The Origin and Evolution of Chinese Lineages*

Graham Noblit

2022-03-20

Abstract

I aim to understand variation in an important and historically novel socio-political institution, the Chinese lineage. There is extensive geographic variation in the historical prominence and relevance of lineages. Using ethnographic and historical-economic evidence, I construct a theory explaining lineages as risk-pooling institutions, which provide lineage members with access to land. More so, variation in regional demand for risk-pooling and/or access to land likely stems from well-studied rice-wheat agro-economic differences. I test this hypothesis by examining whether lineage activity is associated with landholding size, precipitation predictability, and historically documented precipitation disasters and find support for my hypothesis.

Contents

The Origin and Evolution of Chinese Lineages **2**

- Introduction 2
- Theory 6
- Data 10
- Results 22
- Discussion 36

Appendix **37**

- Continuous Suitability Measures 37
- Both Rice and Dry-Crop Suitability 40
- Alternative Rice Suitability Measure 44
- Prefectural Data and Endogeneity 47

References **50**

List of Figures

1 Distribution of household farm-size. 7
2 Spatial distribution of Chinese lineage activity. 12
3 Scatterplot of lineage activity and household arable land. 13
4 Spatial distribution of paddy-rice suitability. 14
5 Spatial variation in precipitation predictability measure. 16
6 Spatial variation in county-level drought and flood disasters between 1644 and 1795 CE. . . . 17
7 Pairwise relationship between lineage activity and precipitation predictability. 25
8 Interaction effect between precipitation predictability and rice suitability on lineage activity. . 29
9 Scatterplot of precipitation predictability measure using data before and after 1961. 30
10 Interaction effect between precipitation variation (standard deviation) and rice suitability on lineage activity. 40

*This work benefitted from comments by Song Chen, Fiery Cushman, Joseph Henrich, Dan Kelly, Nathan Nunn, Christina Rogers, and Thomas Talhelm.

11 Spatial variation in rice suitability (IHS) 41
12 Spatial variation in dry-crop suitability. 42
13 Spatial variation in GAEZ paddy-rice suitability measure. 46
14 Distribution of lineage activity (IHS) when measured at county and prefecture levels. 49

List of Tables

1 County Level Variables 19
2 Community Level Variables 20
3 Precipitation Predictability Verification 21
4 CHIPS Farm Size 22
5 CFPS Lineage Activity 23
6 Lineage Activity and Precipitation Predictability 27
7 Endogeneity Results 31
8 REACHES Historical Disasters 34
9 Continuous Suitability Measure 38
10 Disaggregated Suitability Measure 43
11 GAEZ Rice Suitability Measure 44
12 Endogeneity Results 47

The Origin and Evolution of Chinese Lineages

Introduction

The human organism is the only known species whose social life is normatively structured (Boyd, Richerson, & Henrich, 2011; Searle, 1995). Scientists then cannot understand human minds, bodies, environments, or behaviors without reference to their institutional context (Curtin et al., 2020; Henrich et al., 2005; Henrich, Heine, & Norenzayan, 2010; Hutchins, 1995; Talhelm et al., 2014). A broad body of research has explored how institutions impact phenotype, be it behavioral or psychological. For instance, work my Talhelm and colleagues (Talhelm et al., 2014) suggests that because paddy-rice agriculture requires intensive cooperation and coordination, individuals from historically rice growing regions of China are cognitively and behaviorally distinct from people from historically wheat growing regions. More so, such psychological differences persist even when individuals across cultural populations play laboratory based economic games with identical formal institutional rules (Ahn et al., 2016; Gächter, Herrmann, & Thöni, 2010; Herrmann, Thoni, & Gachter, 2008; Vollan, Landmann, Zhou, Hu, & Herrmann-Pillath, 2017). For example, Herrmann and colleagues exemplify this in their work on public goods where they had participants from across the globe play simple public goods games with punishment in a laboratory setting and identified cultural-level variation both in participants' willingness to punish as well as how they used punishment (Gächter et al., 2010; Herrmann et al., 2008). Such work helps to explain variation in individuals' psychological, cognitive, and behavioral propensities by referencing variation in the socio-cultural or institutional environments that individuals are embedded in. A similar vein of work exists in development-economics and economic history literature, which sometimes attempts to understand how individual behavior or aggregate-outcomes, such as GDP change, relate to institutional variation (Nunn, 2009). Findings such as these have obvious psychological and historical significance but also imply that policy success or outcomes will conditionally depend on a population's institutional and even psychological history (Ashraf, Bau, Nunn, & Voena, 2020; Nunn, 2020).

In this paper, I study how institutions impact human behavior. However, instead of exploring institutional change diachronically across ethno-linguistic groups, I instead attempt to explain the geographic distribution of an institution within a given ethnic group. Essentially, I ask why an institution, specifically the Chinese family structure which I refer to as "lineages", is more or less represented or important across regions of China.

A broad body of work examines the role of family structures on behavior and psychology. Cultural evolutionists (Barrett et al., 2016; Curtin et al., 2020; Schulz, Bahrami-Rad, Beauchamp, & Henrich, 2019), building on foundational work in anthropology (Gluckman, 1955, 1965a, 1965b; Maine, 1861), attempts to link the nature of kinship systems, whether they are more inward-looking and interdependent, labeled "intensive" or outward-looking and labeled "extensive", to a cohesive set of psychological and behavioral traits. These traits include altruistic behavior directed at strangers and the consideration of actors' intentions when making moral judgments. Such intensive kinship systems may be those defined by static social-networks where a focal agents' social partners are inherited and determined by individuals' relative positions within the network. A similar vein of research exists in the economic history literature examining how family structure impact behavior with economic consequences (Alesina & Giuliano, 2010) and political participation (Alesina & Giuliano, 2011). Other work takes a dynamic perspective and considers how distinct social structures — tight kinship institutions, labeled "clans", where cooperation among individuals is largely within a given group or urban institutions, labeled "cities" where cooperation is largely across kinship structures — impact broader institutional trajectories (Greif & Tabellini, 2010). Along a similar vein, political scientists have studied how well-developed kinship institutions impact public goods provisioning through politicians, finding that it can both drive up (Xu & Yao, 2015) and drive down (Cruz, 2020) the allocation of public goods. These divergent results can likely be explained by the fact that these studies are conducted within different national and political contexts and though they both refer to strong kinship institutions as "clans" in actuality are referring to distinct bundles of social norms characterized by unique rules and responsibilities and thus psychologies.

Background

Numerous institutions in Chinese history have ordered social, economic, and political life: Buddhist monasteries, inter-village religious associations, the state, and importantly, village associations (Chen, 2017; McDermott, 1999, 2013). Such institutions were ever-present social realities for life in China from the Song dynasty onward. To varying degrees, all of them organized local militias for defense, invested in public goods, and organized rituals (McDermott, 2013). Kinship institutions similarly have been an ever-present aspect of Chinese social life. Han Chinese kinship institutions have taken various forms across time, but generally been patrilineal, tracking descent through males (McDermott, 2013; Watson, 1982). Early in Chinese history, large communal families organized around patrilineal descent, called *yihu*, were an important kinship institution whereby all kin-members' property was communally held, albeit controlled by the head-of-household; kin-members resided in proximate or shared households; and patrilineal ancestors were worshiped collectively (McDermott, 2013: 116). Life in such large families was incredibly communal —family members dined together in large halls where they sat according to rank and gender, individuals were forbidden private property, and individuals' economic roles and behavior reflected family needs (McDermott, 2013: 117). Historians consider *yihu* to have been "self-contained [communities]," which, because of the "physical and material security" they provided members, flourished in insecure political environments (McDermott, 2013: 117). In more secure political environments where the threat of violence were less prominent, individuals or smaller-family units would likely be less willing to accept the costs of such limited autonomy.

Such communally intensive kinship organizations are rather unlike the lineage institution that emerges later in Chinese history and continues to characterize Chinese social life today. In the 11th century, perhaps because of growing economic and political stability (McDermott, 2013: 118), a prominent Song dynasty official, Fan Zhongyan, instituted reforms in his kinship group. These reforms granted individuals with private property rights, devolved control over decision-making to smaller units at the sub-lineage-, family-, or household-level, and freed lineage-members from obligatory co-residence (Chen, 2017; H. Hu, 1948; McDermott, 2013: 118). By lessening lineage-members' obligations to the group and freeing them from the intensively communal demands of *yihu*, Fan Zhongyan's reforms permitted more flexible political and economical behavior among members. Whereas *yihu* membership provided individuals with and emphasized material and physical security, Fan Zhongyan's lineages made far fewer demands of members but correspondingly offered them fewer benefits, generally "no more than a portion of their food and clothing needs, of their funeral and wedding costs, and of their other living expenses" (McDermott, 2013: 118-121). However, the kinship institution that Fan Zhongyan generated quickly changed form. Researchers note that much of the social and ritual elaboration that we identify today as associated with lineages were initially absent and were only innovated in the later Ming and Qing dynasties (Chen, 2017).

Perhaps the key difference between *yihu* and lineages is the nature of institutional property holdings. As already mentioned, *yihu* entailed communally held property managed by the head-of-household. Fan Zhongyan's lineage revolution should be understood fundamentally as a property revolution whereby, unlike as is the case in *yihu*, individuals were not required to forgo private property claims and were permitted landholdings of their own. At the level of a lineage as a whole or single sub-branch, land was also entrusted, from which individuals would receive dividends. The lineage as compared to the *yihu* then, "refocus[ed] its activity from the retention of labor to the management of a corporate estate" (Chen, 2017: 168). Note that for the several centuries following Fan Zhongyan's innovation, such corporate estate holdings would remain relatively small and mainly exist for sacrificial purposes (Chen, 2017; Ebrey, 1986). During this period, lineages largely spread among the political class to serve political interests rather than economic ones (Chen, 2017: 169). The lineage institution only spreads broadly across economic and political classes in the 15th and 16th centuries (Chen, 2017), perhaps because the Ming government unwittingly provided wealthy non-gentry individuals with an appropriate legal apparatus to construct stronger lineages when it permitted its community pact institutions to be organized around patrilineal kinship (McDermott, 1999). It is during this period, once the institution spread more broadly, that lineages increasingly adopt ritual and social functions, such as the building of ancestral halls and increased ancestral worship (Chen, 2017; Szonyi, 2002).

As lineages spread in the 15th and 16th centuries they did not only elaborate their ritual functions but also quickly came to be something like miniature and localized states. For instance, lineage organizations established systems of rules which members were obligated to conform to (H. Hu, 1948; McDermott, 2013). To stabilize these rules and social order generally, lineages additionally established judicial positions to investigate, adjudicate, and enforce rule systems, also both restricting the state from adjudication rights over lineage members as well as restricting lineage members from using state-courts (McDermott, 1999). In essence, lineages competed with the state for legal rights. Lineages also invested in violence, forming militias and seizing landholdings from surrounding competing institutional forms, such as Buddhist monasteries and temples (Chen, 2017; McDermott, 2013). More so, lineages pursued social and cultural power, heavily investing in education of its members both as a path to political power through the state bureaucracy as well as to inculcate lineage members with specific values and beliefs through moralizing lectures(McDermott, 1999, 2013).

Importantly, there is significant geographic variation in the presence, importance, and role of lineages (Cohen, 1990; Freedman, 1958; Pasternak, 1969). Historical documentation indicates that gentry members attempted to create corporate property holding lineages in North China, but unlike in South China, these attempts failed to take root and spread with a given lineage generally falling apart upon the founder's death (Chen, 2017). Lineages do exist in northern China today in that individuals understand themselves to be members of patrilineally kinship groups defined by common descent (Cohen, 1990), but unlike their southern form, northern lineages are less likely to compose the entirety of a village's population (H. Hu, 1948); to own corporate property holdings (Cohen, 1990; Huang, 1985); are less socially relevant (Huang, 1985); and in the opinion of at least one ethnographer, are less important than other village or inter-village institutions in village members' eyes (Yang, 1945). It is this precisely this geographic variation in the importance or strength of lineage institutions that I seek to characterize and understand.

Anthropologists of China have long discussed geographic variation in lineage importance or activity. Freedman most famously argued that lineages are more common in South China largely because of a connection to paddy-rice agriculture Freedman argues that paddy-rice requires irrigation and because irrigation infrastructure requires a large and cooperative labor investment some farmers may have chosen to corporately hold ensuing, developed lands rather than subdividing them among the cooperating parties (Freedman, 1966: 161). Pasternak perceptively raises the point that even if such historical events did occur they did not need to produce lineages *per se* (Pasternak, 1969). Freedman also argued that inter-ethnic conflict or at the very least a collective need for self-defense may have led agnatic relations to elaborate into a more corporate form (Freedman, 1966: 163). Pasternak again considers this argument and an analogous case of Chinese immigration to Taiwan, arguing that similarly to where Freedman studied, immigrants to Taiwan faced inter-ethnic competition over land and water. This conflict resulted partially in more elaborate corporate lineage forms but also extensive cooperation among geographically close but agnatically unrelated lineages and villages, a far cry from the inwards looking lineages that Freedman describes (Pasternak, 1969). Pasternak instead suggests that it is only after a frontier has been opened and a region initially colonized that individuals may be able to afford the more inwards looking, patrilineally focused lineage institutions Freedman studied (Pasternak, 1969).

As it stands, no successful overarching theory of Chinese lineages exists and perhaps no such theory can explain all variation in lineage importance or its institutional traits. In what follows, I offer a theory to explain geographic variation in lineage importance. This theory is not meant to explain all variation in lineage importance. For instance, I completely ignore how lineages interact with the highly commercialized economy of South China or benefited from state monopolies (McDermott, 2013; Szonyi, 2002, 2017). Instead, I ask why an institution defined by corporate property trusts spread successfully to some regions of China but not others.

Theory

Summary

For brevity and clarity, I first summarize my theory explaining lineage variation before discussing it in more detail. In both North and South China, partible inheritance practices exert a constant downwards pressure on household landholding sizes over time. The resulting small landholdings impose large costs on households, exposing them to increased risk of shocks to production or wealth. Households therefore are incentivized to either acquire additional land and/or to pool risk. But regional agro-technological and economic differences between North and South China indicate that southern households suffered from smaller landholdings than in the north, implying an increased risk of shocks and demand for new land. More so, southern households also were relatively less able to acquire access to land, either by purchasing or renting it, compared to northern households. These differences perhaps help us to understand why lineage activity varies regionally —access to incorporated land and its derivatives was in greater demanded in South China compared to North China making southern households more interested in adopting the costs associated with participating in, conforming to, and supporting lineage institutions in order to gain the benefits of land access.

Full Theory

Both northern and southern China are characterized by partible inheritance norms, whereby offspring all claim a (usually equal) portion of a household's resources upon inheritance as opposed to monogeniture, whereby a single offspring inherits all household resources and wealth (Fei & Chang, 1949; Huang, 1985; Wakefield, 1998). Notably, households split their landholdings among sons regardless of absolute landholding size, i.e. partible inheritance still occurs even if it means that sons inherit insufficiently small farms to support themselves (Wakefield, 1998). Partible inheritance, *ceteris paribus*, exerts a constant downward pressure on landholding size across all of China and therefore small landholdings should characterize both northern and souther regions.

However, all is not in fact equal and northern property holdings are larger on average than those in South China. I show this using farm data in Figure 1 where I plot the distribution of farm-size as a function of region. Though I use 21st century data, historical data from various time periods shows an identical, or more exaggerated, trend with northern landholdings being larger on average than southern ones. This regional difference results from multiple causes. North China is characterized by lower population densities and cheaper land prices (Buck, 1964) permitting households to acquire more land within a single generation than equivalent southern households (Huang, 1985, 1990). North Chinese dry-grain farming can also be centrally managed and does not especially benefit from increasing the skill level of labor inputs, whereas South Chinese paddy-rice farming cannot be centrally managed and requires skillful labor inputs for productivity increases (Bray, 1986). One ethnographer notes that in North China, a lazy male farmer can work 15 *mu* of dry-grain land a day and a hard-working one 20 *mu* (Huang, 1985: 166) far more than in paddy agriculture, which requires the coordination of labor practices among groups to meet labor demands (Bray, 1986). The incentive to centrally manage production benefited from North China's relatively non-commercialized and non-monetized economy where real wages were lower than in South China (Huang, 1985). The possibility of high yields with centrally managed and cheap labor permitted North Chinese families to directly manage large farms when they could afford to acquire land (Huang, 1985; Wakefield, 1998). Such large managerial farms produced higher returns (13 – 14% annually of the price of land) compared to the mere 5% earned renting the land out to tenants (Huang, 1985: 173). The opposite was true in South China where large landowners preferred to keep tenants on so that they would "[shoulder] all or part of the risks of production" rather than pay the large costs associated with centrally supervising production themselves (Bray, 1986: 115).

In South China, farm sizes historically were, and continue to be, smaller because of the above noted reasons as well as the following factors. First, because paddy-rice is so productive per unit area, that when combined with knowledgeable and intensive labor inputs and irrigation practices, a farm of given size will produce more calories than an equivalently sized dry-grained farm (Bray, 1986). In fact, the productivity of a paddy farm can actually benefit from smaller plot sizes as they permit the more precise irrigation and labor-intensive planting practices that release paddy-rice productivity (Bray, 1986). Second, the highly commercialized and monetized southern economy meant that households with farms too small to support their subsistence could still meet their caloric requirements by producing, and exchanging at market, cotton, sugar, and silk (Bray, 1986; Huang, 1990). Third, South China's high population densities, caused by intensive paddy-rice production, meant that land simply was split more on over time through acts of inheritance (Bray, 1986; Huang, 1990).

