Family Clans and Public Goods: Evidence from the New Village Beautification Project in South Korea*

Hyunjoo Yang†

February 11, 2016

Abstract

Ethnic and linguistic heterogeneity are widely studied as determinants of social capital, conflict, and institutional quality. In many cultures, another important dimension of heterogeneity is family clan membership. I study the relationship between family clan diversity in South Korean villages and the voluntary production of public goods and contributions of private resource for village projects. Under the 1970-1971 New Village Beautification Project, the government distributed resources to each village for the production of village public goods. Subsequently, the government systematically evaluated how well these resources were applied. I combine these data with information on village family clan structures collected by the Japanese Colonial Government, as well as records of land donations for village projects between 1970 and 1980. I find an inverted-U-shaped effect of group heterogeneity on the improvement of public goods and on the average amount of donated land per household. I suggest that the concave relationship reflects the trade-off between better coordination among clan members and less accountability of clan leaders as village clan homogeneity increases.

JEL Classification: O12, N25

Keywords: Social Capital, Family Clans, South Korea

*I thank my thesis advisors, Nathaniel Baum-Snow, Stelios Michalopoulos, and David Weil for their advice and suggestions. I also thank Pedro Dal Bo, Andrew Foster, Oded Galor, J. Vernon Henderson, Kanghyock Koh, Sriniketh Nagavarapu, Louis Putterman, and participants at Brown Urban Lunch Seminar, Brown Macro Lunch Seminar, and Brown Macroeconomics Seminar for their comments. I thank Joonwoo Bae, Chia-En Hsieh, Min G. Kim, Pei-fang Wang, and Dahae Yang for their research assistance. I also thank Hyunjoo Lee from Korea Saemaulundong Center, Hoseong Wang from the National Archive, and Taeyoung Ryu for their help on data collection. All errors are mine.

†Brown University, Department of Economics, 64 Waterman Street, Providence, RI 02912, USA (email address: hyunjoo_yang@brown.edu).

1 Introduction

Heterogeneous communities tend to have worse economic outcomes than homogenous ones. Researchers have found that heterogeneous communities have more conflict, and have fewer public goods, and that members exhibit less trust and interact less often (Knack and Keefer, 1997; Collier and Hoeffler, 1998; Alesina and La Ferrara, 2000; Miguel and Gugerty, 2005).[1]

Many scholars have investigated the roles of ethnic, linguistic and religious heterogeneity in communities (Mauro, 1995; Alesina et al., 2003; García Montalvo and Reynal-Querol, 2005). In this paper, I introduce family clans as a different, but complementary dimension of heterogeneity in a society. In many societies, family clans play significant economic and political roles. A family clan is broadly defined as a group of people who share the same paternal lineage. In Iraq, for example, individuals are more strongly bounded by clan and tribal ties than by ethnic or religious background (Hassan, 2007).[2] Although Somalia is the most ethnically homogenous country in Africa, clan warfare devastated the country in the 1990s (Arnold, 2001). In Syria, armed tribes and clans have actively participated in its current civil war (International Crisis Group, 2015). Historically, cleavages between Scottish clans heightened during civil wars, such as the 1689-1745 Jacobite Risings (Barthorp and Embleton, 1982). Likewise, a single clan in Uzbekistan wielded considerable political influence by occupying many important ministry seats (Collins, 2002).

I use family clan organization in South Korean villages to study the effect of village clan heterogeneity on the production of public goods. Studying family clans in Korea offers a number of advantages. First, South Korea has little variation in other important dimensions of heterogeneity explored in the literature such as ethnicity, race, language, and landholding, thereby enabling the investigation of one dimension of group heterogeneity (i.e., family clans).[3] Second, family clans are located in villages with typically comprised of about 100 households on average. Cooperation, social sanctions, and information transmission, some key ingredients of social capital, could be

[1] See Costa and Kahn (2003) for an in-depth review.

[2] Immediately after Iraqi independence, tribes were estimated to have 100,000 rifles in their possession, whereas the government had 15,000 (Marr, 2011).

[3] As an additional advantage is that, unlike ethnicity or race, clan membership is easily identifiable. In general, a Korean individual can easily identify his or her clan by the family name and the ancestor's place of origin.

prevalent in such small communities.[4] Third, as I show in this paper, there exist excellent data on both clan composition and public good outcomes.

To measure family clan heterogeneity, I use family clan data from *Family Clans in Chosun*, part of the 1930 census conducted by the Japanese Colonial Government in Korea. These village-level cross-sectional data include the clan names and the total number of households belonging to every clan in each village in South Korea, as long as the share of clan households exceeded 10% of total village households. Using the share of clan households in each village, I construct heterogeneity measures based on the percentage of the households belonging to the most dominant clan in a village, the Herfindahl Index of clan concentration, and a polarization index suggested by García Montalvo and Reynal-Querol (2005) to capture clan polarization in a village.

The identification strategy of this paper relies on the historically determined settlement of family clans. Family clans settled in Korea hundreds of years ago. By controlling for family clan identities and township fixed effects, I eliminate alternative explanations that the results are driven by the presence of specific family clans or unobserved characteristics that vary across townships. Furthermore, the estimates are robust to various controls for potential determinants of village clan structure, such as distance to town centers, rivers, and roads, terrain ruggedness, altitude and soil type.

In my first empirical analysis, I use the voluntary production of public goods as an outcome variable. This measure originates from government evaluations of how well government resources were used to produce village public goods. Under the New Village Beautification Project (henceforth, NVBP) in 1970, the government distributed 335 bags of cement to every village in South Korea, irrespective of village characteristics. These bags of cement were dropped at the entrance to each village for the production of public goods. In consultation with village elites who were members of village development committees, residents decided how to use cement, and produced public goods. As only cement was provided, village residents contributed voluntary labor and private resources. Unbeknownst to the village residents when they received the bags of cement, the government decided to evaluate how much public goods had been improved with a grade (A, B or

[4] Due to data limitations, U.S. metropolitan statistical areas (MSAs), which have millions of people, are often used in the social capital literature to study community-level phenomena.

C) the following year. I digitized the village evaluations and other village characteristics from the *New Village Comprehensive Survey* (NVCS), which was published by the government in 1972.

I find that there was a robust relationship between family clan heterogeneity and the production of village public goods. The relationship, however, was nonlinear. Using the household share of the largest clan in a village as a heterogeneity measure, I show that there was a concave relationship between the family clan heterogeneity and the probability of getting an A grade. The concave curve peaked when the share of the largest clan in a village is around 40%. Using the Herfindahl Index yields a similar inverse-U-shaped relationship. Additionally, the polarization index is positively associated with improvement of public goods and the coefficient is statistically significant. A change from no polarization to the maximum level increases the probability of getting an A grade by 4 percentage points.

In the second set of empirical analyses, I use land donation as a measure of private contributions for village projects. Village residents donated private land to enable the production of public goods, such as village roads. I digitize *Glorious Footsteps*, a government publication documenting village public projects in detail between 1970 and 1980. It contains a donation list with the name of each donor, and the amount and type of land donated. I show that the amount of land donated per village household during the 1970s also has a concave relationship with the family clan heterogeneity. When the household share of the largest clan in a village is used as a heterogeneity measure, the concave curve peaks around 0.6.[5] I also find a positive association between the polarization index and land donation.

My last findings relate to agricultural mechanization as a consequence of village improvements, particularly village roads. Many anthropologists and sociologists have found that village residents widened village roads in the 1970s to facilitate the use of wheeled agricultural machines in order to saves labor costs.[6] If villages successfully improved village roads through village projects that relied on private contributions such as land donations, it is likely that they would have utilized

[5]As an alternative outcome measure, I use the total amount of donated land as well as the amount of donated land as a percentage of the total amount of cultivated land in a village. The results are qualitatively the same.

[6]It has been widely documented that demand for wider village roads due to increasing wages and the introduction of the power tiller, the most popular agricultural machine in Korea. In a survey, farmers listed road improvement projects as the highest priority at that time (Park, 1998).

more wheeled agricultural machines. Based on township-level agricultural census data from 1970 and 1980, I show that, 10 years after the NVBP, the change in the number of two-wheeled power tillers per agricultural household between 1970 and 1980 has a positive and statistically significant relationship with clan heterogeneity. The polarization index is positively related with changes in the number of power tillers.

The persistent concave relationship between family clan heterogeneity and the production of and contributions toward public goods may reflect the trade-off between coordination and accountability. Scholars have found that in more homogeneous societies, members coordinate better and contribute more to public goods (Miguel and Gugerty, 2005). Yet other studies have shown that highly homogenous communities lack competition, enabling elites to waste resources in the absence of checks and balances (Platteau and Gaspart, 2003; Acemoglu et al., 2013). In traditional societies without strong political institutions, highly homogenous communities may be likely to have both a high level of coordination *and* highly autocratic elites. In Korea, clan elders have considerable power in clan and village matters (Lee, 1997; Bang, 2004; Kim, 2009), and there has been, and still is, no formal government or political organization in villages.[7] Therefore, as clan homogeneity increases, the provision of public goods may improve since such communities have better coordination. However, as communities become highly homogeneous, the provision of public goods may worsen because clan elites could be more prone to despotism with no institutional countervailing force. This reasoning is also consistent with the positive relationship between the polarization and public goods found in my data.

