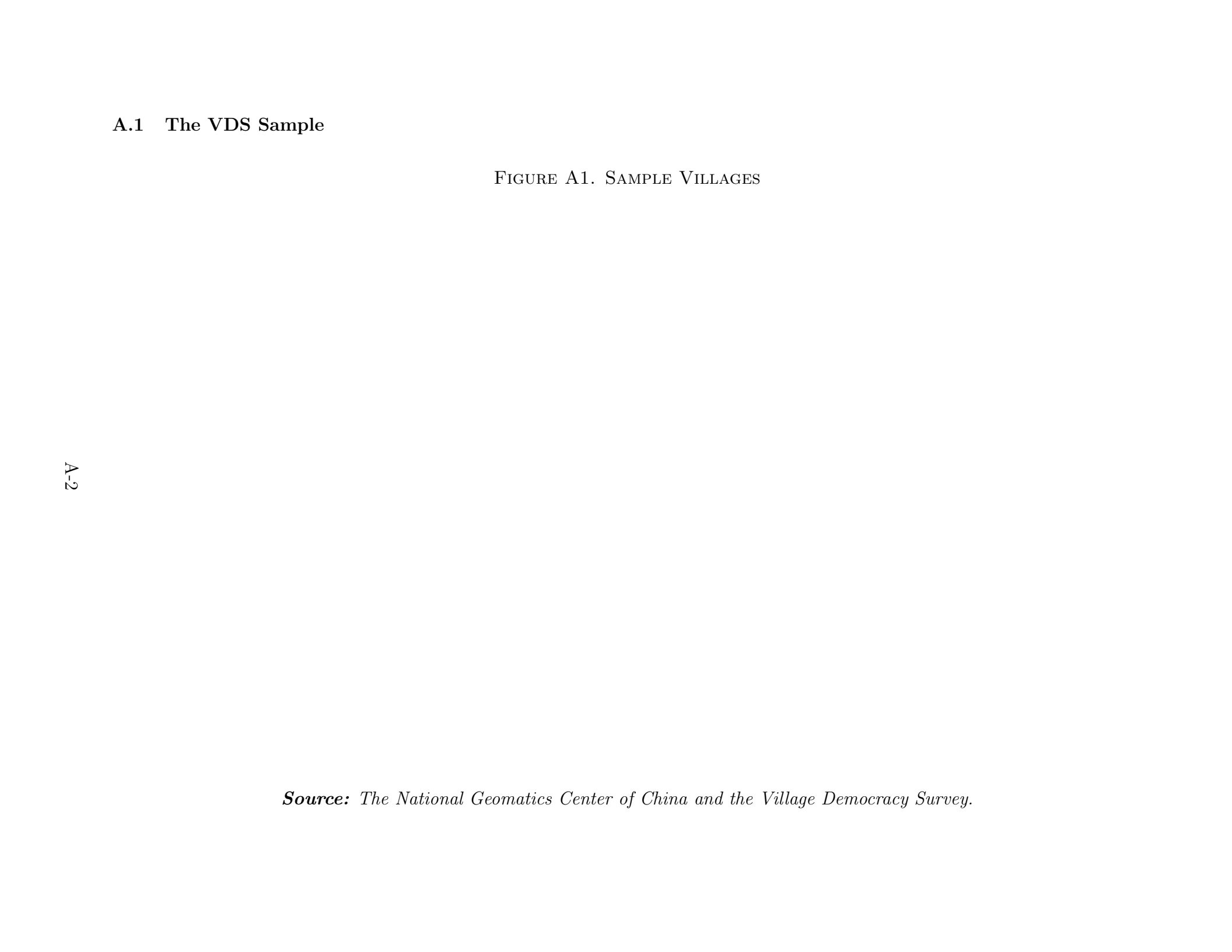
A Online Appendix (Not For Publication)

Contents of the Appendix

A.1 The VDS Sample

FIGURE A1. SAMPLE VILLAGES



Source: *The National Geomatics Center of China and the Village Democracy Survey.*

A.2 Robustness Checks for the Main Results

In the main text, we only use observations in the post-election period. As a result, the panel is imbalanced. If the timing of the introduction of elections were correlated with the presence of a VC of large clans and public goods expenditure, the estimated coefficient of VC of large clans could be biased. O'Brien and Li (2006) report that regional governments did have concerns to introduce elections to villages that were dominated by one large lineage group. The governments were worried that the elected positions would be captured by the dominant clan, which would implement policies for the benefits of its members at the cost of others. To minimize potential biases caused by the onset of elections, we use a subsample of post-1995 observations and re-estimate the models. Since most villages already began elections in 1995, the panel is much more balanced.

Table A1 Columns 1–4 present the results. The estimates are slightly larger than the baseline results and remain statistically significant. Column 5–7 in the same table show that the estimates are stable when we drop observations after 2000, when the rural tax-and-fee reform started to be experimented within some regions. Note that we do not include village-specific time trends when using subsamples because the time series are too short, which results in highly singular variance-covariance matrix; however, the estimated coefficients of the VC dummies are always large and positive.

