Informal Institutions, Collective Action, and Public Investment in Rural China[*]

Yiqing Xu[†]   Yang Yao[‡]

Abstract: Do informal institutions, rules and norms created and enforced by social groups, promote good local governance in environments of weak democratic or bureaucratic institutions? This question is difficult to answer because of challenges in defining and measuring informal institutions and identifying their causal effects. In the paper, we investigate the effect of lineage groups, one of the most important vehicles of informal institutions in rural China, on local public goods expenditure. Using a panel dataset of 220 Chinese villages from 1986 to 2005, we find that village leaders from the two largest family clans in a village increased local public investment considerably. This association is stronger when the clans appeared to be more cohesive. We also find that clans helped local leaders overcome the collective action problem of financing public goods, but there is little evidence suggesting that they held local leaders accountable.

INTRODUCTION

Do informal institutions promote good governance in localities where formal democratic and bureaucratic institutions are weak? Or do they prevent local governments from functioning properly? Scholars find that in Latin America, Eastern Europe, and Central Asia informal institutions often breed clientelism, corruption, and mafia activities (e.g., O'Donnell 1996; Böröcz 2000; Collins 2003) and cause citizens to be excluded from the state's public services (Narayan 1999). However, Sklar (2004) suggests that traditional institutions in Uganda and Nigeria improve government performance and maintain regime stability. Tsai (2007) shows that in the context of rural China, solidary groups, such as temple associations and village-wide lineage groups, hold government officials accountable and motivate them to provide more public goods.

Although a universal answer to the question may not exist, a clearer understanding of the role of informal institutions in specific social contexts deepens our knowledge of what determines good local governance. However, researchers on informal institutions often face challenges in defining and measuring informal institutions and in identifying their causal effects on governance outcomes (Helmke and Levitsky 2004). We attempt to address those challenges. We follow Helmke and Levitsky (2004) and define informal institutions as rules and norms that are created and enforced by social groups rather than the state. In this paper, we specifically focus on the set of informal institutions that could affect local public goods provision.

Public goods provision in environments of weak formal institutions faces two fundamental problems: (1) to convince community members, who are often in poor living conditions and have tight budget constraints, to contribute to public goods expenditure and (2) to motivate local leaders to initiate necessary public projects, while preventing their moral hazard behavior, such as embezzlement and corruption, during the process of providing public goods. The first problem is essentially a collective action problem, while the second one is about local government accountability. If informal institutions are to promote local

public goods provision, it is likely they either help solve the collective action problem among community members, or hold local officials accountable, or both.

Using a panel dataset of 220 Chinese villages from 1986 to 2005, we study the effect of informal institutions embedded in large and organized family clans on public goods provision and investigate the collective action and accountability mechanisms. Family clans are organized based on shared patrilineal ancestors and are regarded as the most important social groups in Chinese villages (e.g., Fei 1946; Freedman 1958; Watson 1982, Duara 1988). Informal institutions of large clans are rules created and enforced by clans and often respected by villagers both inside and outside clans. Large clans may have disproportionate advantage over small clans because they have deeper historical roots in the village and are often better organized. As a result, they may promote local public goods provision by either helping the officials coordinate collective action or by holding them accountable.

The exact outcome variable in our analysis is the amount of public investment the village committee spent each year in the period of 1986-2005 after village elections were introduced. We focus on the post-election period because it is the period during which we have complete data on elected village chairpersons (VCs). The key independent variables are binary indicators of whether a village leader, such as an elected VC or an appointed village party secretary (VPS), came from the village's largest or second-largest clan. Our theoretical premise is that leaders from these two clans have access to richer and stronger informal institutions than leaders from small clans. Specifically, there are two possible channels that a village leader might be affected by his own clan when he attempts to initiate a public investment project: (1) he could get help from his clan and use the clan's social power to mobilize resources from villagers and (2) he might be morally bounded by the rules of his clan such that he would make good use of existing resources. In either case, large clans dominate small clans in terms of public goods provision because of their advantageous positions in the village. To further illustrate that it is the rules and norms of a large clan, rather than the number of its members, that matters, we use the information on whether a large clan kept records of family trees and whether it maintained a lineage hall since the beginning of the observed

2

data period as indicators of clan cohesiveness and investigate how the effect of informal institutions changes due to the changes of the level of cohesiveness.

Because variations of informal institutions in existing quantitative studies are usually cross-sectional, it is often difficult to identify the effect of informal institutions on governance outcomes. One can imagine a situation in which associational activities flourish in places with better infrastructure or rapidly increasing living standards. The positive correlation between associational activities and public goods provision does not necessarily imply that the former, which are sometimes used as proxies for informal institutions, causes the latter. Controlling for time-invariant heterogeneity would alleviate this concern of omitted variables to a great extent. In this paper, we exploit the advantage of the panel data structure and only look at the within-village changes of public goods expenditure due to within-village changes of informal institutions associated with village leaders. To the extent that informal institutions affect local governance, we would then expect to observe systematically different policy outcomes produced by villager leaders of different clans within the same village.

Our research design, therefore, is to compare the level of public goods expenditure during the terms of village leaders who came from the village's two largest clans and the terms of the others within each village (focussing on the largest clan gives qualitatively the same results). We primarily focus on VCs instead of VPSs because we have more complete data on the former than the latter. To address the concern that electoral outcomes might be endogenous to public goods expenditure—for example, villagers expect leaders from large clans to provide more public goods and, therefore, elect them into office—we conduct a regression discontinuity analysis based on elected VCs as a robustness check for our main results.

Setting the study in the context of rural China has several advantages. First, because of the large scale of the country, there is enough variation in the lineage composition of a village. Forms, origins, and functions of the institutions associated with lineage groups are relatively well understood by scholars (e.g., Fei 1946; Freedman 1958; Watson 1975; Wang 1996; Duara 1988; Tsai 2007), which makes it considerably easier to understand how such

informal institutions work than in other less researched contexts. Second, Chinese villages are relatively homogeneous in other aspects and subject to similar social and political shocks at the provincial or national level. This aspect makes them better comparison groups of each other than nations in cross-country studies. Third, Chinese villages were largely autonomous in terms of determining and financing public goods in the period of our study. Fourth, the introduction of village elections in the mid-1980s offers a rare chance to examine the influence of both formal and informal institutions.

Our empirical analysis shows that during the terms of VCs of the two largest clans (hereafter, VCs of large clans), the amount of village public investment increased by more than 35 percent on average as compared with the amount during the terms of other VCs. A VPS of the two largest clans (VPS of large clans) also increased the average level of public investment considerably. We interpret these results as evidence that the informal institutions of lineage groups, rather than village leaders of a certain kind, led to more public goods expenditure and, presumably, better local governance. We show that the association between VCs of large clans and public goods expenditure is stronger in places where large clans appeared to be more cohesive (i.e., clans that had maintained lineage halls since or before the onset of elections). Combined together, these results indicate that it is informal institutions of the clans from which village leaders originated that drive our main finding.

In addition, we explore two mechanisms through which informal institutions of large clans may facilitate public goods provision: (1) the collective action mechanism and (2) the accountability mechanism. The collective action problem has been a central topic in political science since Olson (1965). Recently, researchers have been focusing on how informal institutions, rather than formal ones, help people overcome the collective action problem. For instance, after observing several long-standing, self-governing common property regimes, Ostrom (1990) argues that informal institutions work through (or are) a set of "self-enforcing rules that each community member commits himself or herself to follow" (p. 99). Banerjee and Iyer (2005) suggest that persistent informal institutions may result in different levels of public expenditure because of the nature of collective action embedded in those institutions.

4

In Africa, Habyarimana et al. (2007, 2009) show that ethnic heterogeneity impedes collective action and the provision of public goods; however, the collective action problem can be alleviated by institutional improvements in monitoring, sanctioning, and enforcement.

In rural China, because village committees often lack measures to enforce levies on villagers, successful collection of levies requires villagers' semi-voluntary compliance. Sheer poverty in the countryside makes that difficult for a VC and his associates.[1] If a large proportion of villagers refuse to pay for a public investment project, village leaders' efforts to provide public goods would be in vain without the help of the upper-level governments. We show that when VCs of large clans were in office, villagers paid more levies to the village committee, and the presence of village public investment projects is highly correlated with extra levies paid by villagers at almost all income percentiles. Our results indicate that with the help of the informal institutions of large clans, elected VCs were more able to enforce levies on villagers and to mobilize resources needed for providing public goods.[2]

Sklar (2004) and Tsai (2007), among others, emphasize the mechanism of informal accountability. To test this hypothesis, we study the amount of administrative costs during each VC's term. Administrative costs are mostly spent by the VC and his associates for their own consumption. Embezzlement and other forms of corruption may also be covered in this category of village spending. A decline of those costs, therefore, can be seen as a result of improved accountability imposed on the VC. However, we do not find evidence that the amount of administrative cost spent by VCs of large clans was smaller than that by VCs of smaller clans. Although we cannot entirely rule out the accountability mechanism, this piece of evidence suggests that the positive association between VCs of large clans and a higher level of public goods expenditure is unlikely to be a result of large clans' superior ability to monitor the VCs.

---

[1]In 2005, the median household in our sample lived with an annual budget of 18,507 yuan, or US$1.61 per household member per day (purchasing power not adjusted). In 1986, that number was US$0.44 per household member per day.

[2]Our finding is consistent with Habyarimana et al. (2007)'s finding from experiments in Africa that ethnically more homogeneous communities achieve greater success in collective action because of better communication technology, more transparency, and more cooperative equilibrium strategies.

We investigate two alternative explanations. First, VCs of large clans may be more competent than others. For instance, Munshi and Rosenzweig (2010) show that elected representatives from large castes in rural India exhibit better observed characteristics, such as higher education, and provide more local public goods for their constituents. Second, gradual improvement of formal institutions, such as electoral rules and procedures, may also contribute to the association between VCs of large clans and public goods provision. As elections become more competitive, electoral outcomes are more likely to reflect the preferences of the constituents. Thus, it is possible that the probability of VCs of large clans being elected and increased public goods provision are moving in the same direction. We show that neither of these two possibilities is likely to be driving our results.

Apart from the informal institution literature and the literature on collective action and public goods provision, this paper also adds to a large literature on village elections and grassroots politics in China (e.g., O'Brien 1994; Manion 1996, 2006; Shi 1999; Oi 1999; Oi and Rozelle 2000; O'Brien and Li 1999, 2000; Pastor and Tan 2000). More recently, Luo et al. (2007, 2010) find that the introduction of elections increases total public goods expenditure and provision. Shen and Yao (2008) find that elections reduce village income inequality through the public goods channel. This paper, instead of investigating the effect of elections *per se*, uses variations generated by elections to examine the causal effect of informal institutions on governance outcomes. To the best of our knowledge, this paper is also the first to apply a regression discontinuity design to village elections in China.

INSTITUTIONAL BACKGROUND

Lineage Groups in Chinese Villages

Lineage groups are one of the most important social organizations in rural China. They are usually organized along the paternal line.[3] Fei (1946) suggests that in imperial times

---

[3]Watson (1982) defines a lineage group as "a corporate group which celebrates ritual unity and is based on demonstrated descent from a common ancestor." He distinguishes clans from lineage groups based on membership recruitment. He argues that clans recruit members based on fictionalized descent rather than descent from known ancestors. However, most scholars do not distinguish the two terms. In this paper, we focus on clans that are formed based on known ancestors and use the two terms interchangeably.

lineage groups served as a link between the imperial ruler and the grassroots and were used by the gentry to preserve the social and political power of their families. Fei finds that through lineage networks the gentry administrated charities and provided local public goods to command the moral height in the villages. Freedman (1958) hypothesizes that lineage groups are substitutive social organizations in places where formal bureaucratic institutions are weak. He finds that lineage organizations were more developed in southeastern China than in the north because villages in the southeast were farther away from central political control.

After nearly one hundred years of radical social changes, there has been a startling withdrawing of the gentry from the rural political field.[4] However, researchers believe that there is still space for lineage groups to survive and flourish. Reformers and even revolutionaries had to take advantage of existing resources, including traditional institutions, to achieve their objectives (Perry 2002). Lineage groups have proven resilient and, in many places, have survived extreme social and political changes (Wang 1996).[5] Tsai (2007) reports in her 2001 survey that 14 percent of the villages had one or more lineage halls.

Previous research on the Chinese village focuses on ideal types of social organizations, such as village-wide lineage groups. Sub-village lineage groups are thought to be not as effective in exercising social powers (e.g., Freedman 1958; Wang 1996; Tsai 2007). However, the introduction of village elections may activate some of the functions of sub-village lineage groups. These groups are often organized around surnames or, when the village has only one surname, *fang* (house), that is, households who share the same grand- or great-grandparents. In the absence of political parties or other modern political organizations, lineage groups can become vehicles for political mobilization. A clan can be as large as 100 households containing more than 400 villagers, although it may only constitute 30 percent of the total village

---

[4]The reason behind the change is complicated. The rise of towns and cities since the beginning of the twentieth century attracted the young and wealthy out of the rural areas. The neighborhood administrative system (*baojia zhi*) in the Republican era and endless social movements after 1949 also contributed to the retreat of clan forces.