My study is most closely related to the literature on the effects of group heterogeneity on social capital. Ample evidence shows that group heterogeneity is detrimental to economic and social outcomes such as investment, corruption, the chace of civil war, provision of public goods, group participation and trust (Mauro, 1995; Collier and Hoeffler, 1998; Alesina et al., 1999; Glaeser et al., 2000; García Montalvo and Reynal-Querol, 2005). To my knowledge, this is the first paper in which family clans are used as a measure of group heterogeneity.

The results of this paper also complement economic literature on the effects of kinship groups

[7]Public goods such as schools or medical clinics are all located in township districts.

on economic outcomes such as insurance, information sharing, resource pooling, and credit access (Rosenzweig, 1988; Besley et al., 1993; Munshi, 2003; Fafchamps and Gubert, 2007).[8] Additionally, this paper contributes to the role of group heterogeneity and social capital on agricultural modernization. Most related is the work of Isham (2002), who showed that the adoption of improved fertilizer is positively associated with village level ethnic homogeneity and household participation in village organizations.

The rest of the paper is organized as follows. In the next section, I describe the context of the study; specifically, I provide detailed background information on family clans and the NVBP, a rural intervention program aimed at the improvement of public goods in South Korean villages. I also provide a conceptual framework to explain the relationship between clans and social capital. In section 3, I explain my data collection. In section 4, I discuss empirical strategies before presenting and discussing the regression results in section 5. In section 6, I present a case study before offering some concluding remarks in section 7.

2 Background

2.1 Family Clans

Clans have fulfilled important economic or political functions in many regions of the world. When governments are weak, clans often assume government functions such as settlement of disputes and protection of property and members from outsiders. In South Korea, clan membership is not only fundamental in traditional social interaction, it also offers many benefits such as mutual support in farming and in emergencies. The common ancestor is worshipped collectively. Clan membership also increases emotional security (Song, 1982).

A family clan in Korea is defined as a group of people who share the same paternal lineage. Clan membership can be identified by a family name and the ancestor's place of origin. Every

[8] See Cox and Fafchamps (2007) for a detailed overview of economic literature on kinship networks. Sociologists also have studied the effects of family and relatives on: out-migration (Palloni et al., 2001); protecting property rights (Peng, 2004); technology adoption (Warriner and Moul, 1992); and coping with long-term personal emergencies (Litwak and Szelenyi, 1969).

native born Korean belongs to a clan. Unlike ethnicity or religion in other countries, however, there are no politically dominant clans in national politics or violent inter-clan conflicts. Further, Korean villages have extremely low ethnic or linguistic fractionalizations. According to measures by Alesina et al. (2003), the values of ethnic and linguistic fractionalization are both 0.002, one of the lowest values among the countries across the world (see Figure 1).

Family clans in a village are often highly concentrated in rural villages. Often the name of a village reflects the dominant clan (e.g., Kim's Village) (Yu, 1986). Villages are often classified by the presence of clans: "the most fascinating persistent example of Korean kin organization is the consanguineous village" (Jacobs, 1985, p. 212). Jacobs (1985) classified villages into four types: (a) all residents are from a single lineage; (b) a single lineage dominates; (c) the residents from strongly competitive lineages; and (d) the residents are from a number of weak lineages. In my data, out of 1,298 villages, there are 23 villages (2%) in which all residents were from a single clan, 143 villages (11%) with one dominant clan (i.e., $>$50% of members), 163 villages (13%) with large concentrations of two or more clans (i.e., no dominant clan), and 969 villages (75%) with weak lineages.[9]

A survey performed by the Japanese Colonial Government in the 1930s indicates that out of 1,227 sampled villages with large concentrations of dominant clans, the histories of 17% could be traced back 500 years, 53% between 300 and 500 years, 28% between 100 and 300 years, and just 2% less than 100 years (Korean Studies Advancement Center, 2014).[10] These kinship ties remained strong in the early 1970s, even when non-rural areas were experiencing rapid political and economic changes (Kim, 1985). In 1930, there were more than 660 different family clans.[11]

One explanation for the high concentrations of clans in villages is the change in the inheritance law in the 17th century (Korean Studies Advancement Center, 2014). Prior to 17th century, a father

[9]For calculations, I use villages that had not split between 1930 and 1970. I use the polarization index to identify villages with large concentrations of two or more large clans and no dominant clan. To obtain the number 163, I count villages in which the largest clan share does not exceed 50% and the polarization index is above 0.3, roughly the 75th percentile of the sample.

[10]The long settlement histories of family clans may show that the initial characteristics that attracted founding ancestors 300 years ago may not have been relevant in the early 1970s, the study period of this paper. Banerjee and Somanathan (2007) used similar argument to justify using historical caste compositions as a regressor over contemporary ones. However, the concern of differential geographical endowment remains. Thus I control extensively for geography related variables.

[11]My own calculation based on the family clan data-set used in this paper.

could leave his land equally to both sons and daughters. This enabled daughters to remain in the villages where they were born and their husbands from different clans to move in. Since Korea is a patrilineal society, daughters, once married, were considered to be members of their husbands clans. Hence, different clans could reside in the same village. Once the changed inheritance law excluded daughters from being inheritors, they tended to leave their native villages to live with their husbands.[12] This prevented the inherited land from being owned by outsiders and therefore reduced the inflow of people from different clans.[13]

Major factors influencing clan settlements were historical. While few systematic studies have been performed, the concentrations of clans in certain locations can be attributed to: (a) relocation due to wars, (b) settlement on land gifted by kings, (c) settlement of loyal families near the tombs of ancestors, and (d) settlement of families of retired senior government officials (Jacobs 1985; Korean Studies Advancement Center 2014). The settlement of elite members raises the concern that the higher quality clans may be concentrated in villages that are conducive to better provision of public goods. In the analysis, I control for clan identities to rule out the alternative explanation that differential characteristics of clans are driving my results.

2.2 New Village Beautification Project

In this section, I describe the unique Korean rural development policy that led to the production of village public goods in the early 1970s. The South Korean government distributed 335 bags of cement to every village between 1970 and 1971 as part of the New Village Beautification Project (NVBP). The purpose of this project was to encourage village residents to produce public goods. As each bag of cement weigh about 40 kilograms, each village thus received a total of 13.4 tons of cement (Ministry of Home Affairs, 1983, p. 22). Cement was distributed between October 1970 and June 1971, so that villages could take advantage of labor availability during the agricultural off-season (Hwang, 1980).

[12] A family with no sons often adopted a son from relatives in order to bestow land, and more importantly, continue the family paternal line and ancestor worshipping duties.

[13] In future work, I plan to exploit the timing of the change in the inheritance law to investigate the possibility of exogenous variation in village clan compositions.

An important aspect of this project was that each village could decide how to use the cement, as long as it was used to produce public goods for the village. Cement was to be used for "village projects meeting villagers' common needs based upon their general consensus" (Moore, 1984, p. 587). The government suggested several potential uses for the cement. For example, villages could improve village roads, repair river embankments, build compost/manure collection points, repair public wells or construct common laundry facilities (Kyunghyang, 1970). To decide how the cement would be used, local government officials encouraged villages to create village development committees with five to 10 members each. Since public projects often required substantial land and labor contributions from village members, decisions on village projects, such as the widening of roads, were made in democratic ways, such as voting by a show of hands during the village meetings. Typically, the eldest male from each household attended these meetings (Park, 1998).

In the year following the distribution of the cement under the NVBP, the government systematically evaluated each village on how well the cement had been used to improve village public goods (see Table A1 for the timeline of the NVBP). Each village was given either an A, B, or C grade.[14] Township or county government officials visited villages and assigned grades. Using cement for mostly private projects (e.g., paving kitchen floors, building stone fences around houses) resulted in getting a C grade. An A or B grade was given to villages that used the cement to produce public goods. For example, some villages widened and straightened village roads. Others established new village roads. Some villages fixed sewage and drainage pipes. Some created common laundry facilities or village wells.

A large number of villages chose to improve transportation infrastructure, particularly roads. A government survey of villages shows that improving transportation was considered to be a top priority by village residents. Table A2 shows that the three most desired projects identified by village households related to improving roads and fixing bridges. Farmers wanted to use more agricultural products for farming and transportation. However, the roads were not wide or straight enough to use wheeled machines. Furthermore, villagers wanted to improve roads so that they could access modern modes of transportation such as trucks, buses, taxies or cars (Ministry of Home Affairs,

[14]The original classifications were independent village, self-help village, and basic village. I have replaced these labels with grades A, B, and C, respectively.

1978). Villagers provided private resources to improve village road infrastructure. In 1973, the largest fraction of labor days spent on village projects were dedicated to road improvement. According to national statistics, 29 million out of a total of 36 million labor days (81%) were spent on village roads (Ministry of Home Affairs, 1973, p. 116).