One might also be worried that our results are driven by a few extreme values. In Table A2, we replace the outcome variable with a binary indicator of whether there was any investment in a year and redo the exercises. The results show that on average a VC of large clans is associated with a 6–8 percent increase in the probability of public investment, or 25–35 percent of the dependent variable mean.

Table A3 shows that our main findings hold if we do not include the indicator of VC of the second-largest clan in regressions.

Table A1. VC of Large Clans and Village Public Investment: Subsamples

	Log Public Investment (1,000 yuan)					
	After 1995			Before 2000		
	(1)	(2)	(3)	(4)	(5)	(6)
	FE	FE	FE	FE	FE	FE
VC of the largest clan	0.445	0.511	0.503	0.386	0.354	0.338
	(0.215)	(0.205)	(0.210)	(0.173)	(0.178)	(0.188)
VC of the second-largest clar	0.320	0.432	0.567	0.282	0.280	0.310
	(0.243)	(0.256)	(0.272)	(0.159)	(0.161)	(0.169)
Dependent variable mean	1.328	1.328	1.310	0.916	0.916	0.891
Year fixed effects	x	x	x	x	x	x
Village fixed effects	x	x	x	x	x	x
Provincial linear trends		x	x		x	x
NFS controls			x			x
Observations	2,317	2,317	2,220	2,644	2,644	2,448
Villages	220	220	217	217	217	206

Note: *This table shows that the association between the presence of a VC of large clans and a larger amount of village public investment is robust in post-1995 and pre-2000 subsamples. Columns 1–3 use observations after 1995 while Columns 5–6 use observations before 2000. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) during that year. The independent variables are two dummy variables indicating whether a VC came from the village's largest or second-largest clan, respectively. The sample is based on village-year observations after village elections were introduced. All regressions control for both village and year fixed effects. In addition, Columns 2, 3, 5, and 6 control for provincial linear time trends. Columns 3 and 6 include five time-varying control variables from the NFS dataset, including average household size, arable land per capita, log income per capita, log village assets, and log village population.*

Table A2. VC of Large Clans and Village Public Investment: Binary Outcome

	Binary Outcome: Any Pubic Investment					
	(1) OLS	(2) FE	(3) FE	(4) FE	(5) FE	(6) FE
VC from the largest clan	0.059	0.082	0.078	0.077	0.074	0.094
	(0.024)	(0.029)	(0.029)	(0.038)	(0.030)	(0.038)
VC of the second-largest clan	0.040	0.060	0.062	0.062	0.060	0.060
	(0.032)	(0.031)	(0.031)	(0.045)	(0.030)	(0.044)
Dependent variable mean	0.231	0.231	0.231	0.231	0.228	0.257
Year fixed effects	x	x	x	x	x	x
Village fixed effects		x	x	x	x	x
Provincial linear trends			x		x	x
Village linear trends				x		
NFS controls					x	x
Persons migrating out						x
Taxes/fees to the upper-level government						x
Transfers from the upper-level government						x
Observations	3,742	3,742	3,742	3,742	3,513	2,530
Villages	220	220	220	220	217	208

Note: *This table shows that the presence of a VC of large clans is associated with a higher probability of a village public investment project. Standard errors clustered at the village level are in parentheses. The dependent variable is a dummy variable indicating whether there was any village investment during that year. The independent variables are two dummy variables indicating whether a VC came from the village's largest or second-largest clan, respectively. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. Column 1 controls for year fixed effects only; the rest control for both village and year fixed effects. In addition, Columns 3, 5, and 6 control for provincial linear time trends; Column 4 controls for village linear time trends; and Columns 5 and 6 include five time-varying control variables from the NFS dataset, including average household size, arable land per capita, log income per capita, log village assets, and log village population. Column 6 additionally controls for the number of persons migrating out of the village each year, log total taxes and fees the village committee handed over to the upper-level government and log transfers it received from the upper-level government, all of which are available after 1993 (the data for 1994 are interpolated).*

Table A3. VC of the Largest Clan and Village Public Investment

	Log Public Investment (1,000 yuan)					
	(1) OLS	(2) FE	(3) FE	(4) FE	(5) FE	(6) FE
VC of the largest clan	0.295	0.349	0.310	0.303	0.306	0.400
	(0.121)	(0.145)	(0.144)	(0.180)	(0.152)	(0.193)
Dependent variable mean	1.092	1.092	1.092	1.092	1.083	1.225
Year fixed effects	x	x	x	x	x	x
Village fixed effects		x	x	x	x	x
Provincial linear trends			x		x	x
Village linear trends				x		
NFS controls					x	x
Persons migrating out						x
Taxes to the upper-level government						x
Transfers from the upper-level government						x
Observations	3,742	3,742	3,742	3,742	3,513	2,530
Villages	220	220	220	220	217	208

