] + \alpha \mathbf{Z}_k + \mu_{vk}$$

This is equivalent to the expression in the baseline model (equation 10) without assortative matching up to the positive multiplicative term μ multiplying the coefficient on the interaction term $MLD \cdot MPROP$. Clearly this term does not affect the sign of the coefficient on the interaction term, and therefore does not affect our interpretation of the relative sizes of X and T .

ONLINE APPENDIX D: INDEPENDENCE OF MLD_k AND $MPROP_k$

Our two key measures are Maratha population numbers, $MPROP_k$, and landholdings, MLD_k , both at the village level. In what follows we discuss how these measures were obtained, and argue that both of these measures are historically pre-determined, and importantly not endogenous to our outcome variables.

D1. Dominance Measures

Both Maratha population ($MPROP_k$) and landholdings (MLD_k) were collected in the village surveys. These surveys were conducted on a focus group discussion model, which included key villagers such as members of the GPs and upper level governments (particularly the *Gram Sevak and Talathi*) as well as school teachers and health care workers. The *Gram Sevak* represents the development wing of the state government and is well versed with the village population, since all of the benefit applications go through him. He, or members of the GP, readily provided the population numbers by caste group in the villages. The *Talathi*, who is from the Revenue department, is responsible for keeping and

updating all land records. It was typically the Talathi who provided us with a ranking of total land ownership by caste group (at the sub-caste or *jati* level) in the villages. Both the Talathi and Gram Sevak are members of the higher levels of government and do not usually reside in the surveyed village.

We can create an alternative measure of Maratha land dominance (\widetilde{MLD}_k) using our household surveys, where we collected information on land ownership (refer to Section B3 in Online Appendix B). We can aggregate this data up to the village level to obtain a measure of land distribution by caste group at the village level. Since only 30 households per village were surveyed, these measures are quite noisy. Nevertheless, if we construct a Maratha land dominance variable from this household level data, it matches our village level data (obtained from the Talathi) for 85% of villages. For those 15% of villages which did not match, the total land ownership of the top two ranked (in terms of land ownership) castes was very close using the household level data. In these cases, according to our village level data, Marathas were typically the second ranked caste in terms of land ownership. In other words, these were villages where two castes were fairly close in terms of their total land ownership, and this explains why the noisy household level data did not match up perfectly to the village level data.

In our baseline empirical analysis we use the village level data to construct our measure of Maratha land dominance (MLD_k). Results are robust to instead using the alternative aggregate measure (\widetilde{MLD}_k), constructed from our household level data, as we have reported in Online Appendix B.

D2. Distribution of Caste Groups

We have no direct proof that caste population numbers are historically predetermined at the village level, and not the consequence of any of our subsequent outcome variables, because no historical records reside at the village level on caste population numbers. However, at the district level, others have exploited the temporal invariance of caste numbers, and used caste composition measures from the historical census to predict outcomes today (Banerjee and Rohini Somanathan 2007). The assumption of time invariant caste distributions at the village level has also been exploited in other states of India (see Siwan Anderson (2011) for Uttar Pradesh and Bihar). Similarly, using the historical censuses of India (1891

- 1931), we can compare the relative population distribution of Marathas across the districts in our sample to the distribution in our current data. Despite our sampling of only non-tribal rural areas, the historical census variation closely matches the current variation found in our data. Of particular note is the virtual absence of Marathas in the most eastern part of the state (East Vidarbha). This part of the state was part of the Central Provinces in colonial times, a region where the Rajput caste were traditionally dominant.³⁷ All of the empirical results that we report are robust to excluding this region in our estimations.

In 41% of villages, Marathas form the majority of the population, but we see considerable village level variation in caste population numbers, which is the norm in India. Villages are typically multi-caste and rarely identical in either the number of castes or in the numerical strength (M.N. Srinivas 1987, David Mandelbaum 1970, McKim Marriott 1955). In general, Indian village anthropological studies link the origins of distributions at the village level back hundreds of years (Srinivas 1987, Mandelbaum 1970, Marriott 1955), and the Marathas are no exception.³⁸ The early settlement of the original tribes that grew into the prominent caste groups in Maharashtra dates to the 6th century BC (D.D. Kosambi 1955). The prominence of Marathas in the region dates back to at least the fourth century AD (A.S. Altekar 1927, Kosambi 1969).³⁹

In our Village survey, we asked directly about the historical origin of caste groups in the village. In more than 95% of cases, the caste groups were reported to have resided in the village since well before independence. A possible concern is the possibility of migration in response to contemporary governance and economic outcomes, which would in turn directly alter village level caste composition. At the individual or household level, these concerns are not warranted. Firstly, this

³⁷The present state of Maharashtra came into being in 1960. The state unites the Marathi speaking people (who have existed for centuries). During British rule, Marathi speakers were geographically divided between Bombay Presidency, Central Provinces and Berar, and the Nizam's state of Hyderabad. After Independence (1947) they continued as respective parts of these states until the formation of the bilingual state of Bombay in 1956 (two languages Marathi and Gujarati). The unilingual state of Maharashtra was formed in 1960.

