NBER WORKING PAPER SERIES

THE COSTS AND BENEFITS OF CLAN CULTURE: ELITE CONTROL VERSUS COOPERATION IN CHINA

Shuo Chen
Raymond Fisman
Xiaohuan Lan
Yongxiang Wang
Qing Ye

Working Paper 32414
http://www.nber.org/papers/w32414

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2024

Chen would like to acknowledge the support from the National Natural Science Foundation of China (71933002; 72121002), Zhuoyue Talent Project, Theoretical Economics Peak Program and Legendary Project on Humanities and Social Sciences (XM04221238) at Fudan University. Wang would like to thank National Natural Science Foundation of China (grant No. 72172090) for financial support. Qing Ye would like to thank the National Natural Science Foundation of China (grant No. 72172060, 72132004) and the Major Project of Philosophy and Social Science Research Funds for Jiangsu University (grant No. 2020SJZDA068) for financial support. We thank Rui Rong for excellent RA work, all remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Costs and Benefits of Clan Culture: Elite Control versus Cooperation in China
Shuo Chen, Raymond Fisman, Xiaohuan Lan, Yongxiang Wang, and Qing Ye
NBER Working Paper No. 32414
May 2024
JEL No. N95,P32,Z10

ABSTRACT

Kinship ties are a common institution that may facilitate in-group coordination and cooperation. Yet their benefits – or lack thereof – depend crucially on the broader institutional environment. We study how the prevalence of clan ties affect how communities confronted two well-studied historical episodes from the early years of the People's Republic of China, utilizing four distinct proxies for county clan strength: the presence of recognized ancestral halls; genealogical records; rice suitability; and geographic latitude. We show that the loss of livestock associated with 1955-56 collectivization (which mandated that farmers surrender livestock for little compensation) documented by Chen and Lan (2017) was much less pronounced in strong-clan areas. By contrast, we show that the 1959-61 Great Famine was associated with higher mortality in areas with stronger clan ties. We argue that reconciling these two conflicting patterns requires that we take a broader view of how kinship groups interact with other governance institutions, in particular the role of kinship as a means of elite control.

Shuo Chen
School of Economics
Fudan University
600 Guoquan Street
Shanghai
cs@fudan.edu.cn

Raymond Fisman
Department of Economics
Boston University
270 Bay State Road, 304A
Boston, MA 02215
and NBER
rfisman@bu.edu

Xiaohuan Lan
School of Economics
Fudan University
600 Guoquan Road
Shanghai, China
xhlan@fudan.edu.cn

Yongxiang Wang
Shanghai Advanced Institute of Finance
Shanghai Jiao Tong University
benbenqingzi@hotmail.com

Qing Ye
Nanjing University
yeqing@nju.edu.cn

1 Introduction

The functioning of any society requires cooperation and coordination, which may be fostered through various means. Most notably, Greif and Tabellini (2010) among others emphasize the prominence of kinship or clan associations as a way of facilitating cooperation and coordination, particularly in East Asia. These clans, generally tied to a specific village or geography, were bound by familial ties and engaged in a range of cooperative activities, ranging from collective land ownership to group worship to joint production. Kinship institutions do not exist in isolation, however; rather, they interact with economic conditions and other societal institutions that, we argue, may on net lead to social benefit, or social harm.

In this paper, we illustrate the heterogeneous consequences of kinship institutions amidst two pivotal episodes in post-war Chinese history: collectivization in the mid-1950s and the Great Famine of 1959-61. We see our contribution as using a consistent measure of kinship strength across these two events to emphasize the heterogeneous consequences of reliance on kinship ties as a form of governance.

Prior work has used a range of proxies for kinship strength. Rather than taking a stand on any particular one, we assume that each measures a latent factor – clan prominence – with noise, and capture clan prominence via the first principle component of our four main measures, which include the following: (a) *North versus South* – as emphasized by Hsiao (1960), Fei and Liu (1982) and Greif and Tabellini (2010) among many others, clan-based cultures are far more prevalent in southern China, for a range of reasons that we delineate in Section 2 (throughout, we use the terms clan and kinship interchangeably); (b) *Genealogies* – Cao et al. (2022), for example, note that genealogical records are much more common in clan-dominated areas; (c) *Rice Suitability* – as emphasized by Noblit (2021), cooperation required to grow rice fostered kinship societies; we proxy for rice suitability with the measure of Talhelm et al. (2014); (d) *Ancestral Halls* – finally, following Padró i Miquel et al. (2015), we measure kinship intensity via the presence of a recognized ancestral hall, a cornerstone of Chinese clan culture.

In our first application, we study how the prevalence of kinship ties affected local responses to the large-scale collectivization of agriculture in 1955-57, precipitated by Mao Zedong's 1955 decree that there should be a rapid shift to "advanced" fully socialized cooperatives. Cooperative members were required to contribute their land and livestock to the collective, with minimal compensation.[1] As documented by Chen and Lan (2017) using a difference-in-differences framework, this collectivization led to the widespread killing of livestock by farmers who preferred the short-term benefit of food rather than surrendering their animals. Our first result is to show that this decline in livestock population was largely avoided in clan-dominated areas.

We next explore how clan prevalence affected communities during the Great Famine of 1959-1961, which occurred amidst the social upheaval of the Great Leap Forward and the first stage of the "People's Commune" movement.[2] There is some disagreement over the extent to which agricultural supply was a

[1] More elementary forms of agricultural collectivization were initiated in the early 1950s, with farmers first organized into "mutual aid teams," each consisting of a handful of households, and then organized into "elementary" agriculture cooperatives that included as many as several dozen farmers. In this early stage of collectivization, farmers were able to maintain private ownership over both land and livestock.

[2] Scholars sometimes consider the famine's onset to coincide with the beginning of the Great Leap Forward in 1958, or to end in 1962 (Brown, 2012; Hsiung and Wang, 2019). Our findings are not sensitive to the exact timing.

main driver of the famine.[3] There is a wider consensus that excessive grain procurement by the central government, which impoverished some otherwise productive agricultural areas, played a substantial role (see Kung and Lin (2003), Li and Yang (2005),Padró i Miquel et al. (2015) and Bernstein (1984) among many others).[4] We show that clan-dominated areas experienced *more* severe famine conditions during the Great Famine (though this result is only marginally statistically significant in most specifications), despite having reported grain output that was at least as high as non-clan areas.[5]

Having documented these empirical regularities, we consider potential explanations for what are, at least superficially, conflicting findings. Much prior work has emphasized the role of kinship in promoting cooperation. In our context, the most direct implication is that kinship should promote a socially more efficient response to collective action problems. This explanation applies in a straightforward way to our findings on agricultural collectivization: clans served to reinforce a government policy that shifted organization toward the type of collective ownership that was already relatively prevalent in clan-based villages.

However, the most straightforward kinship-as-cooperation theory is at odds with clans' poor response to the Great Famine. A more complete view of the Great Leap Forward policies that contributed to the famine provide several possible mechanisms through which clan-dominated areas might have suffered more, which are informative with respect to the interaction of clan allegiances and the state-building efforts of the Communist government. While we cannot offer a definitive interpretation, we discuss in Section 5 historical accounts that offer two main possibilities.

Our favored explanation is based on clan organization as a means of social control. Greif and Tabellini (2017) among others observe that clans may also offer an effective means of control by elites, who would deploy clan leaders to ensure that their members adhered to state policies. In this "elite control" view of clans, under advanced collectivization clan leaders coerced their members to surrender their livestock to the collective even if it ran counter to individual members' interests (but may have served the agenda of the clan leader, who needed to stay in the good graces of government officials). Similarly, during the Great Famine, clan leaders may have been more effective in coercing villagers to surrender their food under state procurement policies, even if these ill-guided policies led to severe food shortages.[6] Beyond our qualitative evidence, we also present very tentative findings – based on province-level data from 19 provinces – consistent with higher procurement in clan-dominated areas.

It is also possible that the disruption of village-level governance during the period we study – which in turn upended the traditional role of kinship in clan-dominated areas – could account for the patterns

[3]Li and Yang (2005), Kung and Lin (2003) and Lin and Yang (2000) partly ascribe the Great Famine to a decline in output, while Meng et al. (2015) and Padró i Miquel et al. (2015) argue that China's agricultural output was sufficient to feed its population during this period and that the famine was instead the result of bureaucratic misallocation.

[4]As we discuss further below, there are various explanations for why the central government failed to reduce its grain procurement quotas during the Great Famine. Meng et al. (2015) ascribe it to inflexible and progressive government procurement policies, while Kung and Chen (2011) attribute it to the promotion incentives of provincial leaders.

[5]This contrasts with the result of Cao et al. (2022), who report that clan-dominated areas had *lower* mortality rates. Unfortunately, since we do not have access to their data, we cannot make a direct comparison to better understand how this difference in findings emerges. However, when we describe our results below, we speculate on the reasons we think account for our very different findings, even when we measure clan strength solely based on genealogical records.

[6]In this regard, it is important to note that, while the Communist Party removed or indeed killed many elite clan leaders, the village may still have been clan-controlled to the extent that the new leader came from a local family with acceptable "peasant" origins.

we document. First, as noted above, clan leaders had traditionally served as the interface between the central government and the village. However, this role was disrupted in a number of ways during the Great Leap Forward. Most importantly, outsiders – who tended to be particularly diligent in following central government directives (Meng et al., 2015) – were often brought in to govern communities. To the extent that this occurred more often in clan-dominated areas (in part as a way of reducing clan influence), these regions may have suffered more because of more rigorous enforcement of government procurement. In this alternative "elite control" view, it is disproportionate elite turnover in clan-dominated areas that potentially accounts for our results.

Finally, the government reorganization which accompanied the Great Leap Forward may account for our results. Historically, the village had served as the organizing unit for an individual clan, which made decisions on matters such as production and redistribution.[7] During the Great Leap Forward, decision-making was centralized at the level of the People's Commune (Chen et al., 1993), which were generally comprised of many villages. This centralization process put multiple – potentially competing – clans under a single administrative unit, creating conflicts that were previously absent, which could in turn account for the relatively poor outcomes in clan-dominated areas during the famine years.

It is beyond the scope of our paper to trace out the particular mechanisms that account for the contrasting results during collectivization versus the Great Famine (though in our discussion below, we discuss the relative plausibility of various classes of explanations). Yet the very different outcomes for clan-based communities in response to these two events emphasize the nuanced interplay between relational institutions such as clans and the formal governmental bureaucracies that they are embedded within.