Note: *This table shows that the presence of a VC of large clans is associated with a larger amount of village public investment. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) during that year. The independent variables is a dummy variable indicating whether a VC came from the village's largest clan. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. Column 1 controls for year fixed effects only; the rest control for both village and year fixed effects. In addition, Columns 3, 5, and 6 control for provincial linear time trends; Column 4 controls for village linear time trends; and Columns 5 and 6 include five time-varying control variables from the NFS dataset, including average household size, arable land per capita, log income per capita, log village assets, and log village population. Column 6 additionally controls for the number of persons migrating out of the village each year, log total taxes and fees the village committee handed over to the upper-level government and log transfers it received from the upper-level government, all of which are available after 1993 (the data for 1994 are interpolated).*

A.3 Informal Institutions and Clan Size

In this section, we show that (1) our main results are robust when we control for the VC's clan size, (2) the effect of informal institutions, as we measure them, varies little across clans with different sizes, and (3) our results are robust when we use clan size (with different thresholds) as a measure of the strength of informal institutions. We also discuss why we think the rank order is a better measure for the clan's social power than the clan size.

Does clan size matter? First, we empirically test whether the magnitude of clan size matters. We directly incorporate both relative and absolute size of the VC's clan in two-way fixed-effect models. The results are reported in Table A4. In Column 1, the key independent variable is the relative size of the VC's clan, measured by the number of villagers in the VC's clan divided by the village's total population. The estimate is positive but not statistically significant. In Column 2, we additionally include the original rank order measure, in which case, we essentially treat the relative size of the VC's clan as a confounding factor. The estimated coefficient of the dummy variable is 0.438 and highly significant while the coefficient of relative clan size becomes negative and statistically insignificant. In Columns 3 and 4, we conduct similar tests but replace the relative size of the VC's clan by its absolute size (in 1,000 persons). The results are very similar. The estimated coefficient of the absolute size is positive but not significant. After we add the original rank order measure to the regression, the coefficient of the absolute size becomes almost zero, while the coefficient of the rank order measure is positive and highly significant. These results, taken at face value, show that once conditional on the rank order, the clan size has very limited explanatory power for the amount of public goods expenditure.

Heterogeneous treatment effect. Second, we want to know whether the effect of informal institutions on public goods expenditure is larger when the VC came from a larger clan. In other words, we are interested in the heterogeneous treatment effect of VC of the two largest clans. We then interact the binary indicator VCs of large clans D_{it} with a third-order

polynomial of the size of the VC's clan:

$$y_{it} = \beta D_{it} + \gamma_1 D_{it} \times \omega_{it} + \gamma_2 D_{it} \times \omega_{it}^2 + \gamma_3 D_{it} \times \omega_{it}^3 + + \eta_i + \delta_t + \epsilon_{it}, \tag{3}$$

where ω_{it} is the population share of the VC's clan in village i in year t (we do not control for the level terms ω_{it}, ω_{it}^2, and ω_{it}^3 because they are highly colinear with the interaction terms). The marginal effect of VCs of clans, therefore, is $(\beta + \gamma_1 \omega_{it} + \gamma_2 \omega_{it}^2 + \gamma_3 \omega_{it}^3)$. We are interested in whether the magnitude of the effect of informal institutions is dependent on the size of the VC's clan. The result is depicted in in Figure A2. Figure A2 shows that the effect of VC of large clans as measured by the rank order of VCs' clan size is relatively stable before the population share of the two largest clans reaches 75 percent. In fact, they are close to the baseline estimate of 0.369 when a constant treatment effect is assumed. However, when the two largest clans consist of more than 75 percent of the village population, the estimates decline quickly and turn insignificant. This change occurs because (1) the number of villages with village-wide lineage groups is very small (as Figure A2 itself shows), and (2) there is simply not enough variation in the VC dummy since most of the VCs in these villages came from large clans.

Different thresholds. In the main text, we mainly use the population rank order to measure a clan's social power (and hence, the strength of informal institutions associated with the VC's clan). In the following exercise, we measure the strength of lineage groups solely based on the number of people a clan has. In other words, if the size of a clan goes beyond a certain threshold, we code the group as a large clan, and estimate the effect of VC of large clans given the threshold. Because a threshold can be arbitrarily set, we try 100 thresholds with an interval of 20 persons between 0 to 2,000 persons (an average village in the period had around 1,500 villagers). The results of this analysis is shown in Figure A3. We find that the coefficient of VC of large clans is positive and statistically significant when the threshold is between 680 to 1240 persons, a large and reasonable interval. Moreover, if we exclude VCs from the third- and fourth-largest clans from VCs of large clans, the coefficient

of VC of large clans is significant at almost all thresholds below 1240 persons. This means that even with the same group size, the largest and second-largest clans in a smaller village were fundamentally different from the third- and fourth-largest clans in a larger village in terms of social power.