[5]This finding is consistent with researchers' finding in Central Asia that clans adapt to resist repressive states (Collins 2004).

population. Such a group, if well organized, can have a real impact on village governance. In this paper, variation in informal institutions comes from these groups.

As a social organization, family clans in rural China have several features. First, as mentioned above, households within a clan consciously identify themselves as members of a closely-bonded group. They often reside within geographical proximity and frequently interact with one another. A well-organized clan holds annual rituals and ceremonies, such as paying respect to ancestral tablets and offering sacrifices to ancestral spirits (usually at its lineage hall), to reinforce group identity (e.g., Freedman 1967, Tsai 2007). Second, clan members often cooperate with each other to obtain material benefits. Before the communist revolution in 1949, clans in southern China often owned land, which gave a basis for clan members to cooperate economically. In the collectivization period, collective production teams were often organized by clans in the south. Although this happened mostly because of the geographic proximity of clan members' residencies, economic ties within the clan were preserved (Watson 1982). In the reform period, economic cooperation among clan members has shifted to other areas of shared interests. For example, rural entrepreneurs tend to hire relatives in their own firms (Oi 1999: 69). Third, clan members share a sense of obligation to the group. Traditional ethics place a sacred value on loyalties generated by kinship and dense social ties. Moral standing is conferred to members who make contributions or bring material benefits to the group (Madsen 1984). Finally, leaders of the group, usually respected senior male members, come forward to enforce social norms and mitigate conflicts both within and outside the group.[6]

Large clans may have disproportionate advantage in this regard. This is first and foremost related to the fact that they often have strong historical roots in villages. Many villages were founded by the largest clans in the first place. Watson (1982) reports that "it is common to find villages that contain one or two corporate lineages together with four or five loosely-

---

[6]As in other hierarchical social groups, not all members of a clan enjoy the same social status in the group. One can imagine that if a well-respected member of the group gets elected, he or she can mobilize more group resources than others. Our dataset does not have such information, but we find that the profiles and characteristics of elected VCs remain stable over time.

defined surname groups" (p. 608). Small clans often consist of families that migrated into the village at a later stage. Because of that, large clans are usually better organized than small clans. In the past, they were often managed by a group of senior members led by *zuzhang*, or lineage chief, a position usually inherited by the most powerful family in the clan (often the family of the eldest son of the clans founder). Today, this more formal power structure has vanished. However, senior members still play a significant role. They are responsible for clan rituals and other collective activities (Cohen 1990).

Large clans thus are more likely than small clans to maintain lineage halls, hold clan ceremonies and keep lineage genealogies. This increases their social cohesiveness and members' sense of belonging. In addition, seniors of large clans are more likely to participate in village affairs. Many villages have a seniors association that is recognized by the government as a vehicle to serve the needs of the senior. However, village leaders often consult the members of the association on important village affairs. Seniors from large clans naturally become the leaders of the association. As a result, their influence can reach beyond their own clans. Their social power originates from both their clans' clout and their reputation of looking after village public interest.

Against this background, village leaders from large clans can have significant advantages over ones from small clans because it is likely that they can only mobilize informal organization resources from their own clans. Informal institutions embedded in large clans can facilitate collective action among villagers through both *persuasion* and *social sanctioning*. When contribution from villagers is needed for a public project, a village leader from a large clan can approach to his clans seniors to ask for help. Resorting to their prestigious social status, senior members of the clan are able to persuade villagers both inside and outside the clan to support the village leader's project and to enforce clan rules and norms when it comes to financial contribution.[7] When a non-trivial proportion of the villagers support the project, social pressure forces the rest of the villagers to contribute their fair shares;

---

[7]The norms are, for instance, each household should contribute to the public good according to its own economic condition (*liang li er xing*); households who fail to fulfill their duties will be socially sanctioned by the clan. Liu (1959) asserts that such obligations are often specified in the appendices of genealogies.

otherwise, they may face severe social sanctions.[8]

A piece of anecdotal evidence from Zhejiang Province in eastern China illustrates how the collective action mechanism might work:

> "My father used to be the *zuzhang* (lineage chief) of our clan. The village chief (chairman of the village committee) was also a member of the clan. Whenever the village committee had some great undertakings to accomplish, like collecting money for building a road, he came to my father and other seniors of the clan. If the seniors thought the chief's plan could work, they would convene a meeting of household heads, together with the village chief, to convince the villagers to support the project, either by giving money or donating working hours. Since in our village, the majority of households are from the *Fu* family, the meeting is almost like a villagers' assembly. People took it quite seriously. They trusted my father because they thought he's impartial and experienced. The seniors didn't enjoy formal titles, and they didn't take charge of daily matters, but they were (moral) authorities of the village."[9]

The above discussion suggests that the population rank order of a clan is a good proxy for the strength of informal institutions that a village leader can rely upon. Even if the sizes of clans are not drastically different, large clans (the largest clan in particular) are more likely to be well organized and enjoy greater social power. In the empirical analysis, our key explanatory variables are dummy variables indicating whether a VC or VPS came from the largest or second largest clan, which we believe summarize most of the information relevant to our study.[10]

Village Self-government, Elections, and Public Investment

Village self-government was reorganized by the CCP in the late 1970s after the abolition of the rural commune system. Village committees are designated as a "self-government

---

[8]For clan members, social sanctions can take the form of a break of relationships, contempt, gossip, or even removal from the clan's family tree. For outsiders, non-compliance with the decision of powerful clans may also lead to unequal treatments in situations involving collective distribution.

[9]From authors' interview on August 7, 2012.

[10]The size of lineage groups may also matter. In Appendix Section A.3, we show that (1) our results are robust when we control for the VC's clan size and (2) the effect of informal institutions, as we measure them, varies little across clans with different sizes. We also discuss why we think the rank order is a better measure for the clan's social power than the clan size.

organization" according to the Chinese Constitution. A village has two self-governing bodies: a village committee, which usually consists of three to seven members, and a village party branch, which includes several CCP members in the village. Village leaders are predominantly male. The VC, who has been democratically elected after village elections were introduced in the mid-1980s, leads the village committee. The position is also sometimes called the village chief or village head. The VPS leads the village party branch. Very often the village committee and village party branch overlap. Existing English and Chinese literature suggests that village officials are "sandwiched" between villagers and the township government, the lowest level of government (O'Brien and Li 1999, Oi and Rozelle 2000, Zhang 2007). They are supposed to be accountable to villagers, but they are also expected to fulfill tasks assigned by the township government.

Village elections first took place in Yishan County in Guangxi province as the People's Commune was dismantled in the early 1980s (Tan 2006: 59–63). Inspired by villagers' self-initiated acts, the CCP promoted village elections as an effort to address the information problem of holding local officials accountable and to improve local governance. To minimize risks, such as the state losing control of villages and compromising unpopular government policies, the government's democratization reform was gradual and highly controlled (O'Brien and Li 1999; Unger 2002). In 1987, a temporary version of the *Organizational Law of the Village Committee* (OLVC) was put into effect and village elections began to be formally introduced in most provinces. The formal version of the law was announced in 1998. Since then virtually all the villages have begun elections.

VCs are elected for three-year terms without term limits. Usually a handful of candidates are nominated in each election and a primary is held to reduce the number of candidates to two. The final round is run between these two front runners. Overt campaigning is not common in village elections (Pastor and Tan 2000; O'Brien and Han 2009). When elections were first introduced to villages, the township government maintained control of the nomination process. Only after 1998 when the OLVC was formally adopted were nominations open

11

to all villagers.[11] The timing of the introduction of elections was largely determined by the provincial government's preferences. Martinez-Bravo et al. (2011) show that the introduction of village elections has shifted accountability from the upper-level government towards villagers and worsened the implementation of unpopular policies, such as tax collection and the One Child Policy.

One of the main jobs of the village committee is to provide village public goods (Whiting 1996; Oi and Rozelle 2000). It is responsible for determining public goods investment, as well as raising most of the funds required for the investment. Because the village committee does not have the legal authority to tax people, the only way it can finance public investment is through collecting fees and levies (hereafter, levies, for simplicity). Although levies were allowed by the central government before 2006, their amounts were usually small. Village leaders had to turn to villagers to ask for more levies if the village was to undertake a large public project. Unlike in more institutionalized contexts in which paying local taxes is enforced by law, village leaders had to exert a large amount of effort to convince villagers to pay levies. The collective action problem arises when villagers' semi-voluntary compliance is required for local public goods provision. This problem partly explains why scholars find that public goods were severely under-provided in rural China (e.g., Zhang et al. 2004; Luo et al. 2007, 2010).

DATA AND RESEARCH DESIGN

Data

This paper mainly uses a panel dataset of 220 villages from 1986 to 2005 from the Village Democracy Survey (VDS), a unique retrospective survey conducted by the authors and their collaborators. The villages were selected from the sample of the National Fixed-Point Survey (NFS), a longitudinal survey maintained by China's Ministry of Agriculture.[12] We depict the

---

[11]Nominations open to all villagers are popularly known as *haixuan* in China. It was first adopted in Lishu, Jilin in 1986 (Tan 2009).

[12]The NFS was started in 1986 to survey the same sample of households and villages over time. Except for 1992 and 1994, it provides annual data aggregated from daily household diaries. The NFS sample was first

locations of the sample villages (the counties that they belong to) in Appendix Figure A1. In 2006, the VDS recorded the history of electoral reforms, traditional organizations, and public goods expenditure. In 2011, the VDS team returned to the same villages to collect data on village clan structure and more information on traditional organizations and elected village leaders.

Data of electoral outcomes and public goods expenditure are obtained from village records, hence, concerns from report errors were minimal. Because the VDS only collected information of elected VCs, we only use observations in the post-election period to study the effect of informal institutions. However, focusing on the post-election period gives us the advantage of isolating the effect of informal institutions from that of electoral reforms.[13]

Information of lineage groups, including the identities of clans (surnames), their relative sizes as measured by shares in the village population, facilities they maintained, and activities and ceremonies they held, draw up on the collective responses of current and former living village leaders and elders, who were invited together to respond to the surveyors. The VDS recorded information on the four largest clans. Although there could be measurement errors in the exact number of villagers in each clan, villagers typically had consensus on the rank order of clan size in their villages. Therefore, we believe that the rank order was precisely recorded.[14] Moreover, because the Chinese government strictly prohibits permanent migration from rural areas, radical changes of the village social structure are less of a

---

selected in 1986 according to a stratified random sampling strategy. Sample counties were first randomly selected from a province with the number of counties being proportional to the province's rural population. Then within a sample county, one village was randomly selected. Over the years, some villages dropped out of the survey mainly because they were incorporated into a nearby city, in which case a village in the same province was randomly selected to replace the dropped village. There are about 300 villages in the NFS. Among them, more than 220 villages have been in the sample for the 20-year period covered by this study. The VDS surveyed these villages. Martinez-Bravo et al. (2011) show that the VDS sample and the entire NFS sample are similar for a broad range of attributes.

[13]In Appendix Table A1, we perform a robustness check using data after 1995 to show that the timing of the electoral reform does not induce significant biases for the informal institution estimates. In 1995, most of the villages in our sample had adopted elections.

[14]In the survey, a meeting of the village leaders and elderly was convened in each village. Usually a handful of them came to meeting. The meeting lasted for about an hour, but consensus was very often quickly reached on the population rank order of the four largest clans. More time was spent on collecting information on the exact size of each of the four largest clans.

concern.[15] Since the VDS also recorded information of the elected VCs, we can identify the clan each VC belongs to by matching his surname with that of the clan's.[16] The VDS asked if a large clan kept records of its family trees (genealogies) or maintained a lineage hall. We use this information to form our measure for clan cohesiveness.

The VDS data are supplemented by annual data collected by the NFS. The control variables we use in this paper, such as village population, village household income, and village assets, come from the NFS. Data of levies that households paid to the village committee also come from the NFS.

The data we use have several merits. First, the information contains the most comprehensive data on village-level reform and governance outcomes in China. They cover a large and nationally representative sample and span a long period of time. Second, the panel structure, as well as the relatively large sample size, allow us to control for not only village and year fixed effects, but also time trends at the provincial level and even the village level. Village and year fixed effects account for unobserved time-invariant factors within each village and shocks that affect all villages in a given year, respectively. Time trends at the provincial or village level capture growing social and economic divergences across regions. Controlling for these factors eliminates a large number of potential confounders for the identification of the effect of informal institutions. For example, because village fixed effects allow us to make the comparison within villages, confounding factors associated with geography are effectively controlled for. Third, the quantitative data we have are mostly based on administrative records and, therefore, are comparable across villages and not likely to suffer from recall biases. Moreover, because the electoral outcomes and public investment data were collected directly from village records, they are not likely to be manipulated by

---

[15]Rural to urban migration soared at the beginning of the twenty-first century and has become an increasingly important issue for scholars who want to understand rural politics in China. We control for this factor in the regression analysis.

[16]In villages with only one surname (which are mostly in the south), we treat houses (*fang*) as separate lineage groups. Family names of women VCs did not reflect the clans they belonged to, because, in most cases, Chinese women do not change their family names after getting married. In the VDS, there were only 10 women elected as VCs in more than 1,000 recorded elections. We code them as coming from small clans. Dropping these observations does not affect our main results.

