Online Appendix: "Elite Capture: How Decentralization and Informal Institutions Weaken Property Rights in China"

Daniel C. Mattingly

World Politics, vol. 68, no. 3 (July 2016)

1 Outcomes of Land Expropriations

Figure 1 compares villages that have experienced land seizures with those that have not. The figures show point estimates and confidence intervals for β_1 in a regression $y_i = \beta_1 x_i + \beta_2 Z + \varepsilon$, where for each village i in the dataset, y_i is the outcome of interest, x_i is a dummy variable for whether or not a land expropriation occurred in the village, and Z is a matrix of pre-treatment variables. The conditioning variables are GDP (using a nighttime luminosity proxy), distance from the county seat, terrain roughness, agricultural suitability, and a set of province fixed effects.

Figure 1a shows that villages that have experienced land expropriations have higher levels of non-farm employment and higher incomes than villages that have not experienced a land seizure. The additional village revenue from these land seizures does not evidently go to increased public goods expenditure on schools, teacher salaries, and village infrastructure. Despite this, villagers are not so dissatisfied by land expropriations that in a typical case they petition higher levels of government.

Yet for an important subset of villages — those villages where lineage group leaders have been incorporated into the state — land expropriations appear to be significantly more predatory. Figure 1b shows the results of land expropriations villages where the leaders of lineage groups are village cadres. In these villages, land expropriations are not associated with increases in non-farm employment or income, or increases in public goods expenditure. Rather, they are associated with higher levels of petitioning, indicating dissatisfaction with the government.

2 Results Using Matching and Entropy Balancing

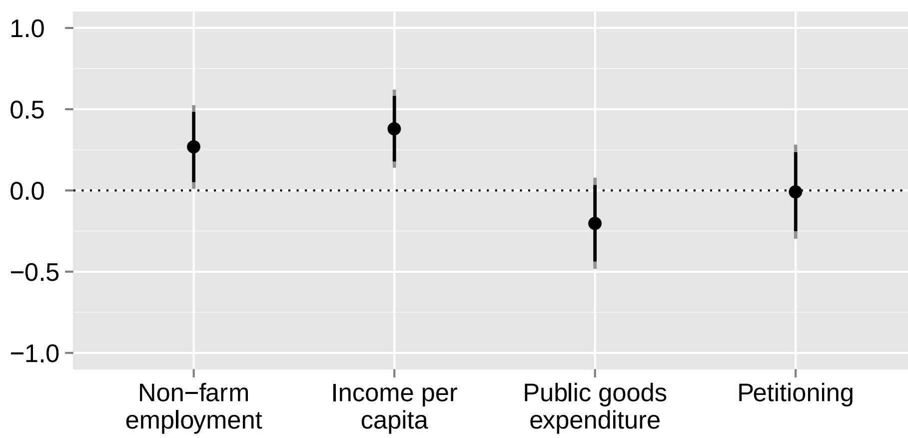
I use three additional techniques to calculate the effect of including lineage leaders in village political institutions on land rights, in each case using the same set of covariates as the regression estimates. First, I use propensity score matching with weighting based on the Mahnalbois distance metric. Propensity score matching matches observations based on the estimated probability of assignment to "treatment" — in this case lineage group leaders becoming village cadres. The observations are then reweighted using the Mahalanobis distance metric. Combining the Mahalanobis metric with propensity score matching is preferable to using propensity score alone.[1] Second, I use genetic matching, which has been demonstrated to improve covariate balance and reduce bias over other matching methods.[2] This technique also matches based on the propensity score, but then uses a machine learning algorithm to find weights for each covariate in order to optimize covariate