Because of the large heterogeneities across the country, clans of the same absolute or relative size may have vastly different levels of social power. For example, a clan of 20 households in a socially fragmented village might be the largest clan of the village and thus more powerful than the largest clan in a village consisted of two clans with more or less equal sizes. Moreover, there can be much bigger measurement errors in the absolute or relative size of clans than in their population rank order, especially when we only took a snapshot in 2011. The size of a clan might have changed substantially over the 20-year period covered by our study, but the population rank order should be more stable. Measures of social cohesiveness, such as lineage halls and ceremonies can provide information about the intensity of within-clan social activities, but may not fully capture clans social power in the village. In the *Main Results* Section of the paper, indeed we see that it is the clan's social power that matters rather than its size.

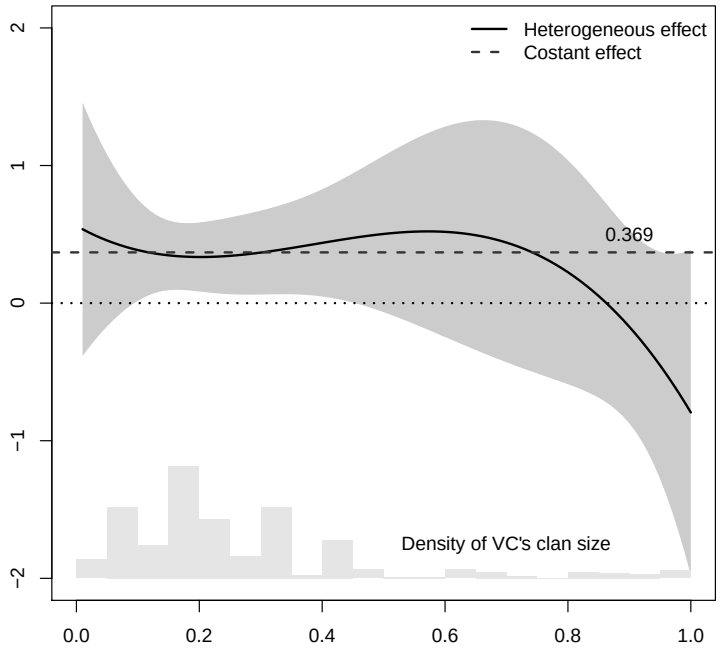
In summary, we find that, the population rank order of clans is controlled for, the clan size has almost no predictive power for the amount of public goods expenditure. These results also indicate that the rank order of a VC's clan is a good proxy for the strength of informal institutions associated with the VC's clan.

TABLE A4. VC OF LARGE CLANS AND PUBLIC INVESTMENT: CLAN SIZE

	Log Public Investment (1,000 yuan)			
	(1)	(2)	(3)	(4)
	FE	FE	FE	FE
Relative size of the VC's clan	0.750	-0.292		
	(0.426)	(0.564)		
Absolute size of the VC's clan			0.381	-0.013
(1,000 persons)			(0.279)	(0.335)
VC of the two largest clans		0.438		0.355
		(0.160)		(0.158)
Dependent variable mean	1.092	1.092	1.077	1.077
Year and village fixed effects	x	x	x	x
Observations	3,742	3,742	3,530	3,530
Villages	220	220	208	208

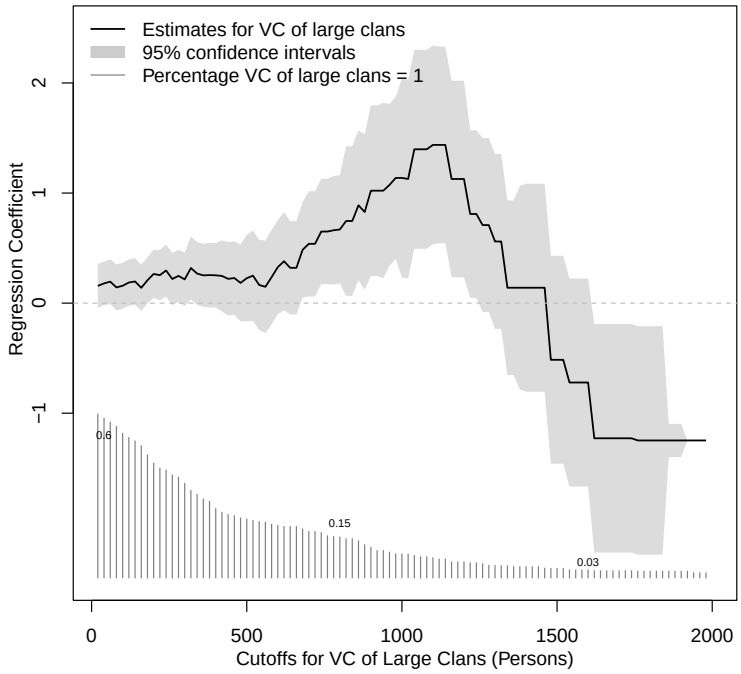
Note: *In this table, we explore the relationship between the VC's clan size, measured by the relative and absolute population share of the VC's clan, and the level of public investment. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) in that year. Note that we only record the size of the four largest clans (surnames) in a village; the size of other kinship groups is coded as 0. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