14

village officials.

(a)

(b)

(c)

(d)
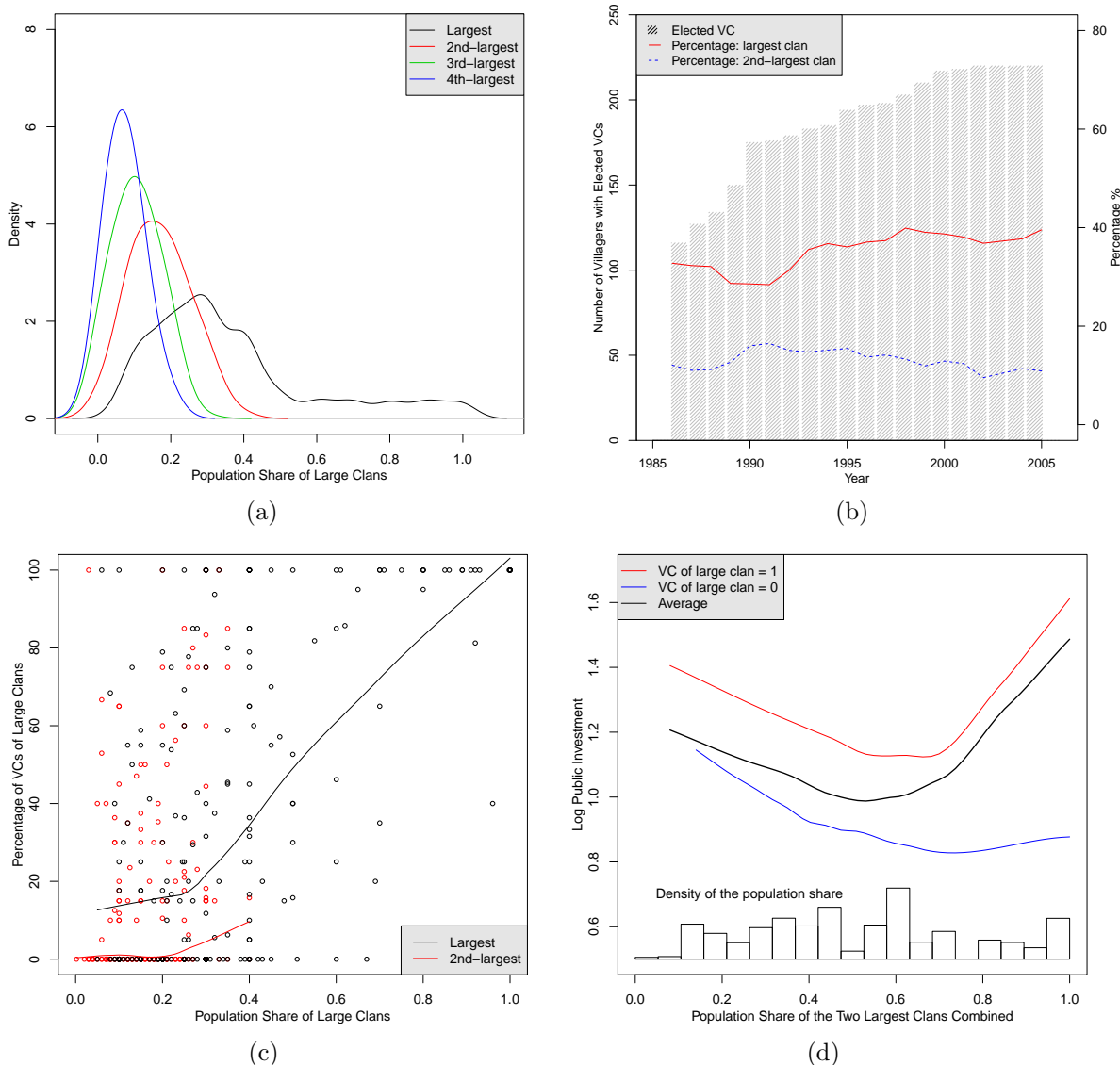
Note: *Figure 1a shows the distributions of the population share of the four largest clans in each village. Figure 1b shows the roll-out of village elections (left axis) and the percentages of elected VCs coming from the largest and second-largest clans (right axis). Figure 1c shows the percentages of VCs from the largest and second-largest clans over their respective (relative) clan size with two loess fits. Figure 1d shows the relationship in raw data between VCs of large clans and the amount of public investment. The x-axis is the combined population share of the largest and second-largest clans in a village.*

Figure 1a plots the distributions of the population shares of the four largest clans in sample villages. The average population shares of the largest and second-largest clans were

15

36 percent and 15 percent, respectively. In 2005, the average village in our sample had around 1,500 permanent residents. The average size of the largest clan in a village was thus around 400 villagers, or 100 households. Also, 81.8 percent of the villages did not have a lineage group that constituted the majority of the village population, and 74.5 percent of the villages had more than 10 surnames on the paternal side.

Figure 1b shows the onset of village elections since 1986. More than half of the villages adopted elections in 1986 and most villages had at least one election by the mid-1990s. The red solid line and the blue dashed line in this figure show the proportions of elected VCs coming from the village's largest and second-largest clans, respectively. On average, 35 percent and 13 percent of the VCs came from these two largest clans. Both numbers remained relatively stable over time. Even though lineage groups might have a big impact on local governance, they did not necessarily dominate village elections. Figure 1c plots the share of VCs of the two largest clans against their respective clan size. It is clear that large clans were not over-represented.[17] There are also large cross-sectional heterogeneities; some villages elected VCs of large clans all the time while others never did. This occurs probably because in some places large clans are well organized, while in other places, the upper-level government has a big influence on putting its favored candidates on ballots or getting them elected.

Public investment falls into six categories: schooling, roads and sanitation (water supply and sewer systems), electric power, irrigation, forestation, and others. Figure 1d plots the average log public investment against the combined population share of the two largest clans. The plot exhibits a non-monotonic relationship between public goods expenditure and the size of the two largest lineage groups. There was more public goods expenditure in the most and least homogenous villages.[18] Such a relationship may be misleading, though, because

---

[17]Su et al. (2011) show that clan networks in rural China mobilize voters to go to voting stations, but there is not enough empirical evidence suggesting that large clans dominate village elections. Anecdotal evidence suggests that the CCP is constantly worried about the possibility that clans capture rural politics and has tried different measures to prevent it from happening (Mattingly 2014).

[18]The fact that more homogenous villages had a higher level of public goods expenditure is consistent with a wealth of literature on ethnic homogeneity and public goods provision, for example, Alesina, Baqir and Easterly (1999), Alesina, Baqir and Hoxby (2004), and Habyarimana et al. (2009), among many others. One

Table 1. Descriptive Statistics

Village-Year Observations	Obs.	Mean	S.D.	Min.	Max.
Any public investment	3,742	0.23	0.42	0	1.00
Log total investment (1,000 yuan)	3,742	1.09	2.15	0	10.60
Log village-average household taxes (yuan)	1,080	4.84	1.50	0	8.94
Log village-average household levies (yuan)	1,080	4.22	1.90	0	7.06
Log village population (persons)	3,513	7.20	0.61	4.67	9.16
Log net income per capita (yuan)	3,513	7.22	0.83	1.86	10.42
Log village asset (yuan)	3,513	9.01	1.62	2.67	15.35
Average household size (persons)	3,513	3.93	0.59	2.00	6.39
Arable land per capita (mu)	3,513	1.75	1.88	0.004	16.20
Number of persons migrating out	2,685	12.70	18.50	0.00	244
Log taxes to the upper-level government (1,000 yuan)	2,530	2.27	1.86	0.00	8.80
Log transfers from the upper-level government (1,000 yuan)	2,530	1.21	1.61	0.00	7.50
Share of administrative expenditure in total expenditure	3,037	0.23	0.22	0.00	1.00
Contested election	3,742	0.77	0.42	0	1
Open nomination	3,742	0.70	0.46	0	1
Secret ballot	3,742	0.38	0.49	0	1
Proxy voting	3,742	0.71	0.45	0	1
Moving ballot	3,742	0.68	0.47	0	1

Village Chairpersons (by Term)	Obs.	Mean	S.D.	Min.	Max.
VCs from the largest clan	1,315	0.36	0.48	0	1
~ from the second-largest clan	1,315	0.13	0.33	0	1
~ of large clans (from either the first or the second)	1,315	0.48	0.50	0	1
Years of education	1,210	6.39	2.30	0	13
Age when running election	1,203	41.56	8.72	19	90
CCP member	1,195	0.75	0.43	0	1
Village cadre before election	1,209	0.56	0.50	0	1
Managerial jobs before election	1,209	0.02	0.14	0	1
Experience of running election	1,205	0.71	0.46	0	1
Family background: poor peasant	1,213	0.79	0.41	0	1
Denounced in Culture Revolution (*pidou*)	1,203	0.05	0.22	0	1
Vote share	1,118	0.76	0.17	0.50	1
Serving as VPSs ("one-shoulder")	1,315	0.08	0.27	0	1
In the village party branch	830	0.62	0.49	0	1

Sample Villages	Obs.	Mean	S.D.	Min.	Max.
No. of surnames	220	27.77	30.73	1	150
Population share of the largest clan	220	0.36	0.23	0.05	1
Population share of the second-largest clan	220	0.17	0.08	0.00	0.40
Population share of the third-largest clan	220	0.10	0.07	0.00	0.30
Population share of the fourth-largest clan	220	0.07	0.05	0.00	0.20
Records of family trees (of the two largest clans)	200	0.48	0.50	0	1
Lineage halls (of the two largest clans)	200	0.17	0.37	0	1

the figure does not control for other variables. For example, many of the least homogeneous villages are located in coastal regions, so more investment in these villages could be because of higher levels of economic development. However, if we compare the amount of public investment during the terms of VCs from large clans and the amount during the terms of VCs from smaller clans, we find a clear gap between the two: when VCs of large clans were in office, there was more investment. Figure 1d illustrates that the gap was relatively stable across villages of different clan structures. Since we are looking at the raw data, though, this gap consists of variations both across and within villages.

As mentioned earlier, among the elected VCs, 35 percent and 13 percent were from the largest and second-largest clans, respectively. The average VC was around 42 years old when he was elected and has received 6.4 years of formal education. Three quarters of them were CCP members and 56 percent were already village cadres when they ran for office. Among the 200 sample villages that have detailed information of large clans, 48 percent had a large clan keeping records of family trees; 17 percent had a large clan maintaining a lineage hall. Table 1 presents descriptive statistics of the variables used in the regression analysis, including the number of observations, mean, standard deviation, minimum, and maximum of each variable.

Our main explanatory variable is constructed on the population rank order of clans because we believe that large clans have richer and stronger informal institutions. Here we present some supporting descriptive evidence. Figure 2 illustrates the relative size and level of organization for the four largest clans in the sample villages. On average, the largest and second-largest clans were more likely to have maintained lineage halls and to hold clan ceremonies on a yearly basis. For example, among all largest and second-largest clans in the sample villages, 15.5 percent and 11.1 percent had lineage halls, respectively. In comparison, the numbers for the third- and fourth-largest clans are around 7.5 percent. Moreover, the probability of having a lineage hall that was established before the reform era is much higher

---

explanation for the non-monotonic relationship between village social fragmentation and overall public goods expenditure is the trade-off between within-group and between-group collective action problems (Esteban and Debray 2001).

for the largest clan than for other clans. In our empirical analysis, we focus on the largest and second largest clans to allow sufficient modelling flexibility. Appendix Table A3 shows that our main findings remain unchanged if we define the key independent variable solely based on the largest clan.