[1] Rosenbaum and Rubin 1985

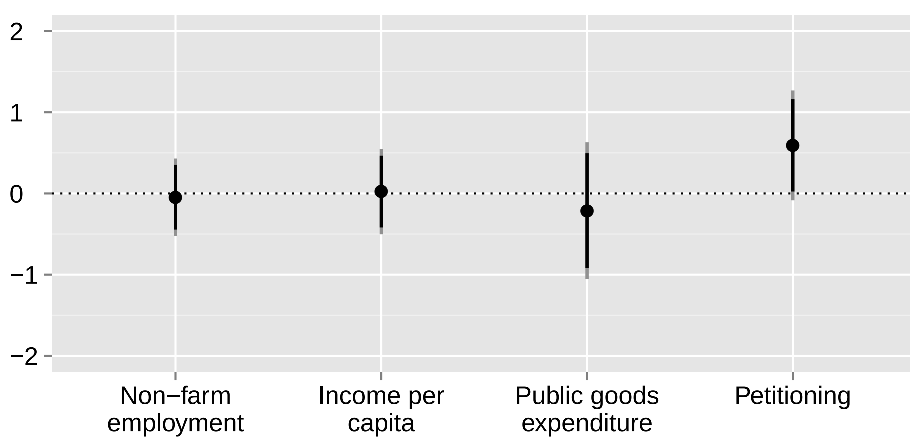
[2] Diamond and Sekhon 2013

Figure 1: Consequences of land seizures. Dependent variables zero-meaned and standardized to aid interpretability. Gray lines are 95 percent confidence intervals, black lines are 90 percent confidence intervals.

(a) All villages in CGSS sample. $N = 408$.



(b) Subset of villages where lineage leaders are incorporated into village political institutions. $N = 63$.



balance. Third, I use Hainmueller's entropy balancing, a technique for preprocessing data which reweights observations without matching. [3] As with matching, the user specifies a set of covariates which form the basis for a reweighting scheme, but in this case an entropy balancing algorithm finds weights for observations in the control group.

Selection into office by lineage leaders appears to be driven by village location and village wealth. Figure 3 shows kernel density plots for the three variables. This is consistent with a story of strategic selection by both higher levels of government and by lineage officials. That is, officials wish to co-opt lineage leaders in villages that are close to population centers and have high levels of economic development, since these leaders can help them facilitate development where land is most valuable. In these villages, the rewards of political office are also likely to tempt lineage leaders, who are more likely to accept nomination when the potential returns to office are high.[4]

Table 1: *Matching and entropy balancing estimates*. Difference in means test. Dependent variable is dummy for village experiencing a land seizure. For the matching estimates, there are 51 treated observations and 51 matched control observations.

Preprocessing Method	Estimate	Standard Error	t-statistic	p-value	N
Propensity Score Matching	0.200	0.082	2.41	0.016	102
Genetic Matching	0.182	0.087	2.09	0.037	102
Entropy Balancing	0.151	0.066	2.28	0.025	106