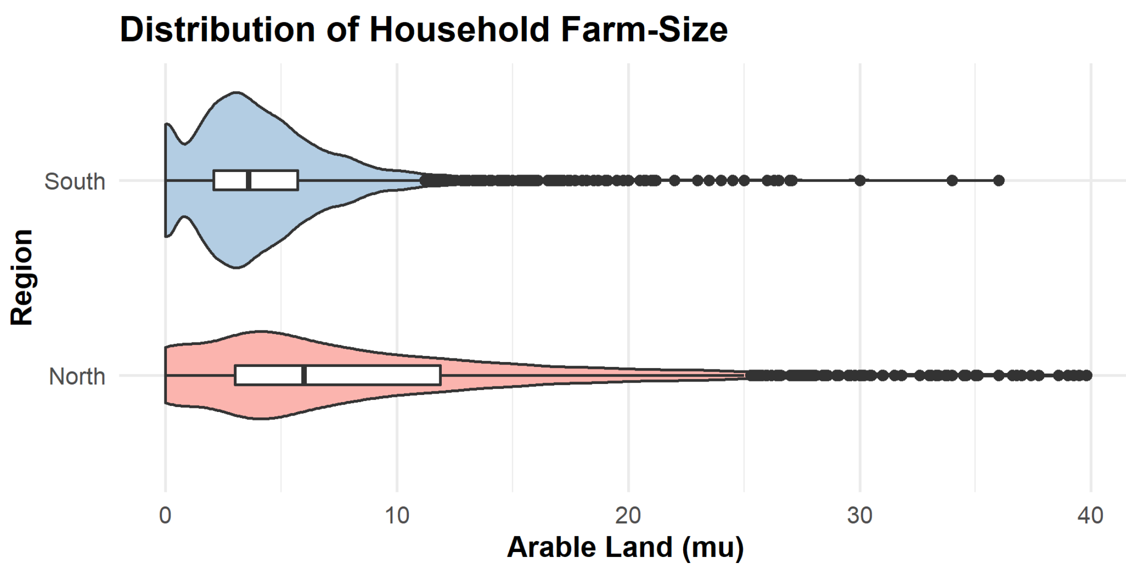

Figure 1: 1 mu is equivalent to 1/6 of an acre

Small landholdings imply significant problems for farmers, the most obvious being decreased incomes. Because plots are rarely contiguous, partible inheritance implies relatively inefficient production as farmers spend an increasing proportion of their time on fixed, non-productive costs, such as time spent traveling between plots (Huang, 1985). Most importantly, both low wealth and smaller landholdings increase households' exposure to shocks to production or wealth. The political and economic anthropologist, James Scott, provides one way to understand this risk exposure. It can be economically rational for individuals who produce at subsistence levels to focus on risk aversion rather than mean yields because such individuals cannot afford to drop below subsistence levels for any prolonged period (Scott, 1976). However, smaller farms will also experience greater risk exposure independent of productivity levels if we consider household-level and geographically circumscribed shocks such as the death of a water buffalo of which small paddy-farms generally only had one (Bray, 1986), if a son and therefore productive member of the household were sent as a soldier to the border and thus likely death (Szonyi, 2017), or certain spatially circumscribed pest infestations. These forms of risk exposure incentivizes small landowners to acquire additional landholdings or to develop institutions that pool risk.

However, regional differences also existed in how whether households could acquire new land and therefore increase farm sizes. Land prices were higher in South China compared to North China (Bray, 1986: 115). Though we should be cautious of extending data backward in time, Buck reports that land prices in the early 1900s were one-half greater (31-yuan per acre) in rice-dominant areas compared to wheat regions (Buck, 1964: 46, 332-4). The hyper-productive nature of paddy agriculture translated to increased land-valuation and rents with high population densities also increasing competition over access to both land usage and property rights (Bray, 1986; Huang, 1990). The high price of land caused a significant problem for land-strapped households. While ethnographic evidence is both rare and documents a historically late period, one ethnographer discussing a village in Southwest China indicates that even the wealthiest village resident, who owned about 25 *mu* or 4.1 acres, simply could not afford to simultaneously purchase new land and absorb any economic shock or ritual cost (marital, funerary, etc.), particularly if he hired out for labor (Fei & Chang, 1949: 128). For the household farm, "land breeds no land" (Fei & Chang, 1949: 129), indicating that in the South, small landholders are constrained in their ability to purchase new land.

Should an individual want to expand their landholdings then they could of course rent new land rather than purchase it directly. However, the nature of paddy-rice agriculture and the southern economy meant that a household could not invest an arbitrary amount of labor into newly rented land, i.e. participate in non-intensive forms of paddy agriculture with lower labor requirements. Land was taxed and landlords bought and sold rental rights seeking profits (Huang, 1990; McDermott, 2013). Landlords consciously brought on tenants who would invest intensive and skillful labor into paddy production in order to access the hyper-productive yields associated with paddy-rice. In the words of one landlord: "There is a proverb which says 'It is better to have good tenants than good land' ... There are three advantages in having good tenants, namely that they are on time with plowing and sowing, they are energetic in fertilizing, and they are resourceful in conserving every drop of water" (Bray, 2013: 70). At least in the Qing dynasty, rental contracts in South China generally were fixed amounts, imposing fixed labor costs on tenants, as opposed to the more common northern pattern of share-cropping (Zhihong, 2017: 430-431). Additional landholdings might be desirable for a small household but if their plots were too dispersed (or "too morcellated and inefficient to cultivate" (Wakefield, 1998, p. 196)), then the labor requirements of renting additional land could exceed a household's maximum labor supply budget, making the new land a net-cost.

Of course, if households could not supply their own labor then they should be able to hire waged labor. As previously discussed, such a strategy was common in North China and critical for the success of managerial farms, whereby a household would supplement its labor pool with cheaply hired and managed wage-laborers once it acquired large amounts of new land (Huang, 1985). However, wage labor for farm production was relatively rare in South China. South Chinese real wages were far higher than in the northern region because of the commercialized nature of the southern economy (Huang, 1990) and possibly because women (even in wealthier households where their feet were bound (Fei & Chang, 1949)) and children absorbed low opportunity cost labor allowing men to specialize in intensive paddy-rice production and thus fetch higher wages (Huang, 1990). For small farms in China, high wages make hired labor inaccessible and farming unprofitable (Bray, 1986: 165). Supplementing household labor with hired labor was not necessarily a feasible strategy for a household farmer considering whether it would be profitable to expand their landholdings.

Because South Chinese farms were smaller on average, we would expect there to be greater demand for access to land and/or risk pooling. However, the above discussion suggests that households in South China were less able than northern households to access new land while also being relatively more in need of that access because of smaller farm size. Correspondingly, I argue that we can understand regional variation in lineage institutional activity by asking why households were willing to accept the constraints associated with lineage participation; what benefits did they gain? Wealthy individuals attempted to instantiate corporate property institutions in both North and South China, indicating that regional variation in lineage activity cannot be explained merely by variation in supply (Chen, 2017). More so, lineages do in fact exist in North China, but are less socially, psychologically, and economically relevant and hold little corporate property (Cohen, 1990; Huang, 1985). I argue that lineages may have been demanded more in South China because they help households to gain access to land, buffering those households from shocks to wealth and production. In other words, because land was so scarce and in high demand in South China, membership in lineages was more valuable to southern households. In the words of Chinese state officials in the 1950s, "[The wealthy] establish lineage land so that the land they own can be maintained a little more securely. Middle and poor peasants do so in part because of the influence of sacrificial ideology, but much more often because their land is so limited. If they divide the land among their heirs, it will not only be insufficient to support a livelihood, but will also result in the land becoming too morcellated and inefficient to cultivate" (Wakefield, 1998, p. 196).

Lineages and corporate property will permit individuals to absorb shocks to production or wealth should lineage members be able to call on corporate resources when they require them or should households which lack land be able to access corporate landholdings. Historical and ethnographic sources indicate that lineage institutions were characterized by exactly these rules. Lineages preferentially take on particularly land-poor lineage members as tenants, often on better rental terms than those granted to non-members of the lineage (H. Hu, 1948). Additionally, lineages often maintained corporate land exclusively designated to provide dividends for poverty-relief in addition to granaries that served to counter local price-depressions, provide famine relief, and to lend grain to lineage members at low interest-rates (with the loans often being completely forgiven) (H. Hu, 1948). Conspicuously, such granaries and need-based aid were completely absent among documented northern lineages (Huang, 1985: 236). While northern lineages do occasionally possess corporate graveyards (Cohen, 1990), and lineage members might be preferentially selected for tenancy, these holdings are so small, being only an acre or two, that they are inconsequential for grain-production (Huang, 1985: 237). Similarly, rents for northern lineage members were still set at market levels (Huang, 1985: 237).

Tenancy on lineage land in South China also appears to have offered increase security of access to land. For instance, lineage tenants who were in arrears faced little risk of dispossession. Lineage treasurers were hesitant to dispossess lineage members who could not meet rent obligations because they feared political repercussions for treating lineage members harshly (Fei & Chang, 1949: 78), suggesting a general norm or ethos that lineage tenants should not be treated on as harsh of terms as non-lineage tenants. Such a norm is surprising considering that in much of South China, rental terms were particularly exploitative (Fei & Chang, 1949). Additionally, lineage land-sale rules appear designed to provide lineage members with access to land, although this may be a by-product of attempting to preserve the strength of a lineage. Lineage members generally had preferential rights to purchase the land another lineage member sought to sell; in fact contracts where an individual sold land to a buyer from outside the lineage without having first granted lineage members an opportunity to purchase the land were legally void (Fei & Chang, 1949; McDermott, 2013).

In summary, individuals across North and South China appear to have responded distinctively to the problem of risk posed by the constant downward pressure on landholding size because of household division. In North China, households appear to have attempted to expand holdings at the household-level, often producing the managerial farms discussed above. They were able to do this because of the relatively low cost of land and the ability to hire and manage labor at low wages. In South China, because land and labor prices were high, households had a difficult time expanding their holdings, exposing them to significant shocks to wealth and production. Lineage rules distribute land and resources to poor lineage members, benefits which appear to be more valuable in rice-suitable regions of the country. These differences paint a picture whereby southern households may have been willing to pay larger costs to participate in lineages, i.e. lineage access was demanded, compared to their northern equivalents.

Predictions

Should lineage function be related to risk pooling and providing access to land, we should expect certain quantitative relationships to hold. First, lineages should be more active and socially relevant in areas where land holdings are smaller. That is, in regions where access to land is constrained, individuals should demand access to lineages and be more willing to support lineages, which hold corporate land and provide lineage members with access to that land. Second, lineages should be more prevalent in regions where households face more risk to agricultural production. More so, because of rice-wheat economy differences, this association should be more extreme for rice-dominant regions because households have smaller landholdings and simultaneously are less able to acquire new land. This relationship can be tested either by examining the association between lineage activity and measures of ecological risk that would impact agricultural production, such as precipitation, or by directly examining the association between lineage activity and the historical occurrence of disasters to production.

I examine precisely these relationships. First, I consider how lineage activity, measured either historically through the presence of genealogical books or contemporaneously through family-level lineage-related behaviors, relates to farm size. Then, I consider how historical lineage activity varies across Chinese counties as a function of precipitation variation or more specifically, predictability. Precipitation disasters, whether in the form of too much or too little rain, was a major determinant of ecological risk and risks to farming productivity in Chinese history (Bray, 1986; Elvin, 2004) and is thus a determinant of the value of risk-pooling. I additionally directly estimate how lineage activity relates to a limited set of documented historical disasters.

Data

Genealogies

In order to measure historical lineage activity, I count the number of genealogical books compiled by lineages within a given region. Note that I only consider regions which consisted of a Han majority by the 19^{th} century and exclude the contemporary provinces of Xinjiang, Qinghai, Inner Mongolia, and Tibet. I take genealogical books as a proxy of the relative importance and relevance of lineages within a region because genealogies were important markers of lineage activity. Genealogies track lineage membership and serve to define lineages not only to outsiders, serving as important markers of a lineage's status, but also how lineage-members relate to one another. In fact, forgers of genealogies, such as those who sought to inflate their own status by claiming membership in a famous lineage, were tracked down by lineages and punished (H. Hu, 1948). Genealogies did not only identify lineage members, but also contained the rule systems defining lineage members' obligations, rights, and responsibilities to one another; breaking such rules earned individuals significant social, economic, and corporal sanctions (H. Hu, 1948; McDermott, 1999). I take the lack of genealogical books in a region to reflect the relative unimportance of lineages as an institution because residents of that region did not assume the costs or efforts of compiling genealogical books.

In order to count genealogies, I use of the "Comprehensive Catalogue of Chinese Genealogies" data set made available by Dincecco and Wang (Dincecco & Wang, 2020; Y. Wang, 2020). This data set provides information about over 50,000 genealogical books compiled across China between 1005 and 2007. Importantly, the data set includes the geocoded coordinates of where and the year when each genealogical book was compiled. In Figure 2, I plot my proxy measure of lineage activity, the raw count of genealogical books from the start of the Ming Dynasty to the end of the Qing Dynasty,at the county level. Note that this measure makes no distinctions between a lineage's first genealogical book and later updating of those books. Both contribute equally to my measure of lineage activity. I transform the count of genealogical books with an inverse hyperbolic sine function ($\operatorname{arsinh}(x) = \ln\left(x + \sqrt{x^2 + 1}\right)$) because the variable is highly skewed while also containing zero values.

Farm Size and Contemporary Lineage Activity

Data on farm size comes from two sources. The first is the 2002 wave of the Chinese Household Income Project (CHIPS) (Shi, 2009; Shi, Chuliang, Zhong, & Ximing, 2008). This data set records information from approximately 10,000 households spanning 1,000 villages. While the CHIPS database does not include information about lineage activity, it does permit me to match county codes which I use to aggregate household and village information to the county level. I use this information to estimate the relationship between the average farm size at the county level and historical lineage activity where farm size is defined as the amount of land a household cultivates that is either irrigated or not. In Figure 3 I plot the geographic distribution of this variable.

I additionally collect micro-level behavioral data on lineage activity. I use the 2010 wave of the China Family Panel Study (CFPS), a nationally representative sample survey funded by 985 Program of Peking University and carried out by the Institute of Social Science Survey of Peking University (described in Xie & Hu, 2014). The CFPS collects data on individuals (approximately 30,000), households (approximately 15,000), and communities and importantly includes behavioral data about ancestral worship, genealogy possession, and farm size. Unfortunately, because of the anonymized nature of these data, I cannot tie them to geographic location and thus I cannot link them to my historical lineage activity measure. In this case, my farm size variable is a community-level variable denoting each village's amount of arable land per capita (*mu* per person). This measure considers only those individuals who are registered to the village under the *hukou* domestic registration system.

Rice Suitability

Rice suitability data comes from Galor and Özak (Galor & Özak, 2016) and takes the form of a caloric suitability index. This measure aims to estimate how suitable a region is for specific crop types under

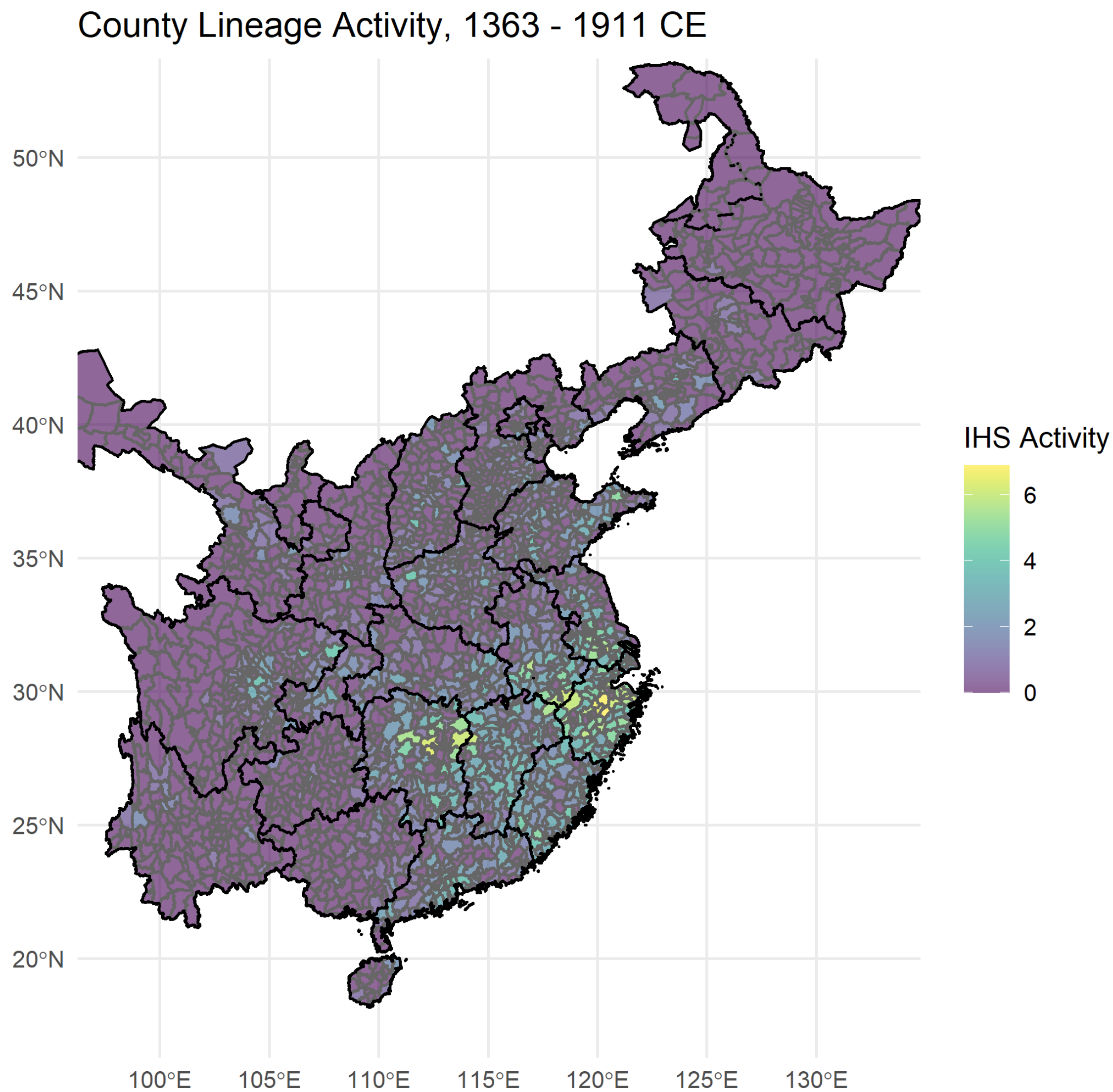

Figure 2: Spatial distribution of lineage activity at the county level. The number of lineage books in each county are transformed by an inverse hyperbolic sine transformation.