FIGURE A2. THE HETEROGENOUS EFFECT OF VCS OF LARGE CLANS ON PUBLIC INVESTMENT



Note: *This figures shows the heterogeneous effect of VCs of large clans on the amount of public investment. The x-axis is the VC's clan size. The y-axis is the marginal effect of VC of large clans. The specification we use is shown in Equation 3.*

FIGURE A3. THE EFFECT OF VCS FROM LARGE CLANS ON PUBLIC INVESTMENT: DIFFERENT THRESHOLDS



Note: *This figure shows the estimated coefficients of VC of large clans using different threshold for large clans. For example, if the threshold is set at 500 persons, the dummy variable VC of large clans would equal one if the VC's clan consisted of more than 500 people and zero otherwise. The bars on the floor of the figure show the percentages of village-year observations when the variable VC of large clans equals one.*

A.4 Clan Cohesiveness and the Role of Village Party Organizations

Figure 4 in the main text is based on the regression results reported in Table A5 Columns 1-3 with each column corresponding to a panel in the figure. In Column 4, when we put all three interaction terms in the regression, the coefficient of the interaction between the VC dummy and lineage halls remains large and significant. The coefficients of the other two interactions are negative but statistically insignificant.

Figure 5 in the main text is based on the regression results reported in Table A6 Columns 2-4 with each column corresponding to a panel in the figure. In Column 1, we only include the dummy variable indicating whether the VPS was from one of the two largest clans (VPS of large clans), as well as its interaction with VC of large clans. We find that the coefficient of VC of large clans is still large and statistically significant. The coefficient of VPS of the large clans is 0.249, slightly smaller than that of VCs of the largest clan, but statistically significant.

TABLE A5. VCs OF LARGE CLANS AND CLAN COHESIVENESS

	Log Public Investment (1,000 yuan)			
	(1)	(2)	(3)	(4)
	FE	FE	FE	FE
VC of large clans	0.433	0.277	0.180	0.301
	(0.147)	(0.164)	(0.134)	(0.165)
× Combined size > 50%	-0.144			-0.181
	(0.242)			(0.256)
× Records of family trees		0.204		-0.107
		(0.261)		(0.256)
× Lineage hall			1.021	1.095
			(0.331)	(0.338)
Dependent variable mean	1.092	1.102	1.102	1.102
Year and village fixed effects	x	x	x	x
Observations	3,742	3,367	3,367	3,367
Villages	220	200	200	200

Note: *This table shows that the association between a VC of large clans and village public investment is stronger in villages with more cohesive large clans, but it is not increasing in the VC's clan size. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) in that year. The independent variables are a dummy variable indicating whether a VC came from the village's largest or second-largest clan and its interactions with (1) whether the combined size of the two largest clans is above 50 percent, (2) whether any of the two largest clans had kept records of family trees, and (3) whether they had maintained any lineage halls since the beginning of the observed time periods. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

Table A6. Large Clan Leaders, Village Party Organizations, and Village Public Investment

	Log Public Investment (1,000 yuan)			
	(1) FE	(2) FE	(3) FE	(4) FE
VC of large clans	0.509	0.581	0.580	0.473
	(0.179)	(0.184)	(0.178)	(0.225)
VPS of large clans	0.333	0.397	0.398	0.620
	(0.183)	(0.189)	(0.191)	(0.258)
VC of large clans × VPS of large clans	-0.172	-0.214	-0.093	-0.514
	(0.274)	(0.288)	(0.335)	(0.417)
VC as the VPS ("one shoulder")		0.370		
		(0.254)		
× VC/VPS of large clans		-0.476		
		(0.367)		
VC and VPS from the same clan			0.215	
			(0.214)	
× VC/VPS of large clans			-0.410	
			(0.340)	
VC in the village party branch				0.219
				(0.220)
× VC of large clans				0.008
				(0.366)
× VPS of large clans				-0.514
				(0.314)
× VC of large clans × VPS of large clans				0.506
				(0.505)
Dependent variable mean	1.083	1.083	1.083	1.092
Year and village fixed effects	x	x	x	x
Observations	2,495	2,495	2,495	2,324
Villages	139	139	139	130