The considerable effort and contributions put toward improving village roads in the 1970s was partly due to increasing agricultural wages and the introduction of a new labor-saving agricultural machine, the two-wheeled power tiller. Prior to the introduction of power tillers–the most popular agricultural machine in the history of Korean agriculture–there was little incentive to build wider, straight roads because traditional technology did not require them. Figure A2 shows the trend of the number of wheeled agricultural machines per agricultural household. Prior to 1970, there were few households that owned these machines. Starting 1970, there was a rapid adoption of power tillers. Power tillers was highly popular: in 2000, there was 0.7 power tillers per every agricultural household in the country. Before 1970, farmers tilled land using animal power, and residents mostly carried goods on their backs (see Figure A3). While I do not have information on the initial quality of village roads, given the nature of traditional agricultural technology in Korean villages, initial road quality, particularly the width of roads, did not seem to differ based on clan structure and other village characteristics.

Based on personal interviews with several village elders who participated in the cement projects, most village residents did not seem to be aware that their projects would be evaluated and that they could earn rewards based on how well they used the cement to produce public goods. In the year following the cement distribution, additional resources were given to villages that had received A or B grades. However, this decision had not been planned in advance. After seeing the grade distribution for the cement projects, President Park Chung-hee, who had initiated the NVBP, suggested providing additional resources only to villages that had received A or B grades (Kim, 2006).

2.3 Family Clans and Social Capital

Why does family clan heterogeneity matter for the production of village public goods and hence for village social capital? I argue that the benefits of participating in village projects and the costs of free riding are both higher if there are more members of the same clan in a village.

First, the incentive of a village resident to participate in village projects may have depended on the increased benefit, either in utility or profit, to the participating individual and on the increased benefit to relatives. Therefore, if more residents of a village were related by blood, each individual receives more marginal utility, ceteris paribus. This may have been especially true in the Korean rural context, because clan members exhibited strong social integration, and had a sense of mutual solidarity (Brandt, 1972). In field interviews, Yi (1981) described the positive role of clans in village projects:

> Of particular importance, according to people interviewed, was the question of clan. The villages dominated by one clan-group evidently had an easier time getting their residents to cooperate in various...[village] projects; where two or more clans were present, this task was more difficult. (pp. 448-449)

Additionally, improving durable public goods by building new roads or bridges benefits not only current clan members, but also future generations, which may have been an incentive to contribute to infrastructure projects. An example from a single clan-dominated village illustrates this point. Among the households comprising Ho-am village in Kyungsang North Province, 95% belonged to the Milyang Park Clan in the 1970s. When land owners were reluctant to donate the lands required to widen village roads, a clan member Kyusam Hong said, "even if we are poor now, let us make our younger generations praise us for giving them better roads" (Ministry of Home Affairs, 1978, p. 552).[15] Clans also provide better coordination for labor-intensive agriculture and resolution of conflict (Seo, 1997), which often are mediated by clan elders before they get out of hand (Yesa Moonhan, 2001).

Second, the cost of free riding is higher when there are more clan members in a community. Social punishment can be strong when members of the same kinship group have semi-permanent relationships and repeated face-to-face interactions in a village, which is the case in clans in South

[15] If being dominated by a single clan is advantageous to public goods improvement, one may expect that by 1970, these villages would already have had better public goods. However, villages had received few government resources prior to the NVBP. Historically the government had provided resources for the production of public goods located primarily in town centers, not in villages. Even irrigation, an important village public good, mostly relied on proximity to rivers and rainfall, instead of planned irrigation systems.

Korea. Collective ancestor worship ceremonies are performed multiple times each year. Additionally, socially unacceptable behavior may damage not only the reputation of the wrongdoer, but also the reputations of immediate family members, such as parents.[16] Even if individuals did not care about the benefits to current and future clan members when making participation decisions, it is still reasonable to believe that the social costs may still have had a strong influence in rural Korean societies.

Social sanctions have been shown to affect various economic outcomes in rural communities in other countries as well. In rural Ghana, La Ferrara (2003) provide evidence that social sanctions influence the loan default rate. Likewise, Miguel and Gugerty (2005) show that ethnically diverse communities in Kenya fail to impose sanctions on parents who do not contribute to school funding.

Lineage groups in other countries also exhibit strong solidarity. For example, in Chinese villages, "lineage groups inculcate a sense of obligation to the group...based on concepts of family and shared patrilineal descent" (Tsai, 2007, p. 359). Mu and Giles (2014) argue that "mutual trust between villagers built through common family lineage may lead to less conflict" (p. 21). In both the U.S. and Hungary, Litwak and Szelenyi (1969) find that relatives are the most helpful group when an individual is dealing with a long-term emergency such as a broken leg that takes 3 months to heal. Among American survey respondents, 73% indicated that relatives would help "very much" in the broken leg scenario, whereas only about 30% indicated that neighbors and friends would help "very much."

In contrast, villages with heterogenous clan membership could face cooperation challenges. Anecdotal evidence suggests that villages with multiple clans may be less cooperative. For example, one village had "forty different family names and it was difficult for them to cooperate" (Ministry of Home Affairs, 1978, p. 625). Likewise, members of another village were described as "selfish because there are numerous families with different clans" (Ministry of Home Affairs, 1978, p. 402).

[16] Posner (1980) described this kind of behavior as *collective responsibility*.

3 Data

3.1 Family Clans and Village Heterogeneity

For clan membership data, I digitized a publication called *Family Names in Korea* which was part of the 1930 population census by the Japanese Colonial Government. The publication lists family clans in villages if the number of households associated with a single family clan comprised more than 10% of total households in the village. In some villages, more than one clan met the 10% threshold. For each village in South Korea, data include the number of households belonging to a specific clan and the name of each clan, including ancestral place of origin and family name.

Figure 3 shows the original version of the cross-sectional data by the Japanese Colonial Government. Another survey on clan membership was conducted in 1985 as a part of a population census. However, clan data in 1985 are available only at the township level, and not at the village level. Townships comprise the lowest administrative division in South Korea, and each village belongs to a township. Each village may also include multiple hamlets, which are concentrated pockets of dwellings within a village.

According to the data, out of 3,124 villages: no single clan comprised at least 10% of the households in 1,708 villages (55%); at least one clan comprised at least 10% of the households in 1,048 villages (34%); and two or more clans each comprised at least 10% of the households in 368 villages (11%). The mean of the share of households belonging to the largest clan in the village is 16%, with a standard deviation of 0.23. The median share of the largest clan is 0, and the 75th percentile is 0.28.

I used multiple measures to capture family clan heterogeneity in a village. First, I used the share of households belonging to the largest clan among the total households in a village (TOPSHARE). Figure 4 shows the distribution of TOPSHARE. Second, I constructed the Herfindahl Index (HERF) to measure the family clan concentration within a village. The index is calculated by the formula

$$HERF_i = \sum_{i=1}^{n} share_i^2, \tag{1}$$

where $share_i$ is the share of clan i in a village, and n is the number of clans exceeding the 10% threshold. A higher index value implies less diversity in clan and higher concentration of a single clan. Figure 7 shows the spatial distribution of the index.

To capture the effect of the presence of multiple clans within a village, I used a polarization index (POLAR) following García Montalvo and Reynal-Querol (2002). The index is computed using the following formula:

$$POLAR_i = 1 - \sum_{i=1}^{n} \left\{ \frac{0.5 - share_i}{0.5} \right\}^2 share_i. \tag{2}$$

This index captures the strength of polarization between clans. For example, a village with two clans that each comprised 30% of households would have a higher index value than a village with a single clan that comprised 60% of households. An index value of 1 implies maximum polarization– two clans each comprising 50% of households in a village. An index value of 0 means there was a single clan comprising 100% of households, or there were multiple clans that each comprised a minimal percentage of households in a village. Figure 8 shows the spatial distribution of the index.

Table 3 shows correlations between homogeneity measures. All three measures are highly correlated. TOPSHARE and HERF have a correlation coefficient of 0.92. TOPSHARE and POLAR have a correlation coefficient of 0.83. HERF and POLAR have the least correlation, at 0.64.

3.2 Economic Outcomes

3.2.1 Public Goods Data

Village characteristics and information on the production of public goods come from the *New Village Comprehensive Survey* (NVCS) published by the Department of Home Affairs in 1972. This government publication includes cross-sectional data from 1971, including names, demographic information, and other characteristics of entire villages in South Korea. Important for this study, the NVCS includes the cement project grades for all villages in South Korea. A total of 16,301 villages received A or B grades, closely matching the official figure of 16,600 villages according to news outlets at that time.[17] Figure 9 shows the original format of the data. Each page of the NVCS has a list of the villages in each township. A map of each township is also included, which shows the village boundaries. Each map also shows the locations of important landmarks such as electrification, rivers, roads, railroads and highways. I digitized these variables from each township map. Village characteristics used in the empirical analysis of this paper are all from the NVCS. Table 1 provides summary statistics.