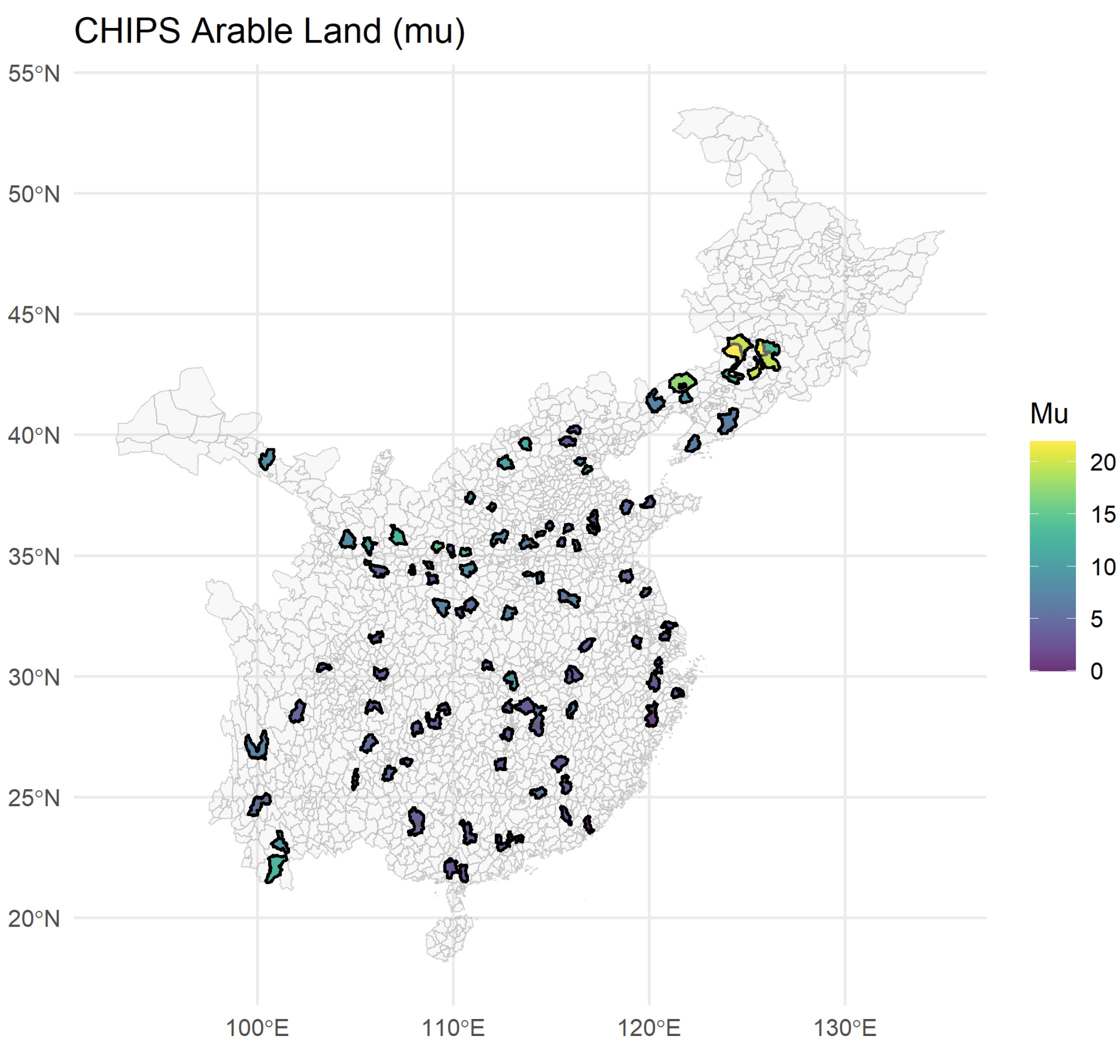

Figure 3: Scatterplot comparing arable land in mu and lineage activity. Excluded counties are shown in white.

various agricultural-input levels and irrigation schemes by estimating the number of calories attainable per hectare-year. These calculations rely on crop-suitability calculations stemming from the GAEZ FAO which consider various soil constraints such as depth, chemical composition, and drainage behavior, as well as terrain-slope constraints (Fischer, van Velthuizen, Shah, & Nachtergaele, 2002). I use Galor and Özak's low-input and rain-fed paddy-rice variables because these are meant to match historical agricultural technologies. Finally, because the suitability measure contains zeros and we expect there to be decaying marginal value in one crop type over another as the measure deviates from zero, I apply an inverse hyperbolic sine transformation. The geographic distribution of this variable is plotted in Figure 4. Finally, in Appendix I demonstrate that including dry-crop suitability metrics has no impact on the results and in Appendix that using an alternative non-caloric metric also does not change my results.

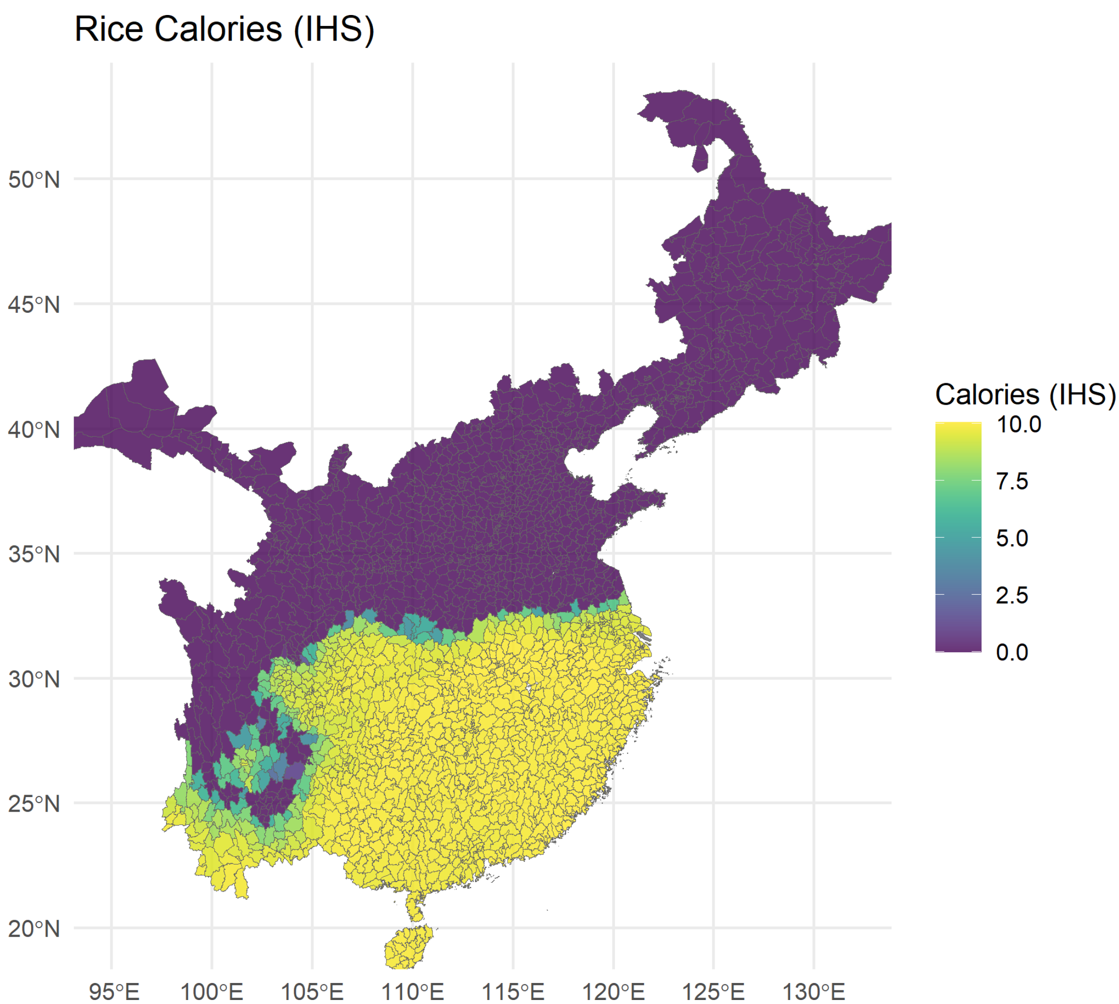

Figure 4: Spatial variation of paddy-rice suitability measured in calories (IHS).

Precipitation Predictability

Measures of precipitation risk can be constructed in a variety of ways. I use a measure developed by an ecologist, Colwell, called "predictability" (Colwell, 1974). Predictability is an information-theoretic measure ranging between 0 (entirely unpredictable) and 1 (entirely predictable), which aims to establish not mere variation in a periodic phenomenon but how well that phenomenon can be predicted. Predictability is the sum of two measures: contingency or seasonality, the degree to which the phenomenon's state varies cyclically, and constancy, the degree to which the phenomenon's state is unchanged across time-divisions (Colwell, 1974). Predictability then is maximized when a phenomenon varies across time but is completely seasonal, or is not seasonal but completely constant and is minimized when all states are equiprobable across time (i.e. when states are highly uncorrelated with temporal periods) (Colwell, 1974). In other words, in regions where precipitation predictability is low, households face an environment where they cannot anticipate what about of rain they will experience.

To construct precipitation predictability, I collect monthly precipitation data from the Global Precipitation Climatology Centre (GPCC) (Schneider, Becker, Finger, Meyer-Christoffer, Anja Rudolf, & Ziese, 2020). The GPCC data set consists of global monthly precipitation totals between 1891 and 2016 at a 0.25° x 0.25° spatial resolution constructed using historically collected rainfall totals alongside spatial interpolation. I consider rainfall amounts only from the period of 1891 to 1960. I select these dates because they are the earliest and highest quality data accessible which precede 20^{th} century weather changes caused by global warming, which are detectable from the 1970s onward (Z.-Z. Hu, Bengtsson, & Arpe, 2000; L. Wang, Yang, Gu, & Li, 2020). Historical data on precipitation would of course be preferable however such data is not available at a sufficiently fine-grained spatiotemporal resolution to study geographic variation in lineages. Therefore, my analysis assumes that predictability patterns in the early 20^{th} century spatially mirror those across the Ming and Qing dynasties, an assumption which may prove too strong. Figure 5 plots spatial variation in precipitation predictability:

Rainfall amounts are binned into "states" before Colwell's transformations are applied. Because bin-number may appear arbitrary, in Appendix I use a continuous measure taken from Giuliano and Nunn (Giuliano & Nunn, 2020) which simply calculates the yearly average rainfall total for each region and then takes the log of the variance across years. While this transformation fails to account for seasonal changes over the course of an individual's life, and thus is better suited for the inter-generational question asked by Giuliano and Nunn, it is perhaps a more intuitive measure. Neither bin-choice nor use of the continuous and discrete predictability measures qualitatively alters my results.

The origins of observed variation in precipitation predictability are hard to infer. Because Chinese precipitation patterns are largely caused by monsoon winds (L. Wang et al., 2020), the most logical source for variation in precipitation predictability is variation in those winds and associated precipitation and tropical cyclone events. Monsoon winds in turn are largely driven by the El Niño Southern Oscillation (ENSO) or sea temperature and pressure differences across the Pacific Ocean (L. Wang et al., 2020), encouraging an interpretation of changes in yearly precipitation patterns as essentially random.

Historical Disasters

To collect data on historical disasters, I use the Reconstructed East Asian Climate Historical Encoded Series (REACHES) data set (P. K. Wang et al., 2018). This data set consists of historical records of natural and social disasters found in the *Compendium of Meteorological Records of China in the Last 3000 Years* (Zhang, 2013), which were then digitized, geocoded, and categorized according to disaster type. At the point of writing, only data for the years 1644 – 1795 has been made available. I construct a historical disaster measure at the county level by counting the number of flood and drought disasters that occured within

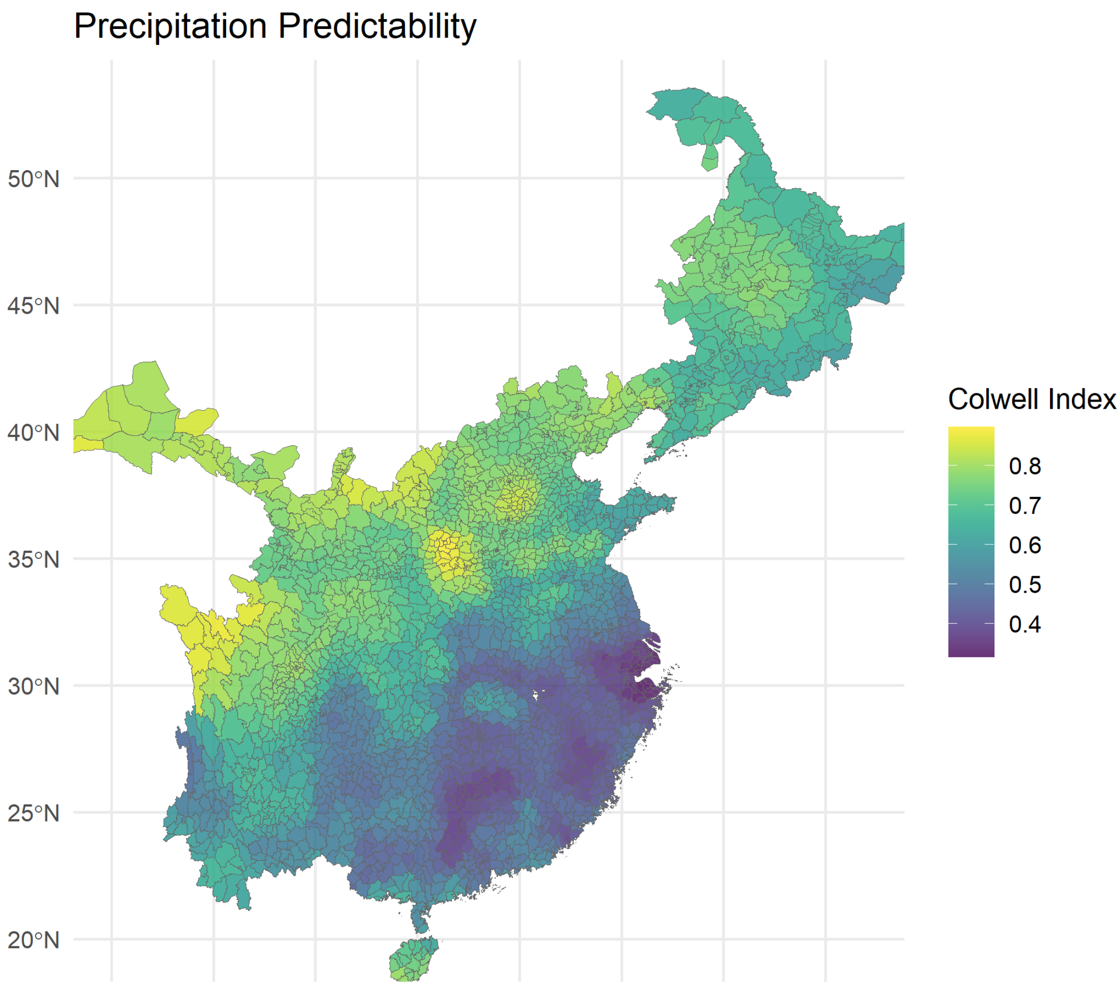

Figure 5: Spatial variation in precipitation predictability measure. Rain amounts are binned into 11 categories before Colwell transformation is applied.

each Chinese county. Again, because these measures are highly skewed and contain zeros, I apply inverse hyperbolic sine transformations. In Figure 6 I plot the geographic distribution of my combined metric for the number of flood and drought events.