FIGURE 2. MEASURING LINEAGE GROUPS
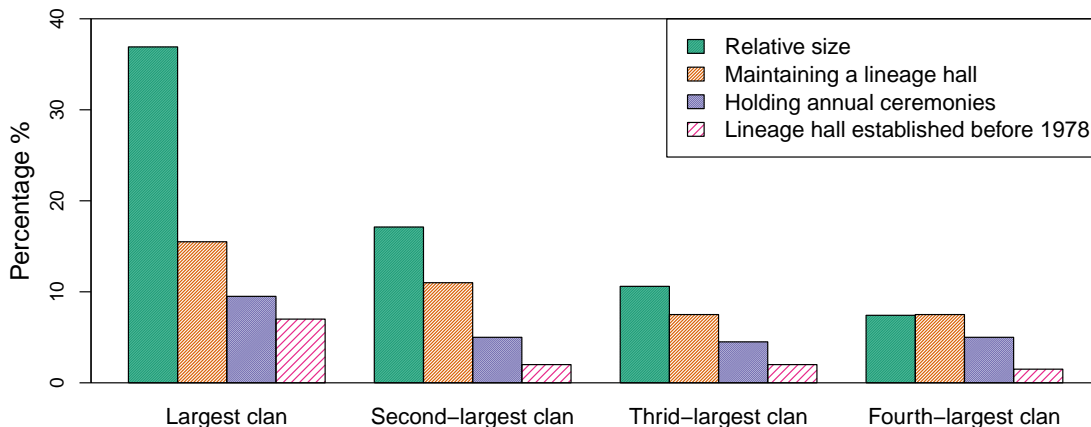
Note: *This figures shows the (1) average size (relative to the village population), (2) percentage of having a lineage hall, (3) percentage of holding annual ceremonies in the past five years, and (4) percentage of having a lineage hall established before 1978, of the four largest clans in the sample villages.*

Main Identification Strategy

Our key independent variables are binary indicators of whether an elected VC came from the largest or second-largest clan in the village. Our baseline specification is the following fixed effects model:

$$y_{it} = \beta_1 D_{it,1} + \beta_2 D_{it,2} + \eta_i + \delta_t + \epsilon_{it}, \tag{1}$$

where y_{it} is the outcome variable (e.g., the log amount of public investment) for village i in year t; $D_{it,1}$ and $D_{it,2}$ are dummy variables indicating whether a VC was from village i's largest or second-largest clan in year t, respectively; η_i and δ_t are village and year fixed effects; and ϵ_{it} represents idiosyncratic shocks. The village and year fixed effects absorb time-invariant heterogeneities across villages and aggregate shocks that affect all villages in a given year, respectively. The identifying assumption is that, conditioning on village and year fixed effects η_i and δ_t, $D_{it,1}$ and $D_{it,2}$ are not correlated with the error terms $\{\epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{iT}\}$. In

other words, we assume that the choice of VCs is quasi-random with respect to the amount of public goods expenditure after both the independent and dependent variables are demeaned within each village and across villages in a given year. The parameters we are concerned about are β_1 and β_2; we expect that they are significantly positive.

We also add four sets of additional controls to the baseline specification. First, we control for provincial linear time trends to capture regional economic divergence. Second, we replace these trends with village-specific linear time trends to take into account trending factors at the village level. Third, we control for time-varying covariates from NFS to show that our finding is not driven by these variables. The covariates will be introduced later when we present the relevant robustness results. Fourth, we control for taxes/fees the village committee paid to the upper-level government and total transfers it received from it to capture the influence of the upper-level government. We also control for the number of villagers migrating out each year. However, these specifications may not rule out the impacts of other unobserved time-varying variables that are correlated with the choice of VCs and public goods expenditure at the same time. We will address this concern later using a regression discontinuity design.

When conducting robustness checks and exploring mechanisms, we also use the following simplified specification:

$$y_{it} = \beta D_{it} + \eta_i + \delta_t + \epsilon_{it}, \tag{2}$$

where D_{it} is a dummy variable indicating whether a VC was from village i's largest clan or second-largest clan in year t. As we will see from the baseline results, both β_1 and β_2 are indeed large and positive, and they are statistically indistinguishable with each other in most cases. Using the above simplification, therefore, does not lose much information.

MAIN RESULTS

This section presents the baseline empirical results and some robustness checks. The main outcome of interest is public goods expenditure. We focus on the association between

VCs of large clans and the amount of public investment during their terms in office. The dependent variable is the log amount of village investment (1,000 yuan).[19] Table 2 shows the baseline results, which are produced by the estimation of Equation 1 (except for Column 1). In Column 1, we show the raw result from an ordinary least squares (OLS) regression without controlling for village fixed effects; the estimated coefficients of both VC dummies are positive. In Column 2, when both year and village fixed effects are controlled for, the coefficients of the two VC dummies are 0.412 and 0.303, respectively. Both are statistically significant at the 5 percent level. This means that a VC from the two largest clans is associated with 35 to 51 percent more expenditure in public investment. In Column 3, we control for provincial linear time trends; the estimates remain stable. In Column 4, provincial linear time trends are replaced by village-specific linear time trends. The estimates of interests are 0.359 and 0.256, similar to the baseline results. The standard errors go up quite a bit, and the dummy for the second-largest clan turns only marginally significant. In Column 5, we go back to provincial linear time trends, but add five time-varying control variables from the NFS, namely, log village population, average village household size, arable land per capita, log income per capita, and log assets owned by the village committee. These controls capture the size, demographics, agricultural endowment, and economic resources of the village. The results are very similar to those in Column 2.

Next, we consider the relationship between the village committee and the upper-level government. Two possibilities might affect the village committee's ability to provide public goods. First, VCs of large clans might have better access to government funds, which were often crucial for investment projects. Second, because the village committee was obligated to follow directives coming from the township government, the amount of money the village committee paid to the township might have a great impact on the village committee's budget constraints. Because of these concerns, in Column 6, we additionally control for log total transfers the village committee received from the upper-level government and log total taxes

---

[19]The dependent variable is generated by `log(x+1)`, in which `x` is the amount of public investment, because investment can be zero in a year.

TABLE 2. VCs of Large Clans 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	0.332	0.412	0.379	0.359	0.378	0.492
	(0.126)	(0.148)	(0.148)	(0.189)	(0.157)	(0.200)
VC of the second-largest clan	0.183	0.303	0.328	0.256	0.367	0.419
	(0.151)	(0.148)	(0.145)	(0.193)	(0.155)	(0.228)
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 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).*

and fees it handed over to the upper-level government each year. Moreover, to account for the impact of high waves of rural to urban migration since the beginning of the twenty-first century, we also add the total number of people migrating out of the village each year in the regression. All three variables are available for 208 villages after 1993. We find that the coefficients of VC of large clans become even bigger.[20]

In summary, the estimated coefficient of VC of the largest clan is very robust, remaining significant and varying only slightly when different controls are added. The coefficient of VC of the second-largest clan is also robust unless village-specific linear time trends are controlled

---

[20]The county government started taking charge of village public goods provision after the agricultural taxes, as well as village levies, were formally abolished in 2006. As shown in Appendix Table A1, our main results are robust when we drop observations after 2000.

for. These results show that the association between VCs of large clans and public goods expenditure is robust and not likely to be driven by trending factors, village-level economic and demographic changes, or differentiated support from the upper-level government. In Appendix Tables A1 and A2, we conduct more robustness checks to show that our finding is robust in different time periods and is not driven by extreme values. Because the coefficients of VC of the largest clan and VC of the second-largest clan are statistically indistinguishable from each other in most cases, in the rest of the paper we use the simplified model of Equation 2.

FIGURE 3. THE DYNAMIC EFFECT OF VCS OF LARGE CLANS
ON PUBLIC INVESTMENT
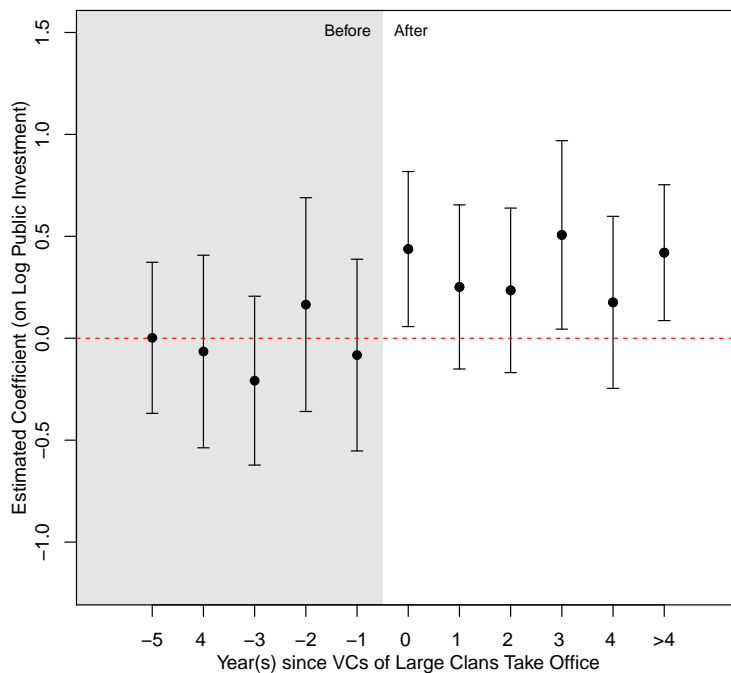
Note: This figure shows the dynamic effect of VC of large clans on the amount of public investment. Each black dot is an estimated coefficient of a dummy variable indicating the year(s) since the most recent VC of large clans took office (or before he took office).

To establish a causal relationship between VCs of large clans and public goods expenditure, we need to be sure that the identifying assumption is valid. We are more confident that this assumption holds if we find that public goods expenditure increases right after VCs of large clans took office. To achieve this, we create a set of dummies d_k,

23

$k = -5, -4, \cdots, 0, 1, \cdots, 4$ where $k = 0$ indicates the year a VC of the two largest clans took office, and other values respectively correspond to a specific year relative to that year. For example, $k = -1$ indicates one year before the closest year when a VC of a large clan replaced a VC of a small clan, and $k = 1$ indicates one year after. All the years that were five or more years before are pooled together as the reference category indicated by $k = -5$, while $k = 4$ includes four or more years after the year a VC of a large clan took office. Then we estimate Equation 2 by substituting this new set of dummies for D_{it}. The estimated coefficients of the dummies are shown in Figure 3. Before VCs of the two largest clans took office, the estimates are mostly negative and statistically insignificant. The coefficients turn positive and statistically significant only after VCs of large clans took office.[21]

Because the VDS recorded the amount of investment by project type, we can check for which types of investment the association between VCs of large clans and public goods expenditure is stronger. The results based on Equation 2 are shown in Table 3. The dependent variable is the log amount of village investment by type. Table 3 suggests that strong associations exist between VCs of large clans and investment in facilities of village primary schools and irrigation infrastructure. Although irrigation infrastructure can be built only for the benefits of large clans, village primary schools are rarely discriminatory in rural China. Therefore, we can at least conclude that having village leaders from large clans also benefited the rest of the villagers in addition to clan members. Moreover, in the long run, we do not observe that the level of income inequality deteriorated more quickly in villages with large lineage groups.[22] The problem of clan capture seems to be less severe than one would otherwise expect.

---

[21]Note that the coefficient is still positive and significant three years after a VC of the largest two clans took office although a VC's term is three years. One possibility is that VCs of the two largest clans stayed in office for more than one term. Another possibility is that the successor also came from the two largest clans.
[22]In Appendix Figure A6, we show that the level of income inequality increased the least in villages where the largest clan comprised of a considerable proportion of the village population.

	Log Public Investment (1,000 yuan)					
	Schooling	Road & Sanitation	Electricity	Irrigation	Foresta-tion	Others
	(1)	(2)	(3)	(4)	(5)	(6)
	FE	FE	FE	FE	FE	FE
VC of large clans	0.161	0.061	0.070	0.148	0.014	0.057
	(0.061)	(0.066)	(0.041)	(0.054)	(0.030)	(0.055)
Dependent variable mean	0.292	0.358	0.185	0.211	0.050	0.176
Year and village fixed effects	x	x	x	x	x	x
Observations	3,742	3,742	3,742	3,742	3,742	3,742
Villages	220	220	220	220	220	220

Note: *This table shows the associations between a VC of large clans and village public investment by project type. Standard errors clustered at the village level are in parentheses. The dependent variable is the log amount of village investment (1,000 yuan) of each type during that year. The independent variable is a dummy variable indicating whether a VC came from the village's largest or second-largest clans. 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.*

Clan Cohesiveness

To provide further evidence that it is the informal institutions of large clans that matter, we examine additional information on large clans in the sample villages. When a clan is more cohesive, it is more likely that it has greater social power in the village, as a result, its rules are more strictly enforced. Therefore, we expect the association between VCs of large clans and public goods expenditure to be stronger in villages with more cohesive large clans.

To test this hypothesis, we look at two indicators of clan cohesiveness: (1) whether the largest or second-largest clan kept records of family trees, and (2) whether they maintained lineage halls. We take these two variables as proxies for clan cohesiveness because they signify how closely clan members were connected with each other and whether a clan had sufficient organizational capacity. Records of family trees and lineage halls are specific to clans and signal a close relationship within the clan and frequent clan activities. Annual sacrificing activities, weddings, funerals and other clan events often take place in lineage halls.

It is possible that in villages with more public investment and better infrastructure,

lineage halls could be more regularly refurbished for reasons that we cannot fully control for. To minimize such biases, the indicator of lineage halls is coded as one if they were built before the observed time periods and zero otherwise. Because maintaining records of the family tree requires persistent efforts of clan members, it is less likely to incur such biases.