Our paper contributes to a sizeable literature that explores the consequences of kinship institutions for a range of outcomes, in China and elsewhere. Earlier contributions tended to rely on either narrowly-bracketed or very small datasets at one extreme (e.g., Peng, 2004) and at the other, broad aggregate cross-country analyses (Enke, 2019). We follow in a more recent tradition, focused most often on studying the role of Chinese kinship organizations, in utilizing within-country variation in the prevalence of kinship ties to study their consequences (e.g., Padró i Miquel et al., 2015; Cao et al., 2022; Cheng et al., 2021). We also contribute to our understanding of the consequences of two major historical episodes in Chinese history that warrant careful analysis in their own right. We build on the agenda set forth by Meng et al. (2015), who find that procurement policies exacerbated the famine's effects, to show that enforcement of these policies may in turn have been impacted by social institutions; we similarly show a role for social institutions in determining the effects of collectivization, thus extending the findings of Chen and Lan (2017). Our work is related most directly to Cao et al. (2022), who report that clan-dominated areas fared *better* during the Great Famine. We use a broader measure of clan strength, which in part accounts for these differences, but given the stark difference in findings we discuss the comparison further below.

The rest of the paper is organized as follows. In Section 2, we provide a brief overview of Chinese clan culture, and how, in particular, clans interacted with the Chinese bureaucracy throughout history. Section 3 provides an overview of the data, including some discussion of potential proxies for clan prevalence. Section 4 provides our results on kinship ties and outcomes under agricultural collectivization

[7]It was possible for a single village to incorporate several clans – see Wang, 2007; Wang and Chen, 2004.

and also the Great Famine, while Section 5 provides a discussion of potential explanations for our Great Famine results as well as concluding comments.

2 Historical and institutional background

We begin with a brief overview of the roots of clan culture in China, as well as their organization in the early years of the People's Republic of China (PRC), followed by historical background on the collectivization movement up to the Great Famine.

2.1 Clan culture in China

Since a basic understanding of the clan's role in rural China is essential to interpreting our results, in this section we provide a brief overview.[8] The *tsu*, or clan, is essentially a group based on kinship, historically rooted in a geographic location. Most often, the locality associated with a clan was a village, settled "at some remote and generally unascertainable time in the dim past" (Smith, 1899). Village households, bound by kinship and a common surname, could trace their ancestry to this common root.

Central features of clan culture

Common ancestry was (and is) the defining feature of clans (Freedman, 1966), which further had several core elements that reflect – in an observable way – the strength of clan connections: ancestral halls, clan-owned land, and genealogies.[9] Ancestral halls, or *zongci*, were constructed as a gathering place where clan members could honor their ancestors, and were most commonly named for a particular individual, for example the founder of the clan's lineage. Ancestral halls served as the center of clan culture and rituals. The construction and maintenance of these often elaborate structures came via collective contribution, and in particular from earnings derived from clan-owned land, a second feature of clan society. The land itself derived from, for example, plots left behind by heirless clan members, or plots that were purchased and then donated by wealthy members.

The joint ownership of landholdings is just one element of clan organization that encouraged a sense of mutual support and cooperation. For example, beyond maintenance of ancestral halls, proceeds from clan-owned land was used to finance subsidies and other forms of assistance (e.g., tuition subsidies) to elderly and otherwise needy clansmen, and for local infrastructure such as road construction and maintenance. Clans also undertook various forms of joint production, building irrigation ditches, water storage reservoirs, and bridges. These undertakings required – and fostered – cooperation and a sense of trust.

Also relevant for our measurement of clan strength, such cooperation may have been more important in geographies where production was a collective rather than individual undertaking. As emphasized

[8] Naturally, it is beyond the scope of our paper to provide a comprehensive discussion of the topic; we refer the interested reader to Chapter 8 of Hsiao (1960) for an introduction that focuses on many of the features of clan culture that are particularly salient for our analysis. This section draws in large part on Hsiao's excellent introduction to the origins of clans and their role in rural life.

[9] Our discussion in this section is far from exhaustive – we focus on clan features that have particular relevance for our empirical settings and interpreting our results.

by Talhelm et al. (2014) and Noblit (2021), among others, because rice production involves irrigation – a public good – and highly intensive labor inputs, it requires greater collective effort relative to wheat farming. As Noblit (2021) in particular argues, this led to stronger clan cultures in areas suitable for rice-growing.

A final feature of clan culture was the compilation of genealogies, which served to trace and record genealogical ties. These records could be quite elaborate, with descriptions and histories of eminent members (as well as a noting of infamous clansmen, who were struck from the genealogy with red ink) and an accounting of clan property, rules, and activities.

Each of these characteristics – whether a cause or result of clan strength – suggest measures that we may use to proxy for clan strength: ancestral halls; genealogies; and rice suitability. We will discuss these measures in more detail in Section 3.

Clan leadership and government control

In clan-dominated villages there was little distinction between clan versus village organization. The clan leadership also led the village. It was thus possible to maintain control of the community via clan leaders, and the imperial government encouraged clan solidarity to facilitate control of rural areas in this way. To take just one example, clans were required to register their members, and were held collectively responsible for disciplining any lawbreakers in their midst. Such delegation of responsibility and control gave the clans legal status, which they enjoyed well into the twentieth century. The relationship certainly had its frictions, as clan interests did not always align with those of imperial governments, with larger clans threatening to undermine imperial control and legitimacy.[10] For the most part, however, clan leaders and imperial governments had a symbiotic existence, with the government conferring legitimacy on local elites, and local elites ensuring local compliance with the law. This latter point will be crucial for interpreting our results.

While in some areas clans were already in decline as early as the eighteenth century, they came under much harsher attack following the creation of the PRC in 1949. Mao Zedong had long been critical of the clan system, which he saw as a vestige of feudalism to be replaced through class struggle. In the early 1950s, ancestral halls and clan-owned lands were confiscated and redistributed to the poor, which in large part put an end to collective worship activities, though many rituals, such as visiting ancestral graves, continued. A further dismantling of the clans occurred under collectivization, with the creation of local administrative units that reported to the Communist Party of China. The new communes were often led by outsiders which served to further disrupt clan rule (see, e.g., Skinner's (1964) classic account of villages in Liaoning Province and Yu (2001) for similar accounts for rural Hunan) – again a point we return to in discussing our empirical findings.

However, kinship networks survived the Great Leap Forward (see, e.g., Xiao, 2001, for a book-length discussion of clans in the period following the Great Leap Forward). Ancestral halls were rehabilitated,

[10]The increased power that came with size led to multiple kinship groups from disparate geographies, which happened to share a surname, to band together as a single "clan." This in turn let to some unfortunate instances of like-named charlatans going from village to village collecting funds from distinct clans – which shared a surname but had different kinship roots – with the object of building a larger temple for the combined entity, and skimming off funds for themselves from both the principal as well as further funds collected for the maintenance of the hall.

and even amidst the Great Famine there is ample anecdotal evidence of within-kinship cooperation, with clan members, for example, favoring their own kin in disbursing food from public canteens.

North versus south, and the potential determinants of clan organizations

Much has been written about why clans dominated villages in some areas and were non-existent in others. The most striking pattern is that clans were far less common in northern China – see Freedman (1966), Wittfogel (1938),Goode (1982), and Huang (1985) among many others for references. Many explanations have been put forth to explain the north-south difference in clan domination, and the topic lies well beyond the scope of our paper. It has been attributed to, among various related factors, some of the geographic and societal features we have already discussed as clan proxies – notably the prevalence of rice versus wheat cultivation (see Noblit, 2021) in the south versus the north respectively, and economic prosperity (which was required to sustain ancestral halls; see, e.g., Faure, 2007). A further distinct factor involved the waves of northern invaders that led to migration from villages in the north, thus disrupting clan-based communities (Hsiao, 1960). Given the range of theories on southern clan prevalence and the empirical observation that clans were indeed far more common in the south, we use this as a fourth and final measure in constructing our overall proxy for clan strength.

2.2 Agricultural policy and agricultural production, 1949-1961

Agricultural Collectivization, 1949-1957

At the founding of the PRC, land and livestock were distributed from landowners to the poor. The country's leaders, however, found that small-scale farms produced insufficient output to fuel their industrial ambitions.[11] These concerns led to collectivization campaigns to better exploit economies of scale in agriculture. We refer the interested reader to Chen and Lan (2017) for additional details on the waves of collectivization that occurred between 1952 and 1957, and provide only a subset of relevant details here. In particular, collectivization began with the organization of so-called Mutual Aid Teams (MATs) in 1952, which were comprised of a handful of households engaged in capital- and labor-sharing arrangements. Individual farmers, however, still owned their land and livestock. Mao Zedong aimed to accelerate production gains through the fuller socialization of agriculture in 1955, via "advanced cooperatives" that were comprised of many more households per cooperative, relative to MATs, and which required, effectively, that land and livestock be surrendered to the collective for minimal compensation. As documented by Chen and Lan (2017), formation of advanced cooperatives occurred rapidly during 1955-1957: in mid-1955, there were only 500 advanced cooperatives nationwide; this number rose to 753,000, covering 96 percent of rural households, by the end of 1957. Chen and Lan (2017) provide empirical evidence showing that such collectivizations led to rural households killing their animals to avoid confiscation (with nominal compensation) by the government.

Kinship was a double-edged sword in the collectivization process. On the one hand, to the extent that production units overlapped with the clan-based aggregations, kinship may have facilitated cooperation –

[11] At the time, Communist Party leaders also aimed to mimic the collectivization that had been followed earlier by the Soviets under Stalin.

after all, clans had operated as a sort of collective for centuries, with mutual aid and common ownership of ancestral land. Thus, villages with stronger clan cultures may have been less concerned with the advent of advanced collectivization and hence be less likely to kill off their livestock. Yet this prediction is far from certain, given the sometimes imperfect overlap between the geographic coverage of individual clans versus the coverage of advanced cooperatives: Tang (2021), in particular, describes how multiple clans were sometimes forced to interact within a single cooperative, causing inter-clan frictions (though as we explain below, this problem became much worse with the advent of People's Communes). And even within clans, collectivization was the source of considerable tension as, by design, collectivization aimed to disrupt class differences that were a dominant feature of clan organization.

The Great Famine (1959-1961) and the initial development of People's Communes

We follow China's official timing of the famine to include the years 1959-1961. There were various factors that are believed to have contributed to the famine, which led to as many as 30 million million deaths (Meng et al., 2015). Most obviously, there were declines in food production. Whether this decline was due to natural disasters or a diversion of resources from agriculture to industrial production is a matter of some debate. According to Meng et al. (2015) and others, however, there was sufficient production to feed the nation. They instead pin the blame on excessive procurement. Specifically, local officials had an incentive to over-report output to please their superiors and advance their careers; this led to excessive procurement from normally productive areas which, when combined with production shortfalls, caused famine in areas of relatively high agricultural output.