Note: *This table shows that the association between a VC of large clans and village public investment is robust when we control for the roles of VPSs and village party organizations. The dependent variable is the log amount of village investment (1,000 yuan) in that year. The key independent variable is dummy variables indicating whether a VC came from the village's largest or second-largest clan, whether the VPS came from a village's largest or second-largest clan, and their interaction. In Addition, in Column 2, we control for whether the VC and VPS were the same person ("one-shoulder", or yijiantiao) and its interaction with VC of large clans. In Column 3, we control for whether the VC and VPS came from the same clan and their interactions with VC of large clans. In Column 4, we control for whether a VC was in the village party branch and its interactions with variables we included in Column 1. Standard errors clustered at the village level are in parentheses. The sample is based on village-year observations from 130-139 villages that report information on VPSs and village party organizations during the period of 1986-2005 after village elections were introduced. All regressions control for village and year fixed effects.*

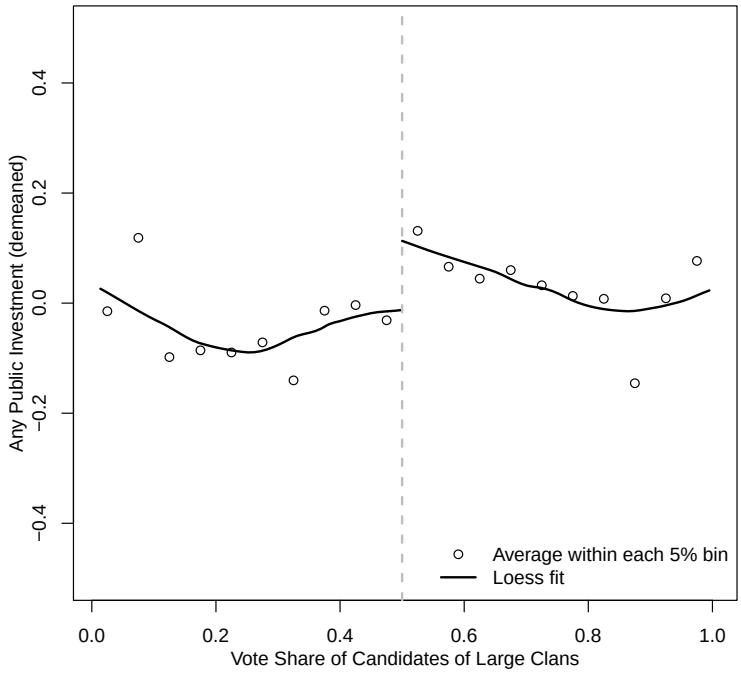
A.5 A Regression Discontinuity Design: Additional Results

Table A7. VC of Large Clans and Village Public Investment: A Regression Discontinuity Design

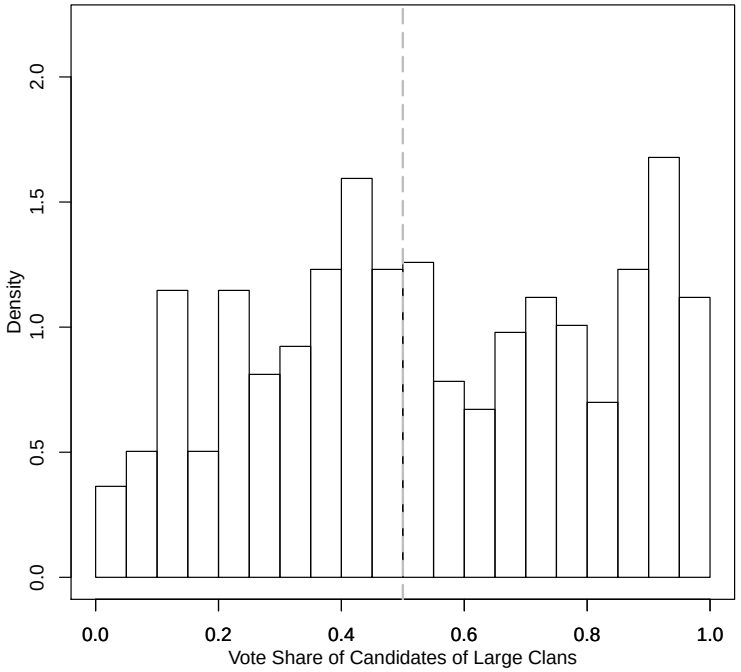
Panel A	Log Investment (1,000 yuan)					
	All with #votes	Vote%≠{0 ,100}	Vote% [40,60]	Vote% [45,55]	1st order poly.	2nd order poly.
	(1)	(2)	(3)	(4)	(5)	(6)
	FE	FE	FE	FE	Loess	Loess
VC of large clans	0.660	0.845	0.731	0.607	0.573	0.521
	(0.189)	(0.355)	(0.847)	(0.820)	(0.301)	(0.435)
Dependent variable mean	1.238	1.189	1.431	1.380	1.189	1.189
Observations	2,296	781	174	89	781	781
Villages	189	132	38	22	132	132
Panel B	Binary Outcome: Any Investment					
	All with #votes	Vote%≠{0 ,100}	Vote% [40,60]	Vote% [45,55]	1st order poly.	2nd order poly.
	(1)	(2)	(3)	(4)	(5)	(6)
	FE	FE	FE	FE	Loess	Loess
VC of large clans	0.125	0.172	0.170	0.166	0.123	0.124
	(0.038)	(0.072)	(0.186)	(0.197)	(0.063)	(0.088)
Dependent variable mean	0.257	0.251	0.310	0.315	0.251	0.251
Observations	2,296	781	174	89	781	781
Villages	189	132	38	22	132	132