FIGURE 4. VCs OF LARGE CLANS, CLAN SIZE, AND COHESIVENESS
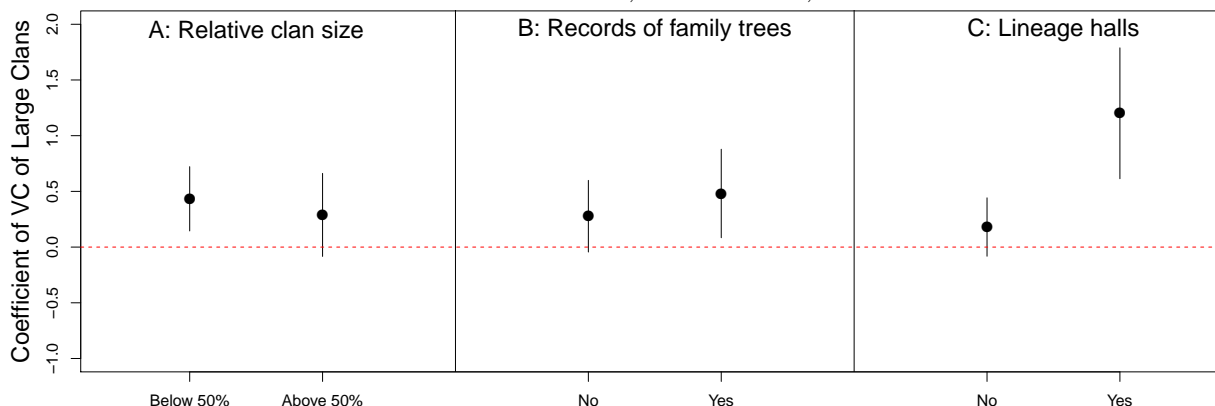
Note: *The above figures show the heterogeneous effects of VC of large clans on public goods expenditure with 95% confidence intervals. From the left to the right, the sub-samples are (1) villages whose combined size of the two largest clans is above or below 50 percent, (2) villages in which any of the two largest clans had kept records of family trees or not, and (3) villages in which the two largest clans had maintained any lineage halls since the beginning of the observed time periods or not.*

We interact VC of large clans with each of the two indicators of clan cohesiveness and put both the VC dummy and the interaction term in regressions using the baseline fixed effects specification. As a comparison, we also use a specification that includes the interaction between the VC dummy and a dummy variable indicating that the combined size of the two largest clans was above 50 percent (roughly the median). Figure 4 visualizes the results.[23] Panel A of Figure 4 shows that the association between public goods expenditure and VCs of large clans is not increasing in the combined size of the two largest clans. However, Panels B and C of Figure 4 show that when large clans appeared to be more cohesive, i.e. having maintained records of family trees and especially lineage halls (in 48% and 17% of the villages, respectively), VCs of large clans are strongly associated with more spending on public investment. These results suggest that what really matters for spending

---

[23]Corresponding regression coefficients can be found in Online Appendix Table A5.

on public investment is not the number of people a large clan had, but the social aspect of the organization, likely the rules and norms it enforced.

The Role of Village Party Organizations

Existing literature shows that VCs can face considerable constraints when exercising power (e.g., Oi and Rozelle 2000, O'Brien and Han 2009). A VC not only receives orders from the township government, but is also subject to checks and even directives of the village party organization, especially the VPS. In fact, studies show that the power struggle between the VC and the VPS paralyses village self-government in some places (e.g., Tan 2010). Would the consideration of VPSs alter our main findings, for instance, would the VPS's clan membership affect the VC's ability to provide public goods? What would happen if the VC was also the VPS, which is called *yijiantiao* (literally, "one-shoulder")? What if the VC and VPS were from the same clan? Or what if the VC was in the village party branch, a sign that he was recognized and supported by the VPS?

Fortunately, the VDS includes data on the VPS and village party organizations for more than 130 villages, roughly 60% of the entire sample.[24] Such information allows us to answer the questions we just posed. Using names of VPSs and data on village clan structure, we define a dummy variable indicating whether the VPS came from the village's largest or second-largest clan.

We first consider how "one-shoulder" affects our results. For that purpose, we define a dummy variable indicating "one-shoulder." We include the VC dummy, the VPS dummy, the "one-shoulder" dummy, as well as the interactions between the VC and VPS dummies and between the VC and "one shoulder" dummies in the baseline two-way fixed-effect model and visualize the result in Panel A of Figure 5. Our specification allows us to compare five scenarios with the reference scenario in which both the VC and the VPS came from small clans and were not the same person: (A1) the VC and VPS were the same person but he

---

[24]Lack of data on VPSs and party organizations for the rest of the VDS sample was due to administrative reasons. Statistical analysis shows that villages with available data are not substantially different from the rest.

was not from one of the two largest clans, (A2) the VPS came from one of the two largest clans while the VC came from a small clan, (A3) the VC came from one of the two largest clans while the VPS came from a small clan, (A4) both the VC and the VPS were from one of the two largest clans yet they were not the same person, and (A5) the VC and VPS were the same person and came from one the two largest clans. Figure 5 shows that the average amounts of public investment of the last four scenarios are significantly higher than that of the reference scenario, after village and year fixed effects have been controlled for. The effect under the first scenario, i.e., "one shoulder" from a small clan, is positive but not statistically significant. Moreover, among the five scenarios, the level of public goods expenditure is highest under the fourth scenario, in which the VC and the VPS, though not the same person, were both from large clans.[25] Those results indicate that "one shoulder" is less important than the clan membership of the VC and the VPS.

FIGURE 5. VILLAGE LEADERS AND VILLAGE PARTY ORGANIZATIONS
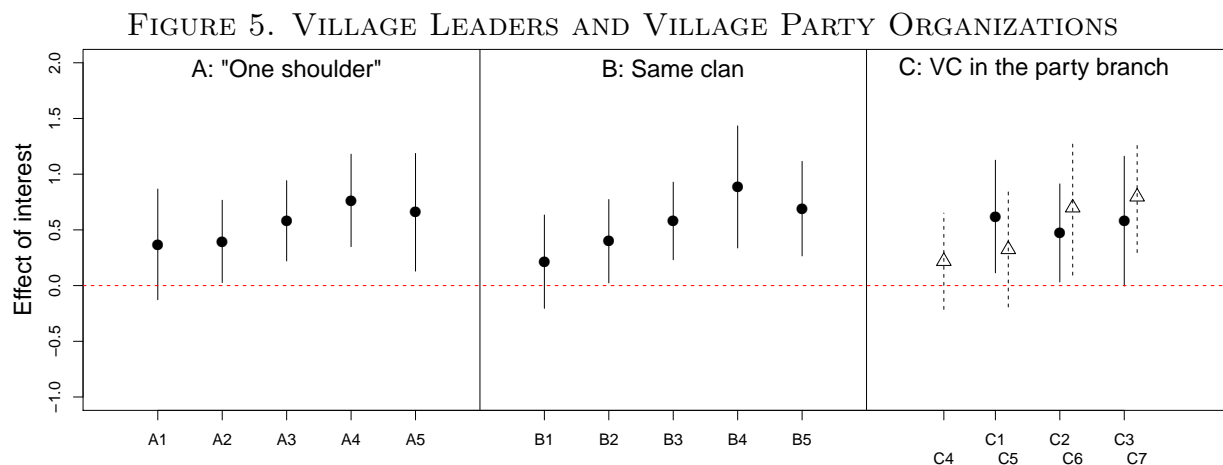
Note: *This figure shows the heterogeneous effects of VCs of large clans on public goods expenditure with 95% confidence intervals. From left to right, we consider three cases: (1) whether the VC and VPS were the same person, (2) whether the VC and VPS came from the same clan; and (3) whether the VC was in the village party branch.*

Next, we consider the effects when the VC and the VPS came from the same clan. The procedure is similar. Again, we estimate a "fully saturated" model using the baseline two-way fixed-effect specification and present the result in Panel B of Figure 5. The reference scenario, in which the VC and VPS were from small and distinct clans, is compared with

---

[25]Corresponding regression coefficients of Figure 5 are presented in Online Appendix Table A6.

the following five scenarios: (B1) the VC and VPS were from the same small clan; (B2) the VPS was from one of the two largest clans while the VC was not; (B3) the VC was from one of the two largest clans while the VPS was not; (B4) the VC and VPS were from large yet distinct clans; and (B5) both the VC and VPS were from the same large clan. The last four comparisons give positive and significant estimates but the first does not. The highest level of public goods expenditure happened when the VC and the VPS came from distinct large clans (Scenario B4). That is, the VC and the VPS did not have to come from the same clan as long as both of them came from a large clan. [26]

Lastly, we investigate whether being in the village party branch enhances a VC's ability to provide public goods. Again, we use a "fully saturated" model controlling for village and year fixed effects. Panel C of Figure 5 shows the results. The reference scenario is that the VC was not in the village party branch and neither the VC nor the VPS was from one of the two largest clans. There are seven scenarios to be compared with. In the first three scenarios, we have the VPS not from one of the two largest clans (the estimated effects of interest are depicted with dots) with one of the following three cases: (C1) the VC was the in the village party branch; (C2) the VC was from one of the two largest clans; and (C3) the VC was from one of the two largest clans and in the village party branch. The other four scenarios, whose estimated effects of interest are depicted with triangles, are when the VPS was from one of two largest clans with one of the following four cases: (C4) the VC was neither from a large clan nor in the village party branch; (C5) the VC was in the village party branch while not from one of two largest clans; (C6) the VC was from one of the two largest clans while not in the village party branch; and (C7) the VC was from one of the two largest clans and in the village party branch. The estimated effects of interest are positive and statistically different from zero in all but Scenarios C1 and C5, in which the VC was in the party branch but not from one of the two largest clans. Also worth noting are that (1) all scenarios in which the VC was from one of the two largest clans have significantly

---

[26]The above two sets of results suggest that it seems a good thing if there existed some competition between the VC and the VPS as long as they came from one of the two largest clans. Further exploration is needed to find out the exact reason behind it.

positive estimates; (2) it is not necessary to require the VPS coming from one of two largest clans to have more public goods expenditure as long as the VC was from one of the largest clans, a clear result when Scenario C1 is compared with Scenario C5; and (3) the effect of VPS coming from one of the two largest clans becomes insignificant if the VC did not, a result shown by Scenario C4.[27]

In summary, we not only show that the strong association between VCs of large clans and a higher level of public investment is robust when we take into account the roles of VPSs and village party organizations, but also find that the level of public investment is higher when the VPS was from a large clan than when he was not. However, we do not find enough evidence that the VC and VPS being the same person or from the same clan brought about additionally more public investment. Nor do we find that the VC being in the village party branch is particularly important for public goods provision once we control for the clan memberships of the VC and VPS. Lastly, the role of the VPS diminished when the VC came from a large clan. In the rest of the paper, we will mainly focus on the role of informal institutions associated with VCs primarily because: (1) we have more complete data on VCs than VPSs, and (2) there are potentially more quasi-exogenous variations in the turnovers of VCs than VPSs—as we will see in the next section, these variations give us more leverage to identify the causal effect of informal institutions of lineage groups.

A REGERSSION DISCONTINUITY DESIGN

In this section, we employ a regression discontinuity (RD) design to address the potential endogeneity of electoral outcomes. Recall that we rely on elections as the source of variations of informal institutions that affect local governance. A natural question is why sometimes the largest clans won the election, while at other times they lost. We admit that the impact factors are complex and mostly beyond our knowledge. One obviously important factor is the CCP. To strengthen its rule in the countryside, the CCP has been trying to demobilize clans

---

[27]These results arise probably because VCs, who were popularly elected, were more able to obtain support from their clans than VPSs, who were appointed by the government.

in elections. Our key identification assumption, though, is that those factors are uncorrelated with public investment, the outcome variable. Our fixed effects approach and additional controls of the provincial and village time trends, as well as other time-varying covariates, buttress up this assumption. However, concerns of reverse causality and unobserved time-varying confounders remain. For example, villagers may expect VCs of large clans to provide more public goods and, therefore, elect them into office. A sharp RD design can address this concern because after conditioning on the forcing variable, the treatment indicator is uncorrelated with time-varying confounders at the cutoff. In our case, the forcing variable is the share of votes of a candidate from the largest or second-largest clan against the share of votes of a contender from a smaller clan.

Several caveats of employing an RD design in this study are worth noting. First, an RD estimate gives the local average treatment effect at the cutoff, which is 50 percent of all the votes. This quantity is not necessarily a quantity of interest and can provide very different estimates from those generated by the benchmark fixed effects models. One might be especially concerned about the external validity of an RD design in the Chinese context. Because formal democratic institutions are weak, elections may not be "allowed" to be close under many circumstances.

Second, although the treatment assignment mechanism is very clear in an RD design, in reality, the assignment mechanism may suffer from manipulation of the forcing variable by interested parties, making the RD design invalid. Evidence exists that an RD design fails regarding elections of US House of Representatives during a certain period of time (Caughey and Sekhon 2011).[28] Vote-buying, electoral frauds, and interference of the upper-level governments were widely observed by scholars for Chinese rural elections (e.g., Shi 1999). Therefore, we may need to worry about the validity of an RD design in the context of rural China.

Third, an RD design typically demands a large amount of data. This problem is especially

---

[28]A follow-up study shows that the problem is not as severe as one might think and RD remains a valid method for causal inference in most situations (Eggers et al. (forthcoming)).

true in our case. To construct an RD design, we need not only information of the elected VCs, but also information of their runoffs, including the lineage groups the latter belonged to and the votes they received in elections. These requirements cut the sample size to 2,230 village-year observations and 871 elected terms, compared with the original 3,742 observations and 1,315 terms. Dropped observations are mostly in early periods of the time series when village elections were not contestable (therefore no runoff information was recorded). Moreover, because we are interested in the effect of VCs of large clans and use VCs from small clans as comparison, only observations in which one of the candidates was from one of the two largest clans while the other was from a smaller clan are useful for constructing the RD design. This requirement further reduces the sample down to 715 observations and 253 terms.[29] Because the identification comes from close elections, the power of our RD analysis is limited.