In September 1958, the Politburo issued, "A decision to establish People's Communes in rural China" as a part of the "Great Leap Forward." Within a month of the edict, almost all township governments in China were replaced by "People's Communes" (*RenMin GongShe* in Pinyin). These changes were a milestone in China's rural history, as for the first time, the central government could directly control rural areas and, more specifically, directly control the villages that had been self-governed for over a thousand years, largely removed from state authority (Huang, 1990). Under the People's Communes administrative system, several villages were now reorganized into a battle company (with the original village now called a platoon).[12] Going along with the battlefield terminology, the organization of agricultural production was similarly "militarized." All laborers within a commune were organized in groupings of regiment—battalion—company—platoon.[13]

While communes were similar to advanced cooperatives in some ways, one crucial distinction was their level of geographic aggregation. Advanced agricultural cooperatives, as discussed above, were largely supervised at the village-level, whereas the People's Communes spanned a number of villages. Whereas the number of advanced cooperatives totaled 753,000 nationwide by the end of 1957 (comparable to the number of villages), the total number of People's Communes was 23,630 in 1958, so that a commune encompassed about thirty times as many members, on average. A commune thus unavoidably introduced multiple clans from multiple villages, which had the potential to create inter-clan conflicts (indeed, some

[12]The battle company became the lowest administrative unit, with the company commander serving as the leader of several villages.

[13]Communal dining was another feature of this system, although it was abandoned soon due to lack of food.

of these conflicts pre-dated the PRC).

The development of People's Communes allowed the state to control villages directly, as a result of their merger into battle companies, which were in turn controlled by the central government. Villages now acted on the instruction of the communes – the lowest level of formal government – which substantially changed the power structure within the village. While clan and/or local elites controlled the village even during the early 1950s, this control was surrendered under the People's Commune system. For example, any villager could be mobilized at any time to work on collective projects like "backyard furnaces" (as a response to Mao's "steel fever") or building dams and irrigation networks (Zhang, 2016). Capital was similarly at the disposal of the communes, which could turn agricultural land into a factory or backyard furnace simply with an executive order. The Great Famine served as a corrective to these changes, as it led, in 1962, to a decision from the Politburo (China's highest authority) to return production and distribution decisions to the village level.

The consequences of a strong clan presence during this time was ambiguous, as in the preceding period of advanced collectivization. As emphasized by Cao et al. (2022), the social capital embedded in clan structures could have mitigated the effects of the food shortages that arose under the commune system. Alternatively, as we describe in the introduction, a number of factors may have led to more severe famine in the presence of clans – most notably, if clans served as a means of tighter social control, compliance with misguided government procurement policies may have been higher in clan-dominated areas, thus leading to even greater starvation and suffering.

3 Data

3.1 Measures of clan culture

We use four variables to proxy for clan culture at the county and prefecture level, motivated by the discussion in the previous section.

The first reflects the number of ancestral halls in a county that have been selected as a "Major Historical and Cultural Site Protected at the National Level" (often abbreviated as *Guobao*, or "nationally protected"), or a "Major Historical and Cultural Site Protected at the Provincial Level" (abbreviated as *Shengbao*, or"provincially protected") by the end of 2014. To receive one of these designations, a site must have significant historical, artistic or scientific value, certified and approved by the State Administration of Cultural Heritage.

Our first step is to compile lists of nationally and provincially protected sites.[14] We then identify ancestral halls as follows. First, we search for "ZongCi" or "JiaMiao" which are ancestral halls in Chinese, as well as words associated with ancestral halls, including "JianZhu" (architectural complex), "Tang" (hall), "GuMinJu" (ancient dwellings), and "ShuYuan" (academy).[15] After the first round of filtering,

[14]The list of "Major Historical and Cultural Site Protected at the National Level" can be found at http://www.wenbozaixian.com/portal/show_room/companylist (accessed on December 15, 2022). Each provincial government has announced the list of "Major Historical and Cultural Site Protected at the Provincial Level" on its official website in batches. For example, Shandong Province announced the sixth batch at http://www.shandong.gov.cn/art/2022/4/7/art_100623_40325.html (accessed on December 15, 2022).

[15]For the last of these, we note that large ancestral halls were often used as educational institutions.

we manually check each item to ensure, via Baidu searches, that they are actually ancestral halls. In total, we identify 722 ancestral halls protected either at the national or provincial level.[16] Just under 17 percent of counties have at least one recognized hall, but conditional on having at least one, nearly 40 percent have two or more.[17] We use the natural log of one plus the number of recognized ancestral halls as our measure of clan strength to account for both intensive and extensive margins, without putting too much weight on extreme values.

We collected genealogy book information at the county level from Chinese Family Tree Database which has been maintained by the Huazhong Normal University since 2019.[18] This database had collected 17,723 genealogies (120,893 volumes in total) by June 2018, which its compilers claim is is the largest Chinese genealogical database in the world (though it grows continuously).[19] We follow Cao et al. (2022) in using genealogies per 10,000 residents (in 1952), also using the log (of one plus) transformation because of the very long right tail in the distribution of genealogies.

The third variable we use to proxy for the strength of clan culture is suitability for growing rice versus wheat. We utilize a prefecture-level variable which captures environmental suitability for growing wetland rice, provided by Talhelm et al. (2014), based on the United Nations Food and Agriculture Organization's Global Agro-ecological Zones database.

Our final measure is an indicator variable for whether a province is located in the south of China. As discussed in Section 2.1, southern China has historically had much stronger clan cultures than the north (for relevant references, see, e.g., Freedman, 1966; Wittfogel, 1938; Goode, 1982; Huang, 1985). We rely on the classification of Tang and Zhao (2023) for demarcating north from south.[20]

All of these variables have been discussed in past work as a correlate (whether as cause or effect) of clan strength, and we see each as having advantages. Rather than taking a position on any particular measure, we think of each as capturing a latent factor, clan strength, with noise, so that a natural way of modeling clan strength is via principle components analysis, to generate *ClanProxy,* which is the first principle component of these four measures, at the county level. We use this as our primary measure of clan strength throughout, though for our main analyses we present our preferred specification with each of the four measures separately.[21]

In Figure 1, we show the geographic distribution of *ClanProxy.* As expected, given that north-south

[16]Note, though, that only 650 are in counties that are in our sample, because our focus is on rural areas – as a result, our data do not include the small number of recognized ancestral halls that are in districts within a prefectural city.

[17]Lanxi county in the southern province of Zhejiang has thirteen, the largest number in our sample; there are 25 counties with five or more recognized halls.

[18]http://gd.ccnu.edu.cn/, accessed on December 15, 2022.

[19]Note that our source is different from Cao et al. (2022), who use the Shanghai Library collection instead. We see each as having distinct advantages and disadvantages. For example, while the Chinese Family Tree Database claims to have the largest collection, this is not precisely right since Shanghai Library data covers more counties in China. The Shanghai Library, however, received a disproportionate share of books from nearby provinces like Jiangsu and Zhejiang (less so from Fujian, a province with a well-recognized clan culture). In practice, the two are very highly correlated ($\rho = 0.76$ in the county cross-section), and it makes no difference for our results which we use, which is unsurprising since genealogies are the clan proxy that on its own is least important in explaining our results, as shown in Tables 2 and 3.

[20]Researchers have also used the Yangtze River as a dividing line. However, this splits some provinces that are historically viewed as "southern" – notably Jiangsu and Anhui – which straddle the Yangtze River.

[21]Unsurprisingly, all four measures are quite highly correlated. The pairwise correlations are all in the range of 0.26-0.35, with the exception of the very well-documented and strong association between rice suitability and latitude – in our data the correlation is 0.82.

is an input into our principle component analysis, there is a very clear north-south difference, but also considerable residual variation within smaller geographic areas.

3.2 Livestock population and other county-level control variables

We use as our point of departure the dataset of Chen and Lan (2017), which includes information on draft animal inventories, grain output, arable land, and population, for 1,720 counties for the years 1952 to 1957. The sample covers 77 percent of Chinese counties and 80 percent of the rural population.[22] Their data were collected from various declassified government files and reports, published compilations of statistics, and county gazetteers (see the online data appendix in Chen and Lan (2017) for details). In each province or prefectural city in China, the bureaus of statistics or of agriculture compiled detailed agricultural statistics. For a centrally planned economy in the 1950s, these statistics were indispensable to the planning committees and governments. In our regressional analysis, we additionally control for precipitation. The historical county-year level weather data for the period of 1952-1979 are taken from the Chinese Academy of Meteorological Sciences (1981).[23] The weather data use a discrete variable for rainfall: 1 for exceptional floods, 2 for limited floods, 3 for normal weather, 4 for limited droughts, and 5 for exceptional droughts. Similar to Chen and Lan (2017), we generate an indicator for counties that are officially recognized as a "revolutionary bases," which account for 10 percent of the counties in our sample, based on a list from the Ministry of Agriculture (1989). Finally, we control for the log of distance to the provincial capital.

Note that for all analyses of cattle stocks, we omit counties with missing information on arable land. Mainly, this effectively drops urban areas, though we observe in practice that none of our results are affected by the inclusion/exclusion of these observations.

3.3 Proxies for mortality during the Great Famine

We use county-year level data for 1952 to 1965 to analyze the impact of clan culture on excess mortality during the Great Famine based on birth data. Excess mortality includes deaths directly (starvation) and indirectly (economic and social impacts) caused by the Great Famine (Meng et al., 2015). Following Meng et al. (2015), we estimate the county-year level abnormal birth rate in two steps. First, we use the non-famine birth cohort sizes (i.e., excluding those born during the Great Famine) observed in the Fourth National Population Census of China in 1990 (1 percent random sample) to interpolate the counterfactual non-famine birth cohort sizes of 1959-1961 via a linear trend regression (i.e., regressing birth cohort size on year). Second, we calculate the abnormal birth cohort size for each year during 1959-1961 respectively as the ratio of the actual famine birth cohort sizes to the interpolated counterfactual non-famine birth cohort sizes. That is, if our abnormal birth cohort size variable has a value of 1.05, it means that the birth cohort size is 5 percent higher than predicted. Similarly, a value of 0.9 means birth cohort size is 10

[22] Draft animals include cows, horses, donkeys, and mules in the north; and cows and water buffalo in the south.

[23] This book provides weather information from 120 weather stations for 1952-1988. Following Chen and Lan (2017), we assign these station records to their closest counties, based on the algorithm of Thiessen polygons. This method creates a polygon around each weather station, and these non-overlapping polygons cover all counties. The counties closest to a station are the counties within the station's polygon.

percent lower than "expected." Thus a *smaller* abnormal birth cohort size value means *higher* mortality rate.

In Table 1, Panels a and b we provide summary statistics for variables used in each of our two historical settings.