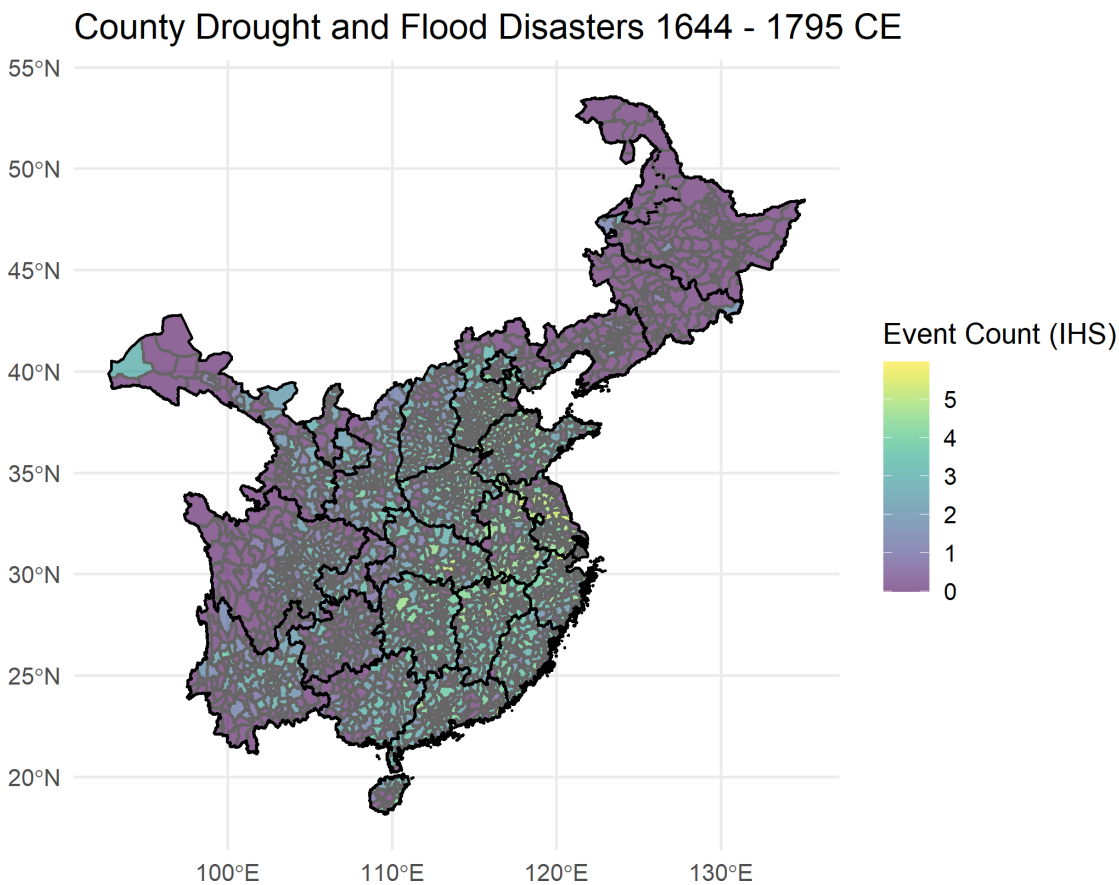

Figure 6: Spatial variation in county-level drought and flood disasters between 1644 and 1795 CE. The raw count of disaster events is transformed with an inverse hyperbolic sine. Black outlines show contemporary provincial borders.

Controls

In order to avoid collider bias and conditioning on mediating variables, which would mask the main effect of interest, I resist kitchen-sink regressions, show all models with and without controls, and use only a core set of reasonable controls. IHS refers to the the inverse hyperbolic sine transformation.

County Level Regressions *County area*: To account for the fact that larger counties should intrinsically have more lineages, I include each county's area in square kilometers (IHS).

Population size: Areas with more people should have more lineages. I control for population size in 1953 (IHS). Historical population size would be better however Ming population records are of notoriously low quality [Peter Bol, personal communication; Ho (1959)].

Economic development: Areas with more economic development should have more lineages as more resources can be allocated towards lineage activities and rituals. To control for access to trade, I include a variable denoting a county-centroid's distance to the nearest coastal point. Following Nunn and Puga's work, I also include a term capturing each county's ruggedness due to its impact on economic development (Nunn & Puga, 2012).

State control: Areas with greater state control may be more dominated by state-actors and less able to produce lineages. Therefore, I include each county-centroid's travel cost to Beijing, the political center of China. Travel cost is calculated using a wheeled-transport cost-function.

Precipitation levels: It is likely that precipitation predictability will correlate with average precipitation levels, therefore when examining the impact of precipitation predictability on lineage development, I include the IHS of average yearly precipitation for each county. This control also helps address endogeneity concerns that rice suitability is correlated with precipitation predictability (see section on *Suitability and Endogeneity* below).

CFPS Regression Because CFPS data cannot be linked to specific counties, all controls are at the household or community/village level.

Community population: Smaller communities may be distinct from larger ones and community size is likely related to the amount of land individuals possess. For that reason I include a count (IHS) of the number of official residents in the community.

Community spatial size: Community size may impact the amount of land available to community residents. Therefore I include a measure of the community's spatial area in square kilometers (IHS).

Economic development: Lineage behaviors may depend on wealth. For instance, rituals or the possession of an ancestral hall requires the capital to fund such a hall and the ritual activities, such as feasts, associated with it. For that reason I attempt to control for community and household variation in wealth, although see the above section for concerns about endogeneity. I control both for household net-assets (IHS) and community net income per capita (IHS). Finally, to account for potential rural-urban differences among communities and their implications for economic variation, I include a measure of the travel time to the nearest town in hours (IHS) and a measure of the percentage of the community engaged in agricultural labor. Similarly, I include a categorical variable whether the community is classified as a rural village, a suburb, a town, or an urban area.

State accessibility: We might expect lineage activity to be a function of state activities in the region. For instance, more rural areas may need to rely on local institutions rather than the state for the allocation of public goods. Therefore, I include a control to account for each community's travel time to the provincial capital in hours (IHS).

Key Variable Summary

In Table 1 and Table 2 I provide summary descriptions of the above variables. "Raw" indicates the raw count of the variable and "IHS" denotes the inverse hyperbolic sine transformation ($\text{arsinh}(x) = \ln\left(x + \sqrt{x^2+1}\right)$) used to reduce variable skew when it includes zero values. Table 2 shows county level data and Table 2 provides the household and community level data from the 2010 wave of the China Family Panel Study. In each table, I provide the overall mean and standard deviation and break up the summary statistics according to North and South China. The number prior to the semi-colon denotes the number of complete observations for that variable if incomplete cases exist.

Table 1: County Level Variables

Characteristic	Overall, N = 2,556	North, N = 980	South, N = 1,576
Lineage Activity (IHS)	0.80 (1.34)	0.36 (0.77)	1.07 (1.53)
Lineage Activity (Raw)	6.13 (29.64)	0.93 (5.05)	9.36 (37.18)
Precipitation Predictability	0.62 (0.14)	0.73 (0.06)	0.54 (0.13)
Rice Caloric Suitability (IHS)	4.60 (4.81)	0.01 (0.23)	7.46 (4.03)
Rice Caloric Suitability (Raw)	4,077.49 (4,689.56)	0.17 (3.98)	6,618.99 (4,344.71)
Arable Land (IHS)	2.42 (0.60)	2.77 (0.59)	2.20 (0.50)
Arable Land (Raw)	6.66 (4.42)	9.38 (5.51)	4.92 (2.24)
Drought Disasters (IHS)	1.05 (1.30)	0.90 (1.20)	1.14 (1.35)
Drought Disasters (raw)	3.33 (6.06)	2.50 (4.60)	3.85 (6.76)
Flood Disasters (IHS)	1.34 (1.53)	1.04 (1.40)	1.53 (1.57)
Flood Disasters (raw)	5.99 (11.43)	4.12 (8.19)	7.16 (12.91)
County Area (IHS)	7.55 (1.48)	7.56 (1.45)	7.55 (1.49)
Population in 1953 (IHS)	13.11 (0.86)	12.94 (0.91)	13.22 (0.80)
Distance to Coast (IHS)	6.21 (1.45)	6.22 (1.37)	6.20 (1.50)
Travel Cost to Beijing (IHS)	11.41 (2.30)	11.07 (2.48)	11.63 (2.15)
Annual Precipitation Mean (IHS)	7.49 (0.53)	7.01 (0.40)	7.79 (0.36)
Terrain Ruggedness Index	102.02 (108.43)	72.68 (74.79)	120.26 (121.36)
Nightime Light (IHS)	1.53 (1.42)	1.77 (1.35)	1.37 (1.44)

[1] Mean (SD)

Table 2: Community Level Variables

Characteristic	Overall, N = 8,908	North, N = 3,762	South, N = 5,146
Lineage Activity == Yes (Binary)	6,891 / 8,888 (78%)	2,929 / 3,757 (78%)	3,962 / 5,131 (77%)
Community Arable Land (IHS)	1.02 (0.57)	1.29 (0.57)	0.83 (0.48)
Community Arable Land (Raw)	1.46 (1.36)	1.97 (1.32)	1.09 (1.28)
Community Paddy Percentage	0.34 (0.39)	0.14 (0.29)	0.50 (0.38)
Community Population Size (IHS)	8.23 (0.80)	7.88 (0.69)	8.50 (0.76)
Community Net Income/Capita (IHS)	8.52 (1.17)	8.50 (1.10)	8.53 (1.22)
Community Travel Time to Town (IHS)	0.33 (0.38)	0.34 (0.26)	0.32 (0.44)
Community Type			
Rural village	7,880 / 8,888 (89%)	3,287 / 3,762 (87%)	4,593 / 5,126 (90%)
Suburb	490 / 8,888 (5.5%)	196 / 3,762 (5.2%)	294 / 5,126 (5.7%)
Town	486 / 8,888 (5.5%)	279 / 3,762 (7.4%)	207 / 5,126 (4.0%)
Urban	32 / 8,888 (0.4%)	0 / 3,762 (0%)	32 / 5,126 (0.6%)
Community Area (IHS)	2.06 (1.43)	1.85 (1.56)	2.22 (1.31)
Community Travel Time to Provincial Capital (IHS)	2.03 (0.69)	2.06 (0.65)	2.01 (0.72)
Community Percentage Agricultural Labor Force	46.93 (23.93)	50.71 (22.89)	44.12 (24.30)
Family Net Assets	11.64 (3.33)	11.40 (3.49)	11.81 (3.19)

[1] n / N (%); Mean (SD)

Verifying Precipitation Predictability

Because predictability is an uncommonly used metric, I first verify it as a meaningful determinant of precipitation events. To do this, I relate flooding and drought disaster outcomes from the REACHES database to my precipitation predictability measure. If precipitation predictability meaningfully impacts precipitation events then it should negatively correlate with precipitation disasters —areas with higher predictability should see fewer flooding and drought events. In Table 3 I show exactly this relationship. Either with or without province fixed effects, more precipitation predictability correlates with fewer drought or flood events across China, suggesting that at least historically, precipitation predictability had tangible consequences for shocks to production and serves as a suitable proxy for historical shocks to production. I also take this analysis to verify the quality of the REACHES data set to a large extent because it is correlated with a separately collected and distinct ecological predictor in a reasonable manner.

Table 3: Precipitation Predictability Verification

	Dependent Variable = Disaster Events			
	Drought 1	Flood 1	Drought 2	Flood 2
Precipitation Predictability	−1.170**	−1.170**	−1.991***	−1.102*
	(0.376)	(0.376)	(0.447)	(0.547)
Area (IHS, km^2)	0.332***	0.332***	0.363***	0.439***
	(0.043)	(0.043)	(0.038)	(0.044)
Ruggedness	−0.002***	−0.002***	−0.002***	−0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Light 1992 (IHS)	0.102*	0.102*	0.105*	0.088+
	(0.051)	(0.051)	(0.041)	(0.047)
Population 1953 (IHS)	−0.016	−0.016	−0.004	0.018
	(0.057)	(0.057)	(0.044)	(0.054)
Distance to Coast (IHS, km)	0.052*	0.052*	0.101**	0.158***
	(0.026)	(0.026)	(0.034)	(0.037)
Travel Cost to Beijing (IHS)	−0.120***	−0.120***	−0.070*	−0.104**
	(0.024)	(0.024)	(0.030)	(0.040)
Mean Annual Precipitation (IHS, mm/m^2)	0.292**	0.292**	0.276*	0.272**
	(0.093)	(0.093)	(0.108)	(0.104)
Num.Obs.	2552	2552	2552	2552
R2	0.155	0.155	0.317	0.332
Province FE			X	X
District Controls	X	X	X	X

Table 3: Precipitation Predictability Verification *(continued)*

	Drought 1	Flood 1	Drought 2	Flood 2
BIC	8225.6	8225.6	7887.3	8649.0
Std.Errors	by: Prefecture	by: Prefecture	by: Prefecture	by: Prefecture

Note:
* p < 0.1, ** p < 0.05, *** p < 0.01

Results

Farm Size

I first examine whether an association exists between historical lineage activity and farm size today, drawn from the CHIPS data set. In Table 4 I estimate the following specification to regress county level lineage activity (IHS) up to the end of the Qing dynasty on average arable land farm-size:

$$y_{i,p} = \beta\Omega_{i,p} + \Gamma X_{i,p} + \delta_p + \mu_{i,p} \tag{1}$$

where $y_{i,p}$ is the historical lineage activity for county i in province p; $\Omega_{i,p}$ is the variable of interest and denotes the county average amount of arable land per household, in *mu*; $X_{i,p}$ is a series of county level controls; and δ_p is a province fixed effect. In all specifications we observe a significant and negative coefficient of interest, indicating that counties with larger arable land averages are associated with less lineage activity. Importantly, this effect is stable to North-South region fixed effects indicating that average differences across North and South China do not drive the results.

Table 4: CHIPS Farm Size

	Dependent Variable = Lineage Activity (IHS)				
	Model 1	Model 2	Model 3	Model 4	Model 5
Arable Land (Mu)	−0.132***	−0.132***	−0.106***	−0.085*	−0.085*
	(0.028)	(0.028)	(0.027)	(0.042)	(0.041)
Rice Suitability				0.234***	0.234***
				(0.059)	(0.060)
Area (IHS, km^2)				−0.256	−0.253
				(0.391)	(0.374)
Population 1953 (IHS)				−0.022	−0.026
				(0.302)	(0.306)
Ruggedness				0.002	0.002
				(0.002)	(0.002)
Light 1992 (IHS)				0.067	0.068
				(0.253)	(0.252)
Distance to Coast (IHS, km)				−0.218	−0.219
				(0.186)	(0.197)
Travel Cost to Beijing (IHS)				−0.038	−0.036
				(0.101)	(0.123)
Latitude				0.125*	0.126*

Table 4: CHIPS Farm Size *(continued)*

	Model 1	Model 2	Model 3	Model 4	Model 5
				(0.052)	(0.061)
North-South FE			X		X
District Controls				X	X
Num.Obs.	105	105	105	105	105
R2	0.126	0.126	0.142	0.343	0.343
BIC	401.1	401.1	403.9	408.4	413.1
Std.Errors	by: Prefecture	by: Prefecture	by: Prefecture	by: Prefecture	by: Prefecture

Note:
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Micro data

The above section identified a negative association between lineage activity and landholding by regressing county level lineage activity against the aggregated county-average for household landholdings. In this section, I examine the same variables but use micro-level data for both lineage activity and farm-size, providing both alternative data sources for the variables of interest, as well as alternative specifications. I use the China Family Panel Study to collect household behavioral data on lineages and landholding size. In Table 5 I predict whether a family participated in ancestor worship and/or tomb-sweeping within the previous year or has a genealogy of their family according to the following full specification:

$$\text{logit}(y_{i,j,p}) = \beta\Omega_{i,j,p} + \Gamma X_{j,p} + \delta_p + \mu_{i,j,p} \tag{2}$$

where logit denotes the logistic transformation, $\text{logit}(x) = \log\frac{x}{1-x}$; i, j, p indexes households, communities or villages, and provinces; $y_{i,j,p}$ denotes whether household i in community j and province p either participated in ancestor worship/tomb-sweeping or has a genealogical book; and $\Omega_{j,p}$ is the variable of interest, denoting the average amount of arable land per capita in a community or village. I drop index i to denote that all households in community j receive the same value for Ω. $X_{i,j,p}$ is a series of household and community district controls; and δ_p denotes a provincial fixed effect.

Table 5 supports the conclusions of Table 4 albeit with different, contemporary, and behavioral data. As predicted, the negative coefficient on community arable land per capita indicates that families from communities with more arable land per capita are less likely to have participated in lineage activities. This relationships holds with the inclusion of provincial fixed effects and household- and community-level controls.

Table 5: CFPS Lineage Activity

	Dependent Variable = Binary Household Lineage Activity			
	Model 1	Model 2	Model 3	Model 4
Community Arable Land PC (Mu; IHS)	−0.431**	−0.417**	−0.619***	
	(0.137)	(0.144)	(0.150)	

Table 5: CFPS Lineage Activity *(continued)*

	Model 1	Model 2	Model 3	Model 4
Community Arable Land PC (Mu)				−0.177**
				(0.057)
Community Paddy Percentage		0.253	−0.705*	−0.702*
		(0.233)	(0.296)	(0.301)
Number of People (IHS)			−0.070	−0.025
			(0.134)	(0.134)
Community spatial size (km2; IHS)			0.061	0.065
			(0.058)	(0.059)
Community Income PC (IHS)			−0.177+	−0.175+
			(0.101)	(0.104)
Travel Time to Provincial Cap (IHS)			−0.219+	−0.209+
			(0.118)	(0.122)
Travel Time to Town (IHS)			0.007	−0.005
			(0.140)	(0.145)
Percentage Labor Force in Agric			0.002	0.001
			(0.004)	(0.004)
Family Net Assets (IHS)			0.027**	0.028**
			(0.010)	(0.010)
Num.Obs.	8888	8625	6718	6718
BIC	9391.6	9063.2	7026.3	7048.8
Province FE			X	X
District Controls			X	X
Std.Errors	County	County	County	County

Note:
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The above analyses demonstrate that lineages, measured either through historical genealogical books or with contemporary behavioral measures (ancestral worship and genealogy possession today), are less prominent in regions characterized by smaller landholdings. The direction of the relationship or even historical precedence of events is ambiguous although scholars do argue that landholding size has remained stagnant in China for at least several centuries (Bray, 1986). Perhaps lineages are not more prominent where land-access is restricted, as I have suggested, and the above relationships instead reflect that lineages purchase and control large amounts of (corporate) land leaving less land available for others. The fact that I use contemporary land data suggests this alternative explanation is unlikely. Land reform in China following the Communist Revolution effectively eliminated the feudal and class relationships which had dominated Chinese society and successfully removed local power-holders from their privileged positions by redistributing land among the peasantry (Noellart, 2020). This makes the observed relationships post-reform puzzling at first because land in China is not easily sold or transferred between individuals and is generally owned by village-collectives, although it can be contractually managed-out to individuals (Dong, 2019)). The presence of land reform is one of the reasons I prefer to use a community-level variable to define access to land rather than a household-level variable. While land itself may change hands over history, its ownership concentrated or diluted over time, the total sum of land to population should stay relatively constant before and after reform.