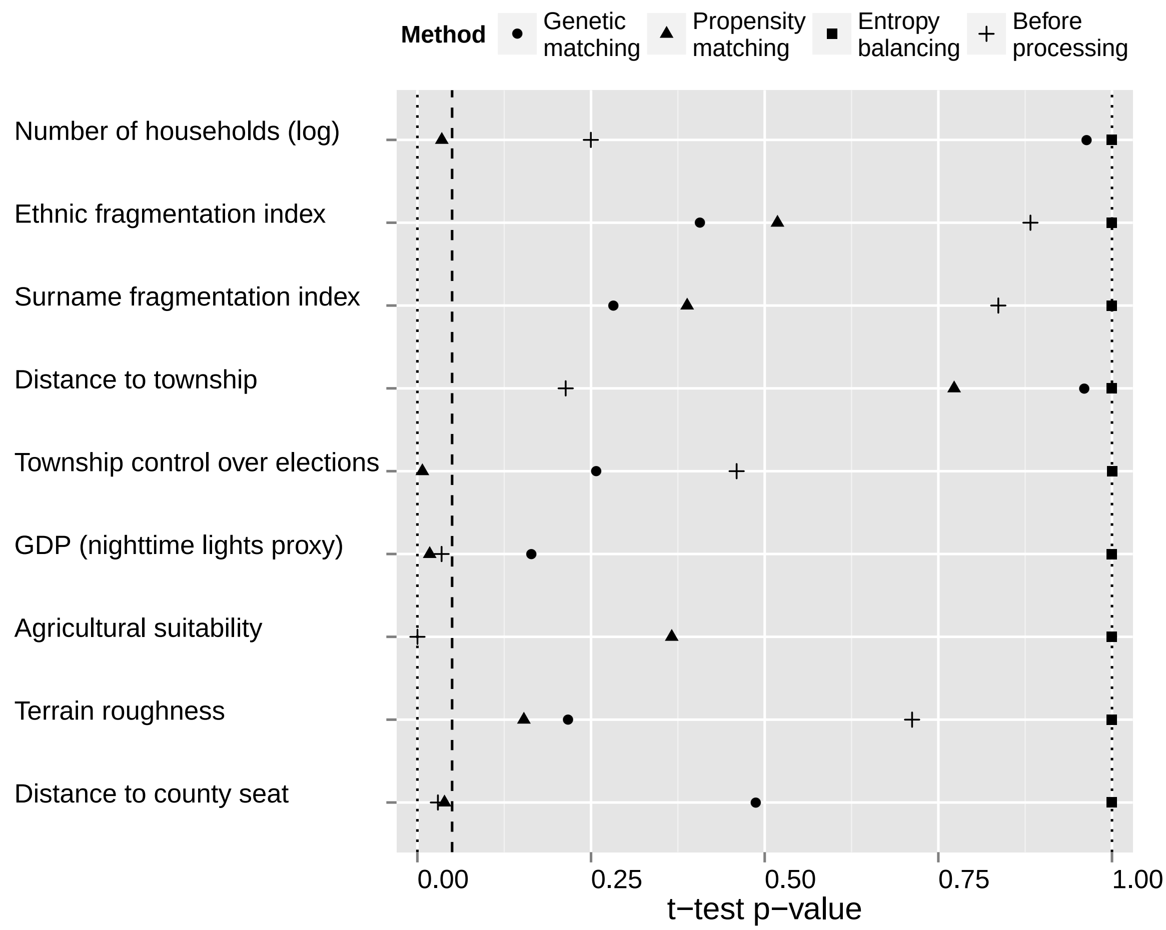
Each preprocessing method produces a different amount of covariate balance, with entropy balancing performing the best. Propensity score matching does not improve balance on the GDP or distance to county seat measures; worse yet, it creates a significant imbalance on the population and elections measures.[5] Genetic matching, on the other hand, improves balance so that there are no significant differences between treatment and control groups on any of the observed covariates. However, the p-values indicate lingering, albeit insignificant, amounts of imbalance. Finally, entropy balancing finds weights that

[3] Hainmueller 2012

[4] The differences in agricultural suitability are probably unrelated to either land expropriation or lineage incorporation, but are an artifact of the fact that population centers historically flourished in areas suitable for intensive agriculture.

[5] In this case, the specified propensity score formula includes all of the listed covariates. Propensity score estimates are included here only for comparative purposes, since both genetic matching and entropy balancing do the equivalent of an iterative search for a propensity score formula that maximizes covariate balance.