4 Clan culture and local policy responses

4.1 Clan culture and collectivization

Before turning to regression analyses, in Figure 2 we show in graphical form how cattle stocks changed during 1952-1957, the period of advanced collectivization, for counties above and below the sample median for *ClanProxy*, our summary measure of clan strength. We normalize livestock to take a value of one in 1952 for each county to facilitate a comparison of pre-trends as well as post-collectivization changes. We use solid lines for counties that experienced collectivization in 1956 and dashed lines for those experiencing collectivization in 1955. In both cases, strong clan (above-median *ClanProxy*) counties are represented by the darker line.

We first observe that we replicate the Chen and Lan (2017) result of a sustained decline in livestock that coincides precisely with the beginning of advanced collectivization, whether the county-specific implementation date was 1955 or 1956. More importantly from our perspective, we also observe that this decline comes entirely from weak clan counties. Furthermore, prior to advanced collectivization, while there is some modest divergence between areas that collectivized in 1955 versus 1956, for both dates the trajectory for weak- versus strong-clan counties is near-identical prior to collectivization. That is, the divergence for regions with differing clan strength occurs in exactly 1955 in counties that initiated advanced collectivization in that year, while the bifurcation occurs a year later for counties where collectivization was initiated in 1956.

While the lack of any pre-collectivization divergence is reassuring, in Appendix Tables A1 and A2 we consider whether other county attributes – most notably proxies for clan prevalence – are correlated with the timing of advanced collectivization, or with livestock levels immediately preceding collectivization (which were normalized out in Figure 2). In specifications both with and without controls, we find that *ClanProxy*, is uncorrelated with the timing of advanced collectivization or with pre-collectivization livestock. (We also show the relationship for each component of *ClanProxy*; genealogies are a strong negative predictor of pre-collectivization livestock populations and also predict timing of collectivization. None of the other measures are at all correlated with these outcomes.)

We now turn to a regression analyses that look at the relationship Figure 2, controlling for a range of other factors that might potentially affect collectivization responses. Our specifications take the following form:

$$log(Livestock_{cy}) = \beta * Collectivization_{cy} * ClanProxy_c + Controls_{cy} + \upsilon_c + \gamma_y + \varepsilon_{cy} \qquad (1)$$

where $Collectivization$ is an indicator variable for whether a county is in the post-collectivization period; we include county and year fixed effects in all specifications. Note that while $ClanProxy$ varies at the

county-level, we think of clan strength 'assignment' as at a higher level of geographic aggregation. We take a conservative approach with standard errors throughout, clustering at the province-level. Naturally, the direct effect of *Clan* is absorbed by county fixed effects. We are interested in how county attributes, possibly correlated with clan strength, might affect the response to collectivization. To control for these county characteristics we thus also include the interaction of *Collectivization* with each of: (a) a set of weather fixed effects (from severe drought to severe flood); (b) whether a county served as a revolutionary base; (c) the fraction of a county's population that is from a minority (i.e., non-Han); (d) distance to the provincial capital.

We present results based on Equation (1) in columns (1) and (2) of Table 2. In the first column, we include only county and year fixed effects. The coefficient on the interaction term *Collectivization*ClanProxy* is 0.025 ($p < 0.001$).[24] The interquartile range for *ClanProxy* is -1.47 – 1.05, so that the coefficient of 0.025 implies that a county at the 25th percentile of clan strength experienced a decline in the stock of cattle of about 6.5 percent relative to a county at the 75th percentile. In the second column we include the full set of controls described above; the point estimate on *Collectivization*ClanProxy* is unchanged. In columns (3) – (6), we present specifications in which we measure clan strengths by each of the four separate clan proxies that are used to construct our summary measure. The interaction of each variable with collectivization is positive, and significant at least at the 10 percent level.

While we cannot rule out the possibility of further unobserved differences across high- versus low-clan areas, we further assess the robustness of our results to using propensity score methods that focus on a sample of "treated" and "control" counties with divergent values of *ClanProxy*, which would nonetheless be predicted to be comparable based on initial (pre-collectivization) attributes. Specifically, we predict *ClanProxy* using baseline values of livestock, population, grain output, and arable land, and use this to create a matched sample of high versus low (above versus below median) clan areas. We present results using this matched sample in Appendix Table A3, and find that the results are very similar to those we report in our main analysis.

4.2 Clan culture and the Great Famine

As in the preceding section, we begin in Figure 3 by showing a mortality proxy over time for counties above and below the sample median for *ClanProxy*, our summary measure of clan strength. As discussed in Section 3, we follow Meng et al. (2015) in proxying for mortality by cohort size. Thus, in the figure, lower values reflect *higher* mortality. We plot *CohortSize* during 1952-1965, normalizing birth cohort to 1 in 1952 for each county. We first observe that, as expected, cohort sizes drop markedly in 1958 and again in 1959, recovering only in 1962 at the famine's end. Of greater relevance for the current discussion, while high-clan counties have marginally lower (normalized) cohort sizes even through the early-to-mid 1950s, a bifurcation between high- and low-clan areas appears only in 1958 – by some accounts the onset of the famine – and further widens in 1959. The gap between the two then disappears in 1962. That is, the famine had a greater adverse impact in clan-prevalent areas.

[24] The coefficient on *Collectivization*, -0.45, is identical to the one presented in Chen and Lan (2017), which is as expected given that the mean of *ClanProxy* is, by construction, close to zero.

It is natural to consider whether this pattern results from food shortfalls in the low- versus high-clan areas. We use reported grain output as the outcome in Appendix Figure A1, and observe that reported harvests are relatively *high* in during the famine in strong-clan areas.[25] As we discuss in Section 2, it is likely that grain output was distorted, perhaps even more so during the Great Leap Forward, as part of local leaders' efforts to curry favor with central government officials. Cao et al. (2022) argue that these reports do convey relevant information about actual harvests. However, we will discuss in the next section several explanations for why leaders in clan-dominated areas may have inflated reported output to a greater degree than their counterparts in areas with weaker clan presence (e.g., as a result of of clan leaders' efforts to maintain their status, which was under threat as a result of the Communist Party's agenda of rooting out clan culture and hierarchy).

As in the previous section, we now turn to a regression framework to explore the determinants of famine severity. The specification we employ is the same as in Equation (1), but with $CohortSize$ as the dependent variable and for the years 1952-1965. The variable of interest is $Famine * ClanProxy$. We report these results in Table 3. As in our earlier regression results, we begin with specifications that include only county and year fixed effects. In both cases the interaction term is significant at the 10 percent level. To provide a sense of magnitude, for low-clan (below-median) areas, *CohortSize* fell from 0.98 during 1952-57 to 0.74 during the famine years of 1959-1961. High-clan areas also had a baseline cohort size of 0.98 in the pre-famine period, so that the coefficient on the interaction term *Famine*ClanProxy* of -0.029 in column (2) implies a that the decline in *CohortSize* would be expected to be about a third greater in strong-clan counties (given that the gap between the 25th and 75th percentile values of *ClanProxy* is about 2.5). We also observe that we have taken a very conservative approach in calculating standard errors, doing so at the province-level; across all specifications, the standard errors are notably smaller if we use prefecture-level clustering.

As in our analysis of collectivization, in the remainder of the table, we show results comparable to those in column (2) for each of our four clan proxies separately. For three of the four, we observe a relative decline in cohort size in strong-clan areas (though the effect is statistically significant only for the north-south comparison). The coefficient on *Famine*Genealogy* is positive (i.e., the opposite sign of other interaction terms), though very small in magnitude (the standard deviation in our genealogy measure is about 0.056) and does not approach statistical significance.[26]

Given the contrast with earlier work, we discuss here some potential explanations for our different findings. Specifically, Cao et al. (2022) find that stronger clan areas suffered less during the famine, using a genealogy-based measure of clan strength. That is in line with our own results when we use genealogy to proxy for clan presence. However, whereas they document a highly robust and statistically significant link, we find only an extremely weak and fragile association. We see our more conservative approach to inference as the main explanation for this difference – when we follow their approach of clustering at the county level, our standard errors are reduced seven-fold, which leads to a highly significant coefficient on *Famine*Genealogy*. However, the fact that standard errors are so drastically affected by the choice

[25] We also show the geographic distribution of average famine severity during 1959-1961 in Appendix Figure A2, for comparison to the distribution of clan strength shown in Figure 1.

[26] Again we repeat our analysis focusing on a sample generated using propensity score matching, and find very similar results to those in our main analysis. See Appendix Table A4.

of clustering is itself an indication that the appropriate level of independent variation, or 'assignment,' is very likely at a higher level of aggregation. It is possible that our approach is overly conservative. If we were to cluster by prefecture, however, we still find that the coefficient on *Famine*Genealogy* is statistically insignificant (though the patterns using our *ClanProxy* measure are then very highly significant, with $p < 0.001$ for the coefficient on the interaction term *Famine*Genealogy*).

This raises the question that we have sidestepped to this point, of whether some clan proxies are more credible than others. Ex ante we would not necessarily have presumed genealogies as any less credible a measure of clan presence. But we can speculate – given the findings here – on why this particular proxy is associated with different (or in the previous analysis, weaker) patterns than the other three. One possible explanation is that, relative to geographic measures or even ancestral halls, genealogies may have been more subject to destruction. While it is true that during the land reforms of 1949-1952, ancestral halls were confiscated, they generally were not destroyed and so later could later be restored and preserved as historical monuments. Genealogies, however, may have been destroyed in larger numbers during 1950-70, in ways that are further correlated with the severity of famine – if an area was harder hit, there are plausibly less likely to be genealogies available today.

In Appendix Table A5 we repeat the analysis from Table 3 column (2) for three alternative measures of famine severity. These include a distinct measure of abnormal birth rate from Chen et al. (2022) as an inverse proxy for mortality ($Cohort\_Alt$); a measure in which we use Meng et al. (2015) to predict birth cohort size, and then use the percentage difference between predicted and actual birth cohort size as a proxy for (abnormal) mortality rate ($Mortality\_1$); and finally a measure in which we use a shorter (6 year) window around the Great Famine (i.e., 1956-1958 and 1962-1964) to estimate the linear trend that is then deployed to generate a predicted mortality rate ($Mortality\_2$). This narrower window avoids the possible influence of other events (e.g., the Cultural Revolution) that might have also impacted cohort size. We obtain results that are very similar to those observed with $CohortSize$ as the dependent variable for $Cohort\_Alt$; for the two mortality measures – which are *increasing* in famine severity – we obtain marginally significant results ($p < 0.10$) in the direction of more severe famine in high-clan areas.

5 Discussion and interpretation

To briefly recap our results, we document (a) that livestock declined less in areas with strong clan ties during collectivization; and (b) deaths were *higher* in areas with strong clan ties during the Great Famine.