Suitability

The previous sections have identified a negative association between lineage activity and landholdings in support of the hypothesis that lineages provide land-poor households with access to land and/or risk-pooling. For an alternative line of evidence connecting lineage activity to the value of land-access, I turn to an arguably exogenous determinant of direct shocks to production and wealth. If lineages serve risk-pooling functions then they should be more common where households are more likely to face risks to production. More so, my theory predicts that the value of pooling risk is larger in paddy-rice dominant regions.

As previously discussed, I use Colwell's precipitation predictability score to quantify risks to production. Figure 7 plots the county level pairwise relationship between lineage activity for the Ming and Qing dynasties (IHS) and Colwell's precipitation predictability score. Points are colored according to whether they are located in North or South China. Clearly, on average, North China is characterized by higher levels of precipitation predictability, demanding the inclusion of geographic fixed effects.

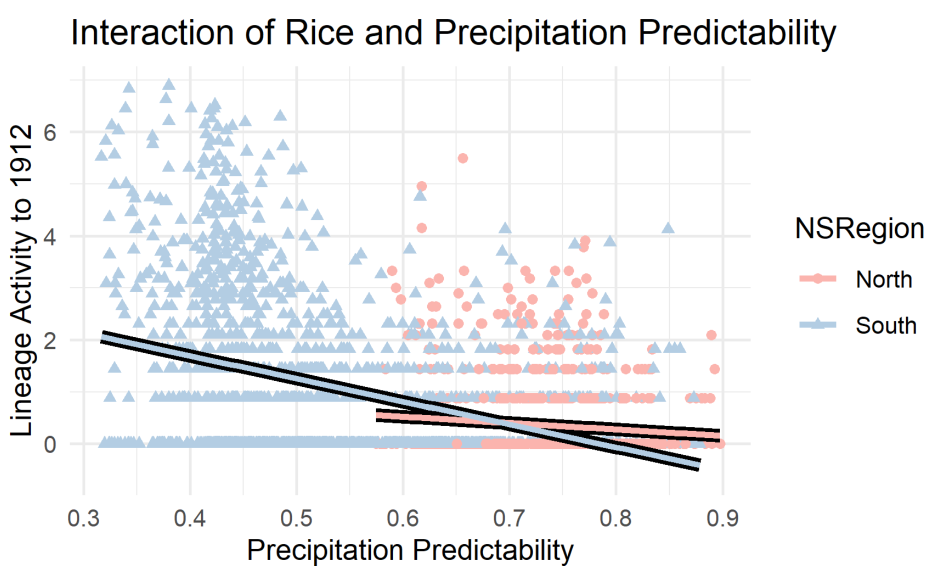

Figure 7: Pairwise relationship between lineage activity and precipitation predictability. Blue triangles denote southern counties and red circles denote northern counties. Regression lines show simple linear models of lineage activity as a function of precipitation predictability.

I estimate the relationship demonstrated in 7 according to two general specifications meant to account for spatial autocorrelation. Rice-suitability, lineage activity, and precipitation predictability are all spatially correlated and such spatial correlation will produce incorrect p-values. The first of these specifications is a spatial lag model which explicitly models a county's lineage activity as a function of the lineage activity of neighboring counties:

$$y_{i,p} = \rho \sum_{i \neq j} w_{i,p} y_{j,p} + \beta_1 \Omega^1_{i,p} + \beta_2 \Omega^2_{i,p} + \beta_3 \Omega^1_{i,p} \Omega^2_{i,p} + \Gamma X_{i,p} + \delta_p + \mu_{i,p} \tag{3}$$

where again $y_{i,p}$ denotes county i in province p's lineage activity (IHS). The term $w_{i,p}$ equals 1 if county j is a neighbor to county i and 0 otherwise. I use a queen neighboring criterion such that neighbors are defined as sharing at least a single point as borders. Thus, $\sum_{i \neq j} w_{i,p} y_{j,p}$ denotes the lineage activity of the counties surrounding county i, p and ρ is the model coefficient capturing this term's effect. There are now two variables of interest as well as their interaction: $\Omega^1_{i,p}$ denotes a county's rice suitability, $\Omega^2_{i,p}$ denotes a county's precipitation predictability, using the 11-bin Colwell measure (see Appendix for models using continuous measures), and their interaction, $\Omega^1_{i,p} \Omega^2_{i,p}$. $X_{i,p}$ denotes a series of county level controls and δ_p provincial fixed effects.

The second specification does not explicitly model spatial dependency but instead captures spatial autocorrelation in the regression error term by permitting errors to be correlated across spatial neighbors (as defined above):

$$\begin{aligned} y_{i,p} = \beta_1 \Omega^1_{i,p} + \beta_2 \Omega^2_{i,p} + \beta_3 \Omega^1_{i,p} \Omega^2_{i,p} + \Gamma X_{i,p} + \delta_p + \mu_{i,p} \\ \mu_{i,p} = \lambda \sum_{i \neq j} w_{i,j} u_{j,p} + \epsilon_{i,p} \end{aligned}$$

where λ denotes how errors are correlated and the term $\sum_{i \neq j} w_{i,j} u_{j,p}$ permits errors to be correlated. The term $\epsilon_{i,p}$ denotes the non-spatial component of the error term. A standard OLS specification sets $\lambda = 0$.

In Table 6 I report OLS, spatial lag, and spatial error models regressing lineage activity (IHS) on precipitation predictability and rice suitability. Because my theory predicts households in paddy-rice regions to generally be more exposed to and less buffered from shocks to agricultural production, I interact predictability with the rice suitability measure measuring how many calories are earned by growing low-input (i.e. with traditional means) paddy-rice. In all specifications, lineage activity is higher in areas more suitable for paddy-rice agriculture. However, how precipitation predictability correlates with lineage activity clearly depends on how rice suitable a given county is. The negative interaction coefficient between rice suitability and precipitation predictability indicates that precipitation predictability impacts lineage activity more strongly among rice suitable counties.

Importantly, because both rice suitability and precipitation predictability strongly correlate with geographical region, I include models with province fixed effects to remove average regional differences. The relationships above are maintained. Finally, I draw the reader's attention to *Model 1*, which reports the baseline specification of lineage activity against precipitation predictability, rice-suitability, and their interaction absent any geographic fixed effects or controls. The R^2 value for *Model 1* indicates that rice suitability, precipitation predictability, and their interaction, without any fixed-effects or control variables, can explain 20% of the variance across over 2500 Chinese counties, a sizable proportion.

Table 6: Lineage Activity and Precipitation Predictability

	Dependent Variable = Lineage Activity (IHS)				
	Model 1	Model 2	Model 3	Model 4	Model 5
Rice x Predictability Interaction	−0.549***	−0.383***	−0.215**	−0.554***	−0.271***
	(0.084)	(0.056)	(0.068)	(0.085)	(0.079)
Precipitation Predictability	−0.259	0.902*	0.163	0.891	0.219
	(0.442)	(0.442)	(0.538)	(0.688)	(0.631)
Rice Suitability	0.360***	0.254***	0.171***	0.369***	0.212***
	(0.051)	(0.035)	(0.042)	(0.053)	(0.049)
Mean Annual Precipitation (IHS)		0.404***	0.483***	0.605***	0.560***
		(0.106)	(0.143)	(0.163)	(0.165)
Area (IHS)		0.282***	0.325***	0.345***	0.343***
		(0.030)	(0.030)	(0.033)	(0.031)
Ruggedness		0.000	−0.001+	0.000	−0.001+
		(0.000)	(0.000)	(0.000)	(0.000)
Light 1992 (IHS)		0.219***	0.264***	0.287***	0.278***
		(0.036)	(0.036)	(0.043)	(0.039)
Population 1953 (IHS)		0.103**	0.099**	0.132**	0.109**
		(0.032)	(0.035)	(0.043)	(0.039)
Distance to Coast (IHS)		0.000	0.066*	−0.008	0.074*
		(0.019)	(0.028)	(0.028)	(0.032)
Travel Cost to Beijing (IHS)		−0.002	0.058**	0.007	0.060**
		(0.014)	(0.020)	(0.018)	(0.021)
Latitude		0.014*	0.006	0.024*	0.010

Table 6: Lineage Activity and Precipitation Predictability *(continued)*

	Model 1	Model 2	Model 3	Model 4	Model 5
		(0.006)	(0.020)	(0.010)	(0.024)
Province FE			X		X
District Controls		X	X	X	X
Model	OLS	Lag	Lag	Spatial Error	Spatial Error
Num.Obs.	2547	2546	2546	2546	2546
R2	0.198	0.354	0.407	0.358	0.406
BIC	8190.2	7801.0	7712.3	7792.8	7717.2

Note:

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 8 makes the conditional impact of precipitation predictability apparent. To produce this figure I create a binary variable classifying each county as either paddy-rice or dry-crop suitable depending on which class of crops provides more calories. I then plot the interaction effect (including provincial fixed effects). Dry-crop suitable regions show essentially no effect of precipitation predictability on lineage activity while paddy-rice suitable regions demonstrate a strongly negative effect. Such a pattern accords with the hypothesis that lineages exist where individuals want to pool risk and that this relationship dominates in rice-suitable regions.

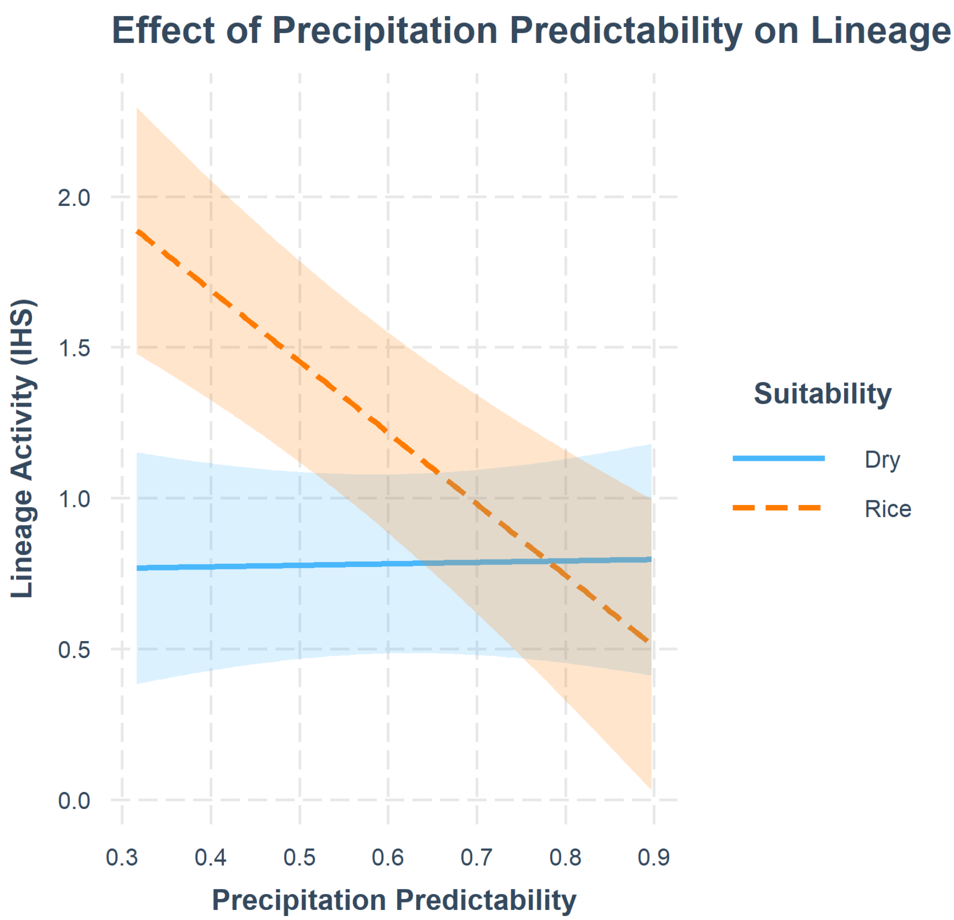

Figure 8: Interaction effect between precipitation predictability and rice suitability on lineage activity. Each county is categorized as either dry-crop or paddy-rice suitability depending on which agricultural scheme provides more calories.

Suitability and Endogeneity

The calculations that construct my measure for rice suitability, caloric potential, rely on the FAO GAEZ rice suitability index. This index introduces an endogeneity concern because it in turn relies on precipitation data (Fischer et al., 2002). If rice suitability is correlated with or a function of precipitation predictability then the interaction term between rice suitability and precipitation predictability that I rely on to support my theory of lineage evolution may simply reflect a non-linear relationship between rice suitability and lineage activity (Balli & Sørensen, 2013). If the interaction term merely indicates a non-linear impact of rice suitability on lineage activity than my above results do not support my theory of Chinese lineages as risk-pooling institutions. Endogeneity then must be accounted for.

Importantly, the rice suitability measure from the FAO GAEZ does not rely on the same temporal period that I do. The GAEZ measure relies on precipitation data between 1961 and 1990, a period that I purposefully avoid because global warming is known to have altered precipitation patterns in China then (Z.-Z. Hu et al., 2000; L. Wang et al., 2020). In Figure 9 I show the pairwise association between my predictability measure for pre-1961 and post-1961 samples. The diagonal line represents the set of points where predictability is equivalent. Clearly while a correlation exists between the two periods they are not equivalent. Interestingly, it appears that predictability has generally decreased in the latter period, in line with my aforementioned concerns about climate change's impact on more recent precipitation data.

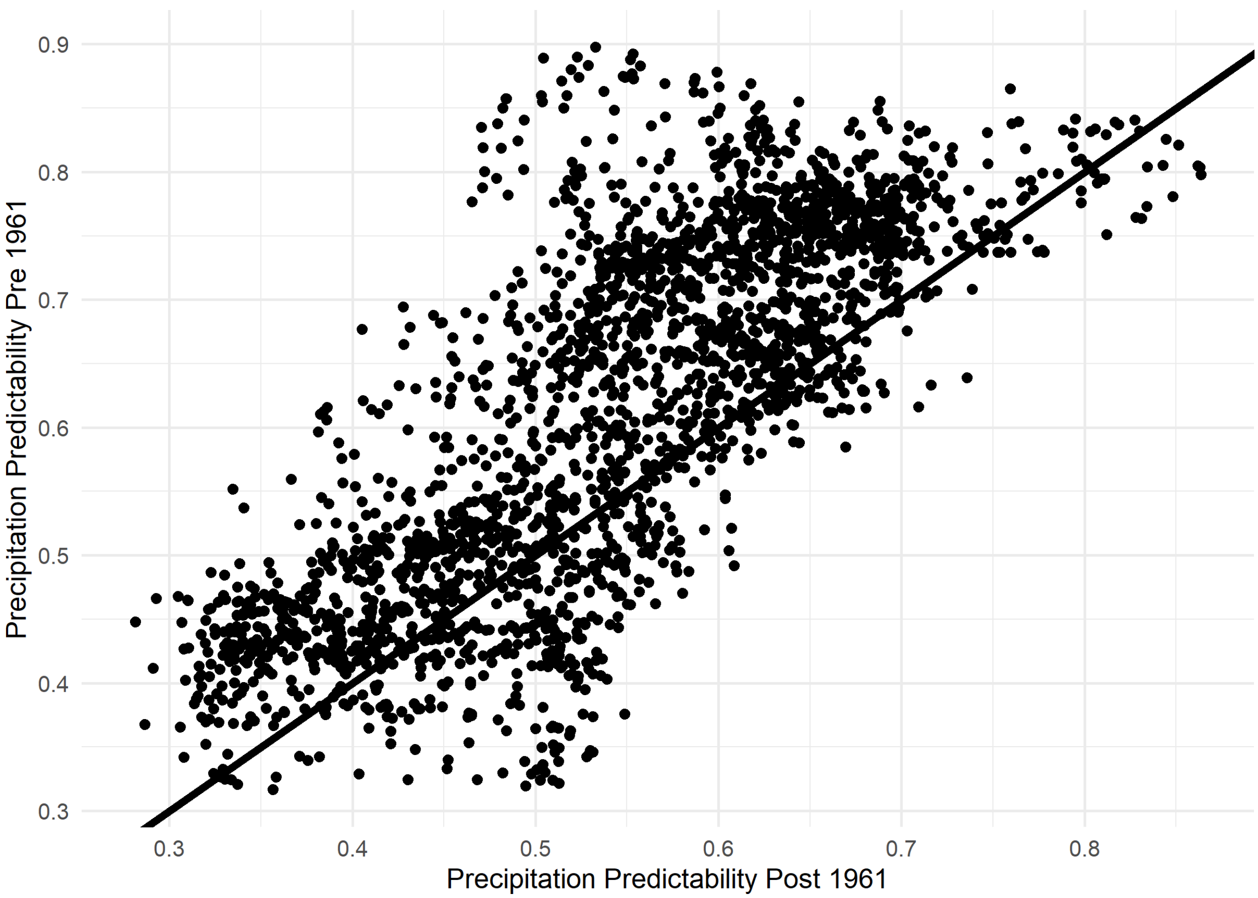

Figure 9: Scatterplot of precipitation predictability measure using data before and after 1961.