While data are available for the universe of villages in South Korea, I restricted the sample to Kyungsang North Province for the analysis (see Figure A1). Kyungsang North Province is the largest province in South Korea with an area of 19,028 square kilometers (7,347 square miles) and a population of 2.6 million in 2010.[18] Kyungsang North Province has the most concentrated settlement of clans and hence the most variations in clan structures in villages (Kim, 2012).

I used the grades assigned to each village under the NVBP as a measure of the production of public goods by constructing a dummy variable that equals 1 if the grade is an A, and 0 otherwise. I also constructed an alternate dummy variable that equals 1 if the grade is either an A or B, and 0 otherwise. Of the 5,539 villages included in the NVCS that meet my sampling criteria (see Section 3.3), 357 villages received an A grade (6%), 2,417 villages received a B grade (44%), and 2,765 villages received a C grade (50%). Figure 10 shows the spatial distribution of the cement project grades.

I used the probability of getting an A cement project grade as a measure of road improvement. There are several reasons why this may be sensible. As discussed in detail in the background section, the most desired village project was road improvement and more than 81% of total labor hours contributed to village projects were related to road improvement (Ministry of Home Affairs, 1973; Park, 1998). Since road improvement required coordinated donation of land among village residents, road related project could be the most difficult village project. Additionally, when more systematic evaluation of the performance on village projects was introduced later, road improvement was one of top five major criteria (Ministry of Home Affairs, 1983).

Since clan data were collected in the 1930s and cement project grades were recorded in the

[17] See, for example Maeil Kyungjae (1971).

[18] Kyungsang North Province is slightly smaller than the state of New Jersey in the U.S.

1970s, I reconstructed the village geographies based on 1930 data. Some village boundaries changed between 1930 and 1970; most frequently, villages split into separate villages. When a village had split into multiple villages between 1930 and 1970, I merged geography data for the villages and used the weighted mean of characteristics and cement project grades based on the number of household. To overcome data limitations, I imputed total village households in 1930 by multiplying the total number of township households in 1930 by a village's share of total township households in 1970.

3.2.2 Data on Land Donation

I compiled data on the amount of land donated for village projects by village residents from *Glorious Footsteps*, a government publication by Ministry of Home Affairs (1978). This publication has detailed information on village projects for about 350 villages across the country between 1970 and 1980. Importantly for this paper, it contains a list of land donors for village projects for each village (see Figure 11 for the original format). For each donor, the list contains information on the type and the amount of land donated. The type of land include land for housing, rice paddies, regular fields, and forest land. For the empirical analysis of the paper, I summed up the total amount of land donated for each village by land type. The main analysis used the total amount of cultivated land donated, which is the sum of rice paddies and regular fields. In robust test, I also tried different combination of summation of each type. The results of the paper were robust to these combinations.

Glorious Footsteps also contain the full distribution of family names of each village. Unlike the family clans data from the Japanese Colonial Government, there is no truncation of missing data at the 10% household share. In the analysis using land donation as an outcome, I use the share of each family names as a measure of lineage heterogeneity.

Additionally, *Glorious Footsteps* contains detailed information on land of a village, such as amount of total land, cultivated land, regular dry field, rice paddies, and forest. Because irrigation in rural area was mostly preformed in rice paddies, I computed irrigation rate using the he proportion of rice paddies out of total cultivated land. These characteristics are used as control variables in the analysis on land donation. Table 4 provides summary statistics.

While *Glorious Footsteps* data have detailed information for each village, the main shortcoming of the data is that there is no explicit information on how villages were selected. While the introduction of the publication indicated that the villages were chosen to be representative, there is possibility that villages that performed relatively better in village projects. Therefore, the results using these data may indicate the relationship between group heterogeneity and the amount of land donated among the villages that were relatively more active in village projects.

3.2.3 Data on Agricultural Machines

I used agricultural census in 1960, 1970, and 1980 for the data on agricultural machines owned by village residents. Agricultural machines and tools in the data include power tillers, ox carts, hand carts, water pumps, sprayers and combines. The most relevant information for the analysis is the number of power tillers. Unlike other agricultural machines, I was able to track the changes in the number of power tillers owned by villagers from 1970 to 1980. Due to changes in survey questions of agricultural census, data on other agricultural machines are not available for both 1970 and 1980.

Since the agricultural census are available at the township level, my empirical analysis on power tillers are also at the township level. Township characteristics are directly from agricultural census. Clan heterogeneity were aggregated from the village level clan heterogeneity measures. I took the weighted average of clan heterogeneity in each village belonging to the same township. The weight was the number of households in each village.

Agricultural census also has various township characteristics. These include the number of villages, literacy, occupations, landholdings, and irrigation rate. I computed a measure of poverty rate from the share of agricultural households which sold harvested crops to the market. This variable captures subsistence farming. I computed the rate of the change in agricultural household between 1960 and 1970 to capture the migration rate. Table 5 provides summary statistics.

3.3 Geographic and Economic Characteristics of Villages

I used various control variables that could be potential determinants of both clan settlements and production of public goods. Michalopoulos (2012) showed that geographic characteristics such as elevation and land quality are important determinants of ethnolinguistic diversity. Similarly, clan diversity could be also explained by geography. Hence the failure to control for these may result in biased estimates as production of public goods using cement may also depend on geographic endowment. For example, to use cement for construction, cement needs mixing with pebbles and water. Therefore, the proximity to river may also an important factor. Additionally, proximity to river may increase productivity of crop field which could attract clans for settlement.

I compiled spatial geography data and created village-level geographical characteristics using ArcGIS software. Data on village boundaries in 2014 are from Geoservice Korea 2014. Terrain Rugged Index was obtained from Nunn and Puga (2012). I also used soil types information from the 2007 Digital Soil Map of the World by the Food and Agriculture Organization. A map of river networks in 2014 is from Water Resource Management Information System, a web portal created by the Ministry of Land, Infrastructure, and Transport in South Korea. Road network data in 2014 comes from the National Transport Information Center in South Korea. Elevation data are from the 2000 Shuttle Radar Topography Mission by the U.S. National Aeronautics and Space Administration. I also used maps showing major battles during the 1950-1953 Korean War, obtained from the U.S. West Point Military Academy website.

Village characteristics in 1970 were also obtained from the digitized data from the NVCS. These include village characteristic of demographics, occupations, distance to township centers, proximity to national transportation infrastructure, and some local geographic characteristics.

4 Empirical Strategy

In this paper, I empirically investigate three questions. First, I use village cement project grades to analyze the relationship between the improvement of public goods and group heterogeneity. Second, I use land donation data to investigate the relationship between the amount of land donations and group heterogeneity. Finally, I evaluate the relationship between agricultural mechanization and group heterogeneity.

While I use separate data for each analysis, they are conceptually related. Since road infrastructure projects were most popular during the 1970s, as described in section 2, private land was needed to be donated in order to widen, straighten existing roads, and build new ones. The objective (and consequence) of road improvement was the utilization of labor saving agricultural machines. Therefore, I present the relationship between clan diversity and cement project grade as evidence of improved road infrastructure. Land donation for village projects is a measure of private contribution for village public goods. The increase in ownership of two-wheeled power tillers 10 years after the NVBP can be interpreted as a consequence of road improvement.

Equation (3) shows the the main empirical specification used for all outcomes:

$$Y_v = \alpha + \beta \ heterogeneity_v + X_v' \gamma + \epsilon_v, \tag{3}$$

where $heterogeneity_v$ is the measure of clan heterogeneity for village v; Y_v is the outcome of interest, such as the improvement of public goods, land donation, and agricultural mechanization; and X_v is a vector of village characteristics used as control variables. The main explanatory variable, $heterogeneity_{m,v}$, measures the heterogeneity of village clan compositions, TOPSHARE, HERF, and POLAR. β is the coefficient of interest which estimates the relationship between clan homogeneity and the outcome variable of interest.

The identification of the empirical analysis relies on the historically determined patterns of the settlement of clans. The identifying assumption is that, once controlling for the potential determinants of clan settlement, the error terms are uncorrelated with $heterogeneity_v$. It is reasonable to assume that the settlement was largely determined by geographic and spatial characteristics such as soil quality, access to water sources, terrain ruggedness, and distance from towns, I extensively control for these relevant variables.

The main empirical analysis of this paper is on the relationship between family clan heterogeneity and cement project grades at the village level. I include township fixed effects, δ_m, to account for across township differences in various unobserved characteristics which also influences the production of village public goods. Additionally, I control for clan identities, τ_c, to rule out that the results are driven by specific clans which may be better at producing public goods. Therefore, the identifying assumption is,

$$E(\epsilon_{t,v} \mid heterogeneity_v,\ X_v,\ \delta_m,\ \tau_c) = 0, \tag{4}$$

that is, clan heterogeneity is exogenous conditional on potential determinants of clan settlement, township fixed effects and clan identities.