Note: *This table reports the estimates from an regression discontinuity design. In Panel A, the dependent variable is the log amount of village investment (1,000 yuan) in that year; in Panel B, it is a dummy variable indicating whether there was any village investment during that year. Both samples are based on village-year observations after village elections were introduced. The independent variable is a dummy variable indicating whether a VC came from the village's largest or second-largest clan. Columns 1–4 report estimates from standard two-way fixed effects models. Standard errors clustered at the village level are in parentheses. In Column 1, observations without vote share data are dropped. In Column 2, observations in which a VC's vote share is either zero or one — neither the VC nor the runoff came from large clans (or both come from large clans) — are further dropped from the sample. Columns 5 and 6 limit the samples to relatively close elections, i.e. vote shares (%) of VCs of large clans are in the range of [40, 60] and [45, 55], respectively. Using the same sample as in Column 2, Columns 5 and 6 fit local linear regressions on both sides of the 50 percent cutoff and report the difference in the loess intercept estimates around the cutoff. Standard errors are produced by bootstraps of 1,000 times. The loess fits in Column 5 control for the level of the vote share (a first-order polynomial) while those in Column 6 control for the second-order polynomial. In Columns 5 and 6, observations are demeaned over time and within villages in advance to reduce dispersion and to account for aggregate shocks during the observed periods and time-invariant village heterogeneities.*

FIGURE A4. ROBUSTNESS CHECK: A REGRESSION DISCONTINUITY DESIGN (CONTINUED)



(a) Log public investment



(b) Histogram of the vote share

Note: *Figure A4a shows the probability of any public investment projects within each 5 percent vote-share bin and two loess fits from locally linear regressions on both sides of the cutoff. Figure A4b plots the density of the vote-share of large-family candidates (values 0 and 1 not included).*

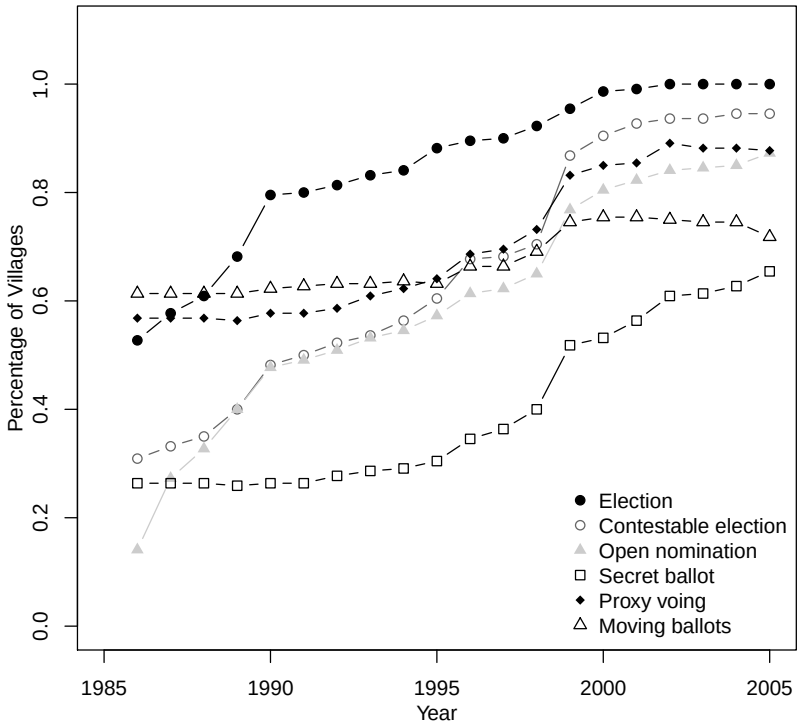
A.6 Alternative Explanations and Additional Results

Table A8. Large Clans and VCs' Characteristics

VC's characteristics	Years of education (1)	Age when running election (2)	CCP member (3)	Village cadre when running election (4)	Managerial jobs when running election (5)	Experience of running election (6)	Family background: poor peasant (7)	Denounced in the Culture Revolution (*pidou*) (8)
VC of large clans	-0.145	-0.163	-0.052	-0.033	-0.001	-0.040	-0.028	-0.020
	(0.225)	(0.946)	(0.047)	(0.031)	(0.008)	(0.040)	(0.049)	(0.024)
Dependent variable mean	6.39	41.6	0.75	0.56	0.02	0.71	0.79	0.05
Year and village fixed effects	x	x	x	x	x	x	x	x
Observations	1,210	1,203	1,195	1,209	1,209	1,205	1,213	1,203
Villages	218	219	216	218	218	216	219	216