FIGURE 6. ROBUSTNESS CHECK: A REGRESSION DISCONTINUITY DESIGN
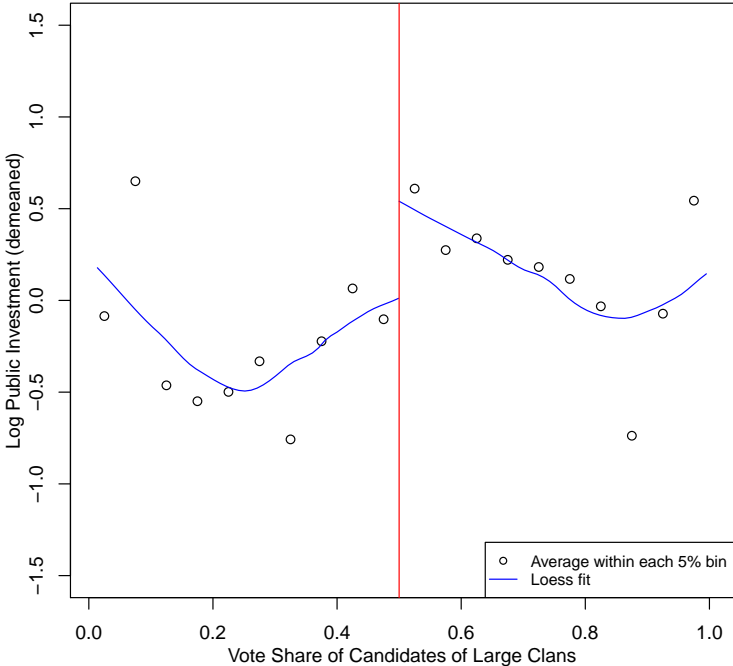
Note: *This Figure shows the averages of log amount of investment within each 5 percent vote-share bin and two loess fits from locally linear regressions on both sides of the cutoff.*

Bearing these concerns and limitations in mind, we present the main result of the RD

---

[29]To remove time invariant heterogeneities and aggregate shocks, we first run a standard fixed effects model controlling for village and year fixed effects and use the residuals in the RD analysis.

design in Figure 6, which shows the averages of log investment within each 5 percent vote-share bin and two loess fits (from locally linear regressions) on both sides of the cutoff. The RD estimate is 0.573 with a standard error of 0.301; both are almost twice as large as the fixed effects estimates.[30] We find that the results from the RD design are consistent with our main finding and offer us more confidence that VCs of large clans causally increased public goods expenditure.

MECHANISMS

We have already shown that the presence of VCs from one of the two largest clans is associated with at least 35 percent more investment in public goods. In this section, we investigate two mechanisms, namely, the collective action mechanism and the accountability mechanism, through which informal institutions could possibly facilitate public goods provision.

First, we test whether the presence of VCs of large clans is connected with easier collective action among the villagers by using household-level data of levies that villagers paid to the village committee. As mentioned before, a VC needed to seek villagers' voluntary compliance to collect levies from them. If VCs of large clans were more likely to collect more levies for public investment than VCs from small clans, we then have a critical piece of evidence to support the collective action mechanism.

We have household-level data for around one-third of the sample villages.[31] Table 4 presents the results based on this sample. Using the baseline model that controls for both village and year fixed effects, Column 1 shows that the presence of VCs of large clans is weakly associated with more levies. When a VC of large clans was in office, villagers on average paid 13.2 percent more levies a year. The estimated coefficient is not statistically significant,

---

[30]In Appendix Table A7 and Figure A4, we present the point estimates from the RD analysis and conduct more validity tests.

[31]The household-level data come from the NFS which only allows researchers to obtain a maximum of one third of its household data. In addition, it does not allow the household-level data to be transported and used directly outside China; the data were first processed in China to generate means and values at each income decile for the variables of interest. Our analysis is, therefore, based on the processed data.

TABLE 4. VCs OF LARGE CLANS AND LEVIES

	Log Levies (yuan)		
	(1) FE	(2) FE	(3) FE
VC of large clans	0.132 (0.192)		0.110 (0.188)
Public investment dummy		0.304 (0.096)	0.321 (0.134)
VC of large clans × public investment dummy			-0.037 (0.174)
Dependent variable mean	4.224	4.224	4.224
Year and village fixed effects	x	x	x
Observations	1,080	1,080	1,080
Villages	69	69	69

Note: *This table shows that (1) the presence of a VC of large clans is weakly associated with more levies villagers paid to the village government and that (2) the presence of village public investment projects is strongly correlated with a higher level of levies. The dependent variable is the log amount of average levies villagers paid to the village government in a particular year. The independent variables include a dummy variable indicating whether a VC came from the village's largest or second-largest clan, a dummy indicating any public investment projects during that year, and their interaction. Standard errors clustered at the village level are in parentheses. The sample is based on village-year observations of 69 villages, of which household level data are available, from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

though, due to the large dispersion of the data. Column 2 regresses the log amount of levies on the public investment dummy when village and year fixed effects are controlled for. It shows that the amount of levies is highly correlated with the presence of public investment projects after time-invariant village heterogeneity and time-varying aggregate shocks are removed; the estimate coefficient is 0.304 and significant at the 1 percent level. In Column 3, we put in both dummies and their interaction.

The result is visualized in Figure 7a, which shows that no matter if VCs of small or large clans were in office, villagers paid more levies when there are public investment projects. On average, though, when VCs were from large clans, villagers paid higher levies to the village committee because higher frequencies of public investment projects were observed during

FIGURE 7. PUBLIC INVESTMENT AND LEVIES ON VILLAGERS
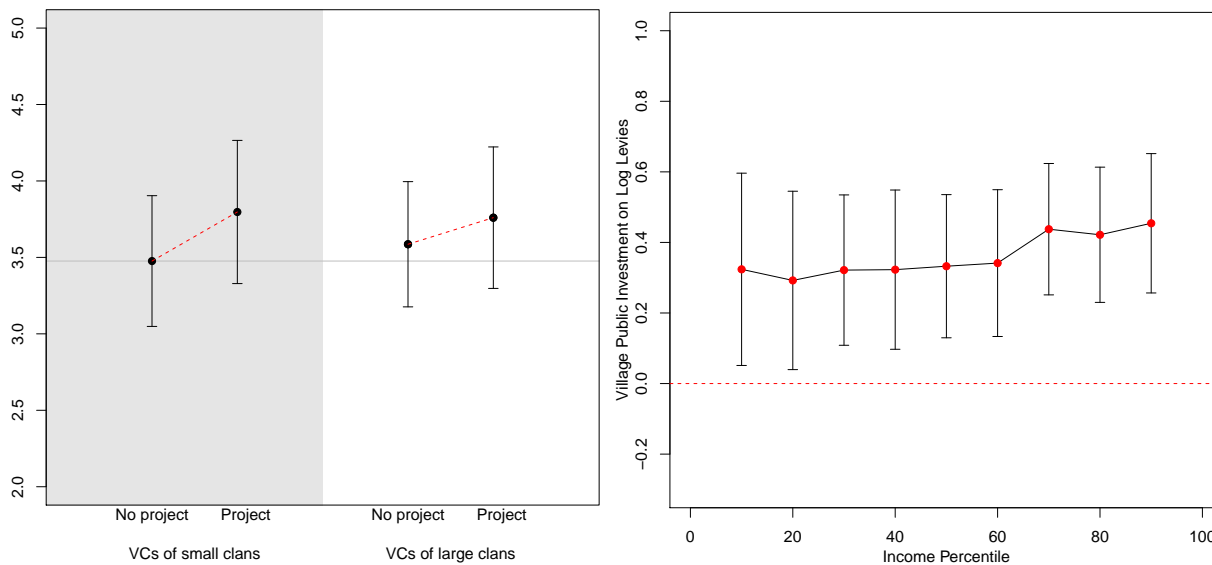
(a) Levies and public investment by VCs

(b) Levies and public investment by villagers' income

Note: *Figure 7a shows that the amount of levies the average household in each village paid to the village committee during terms of VCs of small or large clans under two circumstances: when there was no public investment project during the year and when there was at least one project. Figure 7b shows the correlations between village public investment projects and the amount of levies households at specified income percentiles paid to the village committee. Village and year fixed effects are controlled for in both figures. The microlevel data in both figures come from 69 villages, a subset of the full sample.*

the terms of VCs of large clans. Figure 7b shows the immediate distributive consequences of public investment on levies. Each red dot is an estimated coefficient from a separate fixed effects regression using levies that households at a specified village income percentile paid to the village committee as the dependent variable and the public investment dummy as the independent variable. It shows that both the rich and poor in the villages paid extra levies when there were public projects. Note that the regressions presented in Figure 7 and Column 2 and 3 of Table 4 do not imply a causal relationship between the presence of public investment projects and the amount of levies villagers paid to the village committee, because both variables are likely results of the presence of VCs of large clans. They show, however, that to make a public investment project happen, a VC often needs to convince the majority of the villagers to pay for it. To the extent that VCs from large clans took up more investment projects than VCs from small clans, this allows us to conclude that large

clans help VCs overcome the collective action problem.

Second, we investigate if there is any sign that informal institutions of large clans hold VCs accountable. We look at the amount of administrative expenditure of the village committee. If a VC is subject to close scrutiny when using public funds, non-productive administrative expenditure is most likely to be curbed. Previous studies have shown that electoral reforms in rural China caused a sharp decrease in the share of administrative expenditure in total expenditure of village committees (Wang and Yao 2007). Using the baseline specification (Equation 2) and both the share of administrative expenditure in total expenditure and the log amount of administrative expenditure as outcome variables, we show that VCs of large clans and administrative expenditure have almost zero correlation after village and year fixed effects are controlled for (Table 5). Although we cannot rule out the possibility, we do not find strong evidence for the informal accountability mechanism.

TABLE 5. VCS OF LARGE CLANS AND ADMINISTRATIVE EXPENDITURE

	Share of administrative expenditure in total expenditure (1) FE	Log administrative expenditure (1,000 yuan) (2) FE
VC of large clans	0.006 (0.014)	0.022 (0.073)
Dependent variable mean	0.230	2.315
Year and village fixed effects	x	x
Observations	3,037	3,037
Villages	208	208

Note: *This table shows that the association between VC of large clans and village administrative cost is close to zero after village and year fixed effects are controlled for. Standard errors clustered at the village level are in parentheses. In Column 1, the dependent variable is the share of administrative expenditure in total village expenditure in that year. In Column 2, the dependent variable is the log administrative expenditure (1,000 yuan). Both are from the NFS data. The independent variables are a dummy variable indicating whether a VC came from the village's largest or second-largest clan. The sample is based on village-year observations from 1986 to 2005 after village elections were introduced. Both regressions control for village and year fixed effects.*

ALTERNATIVE EXPLANATIONS

In this section, we discuss two alternative explanations for the observed association between VCs of large clans and more public goods expenditure, including (1) superior ability of VCs of large clans, and (2) improvement of formal institutions.

TABLE 6. LARGE CLANS, VCS' CHARACTERISTICS,
AND VILLAGE PUBLIC INVESTMENT

	Log Public Investment (1,000 yuan)		
	(1) FE	(2) FE	(3) FE
VC of large clans	0.345	0.328	0.331
	(0.127)	(0.129)	(0.129)
Years of education	-0.013	-0.009	-0.009
	(0.021)	(0.024)	(0.025)
Age when running election		-0.000	-0.002
		(0.006)	(0.006)
CCP member		0.008	0.009
		(0.125)	(0.126)
Village cadre when running election		-0.003	0.005
		(0.165)	(0.169)
Managerial jobs when running election		0.019	-0.035
		(0.533)	(0.555)
Experience of running election		0.139	0.139
		(0.125)	(0.128)
Family background: poor peasant			-0.114
			(0.153)
Denounced in the Culture Revolution (*pidou*)			0.183
			(0.325)
Dependent variable mean	1.125	1.146	1.143
Year and village fixed effects	x	x	x
Observations	3,487	3,375	3,347
Villages	218	214	213

Note: *This table shows that the association between a VC of large clans and village public investment is robust when we control for the VC's characteristics. 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 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-year observations from 1986 to 2005 after village elections were introduced. All regressions control for village and year fixed effects.*

First, do large clans select more competent leaders? Munshi and Rosenzweig (2010) find

that in Indian parochial elections, castes with large population shares help select leaders with superior observed characteristics, such as providing more public goods. It is also possible that a successful entrepreneur from a large clan uses his or her resources and expertise to bring increased prosperity to the village.[32] To investigate these possibilities, we compile data of VCs' characteristics, including years of formal education, age, administrative experience, experience of running businesses, CCP membership, historical family background, etc.[33] We control for these characteristics in the regressions. The results are shown in Table 6. The estimated coefficient of the VC of large clans remains almost unchanged. In fact, VCs' observed characteristics, such as education and administrative experience, do not seem to have any predictive power for the amount of public investment. In Appendix Table A8, we show that, compared with others, VCs of large clans did not have higher education or more administrative experience; after controlling for village and year fixed effects, we find that they appeared to be quite similar to the rest of the pool. The evidence does not support that lineage groups in rural China helped select more competent leaders.