³⁸For example, in the case of Palanpur, a village in western Uttar Pradesh, events which took place some two hundred years ago explain the dominance of an upper caste group (Jean Dreze et al. 1999). Another village level study in northwest Uttar Pradesh dates the origins of present caste composition to more than 600 years ago (Ajit Danda 1987).

³⁹Basic elements of the village organization, the *balutedari* system, were developed by the fourth century AD. This system was a reciprocal arrangement between the hereditary farming and artisan castes (OBCs in today's classification), service castes (SCs), and the higher landholding castes.

is almost unheard of in our sample of villagers. Secondly, given the strict rules governing hereditary caste rankings, there is virtually no mobility of individuals across different caste groups. Moreover, there is very little migration in India as a whole; see Munshi and Rosenzweig (2005) for an extensive analysis. This seems to be primarily because of reliance on sub-caste networks of mutual insurance that do not seem to cross village boundaries. At the caste level as a whole, there is no evidence of large scale migrations that could explain the variation in caste population dominance that we observe today.⁴⁰

D3. Land Ownership of Caste Groups

Marathas own the most land in 59% of the villages of our survey. Throughout history, Marathas have been the dominant land owners in Maharashtra and their prominence has been traced back to at least the fourth century AD when major chieftainships were under their control (Altekar 1927, Kosambi 1969). With respect to landholdings, their documented dominance of landowning extends back to at least the fourteenth century. Prior to independence, under either foreign rule or during their own Maratha empire, Marathas were the dominant land owners.⁴¹ Under both Muslim and British regimes, land was allocated to Marathas by outside rulers to buy the loyalty of dominant lineages, and in return for supply of armies (Altekar 1927, Kosambi 1969, Charles Drekmeier 1962, S.M. Dahiwale 1995). During colonial rule, the regions of present-day Maharashtra fell under different administrative units and systems of land revenue collection.⁴² However, irrespective of the land revenue system used, Marathas continued to own the large

⁴⁰With the exception of the movement of a small population of Brahmins from rural to urban areas in the early 20th century. They are less than 1% of our sample, and it is this exodus of Brahmins from rural areas that further solidified the dominance of Marathas in this region.

⁴¹Under the leadership of Chhatrapati Shivaji, the Maratha Empire was founded in 1674. At its height in the 18th century, the empire extended from present-day Pakistan to Orissa in the east and from Punjab to central Karnataka in the south. It also included Tamil Nadu. The vast empire was in decline by 1818 when Maharashtra had fallen to the British East India Company, however remnants of it lasted until Independence in 1947.

⁴²In particular, Western Maharashtra was part of the Bombay Presidency which had a *ryotwari* (cultivator-based) system of land revenue collection. Eastern Vidarbha was part of Central Provinces which had a *zamindari* (landlord-based) system. Western Vidarbha was a part of Berar, formerly part of the princely state of Hyderabad, which was given to the East India Company as a debt payment in 1860 and made into a *ryotwari* region at that time. Marathawada never fell under British rule and remained a part of the princely state of Hyderabad until Independence in 1947. Land there was divided between government and feudal ownership. The former was run similarly to the *ryotwari* system whereas the latter was more similar to the landlord system. Refer to Banerjee and Iyer (2004) who analyse the impact of these different land systems on outcomes today. Our estimation results include regressors which control for these different land revenue systems.

majority of agricultural land. (This is documented in the *Imperial Gazetteers of India* which report the break down of caste land ownership patterns at the district level.⁴³).

Upon Independence, Indian states legislated large scale land reforms. In Maharashtra, the Tenancy and Agricultural Lands Act of 1948 placed a ceiling on all landholdings and transferred ownership rights to tenant cultivators. These acts effectively redistributed land from large land owners to their former permanent tenants (“Other Backward Castes” or OBCs under today’s classification). This led to a dramatic change in ownership (but not cultivation) patterns.⁴⁴ These land reforms thus represent a striking break with the past. They gave rise to a new class of landowners drawn from a previously non-landowning caste. The land reforms thus fully account for villages where a non-Maratha caste are the largest landowning group in our sample.

Since the reforms, other changes in land ownership and distribution have been almost entirely due to the process of inheritance and partition (land is typically divided amongst sons), with the combined ownership of each dynasty remaining fairly constant. Formal sales of land are rare. In our sample less than 2% had bought or sold land within the past 5 years (almost all distress sales) and 86% of our sample of landowners report that they inherited their land. Almost 12% report that they purchased some of their land, but this was, in almost all cases, a purchase from a relative or co-caste member.