By controlling for township fixed effects, I compare villages within the same township which is the lowest administrative unit in the country. The average area of a township in my sample is small. It is roughly similar to the land area of Syracuse in the state of New York (25 square miles). Therefore township fixed effects could account for unobserved variables that vary even in a small spatial scale. Additionally, I control for village-level geographic characteristics such as terrain ruggedness, distance from rivers and public road network.

Using village data, I checked whether mean values of village characteristics differed significantly between two groups of villages: group 1 had TOPSHARE values below the median, and group 0 had TOPSHARE values above the median. Following Kline (2013), I calculated standardized mean differences for village characteristics using the formula $(\mu_1 - \mu_0)/\sigma_0$, where μ_1 and μ_0 are the means of a variable for groups 1 and 0, respectively, and σ_0 is the standard deviation of group 0.

Figure 12 shows standardized mean differences for various village characteristics. The values are mean differences relative to the standard deviation of each variable. Most variables have standardized mean differences that are less than 0.1 of the standard deviation in absolute terms. The average differences in absolute terms are 0.05 of the standard deviations for all village variables shown in the figure. Village characteristics with the largest mean differences are village altitude (0.13) and distance from the town center (0.11), which are not drastic. Potentially important geographical determinants of clan settlements, such as ruggedness, distance from rivers, distance from major roads, and soil type, all show differences less than 0.1. Distance from battle sites during the Korean War, which may be a proxy for differences in initial village public goods due to war time destruction, also show little differentiation (0.01). Other evidence of the exogeneity of the heterogeneity variable is that the estimated coefficients of heterogeneity do not differ whether control variables are included or not.

Based on the context of the study, reverse causality can also be ruled out. Village residents could sort into places with preferred types or quantities of public goods. However, Korean clans–and the ancestors of typical village residents in general–settled in places hundreds of years ago. As farmers often were reluctant to sell farm lands bestowed by ancestors, mobility between villages was limited. Additionally, the government did not provide resources for the production of public goods in villages prior to 1970. Hence, the lack of public goods overall may have weakened villagers' incentives to move based on differences in public goods.

5 Empirical Results

Empirical results show that clan heterogeneity was systematically related to the improvement of public goods, land donation, and agricultural mechanization. In all three sets of empirical results, I find a concave relationship between outcomes and clan diversity. I suggest a possible explanation for the concavity results at the end of this section.

5.1 Effect of Clan Heterogeneity on Cement Grades

Clan heterogeneity is systematically related to the improvement of village public goods, measured by NVBP cement project grades. There is a concave and statistically significant relationship between the group heterogeneity measure and the probability of getting an A grade.

I used the following specification to test the main hypothesis:

$$cement\ grade_{m,v} = \alpha + \beta\ heterogeneity_{m,v} + X'_{m,v}\gamma + \delta_m + \tau_c + \epsilon_i, \quad (5)$$

where $cement\ grade_{m,v}$ is a dummy variable with a value of 1 if village v in township m received an A grade from the government under the NVBP, and 0 otherwise. This is the measure of the improvement of village public goods using provided government resources. δ_m denotes township fixed effects and τ_c is clan identities.[19]

Figure 13 presents the relationship between TOPSHARE, the percentage of households belonging to the largest clan in a village, and the probability of getting an A grade (sub-figure A). Since the outcome variable is a dummy variable, I divided the sample into 40 bins using the percentiles of TOPSHARE and took the average value of the outcome variable for each bin. Each bin has about 80 villages. The graph shows that the relationship was non-linear and concave. At the low level of TOPSHARE, the increase in TOPSHARE resulted in the increase in the probability of getting an A grade. However, at around 0.5, the relationship becomes negative. A further increase in TOPSHARE results in a reduction of the outcome variable. The relationship between HERF, the Herfindahl Index, and the outcome variable is qualitatively similar to TOPSHARE results (sub-figure B). Sub-figure C shows the positive relationship between POLAR, the polarization index, and the outcome variable.

Table 6 reports the estimation results. The dependent variable is the dummy variable indicating whether a village received an A grade or not. Panel A shows estimates without any control variables. Panel B includes full controls and township fixed effects. Specifications in each column are based on different heterogeneity measures. Column 1 in panels A and B shows a linear relationship between cement project grade and TOPSHARE. Column 2 includes a squared term of TOPSHARE, capturing nonlinearity of the TOPSHARE measure. In column 3, HERF is used as the heterogeneity measure. Column 4 includes the squared term of HERF. In column 5 POLAR is used.

Panel B includes various control variables. These include the total number of households, percentage of households engaged in agricultural occupations, average cultivated area per agricultural household, percentage of the population under age 14, distance from the township center, whether the local administrative office was in the village, the number of sub-villages (or hamlets), and the

[19] In an alternate specification, the dummy variable equals 1 if a village received an A or B grade and 0 otherwise.

age of the village leader. Additionally, panel B includes dummy variables indicating whether a village had electricity access, and was located next to the sea, a river or stream, county roads, regional roads, national roads, highways, and railroads. It also includes geographic classifications of each village based on the NVCS. I created dummy variables for each of the eight different classifications: agricultural villages, villages near cities, villages near highways, semi-urban villages, villages in mountainous regions, fishing villages, villages near rivers, and villages near the sea.[20] Following Michalopoulos and Papaioannou (2013, 2014), I included variables to present geographic characteristics that were potential determinants of village clan structures including the Terrain Ruggedness Index (Nunn and Puga, 2012), altitude and soil type based on classifications from the Food and Agricultural Organization as proxies for crop suitability. I also included distance from major battle fields during the 1950-1953 Korean war, to account for differences in initial levels of public goods due to war time destruction.

TOPSHARE does not predict the probability of getting an A grade, with or without control variables. However, as expected from Figure 13 showing a concave relationship, if the squared term of TOPSHARE ($TOPSHARE^2$ in the table) is included, both TOPSHARE and $TOPSHARE^2$ are statistically significant. As shown in Table 7, the estimates are robust to inclusion of control variables. Since $TOPSHARE^2$ has a negative value, the relationship between TOPSHARE and the probability of getting an A grade is concave, and peaks at 0.41. The interpretation is that at low levels of TOPSHARE, an increase in homogeneity is associated with greater improvement of public goods. However, once TOPSHARE reaches 0.41, the relationship becomes negative, and an increase in TOPSHARE is associated with less improvement of public goods. Similarly, HERF has a concave relationship with the outcome variable that peaks at 0.36. Polarization is positively associated with the outcome variable and the estimates are statistically significant. An increase in polarization from 0 to 1 is associated with a 5 percentage point increase in the probability of receiving an A grade, which is about one-fourth of the standard deviation of the outcome variable, or about 70% of the mean of the outcome variable.

[20] Classifications are mutually exclusive. For example, if a village can be classified as a fishing village, then this village is not classified as a village near the sea. Specific criteria (e.g., percentage of the population engaged in fishing as an occupation) were used such as the fraction of fishing population to designate each village as either a fishing village or a village near the sea.

In the Appendix, I also show that the concave relationship persists when the sample is restricted to villages that did not split between 1930 and 1970. Because the available clan data were collected in 1930, I used 1930 geographic data in my analysis. If a village split between 1930 and 1970, I merged the split villages back into one. This introduces measurement errors in group heterogeneity; restricting the analysis to non-split villages might reduce the measurement errors. The results show that among non-split villages, concavity still exists and both TOPSHARE and TOPSHARE^2 are statistically significant (see Table A3).

It is possible that the results are mainly being driven by other characteristics that are correlated with the heterogeneity measures. For example, villages with a high percentage of households belonging to the largest clan already may have had substantial levels of public goods, thus their improvement of public goods might have been low. However, road improvement only gained importance around the time when the cement was distributed. The demand for better roads coincided with the introduction of the power tiller (a two-wheeled, labor-saving agricultural machine), and increases in agricultural wages in the late 1960s. Therefore, it is not likely that some villages already had wider and straighter roads prior to the NVBP, because there was little incentive to create better road infrastructure. Traditional agricultural technology simply did not require wider roads.[21] Furthermore, I controlled for distance from the major battles during 1950-1953 Korean War, to account for initial differences in the levels of village public goods due to destruction. While the cement projects may have complemented to existing village public goods, major public goods such as schools and medical clinics were located at the centers of townships, not in villages.

A higher irrigation rate may result in higher agricultural yield, which in turn may lead to higher demand for village public goods such as roads. During the study period, the majority of irrigation was based on natural sources such as rivers and rainfall, and less than 20% of irrigation was performed using a formal irrigation system. While I do not have data on irrigation for the empirical results shown in this section, I extensively control for geographic variables that could predict irrigation such as river network and terrain ruggedness. In the following two outcome variables, land donation and the adoption of power tillers, I explicitly controlled for irrigation rate in the two other

[21]The most widely used method of transportation was A-frame backpack carriers. For tilling, animal power was most often used. These two technologies did not require wider roads.

empirical analyses. The results are robust to its inclusion.[22]

Another alternative explanation is that clan diversity may be correlated with geographic characteristics that happen to be conducive to cement use. Since cement must be mixed with pebbles and water, the proximity of these materials, not clan diversity itself, may have influenced the improvement of public goods. I extensively controlled for geographic variables such as proximity to rivers, terrain ruggedness and altitude to account for these factors, and the results are robust to their inclusion.