The first of these is readily explained by a very standard view of the clan as an informal governance mechanism that facilitates trust and cooperation – in areas with strong clan ties, communities already in essence operated as a collective, even prior to formal collectivization, so that farmers were more willing to turn their property over to communal ownership. A straightforward extension of this reasoning to the Great Famine would have suggested a better outcome for areas with strong clan connections, which would be better able to shield the vulnerable from starvation. The fact that we observe the opposite suggests, at a minimum, a more nuanced role of clans in local governance, and one that interacted with specific institutions and circumstance at a given point in time. Our analysis does not allow us to focus in on any single explanation for this pattern, but we may speculate on which explanation or collection

of explanations most plausibly contributed to the patterns we document

5.1 Clans as a means of elite control

As we observed in Section 2.1, the clan served as a primary means by which imperial governments exercised control over rural areas. This was partly a function of clan cohesion – clan leadership could more directly implement government directives than any outside functionary. But this power was also amplified by specific government policies, such as the collective responsibility shared by the entire clan for any lawbreakers in their midst. To the extent that the overlap of clan and government leadership led to greater elite control in clan-dominated areas, any government policy – whether collectivization or grain procurement – could be implemented more effectively. Clan leaders may have grown particularly zealous in enforcing government directives during the Great Famine – given Mao's enthusiasm for reducing or eliminating clan influence, they would justifiably be insecure about their positions, and hence eager to follow central government mandates. This may be seen as a particularly stark case of clan leaders' need to satisfy central government officials, whose support had long been the source of legal status and resources for clan leadership (Hsiao, 1960), but had now become increasingly uncertain (Lewis, 1963; Wang and Chen, 2004; Lu, 2008). We view this "clans as elite control" view as the most straightforward unifying theory for our two sets of results.

While it does not necessarily adjudicate amongst clan-as-control versus other possibilities described below, we present some tentative evidence in favor of explanations based on compliance with harmful government directives in general. Specifically, we explore whether grain procurement rates were particularly high in clan-dominated areas. As emphasize by Meng et al. (2015), excessive central procurement was a primary contributor to local famine severity. Unfortunately, such data are only available at the province-level and only for 19 provinces, so our analysis should be viewed as preliminary at best. Following Meng et al. (2015), we use grain retention per capita as our primary measure of procurement compliance. This is defined as the per capita difference between procurement and production in kilograms (we also use their transformed measure, $log(5 + RetentionPC)$ which reduces the influence of outliers, without creating missing values). We take the mean value of $ClanProxy$ by province, and consider the following specification:

$$RetentionPC_{py} = \beta * ClanProxy_p * Famine_y + \gamma_p + \upsilon_y + \varepsilon_{py} \tag{2}$$

where γ_p and υ_y are fixed effects for province and year respectively, thus absorbing the direct effects of $ClanProxy$ and $Famine$. The coefficient β captures the differential effect of famine on retention in high- versus low-clan areas. To account for the small number of clusters, we report in square brackets p-values based on wild bootstrap standard errors. These results appear in Appendix Table A6. We observe a negative coefficient on $ClanProxy * Famine$, indicating that less output was kept on net by provinces with a high clan presence during the Great Famine, consistent with government policy differentially impacting high clan areas during these years. The results are near-identical whether we use $RetentionPC$ or its log transformation as the dependent variable. Note that the wild bootstrap p-values imply that this result is not significant at conventional levels ($p = 0.12$ in both cases), possibly because of

the highly aggregated nature of the analysis; as such, these results should be interpreted with particular caution.[27]

5.2 Clans, outside leadership, and elite control

A variant on the elite control view – one with substantially distinct implications for the role of clan leadership during periods of adversity – is that high-clan areas were more apt to be governed by outsiders during the famine, as a result of deliberate efforts by Mao and the Communist Party to disrupt clan rule (again, see the discussion in Section 2.1). To the extent that cadres brought in from outside dealt more harshly with the local population and/or were more apt to follow government directives as closely as possible, this alternative take on "elite control" could similarly account for worse outcomes in clan-dominated areas.

Anecdotal accounts suggest at least some role for this alternative explanation. Chen (2022), for example, provides a study of excess mortality in a clan-dominated village, Sanbaoli Village of Xuancheng County in Anhui Province. This village had a population of 296 spread amongst 61 households in 1958. By November 1960, 220 villagers had died of starvation and 10 households had lost all of their family members. As is emphasized by Chen (2022), one of the most important reasons for Sanbaoli's tragic experience is that an outsider was sent to lead the (militarily organized) "battle company," which was composed of Sanbaoli and two other villages. The leader acted with indifference toward the villagers and their suffering. Even at the famine's peak, he continued to exaggerate agricultural output, and forbade villagers from leaving to seek a better chance of survival. He was strict in his treatment of citizens within his battle company, threatening harsh punishment for stealing food from the collective (thus reducing output available to be procured by the central government) or for eating unripened crops.[28]

We observe, however, that some strong assumptions are required for this to account for a *disproportionate* suffering in high-clan areas during the famine. Crucially, it requires that outside leaders were more likely to be brought in to manage clan-dominated areas. This is possible, if it were one means of dismantling clan influence, but without systematic and detailed information on the birth places of local leaders before and during the famine, it is impossible to make this assessment.

5.3 Disruption of village-based governance

The state-building efforts of the Communist Party may have had a disproportionately negative impact on clan-dominated areas, particularly to the extent that these efforts targeted clans themselves. First, administrative reorganization often brought multiple clans under the same administrative umbrella, which may have exacerbated extant inter-clan conflicts. As Duara (1991) notes, clan conflicts were amplified when resources increasingly came under control of the state, leading to arguments over the incidence of tax burden as well as control over amenities like schools. This expansion of the state-funded public sphere (and the clan conflicts that were thus generated) began much earlier. However, the conditions of

[27] Note that this result is the opposite of what Cao et al. (2022) find in their analysis of procurement, which focuses on genealogies as a measure of clan strength.

[28] Consumption of unripe crops was an important survival tool during the famine. See, e.g., the account of a village in Henan Province by Thaxton Jr (2008).

extreme scarcity that characterized the Great Famine could surely have exacerbated pre-existing tensions between clans, and the creation of larger administrative units in the form of People's Communes forced larger clan groupings into direct interaction with one another. This shift plausibly led to more and wider-reaching inter-clan conflicts than had existed prior to the Great Leap Forward. Furthermore, to the extent that clans facilitated social cooperation, their dismantling may have led to disproportionate suffering during the famine. Again, however, we note that this view requires that, for example, non-clan areas developed alternative institutions that served a similar function to clans, but were less disrupted by upheaval during the Great Leap Forward.

We conclude by noting that the above arguments by no means rules out a positive role for clans dring the famine, but rather that the negative effects dominated. And overall , we see our results as emphasizing the possibly offsetting consequences of clan-based organization, and the importance of considering the broader social and historical context in considering their impact.

References

Bernstein, T. P. (1984). Stalinism, famine, and chinese peasants. *theory and Society 13*(3), 339–377.

Brown, C. D. (2012). China's great leap forward. *Education About Asia 17*(3).

Cao, J., Y. Xu, and C. Zhang (2022). Clans and calamity: How social capital saved lives during china's great famine. *Journal of Development Economics 157*, 102865.

Chen, J., J. Chen, and X. Yang (1993). *Social and Economic Changes in Rural China: 1949-1989*. Shanxi Economic Press.

Chen, S. and X. Lan (2017). There will be killing: Collectivization and death of draft animals. *American Economic Journal: Applied Economics 9*(4), 58–77.

Chen, S., F. Xinyu, and W. Xuanyi (2022). Starving and deceiving: Famine experience, politician risk preference, and gdp manipulation in china.

Chen, Y. (2022, February). Political power, clan, and survival: Sanbaoli village of xuancheng county in anhui province during the great famine. *Twenty-First Century 189*(2), 79–98.

Cheng, J., Y. Dai, S. Lin, and H. Ye (2021). Clan culture and family ownership concentration: Evidence from china. *China Economic Review 70*, 101692.

Chinese Academy of Meteorological Sciences (1981). *Atlas of Drought and Flood Distribution in China in the Past 500 Years*. China Map Press.

Duara, P. (1991). *Culture, power, and the state: rural North China, 1900-1942*. Stanford University Press.

Enke, B. (2019). Kinship, cooperation, and the evolution of moral systems. *The Quarterly Journal of Economics 134*(2), 953–1019.

Faure, D. (2007). *Emperor and ancestor: State and lineage in South China*. Stanford University Press.

Fei, J. C. and T.-J. Liu (1982). The growth and decline of chinese family clans. *Journal of Interdisciplinary history*, 375–408.

Freedman, M. (1966). *Chinese Lineage and Society :Fukien and Kwangtung*. Athlone Press.

Goode, W. J. (1982). *The Family*. Pearson.

Greif, A. and G. Tabellini (2010). Cultural and institutional bifurcation: China and europe compared. *American economic review 100*(2), 135–40.

Greif, A. and G. Tabellini (2017). The clan and the corporation: Sustaining cooperation in china and europe. *Journal of Comparative Economics 45*(1), 1–35.

Hsiao, K.-c. (1960). Rural china; imperial control in the nineteenth century.

Hsiung, P.-C. and Y. Wang (2019). Unmasking china's great leap forward and great famine (1958-1962) through shunkouliu. *Qualitative Inquiry 25*(8), 811–821.

Huang, P. (1985). *The peasant economy and social change in North China*. Stanford University Press.

Huang, P. C. (1990). *The peasant family and rural development in the Yangzi Delta, 1350-1988*. Stanford University Press.

Kung, J. K.-S. and S. Chen (2011). The tragedy of the nomenklatura: Career incentives and political radicalism during china's great leap famine. *American Political Science Review 105*(1), 27–45.

Kung, J. K.-s. and J. Y. Lin (2003). The causes of china's great leap famine, 1959–1961. *Economic Development and Cultural Change 52*(1), 51–73.

Lewis, J. W. (1963). The leadership doctrine of the chinese communist party: the lesson of the people's commune. *Asian Survey*, 457–464.

Li, W. and D. T. Yang (2005). The great leap forward: Anatomy of a central planning disaster. *Journal of Political Economy 113*(4), 840–877.

Lin, J. Y. and D. T. Yang (2000). Food availability, entitlements and the chinese famine of 1959–61. *The Economic Journal 110*(460), 136–158.

Lu, H. (2008). How "satellites" were launched:grass-roots cadres in rural china and the great leap forward. *Open Times 5*, 116–126.

Meng, X., N. Qian, and P. Yared (2015). The institutional causes of china's great famine, 1959–1961. *The Review of Economic Studies 82*(4), 1568–1611.

Ministry of Agriculture (1989). *Encyclopedia of Statistics of the Rural Economy in China, 1949-1986*. Beijing: Agricultural Publishing House.

Noblit, G. (2021). The origin and evolution of chinese lineages.

Padró i Miquel, G., N. Qian, Y. Xu, and Y. Yao (2015). Making democracy work: Culture, social capital and elections in china. Technical report, National Bureau of Economic Research.