This settlement history, and the fact that land reforms managed the redistribution of large landlord holdings ensures a distinct pattern of caste and land ownership in Maharashtrian villages today.⁴⁵ In villages where few Marathas reside, the dominant land-owning caste *can* be a low caste (OBCs, former tenants). In villages where Marathas are populous, although the lower castes typically also

⁴³The relevant publications are *Imperial Gazetteer of India, Provincial Series (1909)* for *Bombay Presidency; Hyderabad State; Central Provinces; and Berar*.

⁴⁴Maharashtra is one of the few states where the agricultural lands acts were comprehensively and successfully implemented, effectively granting of ownership rights to former tenants. Land ceilings were sometimes circumvented via transfers to extended family members, but land redistributions away from intermediaries and absentee landlords were highly effective. (A.R. Kamat 1980).

⁴⁵Anderson (2011) similarly treats land dominance by caste groups at the village level as predetermined using data from Uttar Pradesh and Bihar. The empirical strategy used here is also related to Besley and Robin Burgess (2000) who estimate the impact of state-level land reforms on outcomes today.

own some land, Marathas are highly likely to constitute the dominant landowning caste.

Marathas may own the highest quality land today because they historically chose to reside in the high quality land villages, and ran the lower quality land as absentee landlords. This is an issue we already discussed in Section I.B. There it was demonstrated, in Table 1, that there are no significant differences in village land use patterns and soil quality measures across Maratha land dominated villages compared to others. Table 1 also demonstrates no significant differences across Maratha and lower (OBC) caste land dominated villages in key demographic and geographic variables. There are no significant differences in terms of total population numbers, proportion SC (the lowest ranked caste group), cultivability of the land, rainfall patterns, and also distance to exogenous (to the GP responsibilities) measures of amenities.⁴⁶ These include distance to a national main road, major rivers, and the nearest town. We checked these differences using our own household and village level data and also using the Village Amenities data from the Census of India 2001.

*

REFERENCES

- Altekar, A. S.** *A History of Village Communities in Western India*. Bombay: Oxford University Press, 1927.
- Anderson, Siwan** "Caste as an Impediment to Trade" *American Economic Journal: Applied Economics*, 3(1), pp. 239-63, 2011.
- Banerjee, Abhijit and Lakshmi Iyer** "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India". *American Economic Review*, 95(4), pp. 1190-1213, 2005.
- Banerjee, Abhijit and Rohini Somanathan** "The political economy of public goods: Some evidence from India," *Journal of Development Economics*, 82(2), pp. 287-314, 2007.

⁴⁶The SC group, Scheduled Castes, are the lowest ranked castes, formerly known as the untouchable castes. They are ranked lower than the OBC category which refers to the backward caste groups. OBCs are the traditional farming and artisan castes while SCs traditionally performed menial tasks not directly related to agriculture, and hence they were not direct recipients of land during reforms.

- Besley, Timothy and Robin Burgess** "Land Reform, Poverty Reduction, And Growth: Evidence From India" *Quarterly Journal of Economics*, 115(2), pp. 389-430, 2000.
- Danda, Ajit K.** "Ethnic Status: Identity and hierarchy" *Journal of Indian Anthropological Society*, 22(1), pp. 52-64, 1987.
- Dahiwale, S. M.** "Consolidation of Maratha Dominance in Maharashtra" *Economic and Political Weekly*, 30(6), pp. 336-342, 1995.
- Drekmeier, Charles** *Kingship and Community in Early India*. Stanford: Stanford University Press, 1962.
- Dreze, Jean, Peter Lanjouw, and Naresh Sharma** "Economic development in Palanpur, 1957-93" in *Economic Development of Palanpur Over Five Decades*, ed. Peter Lanjouw and Nicholas Stern, pp. 114-242. Oxford University Press, 1999.
- Kamat, A. R.** "Politico-Economic Developments in Maharashtra: A Review of the Post-Independence Period" *Economic and Political Weekly*, 15(39), pp. 1627-1630, 1980.
- Kosambi, D. D.** "The basis of ancient Indian History, I" *Journal of the American Oriental Society*, 74(4), pp. 330-337, 1955.
- Kosambi, D. D.** *Ancient India: A History of its Culture and Civilization*. New York: World Publishing Company, 1969.
- Mandelbaum, David, G.** *Society in India*, University of California Press, 1970.
- Marriott, McKim** *Village India: Studies in Little Community*. Bombay, Asia Publishing House, 1955.
- Munshi, Kaivan and Mark Rosenzweig** "Why is Mobility in India so Low? Social Insurance, Inequality, and Growth" BREAD Working Paper No. 97, 2005.
- Srinivas, M.N.** *The Dominant Caste and Other Essays*. New Delhi, Oxford University Press, 1987.