5.2 Effect of Clan Heterogeneity on Land Donations

I turn to the relationship between the amount of donated land per household and clan heterogeneity. Similar to the cement project results in the previous section, I find a concave relationship. I used the following specification for the analysis:

$$land\ donation_{p,v} = \alpha + \beta\ heterogeneity_{p,v} + X'_{p,v}\gamma + \theta_p + \epsilon_i, \tag{6}$$

where $land\ donation_{p,v}$ is the amount of donated land per household in a village v and province p, $homogeneity_{v,p}$ denotes heterogeneity measures in the village, and the vector $X_{v,p}$ contains village level controls. Unlike the previous empirical analysis, here I used province fixed effects denoted by θ_p, since the donation data are from all provinces in South Korea[23]. Data are available for less than 30 villages in each province so there is a limited number of villages from each township.

Figure 14 shows the relationship between TOPSHARE and the amount of cultivated land donated. Similar to cement project grade results, the relationship was non-linear and concave. The relationship between HERF and land donation was similarly concave. There was a positive relationship between POLAR and land donation.

Table 8 shows the results. Panel A presents bivariate relationships between heterogeneity measures and the average amount of donated land per household. Panel B includes village controls and

[22] I collected historical high quality maps produced by the Japanese Colonial Government in 1918. The maps contain information on land types, including irrigated rice paddies and regular dry farm fields, and the locations of villages. I plan to digitize maps and include the irrigation rate variable in my analysis.

[23] A province consists of townships.

province fixed effects. Control variables include the number of village households, the percentage of agricultural households, average cultivated area per household, total amount of cultivated land, and total amount of land.

As expected from the relationship shown by the figures, TOPSHARE has a concave and statistically significant relationship with land donation. The concave curve peaks around 0.6. HERF also has a concave relationship with land donation that peaks at 0.5. There is also a positive relationship between POLAR and the outcome variable. However, HERF and POLAR estimates are not statistically significant when controls are included.

These results are related to the previous analysis on cement project grades and group heterogeneity. The production of public goods requires contributions of private resources such as voluntary labor and land. The results in this section indicates that either low or high clan diversity in a village is associated with a relatively smaller amount of land donated. If land donation has a concave relationship with clan diversity, it is possible that the production of public goods also has a concave relationship with clan diversity.

It is possible that the concavity pattern of the amount of donated land is due to the concavity of the land endowment with respect to clan diversity. That is, villages with a medium level of clan diversity had more cultivated land per household so they tended to donate more. However, I showed in the Appendix that the total cultivated land per household does not have a concave relationship with clan diversity. While the estimates are not statistically significant, it had a convex relationship with clan heterogeneity given the positive sign of the coefficient of quadratic term when control variables are included (Table A4).

Another concern is that the clan diversity measure in this analysis is based on the family name distribution instead of the actual clan distribution. For example, the family name Kim is associated with many different clans. Since I constructed clan diversity based on the share of each family name in the village, clan diversity has measurement errors.

Another concern is that the donation list does not include compensated land contributions. It is possible that some villages had residents with higher incomes who were able to collectively purchase the land needed for village projects. Although the land donation data do not include income or wealth level, I controlled for the rate of irrigation as a proxy for income since a typical village relied on rice crops as a primary source of income. The results shown here are robust to the inclusion of irrigation rate.

5.3 Effect of Clan Heterogeneity on Agricultural Mechanization

One major reason to improve public goods, particularly village roads, was to adopt labor-saving agricultural technology in the 1970s. As discussed in Section 2, once agricultural wages began to increase and power tillers had been introduced, village residents increasingly demanded improvements to village road infrastructure.

As shown in previous results, if clan heterogeneity had effects on the improvement of public goods through contributions of private resources such as land and labor, one would expect that power tiller adoption may also be systematically associated with clan heterogeneity. In this section, I investigates this relationship. I find that the change in the number of tillers owned per household in a township is positively associated with clan homogeneity, and positively correlated with the polarization index.

The econometric specification is:

$$\Delta tiller_{1970-1980,m} = \alpha + \beta \ heterogeneity_m + X_m' \gamma + \epsilon_i, \tag{7}$$

where $\Delta tiller_{1970-1980,m}$ is the change in the number of power tillers owned per household in township m, $heterogeneity_m$ is the heterogeneity measure, and X is a vector containing the control variables.

Figure 15 shows the relationship between TOPSHARE and the changes in the number of power tillers owned per household between 1970 and 1980. While there is some indication that the relationship is concave, it is much less pronounced compared to the previous results on cement project grades and land donation. Further, the most of the values of TOPSHARE is from 0 and 0.4. This is due to using township-level information. By aggregating at township level, the variability of the values of village-level TOPSHARE was much reduced. The relationship between HERF and land donation was also concave. Additionally, there was a positive relationship between POLAR and the outcome variable.

Table 9 shows the regression results. There is a positive and statistically significant linear relationship between TOPSHARE and the changes in the number of power tillers per household between 1970 and 1980. TOPSHARE and TOPSHARE^2 are both statistically significant, suggesting a concave relationship. Since R^2 of the specification using both the level and quadratic terms of TOPSHARE (Column 2) are higher than R^2 of the specification with level terms only (Column 1), the Column 2 specification provides a better fit with the data. Polarization is also positively associated with the outcome variable and the estimates are statistically significant.

These results, however, cannot be interpreted directly as consequences of the production of public goods between 1970 and 1980. If the production of public goods is the main driver of the increase in power tillers, one would expect that heterogeneity measures should not predict the outcome when controlling for cement project grade. In the Appendix, I show that even after including the probability of receiving an A cement project grade as a control, clan diversity still predicts the outcome (Table A5). This implies that other factors related to clan heterogeneity were associated with the increase in the number of tillers. The estimated coefficient of the effect of cement project grade on power tillers has economically large and statistically significant, which suggests that the improvement of roads was also an important driver of the increase in power tillers.

Another concern is that since the analysis is at the township level, the clan heterogeneity measure is also aggregated at the township level, which could lead to measurement error. Since the projects were mostly implemented at the village level, examining township level data may result in some form of aggregation bias. This is a data limitation, since data on agricultural machines from the agricultural census are at the township level.

Additionally, the magnitude of the coefficients drops significantly as more controls are added when estimates from panels A and B are compared. It is possible that some omitted variables could completely explain away the effect of clan heterogeneity on power tiller ownership.

5.4 The Concave Relationship between Clan Heterogeneity and the Production of Public Goods

Empirical findings in the previous sections show a robust concave relationship between family clan heterogeneity and the production of public goods and resource contributions. In the existing literature, scholars have shown that group heterogeneity has a linear relationship with the provision of public goods. The concave relationship found in this study may imply that homogeneity yields not only benefits, as suggested in previous literature, but also some possible costs.

Finding an empirical answer to this question using data is beyond the scope of this paper. However, a possible trade-off between coordination and accountability could explain the concave relationship. The social capital literature provides evidence that more homogeneous societies have better coordination among members and make more contributions to public goods. Yet, another strand of literature has shown that the absence of checks and balances provides incentives for elites to waste resources.[24] For example, Acemoglu et al. (2013) shows that, in Sierra Leone, the number of families in a chiefdom was negatively related to comparative development outcomes, and positively related with social capital, suggesting the control of family organizations by the elite.

It is possible that in traditional societies without strong political institutions, highly homogenous communities are likely to have a high level of coordination as well as highly autocratic elites. This could be a reasonable characterization of Korean rural villages in the 1970s. The strong influence of the Confucian doctrine of obedience to family elders may have resulted in elders having powerful influence in clan and village matters. This may have been especially true in villages, since they lack formal governmental or political organization. Therefore, the potential tradeoff between coordination and accountability could imply that, as clan homogeneity increases, village residents coordinate better. However, clan elites could more easily control the clan organization and use resources for themselves in the absence of checks and balances. At a higher level of clan

[24] Several authors have emphasized the complementarity between social capital and the quality of leaders. Durlauf and Fafchamps (2004) argued that the delivery of public goods depends on local trust and leadership. Krishna (2001) shows that social capital is associated with better social outcomes only in the presence of strong organizational leadership.

homogeneity, public goods provision could be negatively affected.

6 Case Study: Effects of Lineage Homogeneity on Cooperation and Participation

In this section, I provide qualitative evidence on the benefit of a village's homogenous lineage in terms of increased contributions to public goods through better coordination and cooperation. This case study links the three sets of empirical results presented in this paper: clan heterogeneity is systematically related to the improvement of public goods, land donations, and agricultural mechanization.

Moonsung village in Kyungsang North Province provides a concrete example of the willingness of dominant clan members to donate land (Lim and Lim, 2013). The village is located in the southeastern part of the province (see Figure A4). There were 68 households in the village in 1970, and the village was one of the poorest in the region. Approximately 67% of the village was mountainous. The education level was low: just 5% of the population had completed junior high school, and the majority had completed elementary school at most.