Peng, Y. (2004). Kinship networks and entrepreneurs in china's transitional economy. *American Journal of Sociology 109*(5), 1045–1074.

Skinner, G. W. (1964). Marketing and social structure in rural china, part i. *The Journal of Asian Studies 24*(1), 3–43.

Smith, A. H. (1899). *Village life in China: A study in sociology*. Revell.

Talhelm, T., X. Zhang, S. Oishi, C. Shimin, D. Duan, X. Lan, and S. Kitayama (2014). Large-scale psychological differences within china explained by rice versus wheat agriculture. *Science 344*(6184), 603–608.

Tang, C. and Z. Zhao (2023). Informal institution meets child development: Clan culture and child labor in china. *Journal of Comparative Economics 51*(1), 277–294.

Tang, Z. (2021). *Local Clan Communities in Rural China: Revolution and Urbanisation Since the Late Qing Dynasty*. Routledge.

Thaxton Jr, R. A. (2008). *Catastrophe and Contention in Rural China: Mao's Great Leap Forward Famine and the Origins of Righteous Resistance in Da Fo Village*. Cambridge University Press.

Wang, L. (2007). The peopleÊŸs commune movement and the transformation of social base of the chinese countryside. *CPC History Studies*.

Wang, S. and Y. Chen (2004). From kinship to citizenship : A study of rural lineages in anhui under the people's republic. *Social Sciences in China (Chinese) 1*, 180–193.

Wittfogel, K. A. (1938). *New Light on Chinese Society: An Investigation of China's Socio-economic Structure*. International secretariat, Institute of Pacific relations.

Xiao, T. (2001). *Chinese Patriarchal Clans in the Process of Village Governance*. Shanghai Bookstore Publishing House.

Yu, J. (2001). *Yuecun Politics: Changes in China's Rural Political Structure in the Transitional Period*. The Commercial Press.

Zhang, L. (2016). *Say goodbye to ideals: Research on the People's Commune System*. Shanghai People's Publishing House.

Table 1a: Summary Statistics, Sample for Collectivization Analysis

Variable Name	Mean	Min	Max	StdDev	Observations
$log(Livestock)$	10.233	6.375	12.122	0.795	8441
$ClanProxy$	-0.091	-2.157	5.602	1.516	7349
$South$	0.473	0.000	1.000	0.499	8441
$log(1 + Halls)$	0.158	0.000	2.639	0.391	8441
$log(1 + GenePC)$	0.070	0.000	1.762	0.182	8243
$RiceSuitability$	0.360	-0.101	1.092	0.260	7469
$Collectivization$	0.394	0.000	1.000	0.489	8441
$Minority$	0.130	0.000	1.000	0.336	8441
$RevolutionBase$	0.115	0.000	1.000	0.319	8441
$log(DistancetoCapital)$	4.900	0.000	7.201	1.007	8441

Table 1b: Summary Statistics, Sample for Famine Analysis

Variable Name	Mean	Min	Max	StdDev	Observations
$CohortSize$	1.015	0.000	2.880	0.284	28294
$Famine$	0.214	0.000	1.000	0.410	28294
$ClanProxy$	0.008	-2.157	6.248	1.505	21924
$South$	0.522	0.000	1.000	0.500	28294
$log(1 + Halls)$	0.162	0.000	2.639	0.398	28294
$log(1 + GenePC)$	0.077	0.000	1.827	0.208	24444
$RiceSuitability$	0.353	-0.101	1.092	0.243	25144
$Collectivization$	0.740	0.000	1.000	0.439	21686
$Minority$	0.147	0.000	1.000	0.354	21686
$RevolutionBase$	0.105	0.000	1.000	0.307	26236

Notes: Table 1a provides summary statistics by county for the years 1952-1957. Table 1b provides summary statistics by county for the years 1952-1965. Variable definitions for Table 1a: $ClanProxy$ is the first principle component of four clan proxies, also listed here: $South$, which denotes provinces in the south of China based on Tang and Zhao (2023); $log(1 + Halls)$, which measures the number of the presence of a provincially- or nationally-recognized ancestral temple in a county; $log(1 + GenePC)$, which measures the number of genealogy books for clans in the county; $RiceSuitability$, which captures the climatic and soil suitability for growing rice (rather than wheat); Additional variables include $Collectivization$ denotes whether a county is in the post-collectivization period. $Minority$ is the fraction of non-Han population. $RevolutionBase$ denotes that the county served as a revolutionary base. $DistancetoCapital$ is the distance in kilometers from the county to the provincial capital. Additional variable definition for Table 1b: $CohortSize$ is the county's birth cohort size as calculated by Meng et al (2015).

Table 2: Collectivization, Clan Culture, and Livestock Populations

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable			$log(Livestock)$			
Collectivization	-0.044***	-0.161***	-0.217***	-0.166***	-0.168***	-0.215***
	(0.014)	(0.035)	(0.033)	(0.033)	(0.035)	(0.038)
$Collectivization * ClanProxy$	0.026***	0.026***				
	(0.007)	(0.006)				
$Collectivization * South$			0.101***			
			(0.020)			
$Collectivization * log(1 + Halls)$				0.034**		
				(0.015)		
$Collectivization * log(1 + GenePC)$					0.067*	
					(0.033)	
$Collectivization * RiceSuitability$						0.143***
						(0.029)
$Collectivization * Minority$		0.065***	0.060***	0.078***	0.082***	0.061***
		(0.014)	(0.016)	(0.024)	(0.024)	(0.015)
$Collectivization * RevolutionBase$		0.035**	0.039**	0.065***	0.061***	0.041***
		(0.016)	(0.016)	(0.018)	(0.017)	(0.014)
$Collectivization * log(DistancetoCapital)$		0.014**	0.014***	0.010*	0.011*	0.014**
		(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
County FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Weather FEs		Yes	Yes	Yes	Yes	Yes
Weather Interactions		Yes	Yes	Yes	Yes	Yes
Observations	7349	6623	7517	7517	7379	6701
R-Squared	.985	.986	.984	.983	.983	.986

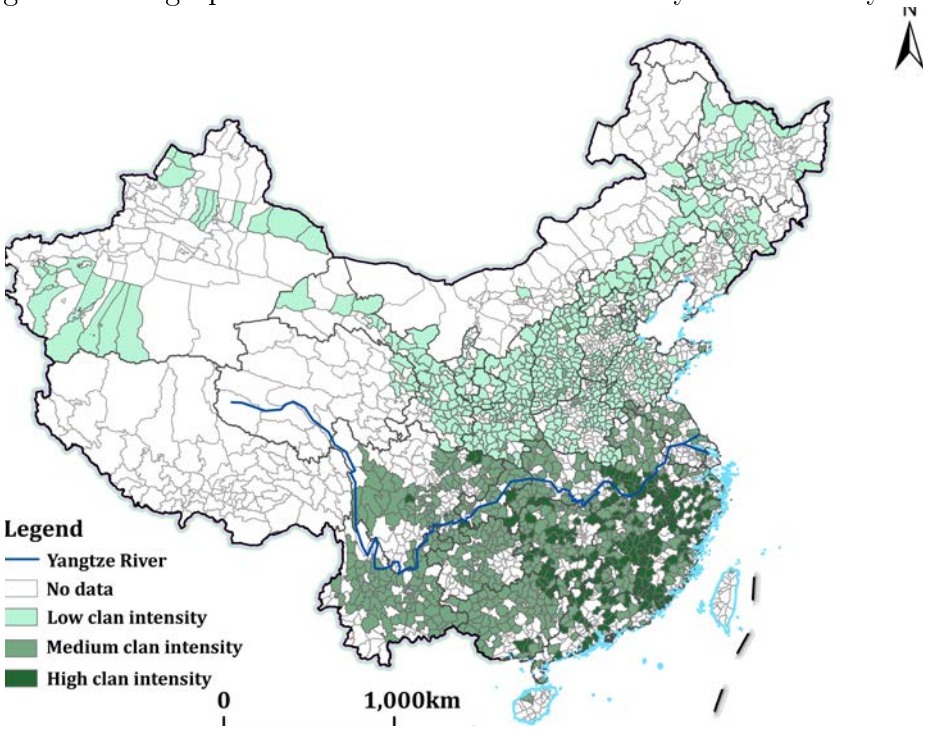
Notes: The sample covers the 1408 counties for which there is data on arable land, for the years 1952 to 1957. The dependent variable in all columns is $log(Livestock)$, the natural logarithm of the county's livestock population. $ClanProxy$ is the first principle component of four clan proxies, listed here: $South$, which denotes provinces in the south of China based on Tang and Zhao (2023); $log(1 + Halls)$, which measures the number of the presence of a provincially- or nationally-recognized ancestral temple in a county; $RiceSuitability$, which captures the climatic and soil suitability for growing rice (rather than wheat). See the notes to Table 1 for detailed definitions of the control variables. Standard errors clustered by province in all regressions.
Significance: * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Famine, Clan Culture, and Birth Cohort Size

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable			*CohortSize*			
$Famine * ClanProxy$	-0.036*	-0.029*				
	(0.018)	(0.017)				
$Famine * South$			-0.144**			
			(0.054)			
$Famine * log(1 + Halls)$				-0.038		
				(0.026)		
$Famine * log(1 + GenePC)$					0.025	
					(0.056)	
$Famine * RiceSuitability$						-0.166
						(0.100)
$Famine * Minority$		-0.026	0.012	-0.025	-0.031	-0.024
		(0.032)	(0.034)	(0.048)	(0.045)	(0.031)
$Famine * RevolutionBase$		0.036	0.036	-0.002	-0.013	0.032
		(0.029)	(0.028)	(0.030)	(0.033)	(0.029)
$Famine * log(DistancetoCapital)$		0.007	0.004	0.008	0.010	0.005
		(0.009)	(0.006)	(0.008)	(0.009)	(0.008)
County FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Weather FEs		Yes	Yes	Yes	Yes	Yes
Weather FE X Famine Interactions		Yes	Yes	Yes	Yes	Yes
Observations	21924	17283	19579	19579	19173	17549
R-Squared	.765	.776	.767	.757	.76	.774