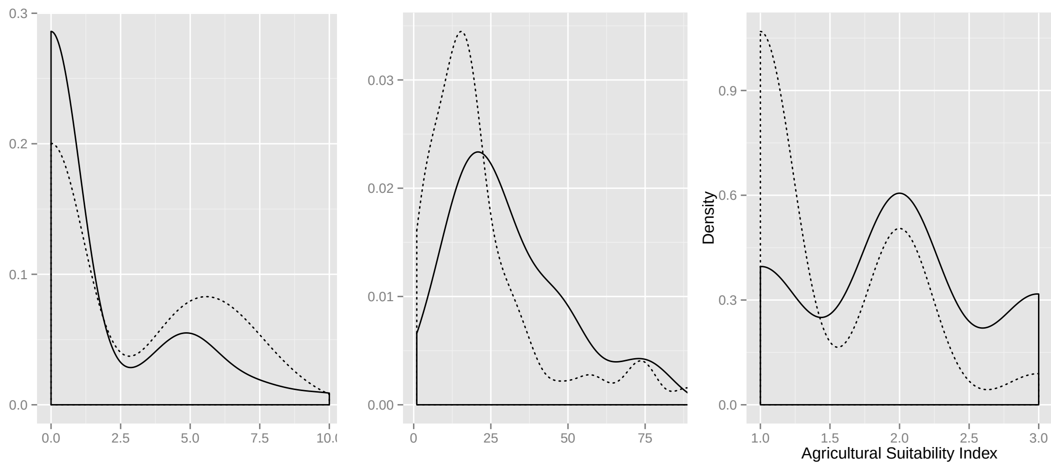
Figure 2: *Covariate balance*. Difference in means t-test p-values comparing the treatment and control groups, using different data preprocessing methods including propensity score matching (Rosenbaum and Rubin, 1985), genetic matching (Diamond and Sekhon, 2013), and entropy balancing (Hainmueller, 2012). Entropy balancing provides the highest degree of observed covariate balance.



result in essentially perfect balance between the treatment and control groups. In this case, estimates using entropy balancing are arguably preferable.

Similar to the regression estimates, these non-parametric techniques find that including

Figure 3: *Selection into treatment*. Kernel density estimates. The dotted line represents villages where lineage leaders are village cadres, and the solid line villages where lineage leaders are not village cadres. The data suggests that lineage leaders select into treatment in villages close to the county seat and with high levels of nighttime luminosity — in other words, where there are high rents from office.



(a) Nighttime luminosity (b) Distance to county seat (c) Agricultural suitability

lineage leaders in village political institutions leads to a 15 to 20 percent higher likelihood of land expropriations. Table 1 shows a difference in means tests using each estimator. The matching estimates find an 18 to 20 percent increase in land expropriations, while rejecting the null hypothesis of no difference between the groups. These estimates are slightly larger in magnitude to the regression estimates. The entropy balancing estimator (which, again, we might prefer over the other estimators because of the plausibility of the counterfactual it creates) finds a 15 percent increase in land expropriations.

3 Subgroup Analysis of Survey Experiment

One observable implication of the overall theory advanced in this article is that the size of the endorsement effect should vary between subsets of villagers. Of the 22 villages in the sample, 12 had announced redevelopment plans that would entail seizing villagers' property. In a survey that followed the experiment, I also asked villagers whether the most influential member of their lineage group was a village cadre.

Consider, first, respondents who live in a village where a property seizure has been announced, and whose actual lineage group leader is a government official. In these villages,

Table 2: Confidence in statement supporting an expropriation plan, endorsed by randomized village leader. Comparison condition is “villager” identity.

Lineage leader endorsement effect in experiment sub-samples

	Lineage leader is cadre	Lineage leader is not cadre
No property seizure announced	0.131 (0.197) N=38	0.189 (0.123) N=104
Property seizure announced	0.359*** (0.133) N=78	0.078 (0.077) N=213

the theory holds, the respondents’ co-opted lineage group leaders should be working to persuade them to accept the property seizure plan. If this is occurring, the endorsement of a hypothetical lineage leader should have a strong effect; after all, it should prime actual endorsements of the plan by actual lineage group leaders. This is precisely what occurs. The bottom row of Table 2 shows that where the respondents’ lineage group leader is also a government cadre, the estimated effect size nearly triples, to an estimate of 36 percent.[6] These results are consistent with a theory of elite-led persuasion and political control.