Another explanation is improvement in formal electoral institutions. As formal institutions improve, it is possible that elected leaders are more likely to implement policies catering to the median voter's interest, such as providing more public goods. It is also possible that under better formal institutions, officials elected into office have preferences that are more in line with preferences of the voters. These preferences might not have been captured by VCs' observed characteristics, but might be correlated with clans where VCs come from. Because our dataset has detailed information of electoral rules and procedures, including contested elections (an election is contested when there are more candidates than positions), open nomination, secret ballots, proxy voting, and moving ballot boxes, we can test if our main results are driven by changes of these indicators.[34] The results are shown in Table 7.

---

[32]O'Brien (1994) reports that successful managers of collective enterprises were more likely to be trusted by villagers. Oi and Rozelle (2000) show that rural industrialization changed elites and other villagers' incentives to participate in grassroots politics.

[33]Historical family background was determined during the land reform in the 1950s by the local CCP authorities. After that, villagers from a poor peasant family background assumed most of the leadership position in the villages, as a legacy of the Communist revolution.

[34]Some of the indicators clearly suggest improvement in the electoral system, such as contested elections, open

	Log Public Investment (1,000 yuan)					Contested Elections	Open Nomination
	Full Sample						
	(1) FE	(2) FE	(3) FE	(4) FE	(5) FE	(6) FE	(7) FE
VC of large clans	0.369 (0.118)	0.364 (0.118)	0.370 (0.117)	0.368 (0.117)	0.368 (0.118)	0.432 (0.147)	0.377 (0.144)
Contested election	0.001 (0.162)						
Open nomination		-0.146 (0.159)					
Secret ballot			0.074 (0.157)				
Proxy voting				0.039 (0.155)			
Moving ballot					0.127 (0.129)		
Dependent variable mean	1.092	1.092	1.092	1.092	1.092	1.144	1.148
Year and village fixed effects	x	x	x	x	x	x	x
Observations	3,742	3,742	3,742	3,742	3,742	2,888	2,631
Villages	220	220	220	220	220	215	196

Note: *This table shows that the association between a VC of large clans and village public investment is robust when we control for formal electoral institutions and procedures and when we use subsamples of contested elections and open nomination. 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 variable is a dummy variable indicating whether a VC came from the village's largest or second-largest clan. Columns 1–5 use the full sample, which includes village-year observations from 1986 to 2005 after village elections were introduced. Columns 6 and 7 use sub-samples in which contested elections and open nomination were introduced, respectively. All regressions control for village and year fixed effects.*

As expected, our main finding is robust when we control for these indicators in the regressions. In fact, the institutional variations over time have very limited explanatory power for the variations in the amount of public investment. Moreover, the estimated coefficients of the VC dummy are slightly bigger in two subsamples where contested elections and open nomination had been introduced, respectively.

---

nomination and secret ballots. The impacts of proxy voting and moving ballot boxes are more ambiguous. They are supposed to increase the turnout of villagers, but they also create plenty of room for corruption and electoral frauds. Appendix Figure A5 shows the overtime changes of these indicators in our sample.

CONCLUSION

In the context of rural China, we find that informal institutions of lineage groups—rules and norms created and enforced by lineage groups—facilitate local public goods provision. Using fixed effects models as the main estimation strategy and a regression discontinuity design as a robustness check, we show that the presence of village chairpersons of large clans increased local public goods expenditure considerably. Such a relationship is stronger in villages where large clans persistently maintained lineage halls. Our finding is robust when we consider the roles of village party secretaries and village party organizations, as well as alternative explanations, such as superior observed characteristics of VCs of large clans and improved formal electoral institutions. This paper is among the first attempts to study the causal effect of informal institutions on governance outcomes.

We explore two possible channels: (1) informal institutions facilitate collective action of financing public goods among villagers, and (2) informal institutions hold VCs accountable to villagers. We show that the collective action channel is better supported by data. We find that villagers at almost all income percentiles paid extra levies to the village committee when there were public investment projects. However, we find little evidence that informal institutions held village officials accountable: on average the amount of administrative cost did not change when VCs of large clans were in office.

Two questions are not fully answered by this paper and require future research. The first is the possibility that large clans capture grassroots politics. The evidence presented in this paper suggests that large clans might have improved local governance in rural China in one specific aspect, namely, spending on public investment. However, it is possible that we do not measure outcomes that deteriorated because of clan power. For example, public goods expenditure as we have measured might have benefited members of large clans much more than the rest of the villagers, or VCs of large clans filled their pockets and those of their clan members' as they provided public goods. Large clans might collude with township officials to capture local politics as well. But because we do not have information on corruption or

who used what public facilities, these consequences are not reflected by our study.

The fact that we do not observe clan capture might be due to that leaders of large clans were under tight control of the CCP. Although open nomination of candidates is the *de jure* procedure in village elections, the CCP, especially its organ at the township level, heavily intervenes in the nomination process. Moreover, as we discuss in the paper, the exercise of power of the VC is constantly checked by the CCP. Such a unique institutional arrangement may limit the generalizability of our finding. For example, in places where local leaders are not closely monitored and controlled by other parties or the upper-level government, informal institutions may enable leaders to extract rents from constituencies or target transfers to a narrow group of supporters.

The second question that requires more research is the co-evolution of formal and informal institutions. How do changes of formal institutions affect the functioning of informal institutions and how do political actors embedded in informal institutions respond to changes of incentives due to formal institutional changes? In this paper, we attempt to identify the effect of informal institutions in the context of rural democracy. Unfortunately, we cannot compare the effect *before and after* the introduction of elections due to data limitations. An equally interesting question is how the role of informal institutions has changed since the tax-and-fee reform deprived villages of their autonomous status of finance.

References

Alesina, Alberto, R. Baqir and C.M. Hoxby. 2004. "Political Jurisdictions in Heterogeneous Communities." *Journal of Political Economy* 112(2):348–96.

Alesina, Alberto, Reza Baqir and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114(4):1243–84.

Banerjee, Abhijit and Lakshimi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review* 95(4):1190–213.

Böröcz, Joszef. 2000. "Informality Rules." *East European Politics and Societies* 14(2):348–80.

Caughey, Devin and Jasjeet S. Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942-2008." *Political Analysis* 19(4):385–408.

Cohen, Myron L. 1990. "Lineage Organization in North China." *The Journal of Asian Studies* 49(3):509–34.

Collins, Kathleen. 2003. "The Political Role of Clans in Central Asia." *Comparative Politics* 35(2):171–90.

Collins, Kathleen. 2004. "The Logic of Clan Politics." *World Politics* 56(2):224–61.

Duara, Prasenjit. 1988. *Culture, Power, and the State: Rural North China, 1900-1942.* Stanford, CA: Stanford University Press.

Eggers, Andrew, Anthony Fowler, Jens Hainmueller, Andrew B. Hall and James M. Snyder. (forthcoming). "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races." *American Journal of Political Science* .

Esteban, Joan and Ray Debray. 2001. "Collective Action and the Group Size Paradox." *American Political Science Review* 95(3):663–72.

Fei, Hsiao-Tung. 1946. "Peasantry and Gentry: An Interpretation of Chinese Social Structure and Its Changes." *The American Journal of Sociology* 52(1):1–17.

Freedman, Maurice. 1958. *Lineage Organisation in Southeast China: Fukien and Kwangtung.* London: University of London and Athlone Press.

Freedman, Maurice. 1967. "Ancestor Worship: Two Aspects of the Chinese Case". In *Social Organization: Essays Presented to Raymond Firth*, ed. Maurice Freedman. Chicago: Aldine.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner and Jeremy M. Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision." *American Political Science Review* 101(4):709–25.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner and Jeremy M. Weinstein. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action.* New York: Russell Sage Foundation.

Helmke, Gretchen and Steven Levitsky. 2004. "Informal Institutions and Comparative Politics: A Research Agenda." *Perspectives on Politics* 2(4):725–40.

Liu, Hui-chen Wang. 1959. *The Traditional Chinese Clan Rules.* New York: J J Augustin.

Luo, Renfu, Linxiu Zhang, Jikun Huang and Scott Rozelle. 2007. "Elections, Fiscal Reform and Public Goods Provision in Rural China." Working Paper, Freman Spogli Institute for International Studies, Stanford University, Stanford, CA.

Luo, Renfu, Linxiu Zhang, Jikun Huang and Scott Rozelle. 2010. "Village Elections, Public Goods Investment, and Pork Barrel Politics, Chinese-style." *Journal of Development Studies* 46(4):662–84.

Madsen, Richard. 1984. *Morality and Power in a Chinese Village.* Berkeley and Los Angeles, CA: University of California Press.

Manion, Melanie. 1996. "The Electoral Connection in the Chinese Countryside." *American Political Science Review* 90(4):736–48.

Manion, Melanie. 2006. "Democracy, Community, Trust: The Impact of Elections in Rural China." *Comparative Political Studies* 39(3):301–24.

Martinez-Bravo, Morica, Gerard Padró i Miquel, Nancy Qian and Yang Yao. 2011. "Do Local Elections in Non-Democracies Increase Accountability? Evidence from Rural China." NBER Working Paper 16948, National Bureau of Economic Research, Cambridge, MA.

Mattingly, Daniel. 2014. "One Party Rule of Divided Societies: Thwarting Ethnic Political Mobilization in China." Presented at the Annual Meeting of the Midwest Political Science Association, Chicago.

Munshi, Kaivan and Mark Rosenzweig. 2010. "Networks, Commitment, and Competence: Caste in Indian Local Politics." Mimeo, Brown University, Providence, RI.

Narayan, Deepa. 1999. *Bonds and Bridges: Social Capital and Poverty.* Washington, DC: World Bank.

O'Brien, Kevin. 1994. "Implementing Political Reform in Chinas Villages." *Comparative Politics Australian Journal of Chinese Affairs* 32:33–60.

O'Brien, Kevin and Lianjiang Li. 1999. "Selective Policy Implementation in Rural China." *Comparative Politics* 31(2):167–86.

O'Brien, Kevin and Lianjiang Li. 2000. "Accommodating 'Democracy' in a One-Party State: Introducing Village Elections in China." *The China Quarterly* 162:465–89.

O'Brien, Kevin and Lianjiang Li. 2006. *Rightful Resistance in Rural China.* Cambridge, UK: Cambridge University Press.

O'Brien, Kevin and Rongbin Han. 2009. "Path to Democracy? Assessing Village Elections in China." *Journal of Contemporary China* 18(60):359–78.

O'Donnell, Guillermo. 1996. "Another Institutionalization: Latin America and elsewhere." Kellog Institute for International Studies, Working Paper 222, Notre Dame, IN.

Oi, Jean. 1999. *Rural China Takes Off: Institutional Foundations of Economic Reform.* Berkeley: University of California Press.

Oi, Jean and Scott Rozelle. 2000. "Elections and Power: The Locus of Decision-Making in Chinese Villages." *The China Quarterly* 162(June):512–39.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, MA: Harvard University Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge, UK: Cambridge University Press.

Pastor, Robert A. and Qingshan Tan. 2000. "The Meaning of China's Village Elections." *The China Quarterly* 162(June):490–512.

Perry, Elizabeth J. 2002. *Challenging the Mandate of Heaven: Social Protest and State Power in China.* New York: M. E. Sharpe.

Shen, Yan and Yang Yao. 2008. "Does Grassroots Democracy Reduce Income Inequality in China?" *Journal of Public Economics* 92(10-11):2182–98.

Shi, Tianjian. 1999. "Village Committee Elections in China: Institutionalist Tactics for Democracy." *World Politics* 51(3):385–412.

Sklar, Richard. 2004. "The Premise of Mixed Government in African Political Studies". In *Tradition and Politics: Indigenous Political Structures in Africa*, ed. Olufemi Vaughan. Trenton: Africa World Press.

Su, Fubing, Tao Ran, Xin Sun and Mingxing Liu. 2011. "Clans, Electoral Procedures and Voter Turnout: Evidence from Villagers Committee Elections in Transitional China." *Political Studies* 59:432–57.

Tan, Qingshan. 2006. *Village Elections: Democratizing the Chinese Countryside.* New York: Edwin Mellen Press.

Tan, Qingshan. 2009. "Building Democratic Infrastructure: Village Electoral Institutions." *Journal of Contemporary China* 18(60):411–20.

Tan, Qingshan. 2010. "Why Village Election Has Not Much Improved." *Journal of Chinese Political Science* (15):153–167.

Tsai, Lily L. 2007. "Solidary Groups, Informal Accountability, and Local Public Goods Provision in Rural China." *American Political Science Review* 101(2):355–72.

Unger, Jonathan. 2002. *The Transformation of Rural China.* New York: M.E. Sharpe.

Wang, Mingming. 1996. "Clans, the Society and the State: Rethinking Freedman's Theory." *Chinese Social Science Quarterly (Hong Kong)* 16.