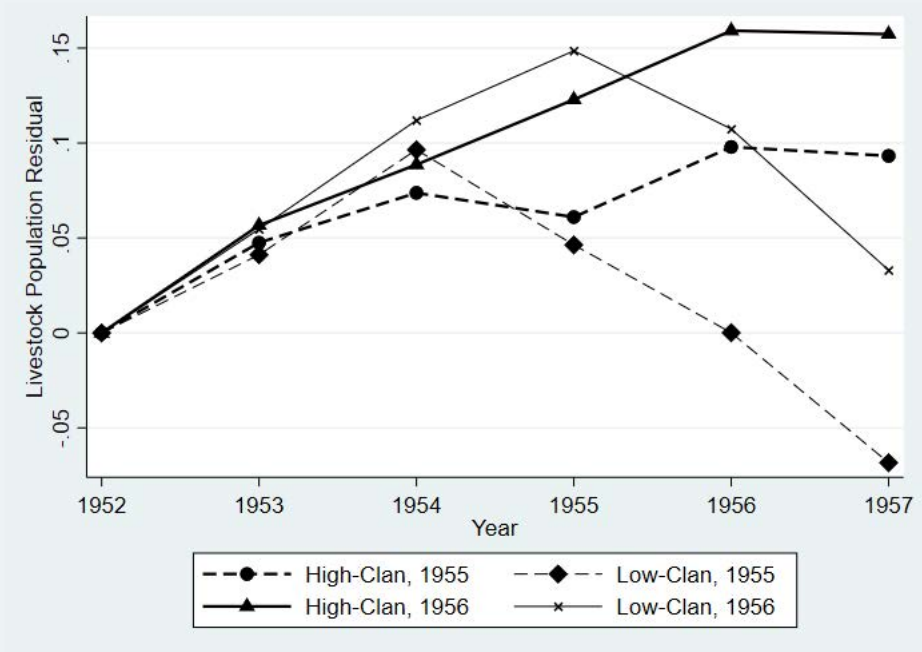
Notes: The dependent variable in all columns is *CohortSize*, which denotes the birth cohort size, as calculated by Meng et al (2015). *ClanProxy* is the first principle component of four clan proxies, listed here: *South*, which denotes provinces in the south of China based on Tang and Zhao (2023); $log(1 + Halls)$, which measures the number of the presence of a provincially- or nationally-recognized ancestral temple in a county; $log(1+GenePC)$, which measures the number of genealogy books for clans in the county; *RiceSuitability*, which captures the climatic and soil suitability for growing rice (rather than wheat). See the notes to Table 1 for detailed definitions of the control variables. Standard errors clustered by province in all regressions.
Significance: * significant at 10%; ** significant at 5%; *** significant at 1%.

Figure 1: Geographical distribution of clan intensity at the county level



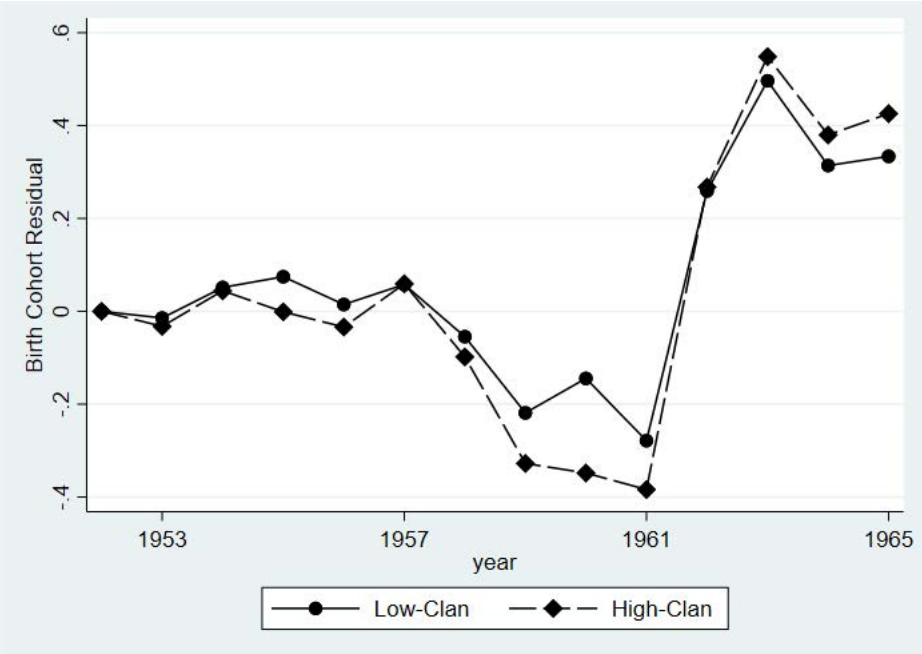
Notes: This figure shows the county-level geographical distribution of clan intensity which is measured by the first principle component of four clan proxies: Whether a province is in the south of China based on the assignment of Tang and Zhao (2023); the log of one plus the number of recognized ancestral halls in the county; the log of the number of genealogy books in the county; and rice-growing suitability. See text for details.

Figure 2: Collectivization and Livestock Population, High- versus Low-Clan Areas



Notes: This figure shows the log of draft animal populations by county during the transition to collectivization, after normalizing livestock for each county to one in 1952 to facilitate cross-county comparisons. The dashed lines are counties that collectivized in 1955 and the solid lines are those that collectivized in 1956. High- versus Low-Clan county assignment is based on a median split of the first principle component of four proxies for clan prominence: Whether a province is in the south of China based on Tang and Zhao (2023); the log of one plus the number of recognized ancestral halls in the county; the log of the number of genealogy books in the county; and rice-growing suitability. See text for details.

Figure 3: Famine and Birth Cohort Size in High- versus Low-Clan Counties



Notes: This figure shows birth cohort size by county, as calculated by Meng et al (2015). We normalize cohort size to one for all counties in 1952, to facilitate a comparison of pre-trends. High- versus Low-Clan county assignment is based on a median split of the first principle component of four proxies for clan prominence: Whether a province is in the south of China based on Tang and Zhao (2023); the log of one plus the number of recognized ancestral halls in the county; the log of the number of genealogy books in the county; and rice-growing suitability. See text for details.

Appendix Table A1: Correlates of collectivization date

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable			*Collectivization* = 1956			
ClanProxy	-0.015	-0.017				
	(0.026)	(0.039)				
South			0.031			
			(0.123)			
$log(1 + Halls)$				-0.031		
				(0.058)		
$log(1 + GenePC)$					-0.269*	
					(0.151)	
RiceSuitability						0.016
						(0.240)
log(Population)		0.062	0.021	0.041	0.047	0.033
		(0.075)	(0.092)	(0.060)	(0.059)	(0.077)
ArableLand		-0.090	-0.049	-0.070	-0.079	-0.060
		(0.069)	(0.081)	(0.058)	(0.056)	(0.078)
Minority		0.211**	0.189**	0.189**	0.176**	0.211**
		(0.090)	(0.076)	(0.075)	(0.078)	(0.091)
RevolutionBase		0.044	0.026	0.039	0.059	0.025
		(0.072)	(0.065)	(0.073)	(0.068)	(0.075)
DirectControl		0.339***	0.131	0.139	0.121	0.315**
		(0.110)	(0.096)	(0.101)	(0.099)	(0.136)
log(DistancetoCapital)		0.015	0.015	0.015	0.011	0.016
		(0.016)	(0.014)	(0.014)	(0.013)	(0.016)
Weather FEs		Yes	Yes	Yes	Yes	Yes
Observations	1350	1039	1165	1165	1162	1039
R-Squared	.00205	.118	.122	.122	.129	.116

Notes: The dependent variable in all columns is an indicator variable denoting that an area collectivized in 1956 (rather than 1955). *ClanProxy* is the first principle component of four clan proxies, listed here: *South*, which denotes provinces in the south of China based on Tang and Zhao (2023); $log(1 + Halls)$, which measures the number of the presence of a provincially- or nationally-recognized ancestral temple in a county; $log(1 + GenePC)$, which measures the number of genealogy books for clans in the county; *RiceSuitability*, which captures the climatic and soil suitability for growing rice (rather than wheat). *DirectControl* indicates where the county is directly governed by province; for other controls, see the notes to Table 1 for further description. Standard errors clustered by province in all regressions. Significance: * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A2: Pre-collectivization livestock populations

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable			$log(Livestock)$			
$ClanProxy$	-0.067	-0.004				
	(0.063)	(0.051)				
$South$			0.037			
			(0.173)			
$log(1 + Halls)$				-0.087		
				(0.083)		
$log(1 + GenePC)$					-0.746***	
					(0.193)	
$RiceSuitability$						0.301
						(0.349)
$log(Population)$		0.322***	0.238**	0.270**	0.293**	0.236**
		(0.108)	(0.101)	(0.125)	(0.118)	(0.111)
$ArableLand$		0.370***	0.403***	0.370***	0.352***	0.465***
		(0.107)	(0.084)	(0.106)	(0.103)	(0.110)
Weather FEs		Yes	Yes	Yes	Yes	Yes
Observations	1350	1038	1166	1166	1164	1038
R-Squared	.015	.401	.392	.393	.414	.406

Notes: The sample includes 1549 counties for the year 1954. The dependent variable in all columns is $log(Livestock)$, which denotes the natural logarithm of the county's livestock population. $ClanProxy$ is the first principle component of four clan proxies, listed here: $South$, which denotes provinces in the south of China based on Tang and Zhao (2023); $log(1 + Halls)$, which measures the number of the presence of a provincially- or nationally-recognized ancestral temple in a county; $log(1 + GenePC)$, which measures the number of genealogy books for clans in the county; $RiceSuitability$, which captures the climatic and soil suitability for growing rice (rather than wheat). Standard errors clustered by province in all regressions.
Significance: * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A3a: Collectivization, Clan Culture, and Livestock Populations: Propensity-score matched sample

	Unmatched	Mean				t-test		
Variable	Matched	Treated	Control	%bias	%reduct |bias|	t	p>|t|	V(T)/V(C)
log(*Livestock*)	U	10.113	10.274	-20.0		-5.91	0.000	1.13*
	M	10.241	10.246	-0.7	96.6	-0.10	0.917	1.16
log(*Population*)	U	12.378	12.083	37.1		10.92	0.000	0.66*
	M	12.315	12.31	0.6	98.3	0.09	0.925	0.97
log(*GrainOutput*)	U	4.3018	3.9267	46.6		13.74	0.000	0.64*
	M	4.1625	4.1374	3.1	93.3	0.49	0.626	0.96
ArableLand	U	10.451	10.883	-57.7		-17.04	0.000	0.92
	M	10.735	10.712	3.0	94.7	0.42	0.677	1.03

Appendix Table A3b: Collectivization, Clan Culture, and Livestock Populations: Propensity-score matched sample

	(1)	(2)
Dependent Variable	$log(Livestock)$	
$Collectivization$	-0.110***	-0.200***
	(0.021)	(0.042)
$Collectivization * HighClan$	0.102***	0.088***
	(0.020)	(0.019)
$Collectivization * Minority$		0.043**
		(0.019)
$Collectivization * RevolutionBase$		0.048
		(0.032)
$Collectivization * DistancetoCapital$		0.013
		(0.008)
County FEs	Yes	Yes
Year FEs	Yes	Yes
Weather FEs		Yes
Weather Interactions	Yes	Yes
Observations	3162	2820
R-Squared	.986	.986