I adopt two strategies in order to confront the concern that the interaction term I take to support my theory merely demonstrates a non-linear impact of rice suitability on lineage activity. In Table 7, I first demonstrate that, even when I model rice suitability's relationship to lineage activity as a second-order polynomial, the significance and coefficient of the interaction term between rice suitability's main effect and precipitation predictability remains the same (Balli & Sørensen, 2013). If rice suitability has a non-linear impact on lineage activity, including the higher-order polynomial term should model this impact and eliminate the precipitation predictability by rice suitability term's significance. That we continue to see the interaction term's significance even when we permit rice suitability to have a nonlinear impact on lineage activity should assuage some endogeneity concerns.

However, the use of a single (or even multiple) higher-order polynomial terms assumes requires assumptions about the functional relationship between rice suitability and precipitation predictability. Therefore, I instrument my rice suitability measure with one of its components, temperature constraints, which does not consider any precipitation data. Unlike wheat, which is a hardier crop and can resist or even benefit from cooler temperatures and frost, rice production is significantly harmed by cold temperatures such that unexpected frosts are a major impediment to rice-yields across the globe (da Cruz et al., 2013). The GAEZ temperature constraints index considers temperature variation and frost risk when assessing how constrained a raster cell is for rice production (Fischer et al., 2002). For this measure, higher numbers demonstrate fewer constraints. In a second instrumental variables regression, I directly block on the precipitation variation measure that the FAO GAEZ suitability index uses, the Fournier index, which measures the amount and distribution of rainfall (Fischer et al., 2002), by interacting it with rice suitability.

Table 7: Endogeneity Results

	Dependent Variable = Lineage Activity (IHS)			
	Model 1	Model 2	Model 3	Model 4
First Order Rice x Precipitation Interaction		−0.317***	−0.354***	−0.314**
		(0.087)	(0.098)	(0.101)
First Order Rice Suitability	−0.147*	0.131	0.260***	0.231***
	(0.065)	(0.091)	(0.059)	(0.063)
Second Order Rice Suitability	0.022**	0.012		
	(0.008)	(0.007)		
Precipitation Predictability		0.473	0.604	18.070*
		(0.583)	(0.647)	(8.192)
Fournier Index				1.862*
				(0.883)
Fournier x Precipitation Interaction				−2.437*
				(1.150)
Mean Annual Precipitation (IHS, mm/m^2)	0.242+	0.359*	0.398*	0.400
	(0.132)	(0.149)	(0.156)	(0.248)
Area (IHS, km^2)	0.148***	0.153***	0.152***	0.150***
	(0.023)	(0.024)	(0.024)	(0.024)
Ruggedness	−0.001*	−0.001*	−0.001**	−0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Population 1953 (IHS)	0.149***	0.162***	0.156***	0.167***
	(0.043)	(0.043)	(0.043)	(0.042)
Distance to Coast (IHS, km)	0.065+	0.047	0.046	0.040

Table 7: Endogeneity Results *(continued)*

	Model 1	Model 2	Model 3	Model 4
	(0.034)	(0.035)	(0.035)	(0.036)
Travel Cost to Beijing (IHS)	0.048+	0.055+	0.057+	0.052+
	(0.029)	(0.029)	(0.029)	(0.028)
Latitude	−0.007	0.004	0.004	0.008
	(0.023)	(0.023)	(0.023)	(0.023)
Province FE	X	X	X	X
District Controls	X	X	X	X
First Stage F Test			148	132
Cluster	Prefecture	Prefecture	Prefecture	Prefecture
Num.Obs.	2546	2546	2546	2546
R2	0.372	0.379	0.379	0.381

Note:

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In the first two regressions I demonstrate that including a higher order rice suitability polynomial term does not eliminate the effect of interacting precipitation predictability and the first-order rice suitability variable. The inclusion of higher order terms beyond 2 (models not shown) also does not alter the sign or significance of the key interaction term. This analysis however assumes that the nonlinear effect of rice-suitability is well-modeled by a quadratic function. The last two models in 7 demonstrate that instrumenting the caloric suitability metric with the temperature constraint component produces identically signed and similar-in-magnitude coefficients to those in Table 6. The second instrumental variable model blocks on two precipitation variability channels, the Fournier Index described above and its interaction with precipitation predictability. The coefficients of interest are unchanged. Reassuringly, including the Fournier measure (1) causes the direct effect of mean annual precipitation to drop out and (2) shows an interaction effect with precipitation predictability, areas more paddy-rice suitable with respect to precipitation specifically display more lineages on average, and this effect weakens in the presence of greater precipitation predictability. Similarly, these effects, the main Fournier effect and it's interaction with precipitation predictability, are larger than instrumented effects which only consider temperature. This is also reassuring because paddy-rice should show a greater relationship with respect to precipitation components as opposed to temperature. This is not to say that temperature is unimportant for paddy-rice but rather that in regions where temperature is erratic and frosts are frequent, individuals are unlikely to fully invest in rice-agriculture (Bray, 1986), thus we would expect a weaker relationship on the main coefficients of interest, First Order rice Suitability and First Order Rice x Precipitation Interaction. In Appendix section I show that using prefectural data produces qualitatively similar results.

While I have purposefully avoided causal language, the above regressions indicate that a causal interpretation of my results is plausible. Both precipitation predictability and rice suitability are likely exogenous and as-if random to county-inhabitants. The above models, particularly the instrumental regressions, indicate that the reader need not be overly concerned that rice suitability and precipitation predictability are correlated and that therefore a causal interpretation of their interaction term is inappropriate.

Historical Disasters

Finally, it is entirely possible that even if Chinese counties face variable precipitation predictability, that this variation is not directly experienced. As niche constructors, humans extensively modify their environment and the ecological pressures they are subject to (Laland & O'Brien, 2011) and cultural learning permits populations to adapt to environmental pressures where individuals could not (Boyd et al., 2011). Such adaptation and niche construction through environmental modification, i.e. water management technologies, could permit communities facing environmental shocks to not directly experience those shocks. Alternatively, it is entirely possible that my precipitation predictability metric, which uses rainfall data from the early 20$^{\text{th}}$ century, says little about the occurrence of historical precipitation predictability and precipitation shocks, which are what would actually impact lineage activity. To assuage both of these legitimate concerns, I re-estimate the previous relationship between lineage activity and precipitation predictability but I replace my likely exogenous environmental predictor with historical records of disasters during the Qing dynasty. These disaster records are meant to measure actual risks to production. In Table 8, I use the digitized REACHES historical records from 1644 to 1795 to observe the relationship between my historical measure of lineage activity and documented precipitation disasters with the following general spatial error model specification:

$$y_{i,p} = \beta_1\Omega^1_{i,p} + \beta_2\Omega^2_{i,p} + \beta_3\Omega^1_{i,p}\Omega^2_{i,p} + \Gamma X_{i,p} + \Gamma X_{i,p} + \delta_p + \mu_{i,p}$$
$$\mu_{i,p} = \lambda\sum_{i\neq j} w_{i,j}u_{j,p} + \epsilon_{i,p}$$

All variables are conserved from the previous section except for the variable of interest, $\Omega^1_{i,p}$. This term denotes a vector containing a county's number of historical disasters (IHS), either combined (flood and drought) or decomposed (flood, drought, or both), which I again interact with precipitation predictability ($\Omega^2_{i,p}$).

Referring to Table 8, whether my disaster measure is aggregated or not, we see the same pattern as that established by the precipitation predictability measure. Areas with more disaster records are associated with higher lineage activity. The interaction terms indicate that the positive association between lineage activity and disasters is more extreme and positive in rice-suitable areas. In other words, disasters in rice-suitable areas are associated with more lineage activity than those in dry-grain suitable areas, even after removing average provincial differences.

Table 8: REACHES Historical Disasters

	Dependent Variable = Lineage Activity (IHS)			
	Model 1	Model 2	Model 3	Model 4
Rice Suitability	0.026*	0.030*	0.029*	0.028*
	(0.012)	(0.012)	(0.012)	(0.012)
Combined	0.072***			
	(0.020)			
Rice x Combined Interaction	0.022***			
	(0.003)			
Flood		0.074***		0.054+
		(0.023)		(0.033)
Rice x Flood Interaction		0.024***		0.011*
		(0.003)		(0.005)
Drought			0.085**	0.040
			(0.027)	(0.040)
Rice x Drought Interaction			0.027***	0.017**
			(0.004)	(0.006)
Mean Annual Precipitation (IHS, mm/m^2)	0.431**	0.431**	0.447**	0.434**
	(0.149)	(0.150)	(0.150)	(0.149)
Area (IHS, km^2)	0.243***	0.247***	0.254***	0.239***
	(0.031)	(0.031)	(0.031)	(0.031)
Ruggedness	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Light 1992 (IHS)	0.270***	0.264***	0.262***	0.259***

Table 8: REACHES Historical Disasters *(continued)*

	Model 1	Model 2	Model 3	Model 4
	(0.037)	(0.037)	(0.037)	(0.037)
Population 1953 (IHS)	0.092*	0.093*	0.098**	0.096**
	(0.037)	(0.037)	(0.037)	(0.037)
Distance to Coast (IHS, km)	0.052+	0.053+	0.066*	0.056+
	(0.030)	(0.030)	(0.030)	(0.030)
Travel Cost to Beijing (IHS)	0.064**	0.064**	0.066**	0.067***
	(0.020)	(0.020)	(0.020)	(0.020)
Latitude	0.006	0.005	0.004	0.006
	(0.022)	(0.022)	(0.022)	(0.022)
Province FE	X	X	X	X
District Controls	X	X	X	X
Model	Spatial Error	Spatial Error	Spatial Error	Spatial Error
Num.Obs.	2546	2546	2546	2546
R2	0.450	0.447	0.448	0.452
BIC	7517.0	7530.3	7527.5	7521.7

Note:

* p < 0.1, ** p < 0.05, *** p < 0.01

Discussion

In this paper I have offered a new theory to explain variation in lineage presence across China. This theory states that lineages exist largely to pool risk or provide individuals with access to land through corporate land holdings and predicts that in regions where individuals can ameliorate risk by expanding their landholdings or where risk is intrinsically low, lineages should be rarer. The analyses I have presented here have largely supported my hypothesis. Lineage activity, when measured through historical genealogical books, is positively associated with precipitation predictability, a measure of risk to production, and historically recorded flood and drought events. More so, this relationship is more extreme in paddy-rice dominant areas. This latter fact is important to note because, unlike previous theories of lineage formation and evolution (Freedman, 1958, 1966), it indicates that rice production itself is insufficient to explain variation in lineage prevalence. By referencing the particularities of rice production (Bray, 1986) and the historical and ethnographic nature of the rice-economies that farmers were immersed in (Fei & Chang, 1949; Huang, 1985, 1990), we can observe how paddy-rice farmers faced different economic pressures from those dry-grain farmers confronted —specifically smaller farm sizes and a reduced ability to purchase or hire both land and labor. Because paddy-rice farms are smaller, individuals are more exposed to geographically constrained shocks to production, such as precipitation shocks. Examining the association between farm size and either the historical presence of lineages, as measured through genealogical books or behavioral data documenting contemporary lineage activity supports this relationship. Areas with smaller farm sizes are characterized by more historical lineage activity and families from villages with less arable land per capita are more likely to participate in lineage activities. I take the above results to indicate support for my theory of lineage development and conclude that Chinese lineages serve risk pooling functions and that variation in lineage importance across regions of China can at least partially be understood through variation in how households respond to risks to production.

Note that, in line with previous theories (Freedman, 1958), I do detect a positive association with historical lineage activity and rice agriculture. The cause of this relationship however is difficult to infer. Rice production is notoriously productive such that we might expect higher population densities in rice-suitable areas causing more pressure on land availability. Land reclamation is a known economic activity that lineages invested in because they demand a relatively large initial capital investment with delayed returns [McDermott (2013); Song Chen, personal communication]. We may observe a rice-lineage relationship because lineages possess not only the required capital but also the necessary time horizon to convert wasteland to paddy-land. Alternatively, an association between rice and lineage activity may arise because paddy-rice irrigation infrastructure is a public good and requires extensive cooperation and coordination to build, maintain, and effectively use (Bray, 1986; Freedman, 1958). However, both of these explanations make it difficult to explain why lineage institutions emphasize horizontal breadth in kinship and not merely vertical depth. Irrigation infrastructure is highly localized and alternative village-based institutions which were already prominent in China and which lineages out-competed only in specific regions, such as worship or religious associations (Chen, 2017; McDermott, 2013), would similarly facilitate such cooperation.

Important work by Talhelm and colleagues (Talhelm et al., 2014) offers a very distinct and socio-psychological potential explanation for a rice-lineage association. Rice production requires extensive interpersonal coordination and cooperation in order to build, maintain, and effectively use the irrigation infrastructure and labor exchange required to benefit from paddy-rice productivity (Bray, 1986). Talhelm and colleagues demonstrated that Chinese students from historically paddy-rice dominant provinces are more collectivistic and holistic in their cognitive orientation than students from historically dry crop dominant provinces who are more individualistic and analytic in their cognitive orientation. While such psychological differences are unlikely to lead individuals from paddy-rice dominant regions to spontaneously form lineages (rice agriculture in China far precedes Fan Zhongyan's 11^{th} century institutional innovations) and it is important to remember that northern Chinese villages have many cooperative institutions (Chen, 2017), we might posit that a more collectivistic orientation facilitates lineage formation (Talhelm & English, 2020) once it is encouraged by historically specific and likely economic factors. I again refer to my discussion of Freedman's ideas above. Lineage institutions are a specific institutional form which emphasize kinship relationships beyond the village context of an individuals' day to day life. I take a strictly collectivistic psychology to not be agreeable to such relationships and that relationships of immediate economic and political consequence, e.g. village-level relationships perhaps more akin to the previously discussed *yihu*, which permit irrigation infrastructure maintenance and use and labor coordination, would quickly take precedence and obstruct lineage breadth elaboration.

Finally, further work needs to be done. First of all, my lineage activity measure could be improved upon. For my purposes, land records would prove a superior index of lineage importance, specifically the percentage of land entrusted in lineages. As far as I am aware however, such records have not been published. Finally, much of my theory of lineage activity relies on the aggregation and cross-referencing of historical quantitative or qualitative ethnographic data. While all quantitative work relies on and in fact benefits from qualitative insights, my theory would benefit from statistical support. For instance, historical longitudinal rice production data could be used to examine how precipitation shocks harm production. I have not collected such statistics because as far as I am aware no such data exists and in fact such data was likely never collected.

Appendix

Continuous Suitability Measures

I repeat the precipitation predictability analysis by replacing my Colwell predictability measure with a continuous measure, the variance of each county's yearly average rainfall amount over the first half of the 20th century (precipitation variability). Unlike Colwell's measure, more positive values indicate more variation across years. Because the variance is highly skewed I both log (LN) and take the square root (SD) of the measure and log it (LNSD).

Table 9: Continuous Suitability Measure

	Dependent Variable = Lineage Activity (IHS)					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Rice x Predictability Interaction (LN Var)		0.038***	0.034***	0.061***	0.035**	
		(0.007)	(0.010)	(0.012)	(0.012)	
Rice x Predictability Interaction (LN SD)						0.071**
						(0.024)
LN Variance Predictability	0.437***	−0.066	−0.031	−0.034	−0.002	
	(0.051)	(0.054)	(0.069)	(0.085)	(0.081)	
LN SD Predictability						−0.004
						(0.162)
Rice Suitability		−0.163***	−0.135*	−0.267***	−0.133*	−0.133*
		(0.042)	(0.058)	(0.068)	(0.067)	(0.067)
Mean Annual Precipitation (IHS)		0.363**	0.183	0.537*	0.460*	0.460*
		(0.139)	(0.157)	(0.214)	(0.190)	(0.190)
Area (IHS)		0.285***	0.137***	0.350***	0.347***	0.347***
		(0.030)	(0.016)	(0.034)	(0.031)	(0.031)
Ruggedness		0.000		0.000	−0.001+	−0.001+
		(0.000)		(0.000)	(0.000)	(0.000)
Light 1992 (IHS)		0.238***		0.307***	0.286***	0.286***
		(0.036)		(0.044)	(0.039)	(0.039)
Population 1953 (IHS)		0.106**	0.157***	0.140**	0.104**	0.104**
		(0.033)	(0.033)	(0.046)	(0.040)	(0.040)
Distance to Coast (IHS)		0.020		0.019	0.097**	0.097**

Table 9: Continuous Suitability Measure *(continued)*

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		(0.020)		(0.031)	(0.032)	(0.032)
Travel Cost to Beijing (IHS)		−0.022		−0.009	0.056**	0.056**
		(0.015)		(0.019)	(0.021)	(0.021)
Latitude		0.017**	0.005	0.029**	0.006	0.006
		(0.006)	(0.020)	(0.011)	(0.024)	(0.024)
Province FE			X		X	X
District Controls		X	X	X	X	X
Model	OLS	Lag	Lag	Spatial Error	Spatial Error	Spatial Error
Num.Obs.	2556	2546	2546	2546	2546	2546
R2	0.109	0.349	0.389	0.353	0.404	0.404
BIC	8475.8	7842.1	7760.0	7833.4	7725.3	7725.3

Note:

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Again, I plot the interaction effect of interest (Appendix Figure 10) to facilitate understanding the interaction terms in Appendix Table 9. In Appendix Figure 10, I use a model which measures precipitation variability by taking the standard deviation of annual rainfall totals, which I then take the logarithm of. Appendix Figure 10 demonstrates the same relationship as shown in the main text —we again see essentially no effect of increasing rainfall variability on lineage activity in dry regions but a strongly positive effect in paddy-rice dominant regions.