On the other hand, where lineage leader is a village cadre but no expropriation has been announced, the effect size is comparatively small, just 13 percent. One important alternative explanation for the observational results presented below is that the incorporation of lineage elites is epiphenomenal — particularly, that it reflects varying degrees of top-down control by higher levels of governments. Were this the case, we would expect a large effect size in this subgroup, especially given the widespread media coverage of government approval of the land expropriations. Instead, the effect size in this subgroup is small, weakening support for this rival explanation.

Next, consider respondents whose lineage group leaders are not government officials,

[6]The results are even stronger when I subset on the narrower question of whether the village chief in particular is a member of the respondent’s lineage group. This aside, a separate concern might be that the experimental manipulation itself contaminates later answers about whether the lineage group leader is a cadre. The results from villages where no seizure have been announced should alleviate this concern.

but whose property is under threat of expropriation. The theory suggests that in these cases, lineage group leaders are more likely to attempt to organize villagers in opposition to property seizures. If this is occurring, we would expect to observe a diminished effect for the endorsement of a hypothetical lineage group leader. Their real lineage leaders will have been weighing in against the offer, not supporting it as the experimental prompt suggests, thus diminishing the credibility of the experimental manipulation. The results presented in Table 2 are consistent with a diminished effect in this subgroup.

Taken together, the results of the experiment suggest that villagers are more likely to have confidence in information about property takings from lineage group elites. It also provides suggestive though by no means conclusive evidence that a process of elite-led persuasion is in fact at work where expropriation plans have been announced, consistent with the qualitative process-tracing evidence.

4 Surname Fragmentation

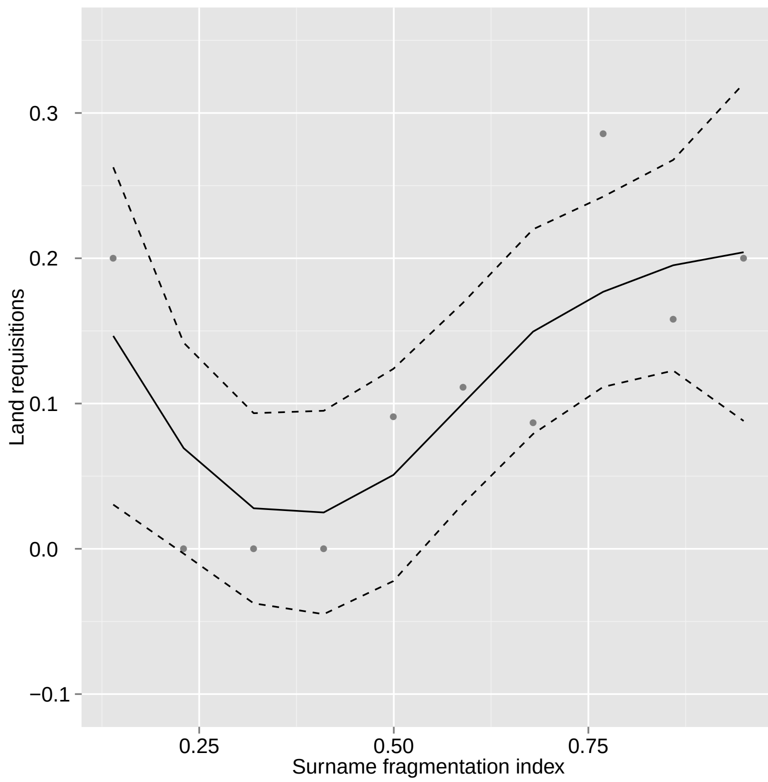
In this section, I discuss two potential issues that go to the heart of the argument that informal institutions are used by elites to control lineage groups and increase compliance with extractive land policies. First, and most importantly, it may be the case that lineage group leaders do not target members of their own lineage group for expropriation.

If lineage leaders do not target their own groups for expropriation, we should expect that low levels of lineage group fragmentation leads to low levels of expropriation. That is, in villages dominated by a single lineage group there would be very few land seizures. On the other hand, if lineage group leaders expropriate from their own group, we would see high levels of expropriation in villages with one lineage group.