Notes: The sample covers the years from 1952 to 1957. The analyses are run on a propensity-score matched sample, where we match each high-clan county (defined as above the median value of $ClanProxy$) to a low-clan county using the nearest neighbor one-to-one matching technique without replacement and setting the caliper to 0.002 so that southern counties are similar to northern counties along observable dimensions. The dependent variable in all columns is $log(Livestock)$, the natural logarithm of the county's livestock population. $HighClan$ denotes counties with above-median values of $ClanProxy$, our main measure of county-level clan strength. See the notes to Table 1 for detailed definitions of the control variables. Standard errors clustered by province in all regressions.
Significance: * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A4a: Famine, Clan Culture, and Livestock Populations: Propensity-score matched sample

	Unmatched	Mean				t-test		
Variable	Matched	Treated	Control	%bias	%reduct |bias|	t	p>|t|	V(T)/V(C)
log(*Livestock*)	U	10.154	10.279	-15.5		-6.48	0.000	1.13*
	M	10.305	10.293	1.5	90.2	0.35	0.726	1.32*
log(*Population*)	U	12.407	12.118	36.5		15.20	0.000	0.67*
	M	12.355	12.382	-3.4	90.6	-0.79	0.431	1.04
log(*GrainOutput*)	U	4.3686	3.9719	49.6		20.66	0.000	0.63*
	M	4.2397	4.2439	-0.5	98.9	-0.13	0.899	0.98
ArableLand	U	10.466	10.891	-56.9		-23.77	0.000	0.93*
	M	10.748	10.751	-0.4	99.3	-0.09	0.927	1.09
CohortSize	U	0.9884	0.9822	6.0		2.51	0.012	0.92*
	M	0.9923	0.9936	-1.2	79.9	-0.27	0.785	1.17*

Appendix Table A4b: Famine, Clan Culture, and Birth Cohort Size: Propensity score matched sample

	(1)	(2)
Dependent Variable	*CohortSize*	
$Famine * HighClan$	-0.092	-0.107***
	(0.056)	(0.013)
$Famine * Minority$		-0.011
		(0.019)
$Famine * RevolutionBase$		0.029
		(0.021)
$Famine * DistancetoCapital$		-0.004
		(0.006)
County FEs	Yes	Yes
Year FEs	Yes	Yes
Weather FEs	Yes	Yes
Weather FE X Famine Interactions	Yes	Yes
Observations	10136	9309
R-Squared	.776	.78

Notes: The analyses are run on a propensity-score matched sample, where we match each high-clan county (defined as above the median value of $ClanProxy$) to a low-clan county using the nearest neighbor one-to-one matching technique without replacement and setting the caliper to 0.008 so that southern counties are similar to northern counties along observable dimensions. The dependent variable is $CohortSize$, which denotes the birth cohort size, as calculated by Meng et al (2015). $HighClan$ denotes counties with above-median values of $ClanProxy$, our main measure of county-level clan strength. See the notes to Table 1 for detailed definitions of the control variables. Standard errors clustered by province in all regressions.
Significance: * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A5: Alternative measure of famine severity

	(1)	(2)	(3)
Dependent Variable	$CohortSize\_Alt$	$Mortality\_1$	$Mortality\_2$
$Famine * ClanProxy$	-0.030	0.024*	0.026*
	(0.018)	(0.014)	(0.013)
$Famine * Minority$	-0.048	0.019	0.016
	(0.036)	(0.026)	(0.029)
$Famine * RevolutionBase$	0.030	-0.030	-0.028
	(0.031)	(0.024)	(0.023)
$Famine * log(DistancetoCapital)$	0.007	-0.006	-0.006
	(0.009)	(0.007)	(0.007)
County FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
Weather FEs	Yes	Yes	Yes
Weather FE X Famine Interactions	Yes	Yes	Yes
Observations	17283	17283	17283
R-Squared	.768	.727	.866

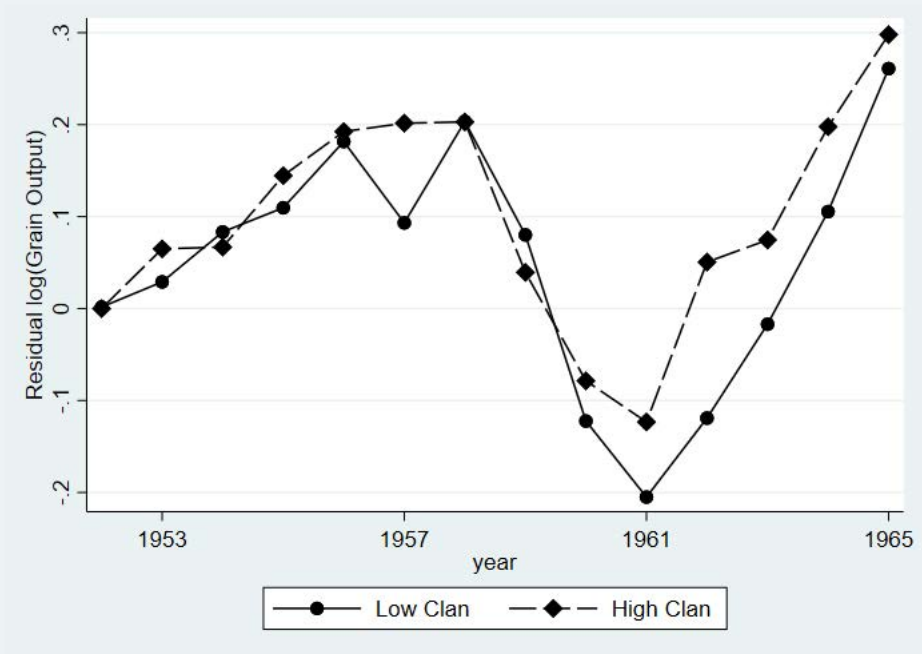
Notes: We use three dependent variables, each of which reflects famine severity. $CohortSize\_Alt$ is an alternative proxy for birth cohort size, from Chen et al (2022). $Mortality\_1$ and $Mortality\_2$ are proxies for excess mortality–see text for additional discussion.$ClanProxy$ is the first principle component of four clan proxies, listed here: $South$, which denotes provinces in the south of China based on Tang and Zhao (2023); $log(1 + Halls)$, which measures the number of the presence of a provincially- or nationally-recognized ancestral temple in a county; $log(1 + GenePC)$, which measures the number of genealogy books for clans in the county;$RiceSuitability$, which captures the climatic and soil suitability for growing rice (rather than wheat).See the notes to Table 1 for detailed definitions of the control variables. Standard errors clustered by province in all regressions.
Significance: * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Table A6: Clan Strength and Grain Procurement during the Great Famine

	(1)	(2)
Dependent Variable	$RetentionPC$	$log(5 + RetentionPC)$
$ClanProxy * Famine$	-31.051	-0.090
	(19.111)	(0.066)
Bootstrap P	[.1]	[.135]
County FEs	Yes	Yes
Year FEs	Yes	Yes
Weather FEs	Yes	Yes
Weather FE X Famine Interactions	Yes	Yes
Observations	234	234
R-Squared	.825	.47

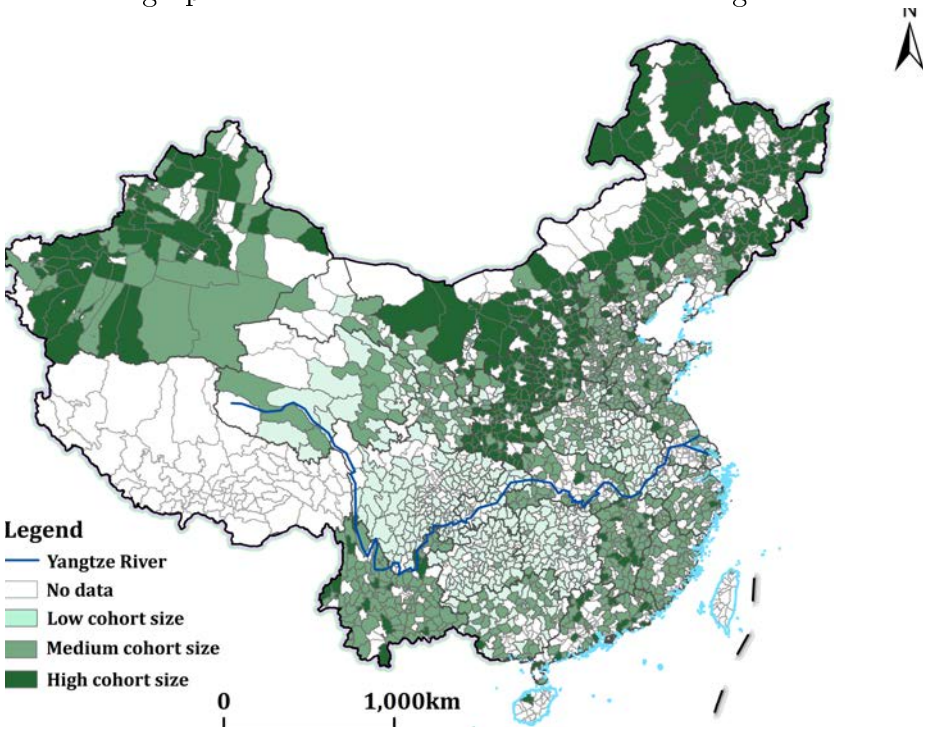
Notes: Sample includes 19 provinces for the years 1952-1965. $RetentionPC$ is the difference in procurement and production in kilograms per person; see Meng et al, 2015 for details. Both specifications include province and year fixed effects. Standard errors clustered by province in curved brackets. Square brackets contain p-values calculated using Wild bootstrap-t to account for the small number of clusters. $ClanProxy$ is the first principle component of four clan proxies, listed here: $South$, which denotes provinces in the south of China based on Tang and Zhao (2023); $log(1 + Halls)$, which measures the number of the presence of a provincially- or nationally-recognized ancestral temple in a county; $log(1 + GenePC)$, which measures the number of genealogy books for clans in the county; $RiceSuitability$, which captures the climatic and soil suitability for growing rice (rather than wheat);
Significance: * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix Figure A1: Famine and Grain Output in High- versus Low-Clan Counties



Notes: This figure shows the log of grain production by county. We normalize output to one for all counties in 1952, to facilitate a comparison of pre-trends. High- versus Low-Clan county assignment is based on a median split of the first principle component of four proxies for clan prominence: Whether a province is in the south of China based on Tang and Zhao (2023); the log of one plus the number of recognized ancestral halls in the county; the log of the number of genealogy books in the county; and rice-growing suitability. See text for details.

Figure A2: Geographical Distribution of Cohort Size during the Great Famine



Notes: This figure shows the county-level geographical distribution of the mean value of *CohortSize* during the Great Famine period, which is the county's birth cohort size as calculated by Meng et al (2015). See text for details.