Wang, Shuna and Yang Yao. 2007. "Grassroots Democracy and Local Governance: Evidence from Rural China." *World Development* 29(10):1635–49.

Watson, James. 1975. *Emigration and the Chinese Lineage.* Berkeley, CA: University of California Press.

Watson, James. 1982. "Chinese Kinship Reconsidered: Anthropological Perspectives on Historical Research." *The China Quarterly* 92:589–622.

Whiting, Susan. 1996. "Contract Incentives and Market Discipline in China's Rural Industrial Sector". In *Reforming Asian Socialism, the Growth of Market Institutions*, ed. John McMillan and Barry Naughton. Ann Arbor: University of Michigan Press.

Zhang, Jing. 2007. *Political Power on the Grassroots: Several Questions on Institutions in the Countryside (Jiceng Zhengquan: Xiangcun Zhidu Zhu Wenti)*. Shanghai: Shanghai Century Publishing Group.

Zhang, Xiaobo, Shenggen Fan, Linxiu Zhang and Jikun Huang. 2004. "Local Governance and Public Goods Provision in Rural China." *Journal of Public Economics* 88:2857–71.

A Online Appendix (Not For Publication)

Contents of the Appendix

A.1 The VDS Sample

FIGURE A1. SAMPLE VILLAGES

Counties where sample villages are located

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 clan	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 (0.121)	0.349 (0.145)	0.310 (0.144)	0.303 (0.180)	0.306 (0.152)	0.400 (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:

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

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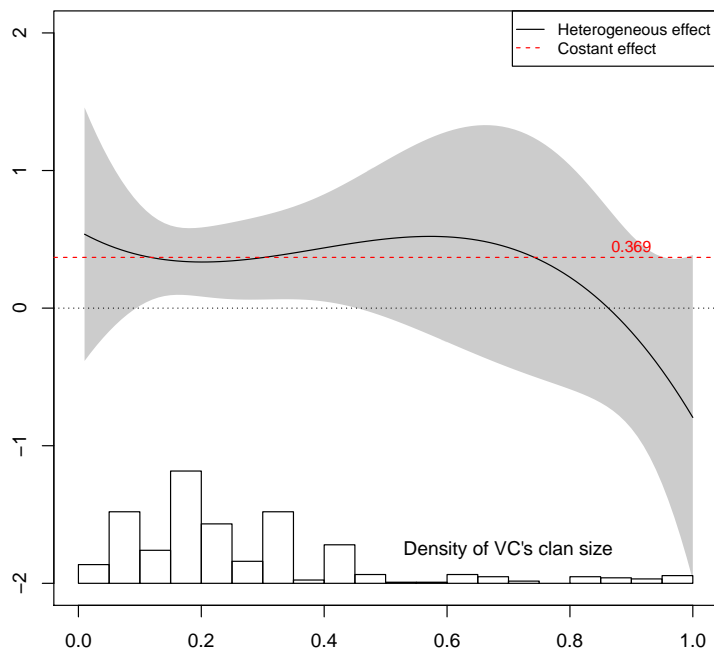
Note: *This figures shows the heterogeneous effect of VCs of large clans on the amount of public investment. The x-axis is the VC's clan size. The y-axis is the marginal effect of VC of large clans. The specification we use is shown in Equation 3.*

FIGURE A3. THE EFFECT OF VCs FROM LARGE CLANS ON PUBLIC INVESTMENT: DIFFERENT THRESHOLDS
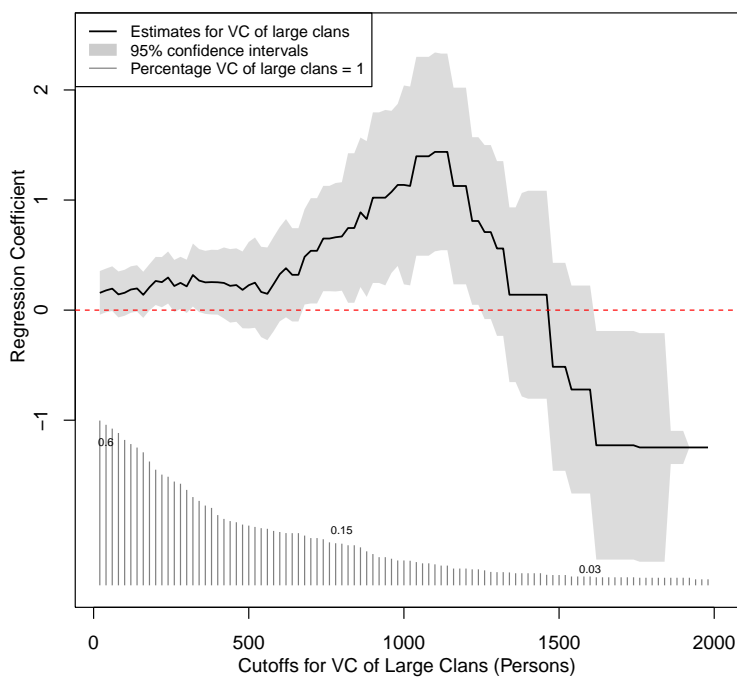
Note: This figure shows the estimated coefficients of VC of large clans using different threshold for large clans. For example, if the threshold is set at 500 persons, the dummy variable VC of large clans would equal one if the VC's clan consisted of more than 500 people and zero otherwise. The bars on the floor of the figure show the percentages of village-year observations when the variable VC of large clans equals one.

A-12

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.

	Log Public Investment (1,000 yuan)			
	(1) FE	(2) FE	(3) FE	(4) FE
VC of large clans	0.433 (0.147)	0.277 (0.164)	0.180 (0.134)	0.301 (0.165)
$\times$ Combined size > 50%	-0.144 (0.242)			-0.181 (0.256)
$\times$ Records of family trees		0.204 (0.261)		-0.107 (0.256)
$\times$ Lineage hall			1.021 (0.331)	1.095 (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.*

	Log Public Investment (1,000 yuan)			
	(1)	(2)	(3)	(4)
	FE	FE	FE	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.*
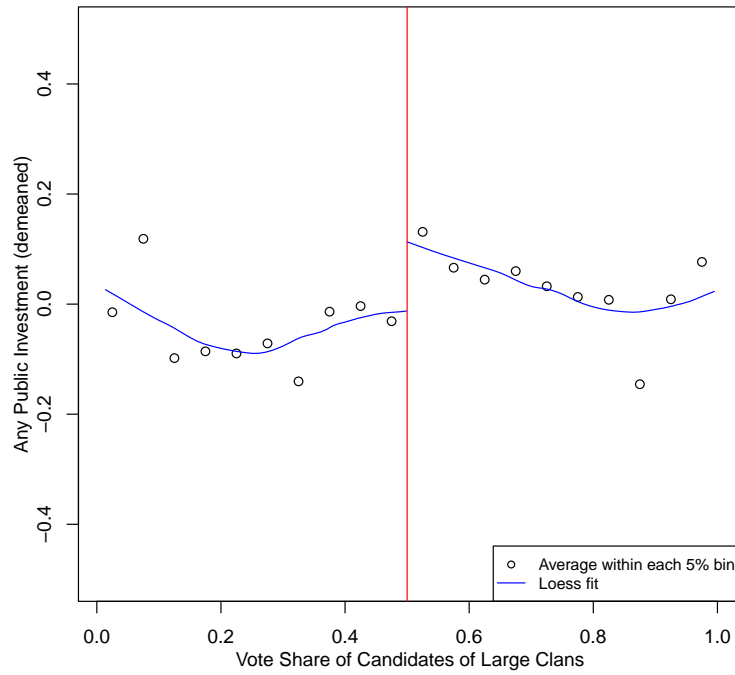
(a) Log public investment
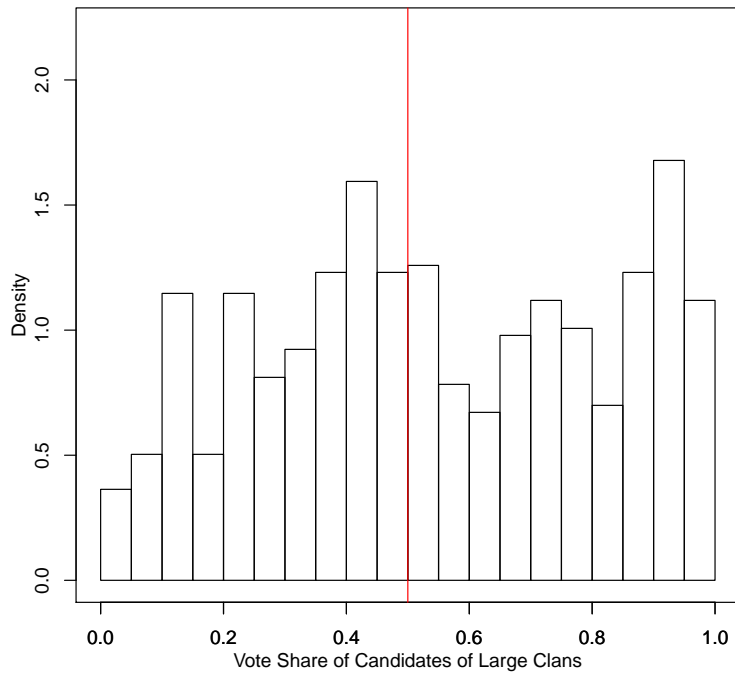
(b) Histogram of the vote share

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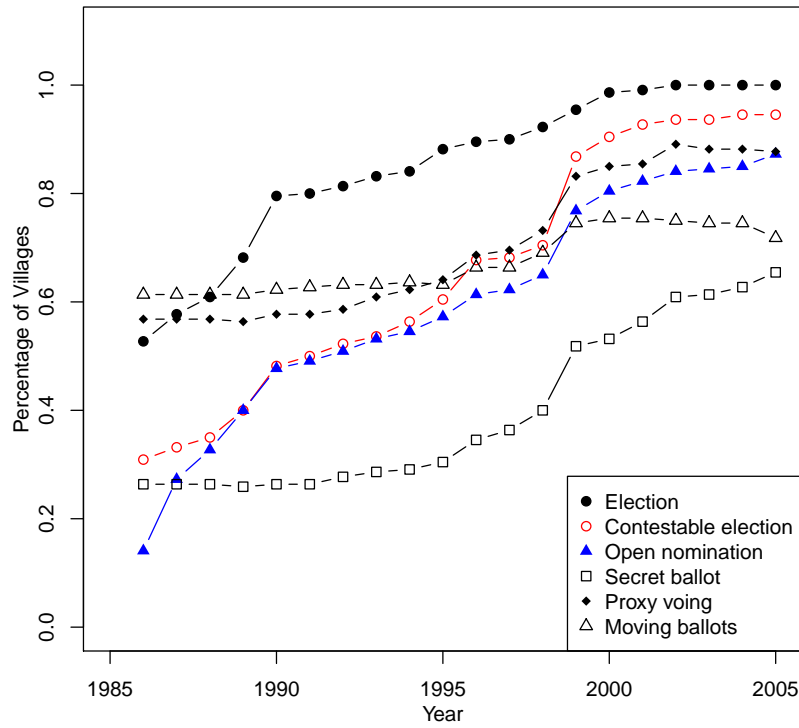
Note: *This figure shows the changes of electoral rules and procedure from 1986 to 2005 in the sample villages.*

FIGURE A6. LARGE CLANS AND INCOME INEQUALITY
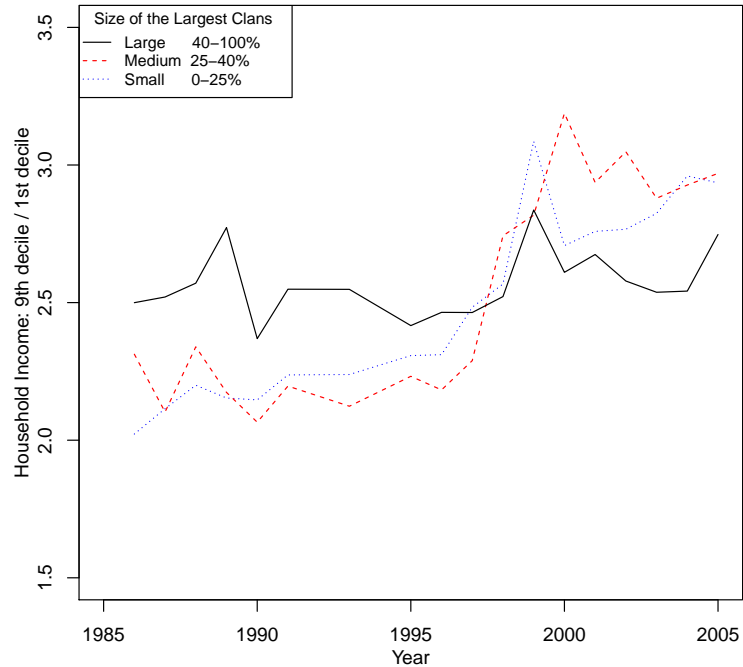
Note: *This figure shows the level of income inequality from 1986 to 2005 for three groups of villages: (1) villages with very big largest clans, (2) villages with medium-sized largest clans, and (3) villages with relatively small largest clans. Income inequality is measured by the ratio of household 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.*