In October 1970, the village decided to widen 1,800 meters of village access road using the cement from the NVBP. The problem was that 0.25 hectares of land had to be donated. Since agricultural land was the most important asset and the source of income, it was difficult to persuade landowners to donate their precious land.

The Namyang Hong clan was the dominant clan in the village, and more than 50% of total households were members. The cohesiveness of the Namyang Hong clan was considered to be very strong. At village meetings, Soon-rak Hong and Seon-pyo Hong volunteered to donate 0.06 hectares and 0.05 hectares, respectively. These two were probably from the Namyang Hong Clan since the family name Hong is not common. Another 17 land owners soon donated their land as well. Since the majority of households was from the same clan, perhaps the two Hongs were more willing to donate their land. If the lack of willingness to donate land was the most binding constraint in improving roads, then the homogeneity in lineage groups in a village could be beneficial for road improvements.

The road improvements created benefits. Better roads enabled the villagers to transport goods using hand carts on wider roads instead of carrying them on their backs. Further, the improved roads facilitated agricultural mechanization in Munsung village in the late 1970s.

7 Conclusion

In South Korean rural villages, village clan heterogeneity had a systematic relationship with the production of village public goods, land donation for public projects, and changes in the number of power tillers owned per household. In contrast to the existing literature on the effects of ethnic diversity on the provision of public goods, I found that the relationship between clan diversity and public goods is not linear, but concave, due to the trade off between coordination and leader accountability.

If this trade-off is a salient feature of the provision of public goods, then strengthening political institutions to prevent clan leaders from dominating communities could be an important policy implication. By providing checks and balances, improved political institutions could prevent elites from controlling community members and the high level of social capital could be better utilized for more economically productive uses.

The New Village Beautification Project, which this paper is based on, was a pilot project for a following full-blown rural development program called the New Village Movement or Saemaul Undong (SU). SU has recently gained international attention from agencies such as the United Nations Development Programme (UNDP) and the Asian Development Bank as a model for rural development. For example, the UNDP, in partnership with the Korean government, plans to apply lessons from SU to development projects in six developing countries: Uganda, Rwanda, Vietnam, Bolivia, Lao People's Democratic Republic, and Myanmar. However, while there is an abundance of case studies and qualitative research on SU, few empirical research studies have been conducted on the factors that influenced the success of SU. To my knowledge, this is the first empirical paper in which an economist has systematically evaluated the effect of group heterogeneity on the successful uses of public resources provided under the NVBP using large-scale village-level data.

The paper has several limitations. In the absence of clear natural experiments on group heterogeneity, identification concerns remain. While I extensively controlled for potential determinants of clan settlement and production of village public goods, it is possible that some omitted variables may have biased the results. Additionally, the outcome measure of production of public goods in this paper was based on evaluation grades given by the government. More detailed data on actual improvements to different types of public goods would provide further insights. Finally, by collecting relevant data, researchers could directly analyze the relationship between group heterogeneity and leader accountability in future studies.

References

Acemoglu, Daron, Tristan Reed, and James A Robinson (2013) "Chiefs: elite control of civil society and economic development in Sierra Leone," Working Paper 18691, National Bureau of Economic Research.

Alesina, A., R. Baqir, and W. Easterly (1999) "Public goods and ethnic divisions," *Quarterly Journal of Economics*, Vol. 114, pp. 1243–1284.

Alesina, A. and E. La Ferrara (2000) "Participation in heterogeneous communities," *Quarterly Journal of Economics*, Vol. 115, pp. 847–904.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003) "Fractionalization," *Journal of Economic Growth*, Vol. 8, pp. 155–194.

Arnold, Guy (2001) *Guide to African political and economic development*: Routledge.

Bang, Ki-joong (2004) *Fascist governing policy of Japanese colonial power and peasant lives*: Hye-an.

Barthorp, Michael and Gerry Embleton (1982) *The Jacobite Rebellions 1689-1745*, No. 118: Osprey Publishing.

Besley, Timothy, Stephen Coate, and Glenn Loury (1993) "The economics of rotating savings and credit associations," *American Economic Review*, Vol. 83, pp. 792–810.

Brandt, Vincent SR (1972) *A Korean village: Between farm and sea*: Harvard University Press.

Collier, Paul and Anke Hoeffler (1998) "On economic causes of civil war," *Oxford Economic Papers*, Vol. 50, pp. 563–573.

Collins, Kathleen (2002) "Clans, pacts, and politics in Central Asia," *Journal of Democracy*, Vol. 13, pp. 137–152.

Costa, Dora L and Matthew E Kahn (2003) "Civic engagement and community heterogeneity: An economist's perspective," *Perspective on Politics*, Vol. 1, pp. 103–111.

Cox, Donald and Marcel Fafchamps (2007) "Extended family and kinship networks: economic insights and evolutionary directions," *Handbook of Development Economics*, Vol. 4, pp. 3711–3784.

Durlauf, Steven N and Marcel Fafchamps (2004) "Social capital," Working Paper 10485, National Bureau of Economic Research.

Fafchamps, Marcel and Flore Gubert (2007) "The formation of risk sharing networks," *Journal of Development Economics*, Vol. 83, pp. 326–350.

García Montalvo, José and Marta Reynal-Querol (2002) "Why ethnic fractionalization? Polarization, ethnic conflict and growth," Economics and Business Working Paper 660, Universitat Pompeu Fabra.

——— (2005) “Ethnic polarization, potential conflict, and civil wars,” *American Economic Review*, Vol. 95, pp. 796–816.

Glaeser, Edward L, David I Laibson, Jose A Scheinkman, and Christine L Soutter (2000) “Measuring trust,” *Quarterly Journal of Economics*, Vol. 115, pp. 811–846.

Hassan, Hussein D (2007) “Iraq: Tribal structure, social, and political activities,” Report for Congress 22626, Congressional Research Service.

Hwang, In-jeong (1980) *Comprehensive agricultural development in South Korea: Assessment and outlook*: Korea Rural Economic Institute.

International Crisis Group (2015) “Chiefs: elite control of civil society and economic development in Sierra Leone,” Middle East Report 163, International Crisis Group.

Isham, Jonathan (2002) “The effect of social capital on fertiliser adoption: Evidence from rural Tanzania,” *Journal of African Economies*, Vol. 11, pp. 39–60.

Jacobs, Norman (1985) *The Korean road to modernization and development*: University of Illinois Press.

Kim, Cheongryum (2006) *From a poor country to an advanced country*: Random House Joongang.

Kim, Kyong-dong (1985) *Man and society in Korea’s economic growth: Sociological studies*: Seoul National University Press.

Kim, Sangkyu (2012) “Study in the mass migration of lineage villages and ethos,” *Cultural Study of the Northeast Asia*, Vol. 32, pp. 5–17.

Kim, Young-mee (2009) *Their New Village Movement*: Blue History Press.

Kline, Rex B (2013) *Beyond significance testing: Statistics reform in the behavioral sciences*: American Psychological Association.

Knack, Stephen and Philip Keefer (1997) “Does social capital have an economic payoff? A cross-country investigation,” *Quarterly Journal of Economics*, pp. 1251–1288.

Korean Studies Advancement Center (2014) “Culture of aristocracy in Andong region,” `http://www.ugyo.net/cf/frm/tuFrm.jsp?CODE1=02&CODE2=03&CLSS=6&sBookNmbr=F003&sMok_Nmbr=4`, Retrieved November 10, 2014 .

Krishna, Anirudh (2001) “Moving from the stock of social capital to the flow of benefits: the role of agency,” *World Development*, Vol. 29, pp. 925–943.

Kyunghyang (1970) “Saemaul Undong in every village,” Kyunghyang Newspaper, p.1 Oct. 2 1970.

La Ferrara, Eliana (2003) “Kin groups and reciprocity: A model of credit transactions in Ghana,” *American Economic Review*, Vol. 93, pp. 1730–1751.

Lee, Kwang-kyu (1997) *Psychological anthropology of Korean culture*: Jibmoondang.

Lim, Han-Sung and Jae-Kang Lim (2013) “A case study on success factors of Saemaeul Undong,” *Korean Journal of Local Government Studies*, Vol. 17, pp. 153–176.

Litwak, Eugene and Ivan Szelenyi (1969) “Primary group structures and their functions: Kin, neighbors, and friends,” *American Sociological Review*, Vol. 34, pp. 465–481.

Maeil Kyungjae (1971) “Three billion Won investment in Saemaul Undong,” Maeil Kyungjae Newspaper, September. 17 1971.

Marr, Phebe (2011) *The modern history of Iraq*: Westview Press.

Mauro, Paolo (1995) “Corruption and growth,” *Quarterly Journal of Economics*, Vol. 110, pp. 681–712.

Michalopoulos, Stelios (2012) “The origins of ethnolinguistic diversity,” *American Economic Review*, Vol. 102, p. 1508.