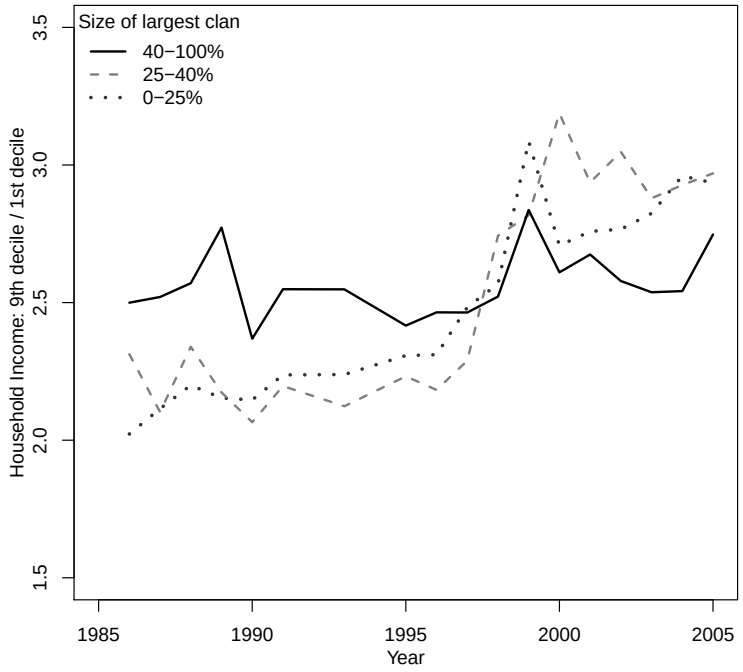
Note: *This table shows that VCs of large clans were not significantly different from those from small clans in terms of observed characteristics. Standard errors clustered at the village level are in parentheses. The dependent variables are observed characteristics of elected VCs. The independent variable is a dummy variable indicating whether a VC came from the village's largest or second-largest clan. The sample is based on village-term observations from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

FIGURE A5. EVOLUTION OF ELECTORAL INSTITUTIONS IN THE SAMPLE VILLAGES



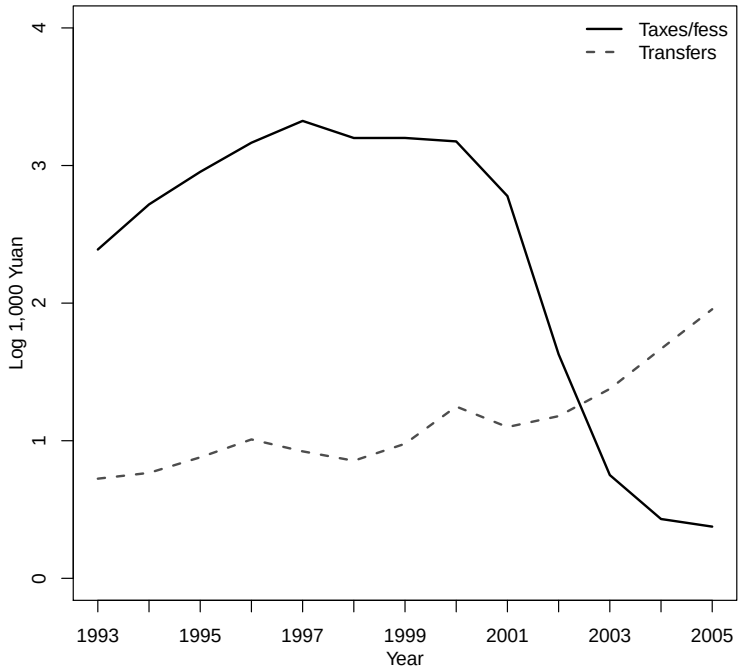
Note: *This figure shows the changes of electoral rules and procedure from 1986 to 2005 in the sample villages.*

FIGURE A6. LARGE CLANS AND INCOME INEQUALITY



Note: *This figure shows the level of income inequality from 1986 to 2005 for three groups of villages: (1) villages with very big largest clans, (2) villages with medium-sized largest clans, and (3) villages with relatively small largest clans. Income inequality is measured by the ratio of household income at the 9th decile over household income at the 1st decile. Household level data are from 69 villages, a subset of the full sample. The data for 1994 are interpolated. The change of income inequality was the smallest in the first group.*

FIGURE A7. AVERAGE LEVELS OF TAXES/FEES AND TRANSFERS: 1993-2005



Note: *This figure shows the average levels of taxes/fees the sample villages paid to the upper-level government and transfers they received from the upper-level government from 1993 to 2005. The data for 1994 are interpolated.*