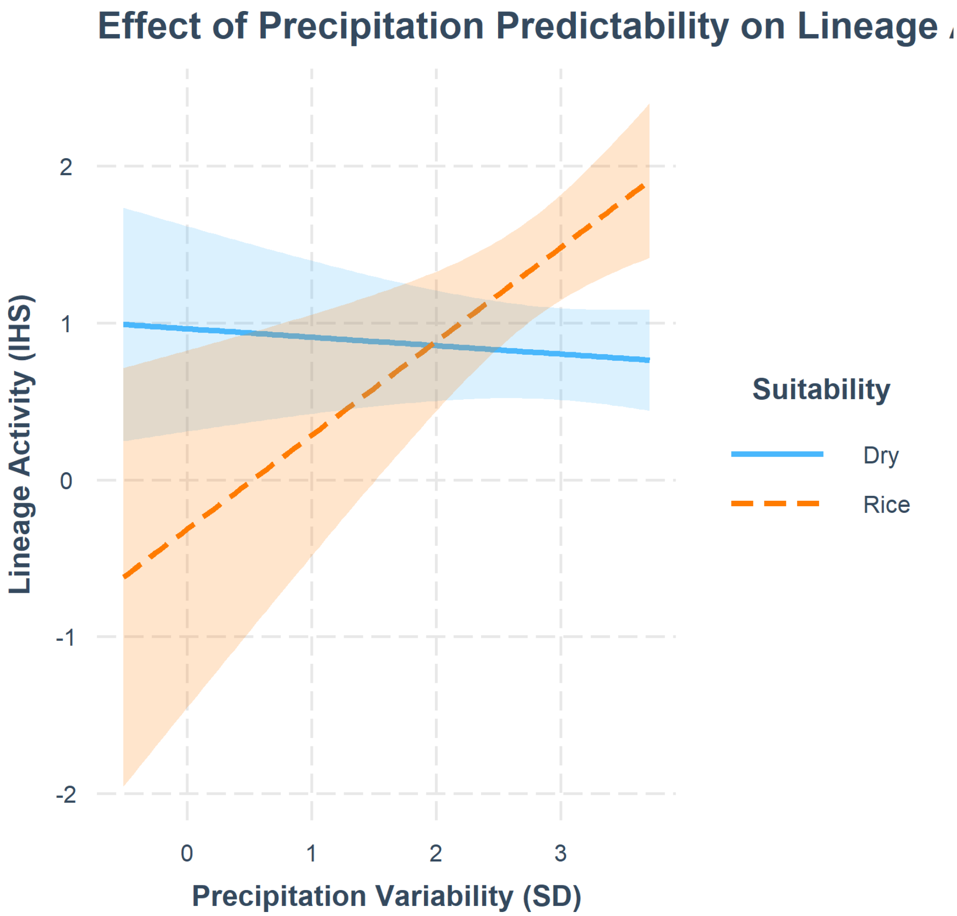

Figure 10: Interaction effect between precipitation variation (standard deviation) and rice suitability on lineage activity. Each county is categorized as either dry-crop or paddy-rice suitability depending on which agricultural scheme provides more calories.

Both Rice and Dry-Crop Suitability

In this section I include both paddy-rice and dry-crop suitability, again in the form of potential calories (IHS), as regressors. I plot rice and wheat suitability maps in Appendix Figure 11 and Appendix Figure 12 respectively. Figure 11 is identical to that Figure 4 in the main text.

In Appendix Table 10 I estimate the following specification:

$$\begin{aligned} y_{i,p} = \beta_1 \Omega^1_{i,p} + \beta_2 \Omega^2_{i,p} + \beta_3 \Omega^1_{i,p} \Omega^2_{i,p} + \Gamma X_{i,p} + \delta_p + \mu_{i,p} \\ \mu_{i,p} = \lambda \sum_{i \neq j} w_{i,j} u_{j,p} + \epsilon_{i,p} \end{aligned}$$

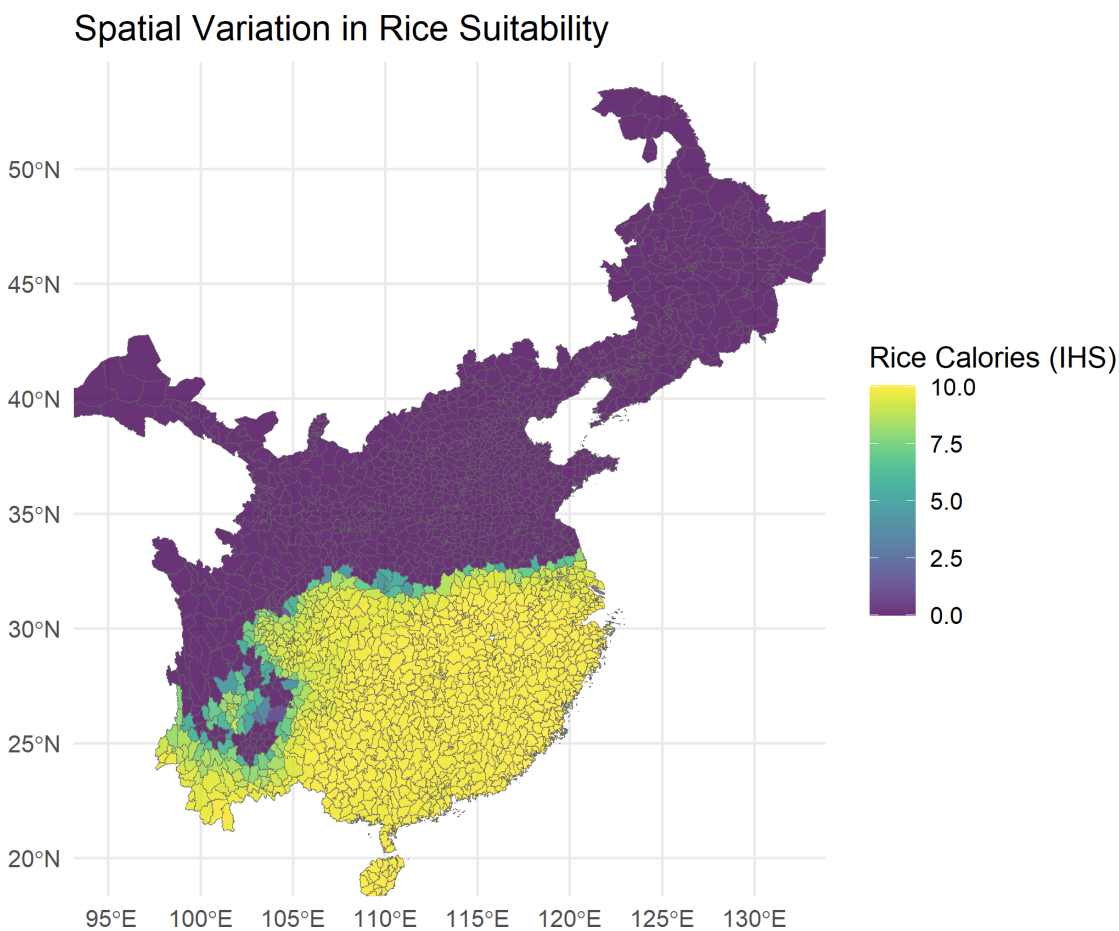

Figure 11: Spatial variation in rice suitability

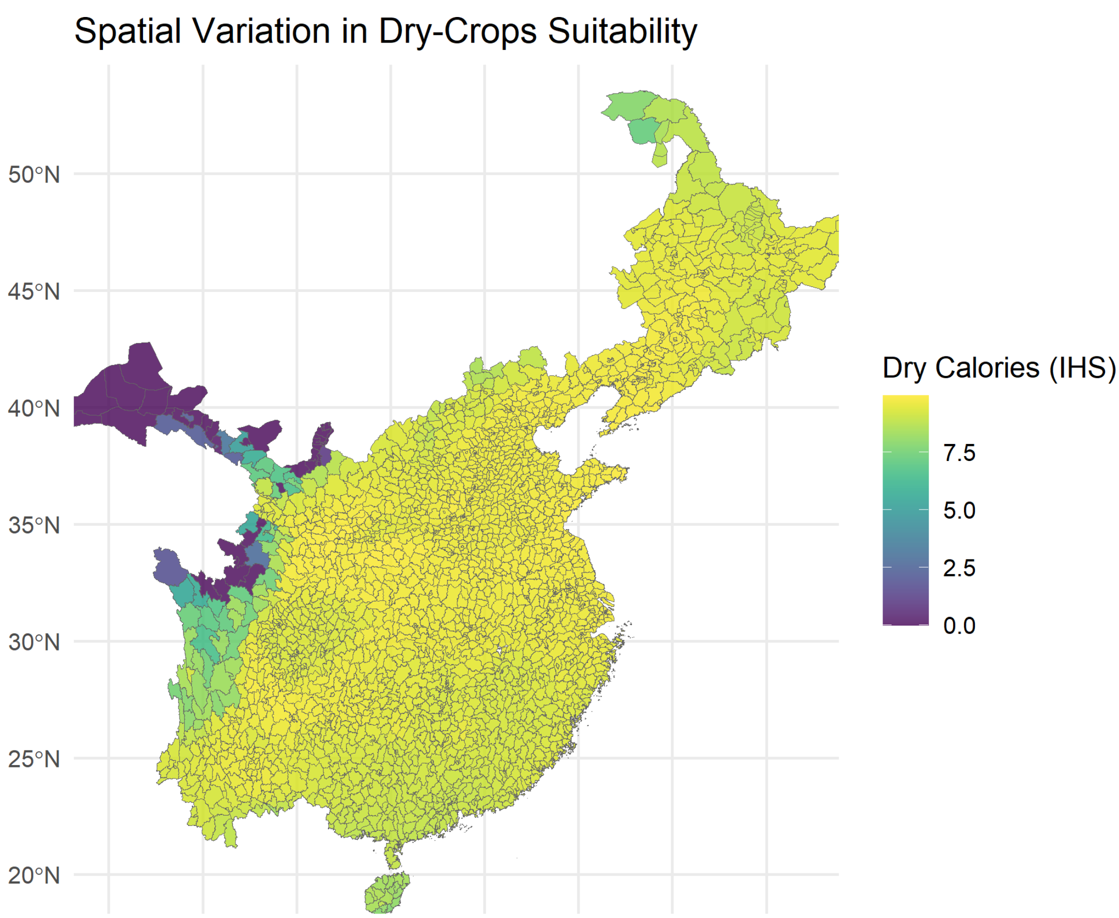

Figure 12: Spatial variation in dry-crop suitability

where $\Omega^1_{i,p}$ represents a vector of length two containing rice and wheat suitability variables (IHS) and $\Omega^2_{i,p}$ denotes precipitation predictability. In Appendix Table 10 we see the same pattern as that identified in the main text. Rice suitability positively correlates with lineage activity and the interaction of rice suitability and precipitation predictability shows a negative effect (precipitation predictability is more negatively associated with lineage activity in rice areas). For wheat there is no main effect or interaction effect with precipitation predictability.

Table 10: Disaggregated Suitability Measure

	Dependent Variable = Lineage Activity (IHS)
	Model 1
Rice x Predictability Interaction	−0.270***
	(0.079)
Dry Crop x Predictability Interaction	−0.003
	(0.434)
Precipitation Predictability	0.244
	(4.107)
Rice Suitability	0.211***
	(0.049)
Dry Crop Suitability	−0.001
	(0.352)
Mean Annual Precipitation (IHS, mm/m^2)	0.568**
	(0.182)
Area (IHS, km^2)	0.343***
	(0.031)
Ruggedness	−0.001+
	(0.000)
Light 1992 (IHS)	0.278***
	(0.039)
Population 1953 (IHS)	0.110**
	(0.040)
Distance to Coast (IHS, km)	0.074*
	(0.032)
Travel Cost to Beijing (IHS)	0.060**
	(0.021)

Table 10: Disaggregated Suitability Measure *(continued)*

	Model 1
Latitude	0.010
	(0.024)
Province FE	X
District Controls	X
Model	Spatial Error
Num.Obs.	2546
R2	0.406
BIC	7732.8

Note:

* p < 0.1, ** p < 0.05, *** p < 0.01

Alternative Rice Suitability Measure

In this section I replace my caloric suitability measure and instead use the FAO GAEZ paddy-rice suitability index directly with no caloric modification (Fischer et al., 2002). I conduct this analysis because as demonstrated in the main text, the caloric-suitability measure for paddy-rice indicates little variation of paddy-rice suitability in South China. The entire region is essentially graded as suitable for paddy-rice. While such a lack of variation reflects the true distribution of rice agriculture in South China, as a sensitivity check, I use an alternative paddy-rice suitability measure in this section. I plot this paddy-rice suitability index (IHS) in Appendix Figure 13.

In **Appendix Table 3** I estimate the following specification:

$$
\begin{aligned}
y_{i,p} = \beta_1 \Omega^1_{i,p} + \beta_2 \Omega^2_{i,p} + \beta_3 \Omega^1_{i,p} \Omega^2_{i,p} + \Gamma X_{i,p} + \delta_p + \mu_{i,p} \\
\mu_{i,p} = \lambda \sum_{i \neq j} w_{i,j} u_{j,p} + \epsilon_{i,p}
\end{aligned}
$$

where now $\Omega^1_{i,p}$ represents a vector of length two containing GAEZ, non-caloric rice and wheat suitability variables (IHS), again disaggregated, and $\Omega^2_{i,p}$ denotes precipitation predictability. We see the same pattern as that identified in the main text and Appendix Table 11.

Table 11: GAEZ Rice Suitability Measure

	Dependent Variable = Lineage Activity (IHS)	
	Model 1	Model 2
Rice x Predictability Interaction	−1.280***	−0.517**
	(0.215)	(0.201)
Dry Crop x Predictability Interaction	0.384	0.326

Table 11: GAEZ Rice Suitability Measure *(continued)*

	Model 1	Model 2
	(0.287)	(0.273)
Precipitation Predictability	−0.362	−1.453
	(1.167)	(1.107)
Rice Suitability	0.987***	0.436**
	(0.143)	(0.136)
Dry Crop Suitability	−0.255	−0.198
	(0.200)	(0.191)
Mean Annual Precipitation (IHS, mm/m^2)	0.399*	0.460**
	(0.171)	(0.175)
Area (IHS, km^2)	0.339***	0.339***
	(0.033)	(0.032)
Ruggedness	0.000	0.000
	(0.000)	(0.000)
Light 1992 (IHS)	0.279***	0.273***
	(0.043)	(0.040)
Population 1953 (IHS)	0.112**	0.097*
	(0.043)	(0.041)
Distance to Coast (IHS, km)	−0.023	0.088**
	(0.029)	(0.032)
Travel Cost to Beijing (IHS)	0.013	0.068**
	(0.018)	(0.021)
Latitude	0.029*	−0.002
	(0.011)	(0.024)
Province FE		X
District Controls	X	X
Model	Spatial Error	Spatial Error
Num.Obs.	2546	2546
R2	0.359	0.404
BIC	7802.7	7744.2

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

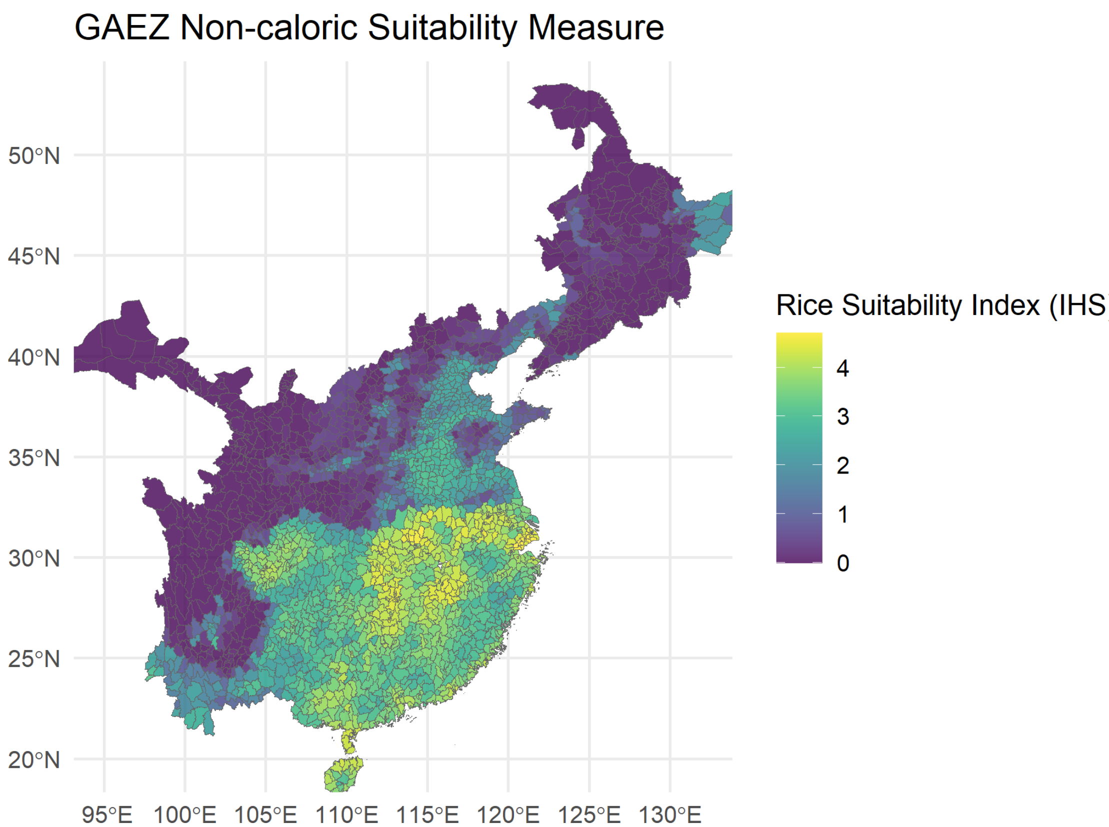

Figure 13: Spatial variation in GAEZ paddy-rice suitability measure.