The results provide the most support to the interpretation that leaders target their own groups, and that informal institutions are largely being used as channels of political control. Figure 4 plots the likelihood of expropriation against the index of surname group fragmentation, which is a reasonable if imperfect proxy for lineage group fragmentation. The figure shows a U-shaped relationship between lineage group fragmentation and land seizures.

Of particular interest is the fact that the results show high levels of expropriation with low levels of surname fragmentation. It shows that in villages with a single surname group, expropriations are likely to occur. This is difficult to reconcile with the potential objection that lineage group leaders exclusively target other groups for expropriation. It also does not rule out the likelihood of some degree of targeting in villages with multiple lineage groups. However, the structure of landholdings in villages — where lineage groups are often deliberately mixed into different landholding collectives — makes targeted expropriation somewhat difficult to accomplish.

Figure 4: *Surname fragmentation and land expropriation*. Points are binned means and the solid line is a loess estimate using weighted least squares; the dotted lines are the 95 percent confidence intervals.



5 Descriptive Statistics, Variable Creation, and Alternative Variables,

The subsequent tables include descriptive statistics, descriptions of the main variables, and alternative operationalizations.

Table 3: Full sample descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
Distance to township (km)	393	5.712	5.300	0	38
Distance to county seat (km)	393	29.489	21.872	0	115
Terrain roughness (meters)	407	231.042	254.248	1	1,204
Agricultural suitability (1-3 index)	406	1.628	0.836	1	7
Wealth, 1992 nighttime luminosity proxy	408	4.025	7.267	0	61
Surname fragmentation (0-1 index)	378	0.713	0.229	0.08	1
Ethnic fragmentation (0-1 index)	408	0.033	0.099	0	0.50
Ethnic minority population (percent)	408	0.080	0.240	0	1
Log number of households	393	6.032	0.717	3.64	7.87
Implemented village elections (dummy)	408	0.856	0.202	0	1
Township controls village elections (0-1 index)	408	0.159	0.209	0	1
Participation rate in elections (percent)	408	0.696	0.461	0	1
Elections have more than one candidate (dummy)	408	0.616	0.305	0	1
Village has active lineage group (dummy)	408	0.279	0.449	0	1
Village lineage group has *citang* or *zupu* (dummy)	408	0.196	0.398	0	1
Lineage leader is village cadre (dummy)	408	0.154	0.362	0	1
Land seizure during leader tenure (dummy)	392	0.151	0.358	0	1
Land seizures prior to tenure (dummy)	392	0.048	0.215	0	1
Villagers petitioned higher levels (dummy)	408	0.115	0.320	0	1
Average village income (yuan)	390	1,899	1,242	50	7,400
Non-agricultural employment (percent)	408	0.044	0.104	0	0.80
Average education level (years)	408	5.697	1.742	0.70	9.91
Participation in local religious festivals (percent)	408	0.119	0.195	0	1

Table 4: Variable descriptions. Abbreviation key at the bottom of the table.

Variable	Source	Variable Description
Village coordinates	Baidu/Google Maps	Latitude and longitude.
Nighttime luminosity	DMSP	Luminosity values from 1992 DMSP stable lights raster.
Agricultural suitability	UN-FAO	1-3 scale. Nutrient availability from the Harmonized World Soil Database v1.2.
Altitude	SRTM	Meters. Mean within 10 km of village centroid.
Terrain roughness	SRTM	Meters. Difference between minimum and maximum altitude in 10 km radius.
Land expropriation	CGSS-A: G2b	Land reallocation due to government land expropriation or confiscation for use by enterprises; limited to those occurring after current village leadership took office (from CGSS-B: C3 and C4).
Village land revenue	CGSS-B: B1f, i	Log revenue in 2004 from land sales and rental of land and other village property.
Village cultivated area	CGSS-B: A12a	Log *mu* of cultivated land in 2004 in the village.
Lineage group is active	CGSS-A: F10	Dummy variable. "Does your village have an active lineage network or organization?" Please see main text and appendix for, respectively, main and alternate coding.
Lineage elite is cadre	CGSS-A: F11	Dummy variable. "Are the leaders or most influential members of the lineage current village cadres?" Dichotomized from Likert response scale to be 1 if response is entirely, mostly, or generally. 0 if never, or somewhat. Likert scale version shown in appendix.
Distance to county	CGSS-B: A4a	Kilometers. Village cadre estimate.
Distance to township	CGSS-B: A4b	Kilometers. Village cadre estimate.
Township electoral control	CGSS-A: G14-1	Percent indicating township control over elections.
Surname fragmentation	CGSS-B: A10	0-1 index. Please see main text for formula.
Ethnic fragmentation	CGSS-A: A4	0-1 index. Please see main text for formula.
Number of households	CGSS-B: A2a	Number of households.
Public goods spending	CGSS-B: Bc, d, f	Log per capita spending on teacher salaries, welfare, and public construction.
Non-farm employment	CGSS: A7	Percent of respondents indicating non-farm employment.
Petitioning	CGSS: F8b-c	Indicator for petitioning government (*shangfang*).

Key: CGSS-A: Household survey; CGSS-B: Cadre survey; SRTM: Shuttle Radar Topography Mission; DMSP: Defense Meteorological Satellite Program; UN-FAO: U.N. Food & Agriculture Office

Table 5: *Alternate dependent variable: Log village revenue from land*. Least squares regression estimates.

	(1)	(2)	(3)	(4)
Lineage leader is cadre	1.021**	0.830**	0.787*	0.853**
	(0.425)	(0.419)	(0.419)	(0.406)
Active lineage with ancestral hall or zupu	−0.174	−0.007	0.010	0.006
	(0.307)	(0.303)	(0.304)	(0.294)
Distance to county seat (km)		−0.009**	−0.006	−0.004
		(0.004)	(0.005)	(0.004)
Terrain roughness		0.0001	0.0001	−0.0001
		(0.0004)	(0.0004)	(0.0004)
Agricultural suitability index		0.132	0.118	0.114
		(0.108)	(0.108)	(0.113)
Wealth (nighttime lights proxy)		0.052***	0.051***	0.052***
		(0.014)	(0.014)	(0.014)
Township control over elections			0.102	0.341
			(0.422)	(0.424)
Distance to township (km)			−0.037**	−0.034*
			(0.018)	(0.018)
Surname fragmentation index				0.354
				(0.399)
Ethnic fragmentation index				−0.177
				(0.865)
Log number of households				0.204
				(0.133)
Constant	0.609***	0.441*	0.561**	−1.000
	(0.101)	(0.249)	(0.266)	(0.828)
Observations	375	374	374	361
R^2	0.019	0.076	0.087	0.109
Adjusted R^2	0.014	0.061	0.067	0.081
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6: *Alternate dependent variable: Log cultivated land in village*. Least squares regression estimates.

	(1)	(2)	(3)	(4)
Lineage leader is cadre	-0.437^{**}	-0.474^{**}	-0.494^{**}	-0.347^{*}
	(0.212)	(0.212)	(0.213)	(0.181)
Active lineage with ancestral hall or zupu	0.118	0.176	0.192	-0.001
	(0.152)	(0.153)	(0.154)	(0.132)
Distance to county seat (km)		0.0005	0.001	0.004^{*}
		(0.002)	(0.002)	(0.002)
Terrain roughness		-0.001^{***}	-0.001^{***}	-0.001^{***}
		(0.0002)	(0.0002)	(0.0002)
Agricultural suitability index		-0.053	-0.054	-0.121^{**}
		(0.055)	(0.055)	(0.051)
Wealth (nighttime lights proxy)		0.005	0.005	-0.001
		(0.007)	(0.007)	(0.006)
Township control over elections			-0.156	-0.025
			(0.219)	(0.193)
Distance to township (km)			-0.007	-0.002
			(0.009)	(0.008)
Surname fragmentation index				0.185
				(0.180)
Ethnic fragmentation index				0.570
				(0.393)
Log number of households				0.705^{***}
				(0.060)
Constant	7.482^{***}	7.669^{***}	7.718^{***}	3.311^{***}
	(0.051)	(0.125)	(0.134)	(0.372)
Observations	391	389	389	374
R^2	0.012	0.045	0.048	0.333
Adjusted R^2	0.007	0.030	0.028	0.313