Prefectural Data and Endogeneity

In this section I again conduct an instrumental variables analyses conducted in the main text in Section however instead of using data at the county level I use prefectural data. Prefectures entail multiple counties and I include this analysis as a simple robustness check. As Figure 14 demonstrates, while the distributions of county and prefectural lineage activity are similar in that they are both right skewed, the county distribution shows a greater proportion of 0 values. In Table 12 I again show that instrumenting the rice suitability measure with its temperature component produces the effects of interest. However, note that when I include the Fournier index and interact it with precipitation predictability, the coefficient of interest loses its effect. As suggested by the erratic and unreasonable coefficients in the second model of Table 12, this is likely because of small sample size and reduced-variation issues due to multicollinearity, the difficulty of detecting the effect of interaction terms, and limited sub-provincial variation. Thus the final regression in Table 12 implements regional instead of provincial fixed effects.

Table 12: Endogeneity Results

	Dependent Variable = Lineage Activity (IHS)		
	Model 1	Model 2	Model 3
First Order Rice x Precipitation Interaction	−0.869*	−0.718	−1.260*
	(0.394)	(0.534)	(0.491)
First Order Rice Suitability	0.651**	0.527	0.832**
	(0.248)	(0.326)	(0.308)
Precipitation Predictability	1.116	42.544+	3.306
	(3.056)	(22.036)	(20.270)
Fournier Index		2.852	−2.851
		(2.279)	(2.000)
Fournier x Precipitation Interaction		−5.721+	0.343
		(3.196)	(2.976)
Mean Annual Precipitation (IHS, mm/m^2)	1.105*	2.425**	3.223***
	(0.514)	(0.911)	(0.781)
Area (IHS, km^2)	0.396***	0.380***	0.169
	(0.109)	(0.109)	(0.127)
Ruggedness	−0.001	−0.001	0.002
	(0.001)	(0.001)	(0.001)
Population 1953 (IHS)	0.640***	0.690***	0.996***
	(0.172)	(0.174)	(0.169)
Distance to Coast (IHS, km)	0.173+	0.121	−0.106
	(0.102)	(0.103)	(0.085)
Travel Cost to Beijing (IHS)	0.027	0.026	−0.116*

Table 12: Endogeneity Results *(continued)*

	Model 1	Model 2	Model 3
	(0.053)	(0.053)	(0.053)
Latitude	0.039	0.049	0.034
	(0.076)	(0.075)	(0.034)
Province FE	X	X	
Region FE	X	X	X
First Stage F Test	18	17	20
Num.Obs.	302	302	302

Note:

* p < 0.1, ** p < 0.05, *** p < 0.01

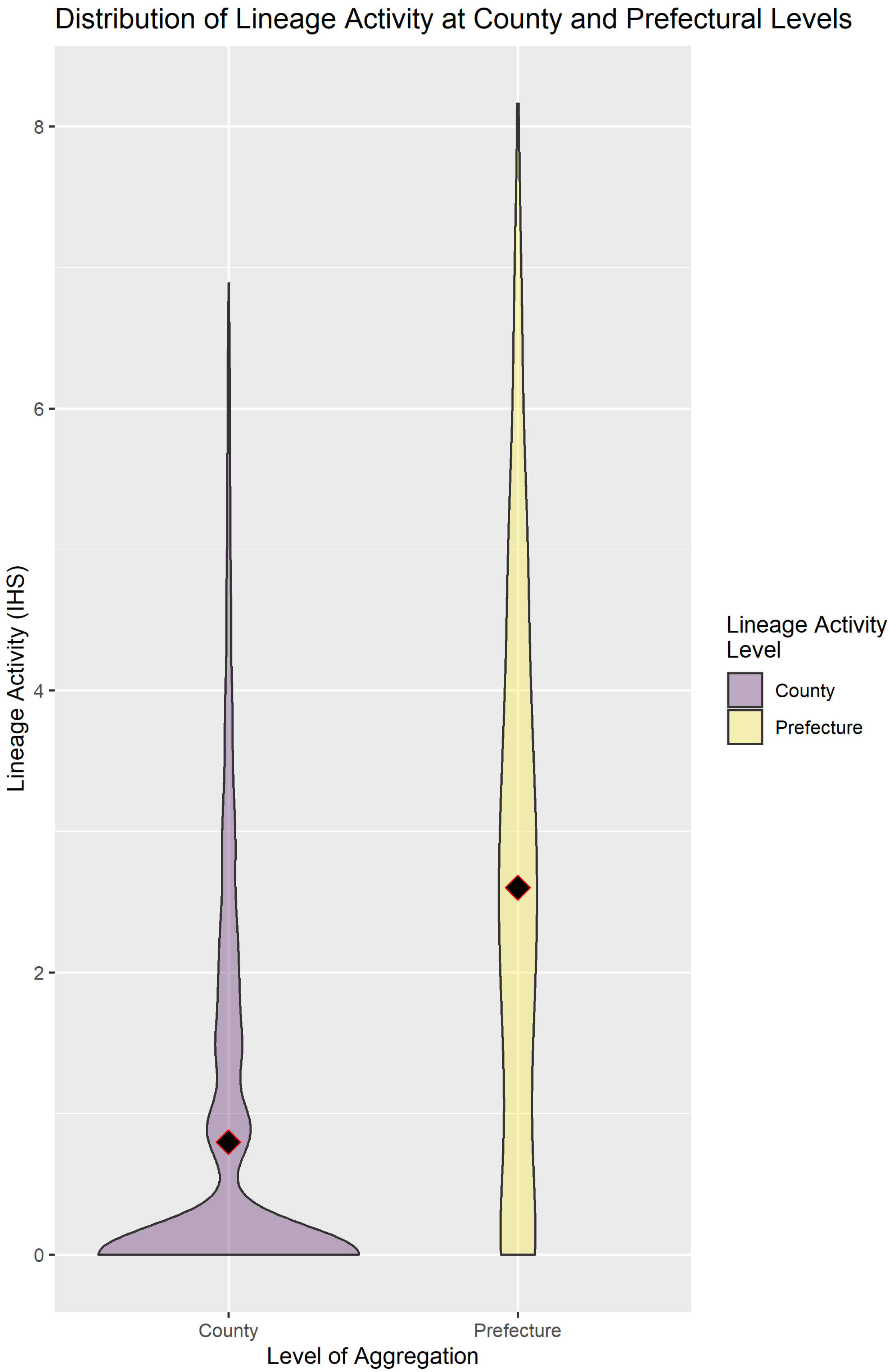

Figure 14: Distribution of lineage activity (IHS) when measured at county and prefecture levels. The red rhombus denotes the mean of each distribution.

References

Ahn, T. K., Balafoutas, L., Batsaikhan, M., Campos-Ortiz, F., Putterman, L., & Sutter, M. (2016). Securing property rights: A dilemma experiment in Austria, Mexico, Mongolia, South Korea and the United States. *Journal of Public Economics*, *143*, 115–124. http://doi.org/10.1016/j.jpubeco.2016.08.015

Alesina, A., & Giuliano, P. (2010). The power of the family. *Journal of Economic Growth*, *15*(2), 93–125. http://doi.org/10.1007/s10887-010-9052-z

Alesina, A., & Giuliano, P. (2011). Family ties and political participation. *Journal of the European Economic Association*, *9*(5), 817–839. http://doi.org/10.1111/j.1542-4774.2011.01024.x

Ashraf, N., Bau, N., Nunn, N., & Voena, A. (2020). Bride price and female education. *Journal of Political Economy*, *128*(2). http://doi.org/10.1086/704572

Balli, H. O., & Sørensen, B. E. (2013). Interaction effects in econometrics. *Empirical Economics*, *45*(1), 583–603. http://doi.org/10.1007/s00181-012-0604-2

Barrett, C., Bolyanatz, A., Crittenden, A., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., … Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*, *113*(17), 4688–4693. http://doi.org/10.1073/pnas.1522070113

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(Supplement 2), 10918–10925. http://doi.org/10.1073/pnas.1100290108

Bray, F. (1986). *The rice economies: Technology and development in Asian societies*. New York: Blackwell.

Bray, F. (2013). *Technology, gender and history in Imperial China: Great transformations reconsidered*. New York, NY: Routledge.

Buck, J. L. (1964). *Land utilization in China: A study of 16,786 farms in 168 localities, and 38,256 farm families in twenty-two provinces in China, 1929-1933*. New York: Paragon Book Reprint Corp.

Chen, S. (2017). The state, the gentry, and local institutions: The Song Dynasty and long-term trends from Tang to Qing. *Journal of Chinese History*, *1*(1), 141–182. http://doi.org/10.1017/jch.2016.30

Cohen, M. L. (1990). Lineage organization in North China. *The Journal of Asian Studies*, *49*(3), 509–534. http://doi.org/10.2307/2057769

Colwell, R. K. (1974). Predictability, constancy, and contingency of periodic phenomena. *Ecology*, *55*(5), 1148–1153. http://doi.org/10.2307/1940366

Cruz, C. (2020). Social network structures and the politics of public goods provision: Evidence from the philippines. *American Political Science Review*, *114*, 486–501. http://doi.org/10.1017/S0003055419000789

Curtin, C. M., Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., … Henrich, J. (2020). Kinship intensity and the use of mental states in moral judgment across societies. *Evolution and Human Behavior*, *41*(5), 415–429. http://doi.org/10.1016/j.evolhumbehav.2020.07.002

da Cruz, R. P., Sperotto, R. A., Cargnelutti, D., Adamski, J. M., de FreitasTerra, T., & Fett, J. P. (2013). Avoiding damage and achieving cold tolerance in rice plants. *Food and Energy Security*, *2*(2), 96–119. http://doi.org/10.1002/fes3.25

Dincecco, M., & Wang, Y. (2020). *Internal conflict and state development: Evidence from Imperial China*.

Dong, J. (2019). Seventy years of changes in China's rural land system: Review, enlightenment and outlook. *International Journal of Rural Law and Policy*, *9*(1). http://doi.org/10.5130/ijrlp.1.2019.6156

Ebrey, P. B. (1986). The early stages in the development of descent group organization. In P. B. Ebrey & J. L. Watson (Eds.), *Kinship organization in late imperial china, 1000–1940* (pp. 16–61). Berkeley: University of California Press.

Elvin, M. (2004). The retreat of the elephants: An environmental history of China. In *Yale university press*. New Haven, Connecticut: Yale University Press.

Fei, H.-T., & Chang, C.-I. (1949). *Earthbound China: A study of rural economy in Yunnan*. London: Routledge. Retrieved from https://arxiv.org/abs/1011.1669v3

Fischer, G., van Velthuizen, H. T., Shah, M. M., & Nachtergaele, F. O. (2002). *Global agro-ecological assessment for agriculture in the 21st century: Methodology and results* ({{IIASA}} Research Report). IIASA, Laxenburg, Austria: RR-02-02.

Freedman, M. (1958). *Lineage organization in southeastern China*. London: Athlone Press.

Freedman, M. (1966). *Chinese lineage and society: Fukien and Kwangtung*. New York: The Humanities Press Inc.

Gächter, S., Herrmann, B., & Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2651–2661. http://doi.org/10.1098/rstb.2010.0135

Galor, O., & Özak, Ö. (2016). Caloric suitability index.

Giuliano, P., & Nunn, N. (2020). Understanding cultural persistence and change. *The Review of Economic Studies*, (November 2020), 1–41. http://doi.org/10.1093/restud/rdaa074

Gluckman, M. (1955). *Custom and conflict in Africa* (pp. ix, 173 p.). Glencoe: The Free Press. http://doi.org/10.1525/aa.1982.84.3.02a00120/abstract

Gluckman, M. (1965a). *Politics, law, and ritual in tribal society*. Oxford: Basil Blackwell.

Gluckman, M. (1965b). *The ideas in Barotse jurisprudence*. New Haven: Yale University Press.

Greif, A., & Tabellini, G. (2010). Cultural and institutional bifurcation: China and Europe compared. *Source: The American Economic Review*, *100*(2), 135–140. http://doi.org/10.1257/aer.l00.2.135

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(6), 795–815. http://doi.org/10.1017/S0140525X05000142

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83. http://doi.org/10.1017/S0140525X0999152X

Herrmann, B., Thoni, C., & Gachter, S. (2008). Antisocial punishment across societies. *Science (New York, N.Y.)*, *319*(5868), 1362–1367. http://doi.org/10.1126/science.1153808

Ho, P. (1959). *Studies on the population of China, 1368-1953*. Cambridge: Harvard University Press.

Hu, H. (1948). *The common descent group in China and its functions*. New York, NY: Columbia University.

Hu, Z.-Z., Bengtsson, L., & Arpe, K. (2000). Impact of global warming on the Asian winter monsoon in a coupled GCM. *Journal of Geophysical Research: Atmospheres*, *105*(D4), 4607–4624. http://doi.org/10.1029/1999JD901031

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

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

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Laland, K. N., & O'Brien, M. J. (2011). Cultural niche construction: An introduction. *Biological Theory*, *6*(3), 191–202. http://doi.org/10.1007/s13752-012-0026-6

Maine, S. H. S. (1861). *Ancient law: Its connection with the early history of society, and its relation to modern ideas*. London: John Murray.

McDermott, J. P. (1999). Emperor, elites, and commoners: The community pact ritual of the late Ming. In *State and court ritual in China* (pp. 299–351). Cambridge, U.K.: Cambridge University Press.

McDermott, J. P. (2013). *The making of a new rural order in South China: I. Village, land, and lineage in Huizhou, 900-1600* (Vol. 1). Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9781107070455

Noellart, M. (2020). *Power over property: The political economy of communist land reform in China*. Ann Arbor, MI: University of Michigan Press.

Nunn, N. (2009). The importance of history for economic development. *Annual Review of Economics*, *1*(1), 65–92. http://doi.org/10.1146/annurev.economics.050708.143336

Nunn, N. (2020). The historical roots of economic development. *Science*, *367*(6485), eaaz9986. http://doi.org/10.1126/science.aaz9986

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *The Review of Economics and Statistics*, *94*(1), 20–36.

Pasternak, B. (1969). The role of the frontier in Chinese lineage development. *The Journal of Asian Studies*, *28*(3), 551–561. http://doi.org/10.2307/2943178

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, Anja Rudolf, B., & Ziese, M. (2020). GPCC full data reanalysis version 2020 at 0.25°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. http://doi.org/10.5676/DWD_GPCC/FD_M_V2020_025

Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science*, *366*(6466). http://doi.org/10.1126/science.aau5141

Scott, J. C. (1976). *The moral economy of the peasant: Rebellion and subsistence in Southeast Asia*. New Haven, Connecticut: Yale University Press.

Searle, J. R. (1995). *The construction of social reality*. New York: The Free Press.

Shi, L. (2009). Chinese household income project, 2002. Inter-university Consortium for Political and Social Research [distributor]. http://doi.org/10.3886/ICPSR21741.v1

Shi, L., Chuliang, L., Zhong, W., & Ximing, Y. (2008). Appendix: The 1995 and 2002 household surveys: Sampling methods and data description. In B. A. Gustafsson, L. Shi, & T. Sicular (Eds.), *Inequality and public policy in China* (pp. 337–353). New York: Cambridge University Press.

Szonyi, M. (2002). *Practicing kinship: Lineage and descent in late imperial China*. Stanford, California: Stanford University Press.

Szonyi, M. (2017). *The art of being governed: Everyday politics in late imperial China*. Princeton, New Jersey: Princeton University Press.

Talhelm, T., & English, A. S. (2020). Historically rice-farming societies have tighter social norms in China and worldwide. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(33), 19816–19824. http://doi.org/10.1073/PNAS.1909909117

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science*, *344*(6184), 603–608. http://doi.org/10.1126/science.1246850

Vollan, B., Landmann, A., Zhou, Y., Hu, B., & Herrmann-Pillath, C. (2017). Cooperation and authoritarian values: An experimental study in China. *European Economic Review*, *93*, 90–105.

http://doi.org/10.1016/j.euroecorev.2017.01.007

Wakefield, D. (1998). *Fenjia: Household divison and inheritance in Qing and Republican China*. Honolulu, Hawaii: University of Hawai'i Press.

Wang, L., Yang, Z., Gu, X., & Li, J. (2020). Linkages between tropical cyclones and extreme precipitation over China and the role of ENSO. *International Journal of Disaster Risk Science*, *11*(4), 538–553. http://doi.org/10.1007/s13753-020-00285-8

Wang, P. K., Lin, K. H. E., Liao, Y. C., Liao, H. M., Lin, Y. S., Hsu, C. T., … Ting, T. T. (2018). Construction of the reaches climate database based on historical documents of China. *Scientific Data*, *5*, 1–14. http://doi.org/10.1038/sdata.2018.288

Wang, Y. (2020, April). Comprehensive catalogue of chinese genealogies. Harvard Dataverse. http://doi.org/10.7910/DVN/PO0VF6

Watson, J. L. (1982). Chinese kinship reconsidered: Anthropological perspectives on historical research. *The China Quarterly*, *92*, 589–622.

Xie, Y., & Hu, J. (2014). An introduction to the China family panel studies (CFPS). *Chinese Sociological Review*, *47*(1), 3–29. http://doi.org/10.2753/CSA2162-0555470101.2014.11082908

Xu, Y., & Yao, Y. (2015). Informal institutions, collective action, and public investment in rural China. *American Political Science Review*, *109*(2), 371–391. http://doi.org/10.1017/S0003055415000155

Yang, M. C. (1945). *A Chinese village: Taitou, Shantung province*. New York, NY: Columbia University Press.

Zhang, D. (2013). *A compendium of Chinese meteorological records of the last 3,000 years*. Nanjing: Phoenix Publisher House.

Zhihong, S. (2017). *Agricultural development in Qing China: A quantitative study, 1661-1911*. Leiden: Brill.