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 7: *Alternate explanatory variable: Likert scale version of degree to which lineage leaders are also cadres*. Least squares regression estimates.

	(1)	(2)	(3)	(4)
Lineage leader is cadre (scaled 0-1)	0.167*	0.177*	0.172*	0.178*
	(0.096)	(0.098)	(0.097)	(0.095)
Active lineage group	−0.086	−0.091	−0.087	−0.077
	(0.062)	(0.064)	(0.063)	(0.063)
Distance to county seat (km)		−0.001	0.0003	0.001
		(0.001)	(0.001)	(0.001)
Terrain roughness		−0.0001	−0.0001	−0.0001
		(0.0001)	(0.0001)	(0.0001)
Agricultural suitability index		0.036	0.032	0.026
		(0.022)	(0.022)	(0.023)
Wealth (nighttime lights proxy)		0.001	0.001	0.001
		(0.003)	(0.003)	(0.003)
Township control over elections			0.113	0.115
			(0.086)	(0.088)
Distance to township (km)			−0.011***	−0.011***
			(0.004)	(0.004)
Surname fragmentation index				0.160*
				(0.083)
Ethnic fragmentation index				−0.266
				(0.180)
Log number of households				0.006
				(0.027)
Constant	0.151***	0.133***	0.150***	0.004
	(0.021)	(0.051)	(0.054)	(0.170)
Observations	392	390	390	376
R^2	0.008	0.022	0.049	0.056
Adjusted R^2	0.003	0.007	0.029	0.027
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 8: *Alternate explanatory variable: Percent of respondents indicating lineage group is active*. Least squares regression estimates.

	(1)	(2)	(3)	(4)
Lineage leader is cadre	0.129**	0.133**	0.139**	0.141**
	(0.057)	(0.058)	(0.057)	(0.057)
Percent of villagers indicating active lineage	−0.217	−0.213	−0.240*	−0.221*
	(0.132)	(0.134)	(0.133)	(0.134)
Distance to county seat (km)		−0.001	0.0004	0.001
		(0.001)	(0.001)	(0.001)
Terrain roughness		−0.0001	−0.0001	−0.0001
		(0.0001)	(0.0001)	(0.0001)
Agricultural suitability index		0.036*	0.033	0.027
		(0.022)	(0.022)	(0.023)
Wealth (nighttime lights proxy)		0.001	0.001	0.001
		(0.003)	(0.003)	(0.003)
Township control over elections			0.127	0.129
			(0.086)	(0.089)
Distance to township (km)			−0.011***	−0.011***
			(0.004)	(0.004)
Surname fragmentation index				0.154*
				(0.083)
Ethnic fragmentation index				−0.262
				(0.180)
Log number of households				0.011
				(0.027)
Constant	0.145***	0.119**	0.135**	−0.039
	(0.020)	(0.051)	(0.054)	(0.171)
Observations	392	390	390	376
R^2	0.014	0.028	0.057	0.063
Adjusted R^2	0.009	0.013	0.037	0.035
Note:		*p<0.1; **p<0.05; ***p<0.01		

References

Diamond, Alexis and Jasjeet S Sekhon. 2013. "Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies." *Review of Economics and Statistics* 95(3):932–945.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(1):25–46.

Rosenbaum, Paul R and Donald B Rubin. 1985. "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." *The American Statistician* 39(1):33–38.