Michalopoulos, Stelios and Elias Papaioannou (2013) “Pre-colonial ethnic institutions and contemporary African development,” *Econometrica*, Vol. 81, pp. 113–152.

——— (2014) “National institutions and subnational development in Africa,” *Quarterly Journal of Economics*, Vol. 129, pp. 151–213.

Miguel, E. and M.K. Gugerty (2005) “Ethnic diversity, social sanctions, and public goods in Kenya,” *Journal of Public Economics*, Vol. 89, pp. 2325–2368.

Ministry of Home Affairs (1973) *The Saemaul Movement*: Ministry of Home Affairs.

——— (1978) *Glorious Footsteps*: Ministry of Home Affairs.

——— (1983) *Saemaul Undong in Korea*: Ministry of Home Affairs.

Moore, M. (1984) “Mobilization and disillusion in rural Korea: The Saemaul movement in retrospect,” *Pacific Affairs*, Vol. 57, pp. 577–598.

Mu, Ren and John Giles (2014) “Village political economy, land tenure insecurity, and the rural to urban migration decision: Evidence from China,” Policy Research Working Paper 7080, The World Bank.

Munshi, Kaivan (2003) “Networks in the modern economy: Mexican migrants in the US labor market,” *Quarterly Journal of Economics*, Vol. 118, pp. 549–599.

Nunn, Nathan and Diego Puga (2012) “Ruggedness: The blessing of bad geography in Africa,” *Review of Economics and Statistics*, Vol. 94, pp. 20–36.

Palloni, Alberto, Douglas S Massey, Miguel Ceballos, Kristin Espinosa, and Michael Spittel (2001) “Social capital and international migration: A test using information on family networks,” *American Journal of Sociology*, Vol. 106, pp. 1262–1298.

Park, Jin-Hwan (1998) *The Saemaul Movement: Koreas approach to rural modernization in 1970s*: Korea Rural Economic Institute.

Peng, Yusheng (2004) “Kinship networks and entrepreneurs in China’s transitional economy,” *American Journal of Sociology*, Vol. 109, pp. 1045–1074.

Platteau, Jean-Philippe and Frederic Gaspart (2003) “The risk of resource misappropriation in community-driven development,” *World Development*, Vol. 31, pp. 1687–1703.

Posner, Richard A (1980) “A theory of primitive society, with special reference to law,” *Journal of Law and Economics*, Vol. 23, pp. 1–53.

Rosenzweig, Mark R (1988) “Risk, implicit contracts and the family in rural areas of low-income countries,” *Economic Journal*, Vol. 98, pp. 1148–1170.

Seo, Bongdo (1997) *Family clans*: Suwon Sisa.

Song, Sonhui (1982) *Kinship and lineage in Korean village society*: Indiana University.

Tsai, Lily L (2007) “Solidary groups, informal accountability, and local public goods provision in rural China,” *American Political Science Review*, Vol. 101, pp. 355–372.

Warriner, G Keith and Trudy M Moul (1992) “Kinship and personal communication network influences on the adoption of agriculture conservation technology,” *Journal of Rural Studies*, Vol. 8, pp. 279–291.

Yesa Moonhan (2001) “Twenty clan villages,” `http://news679.ndsoftnews.com/news/articleView.html?idxno=14833`, Yesa Moonhan Newspaper, Oct. 1 2001, Retrieved July 2, 2014 .

Yi, Man-gap (1981) *Toward a new community*: Institute of Saemaul Undong Studies, Seoul National University.

Yu, Tae-yeung (1986) *The patterns of rural development in Korea in the 1970s*: Saemaul Institute, Kon-Kuk University.

Table 1: Summary statistics for village level analysis

	mean	sd	min	max
Dependent Variables				
Village with an A grade (dummy)	0.07	0.21	0	1
Heterogeneity Measures				
TOPSHARE	0.16	0.23	0	1
HERF	0.08	0.17	0	1
POLAR	0.14	0.21	0	1
Village Characteristics				
Leader age	39.7	6.07	19	73
Average cultivated area per household (ha)	0.94	0.66	0	11.5
Total number of households	180	201	11	4,527
Fraction of agricultural households	0.86	0.17	0.05	1
Fraction of population with age below 14	0.39	0.04	0.12	0.66
Number of hamlets	2.09	1.05	1	9
Distance from town center (km)	4.61	3.55	0	30
Local admin office in village (dummy)	0.05	0.17	0	1
Electricity access in village (dummy)	0.30	0.44	0	1
Sea next to village (dummy)	0.03	0.16	0	1
River passes in village (dummy)	0.37	0.44	0	1
National road passes village (dummy)	0.15	0.33	0	1
Regional road passes village (dummy)	0.19	0.36	0	1
County road passes village (dummy)	0.16	0.34	0	1
Highway passes village (dummy)	0.02	0.12	0	1
Railroad passes village (dummy)	0.06	0.23	0	1
Fraction of improved roofs	0.40	0.25	0	1
Sample Size	3,124			

Table 2: Summary statistics for village level analysis (continued)

	mean	sd	min	max
Village Classifications by NVCS				
Agricultural village (dummy)	0.78	0.40	0	1
Village near urban center (dummy)	0.07	0.25	0	1
Village near highway (dummy)	0.05	0.20	0	1
Village belong to town center district (dummy)	0.05	0.19	0	1
Village near mountainous area (dummy)	0.02	0.15	0	1
Village with more than 50% of people with fishing occupation (dummy)	0.01	0.11	0	1
Village located less than four kilometers from coast (dummy)	0.02	0.12	0	1
Geographical and Spatial Variables				
Area (km^2)	6.4	6.2	0	94.2
Terrain Ruggedness Index	169	95	2	549
Distance from major roads (km)	3.7	9.4	0	219.5
Distance from major river (km)	1.8	9.2	0	219.6
Distance from major battles during Korean War (km)	27.8	18.4	0.1	248.7
Altitude (m)	212	144	2	955
FAO Soil Type 3964 (dummy)	0.25	0.44	0	1
FAO Soil Type 4295 (dummy)	0.26	0.44	0	1
FAO Soil Type 4352 (dummy)	0.27	0.45	0	1
FAO Soil Type 4290 (dummy)	0.09	0.29	0	1
FAO Soil Type Other (dummy)	0.12	0.32	0	1
Sample Size	3,124			

Table 3: Correlation coefficients between heterogeneity measures

	TOPSHARE	HERF	POLAR
TOPSHARE	1		
HERF	0.92	1	
POLAR	0.83	0.67	1

Table 4: Summary statistics for land donation analysis

	mean	sd	min	max
Dependent Variable				
Amount of cultivated land donated per household (ares)	0.36	0.49	0	3.4
Explanatory Variables				
TOPSHARE	0.35	0.19	0.09	1
HERF	0.22	0.17	0.02	1
POLAR	0.48	0.17	0	0.89
Village Characteristics				
Number of village households	96.5	55.6	20	360
Fraction of agricultural households	0.79	0.23	0.03	1
Irrigation rate	0.60	0.49	0	1
Cultivated area per household (hectares / household)	1.09	0.94	0.07	11.2
Total amount of land (hectares)	195.3	228.2	7.1	1,906
Total amount of cultivated land (hectares)	66.1	32.1	3	193
Sample Size	207			

Table 5: Summary statistics from power tiller analysis

	mean	sd	min	max
Dependent Variable				
Number of power tillers per agricultural households in 1980	0.19	0.09	0	0.43
Explanatory Variables				
Clan Share	0.17	0.12	0	0.72
Herfindahl Index	0.07	0.06	0	0.25
Polarization Index	0.12	0.1	0	0.57
Township Characteristics				
Number of villages	12.46	4.54	3	33
Illiteracy rate	0.14	0.04	0.02	0.3
Number of agricultural households	1659	595	398	4,433
Total cultivated area	1442	517	331	3,476
Irrigation rate	0.54	0.15	0.03	0.82
Average cultivated area per agricultural household	0.89	0.15	0.48	1.3
Fraction of harvested agricultural household which sold harvest to market	0.76	0.17	0.1	1
Number of tillers per agricultural household	0.005	0.01	0	0.13
Number of ox carts per agricultural household	0.1	0.09	0	0.44
Number of hand carts per agricultural household	0.23	0.16	0	0.7
Rate of change of agricultural household between 1960 and 1970	0.05	0.08	-0.33	0.4
Sample Size	239			

Note: All control variables are based in 1970.

Table 6: Effects of family clan heterogeneity on public good production

	Dependent variable: (Prob. Getting an A grade)				
	(1)	(2)	(3)	(4)	(5)
PANEL A: No controls					
TOPSHARE	0.02	0.13***			
	(0.02)	(0.04)			
TOPSHARE^2		-0.16***			
		(0.05)			
HERF			-0.00	0.15**	
			(0.02)	(0.06)	
HERF^2				-0.21***	
				(0.06)	
POLAR					0.04**
					(0.02)
Peak of Concavity		0.41		0.36	
Observations	3,124	3,124	3,124	3,124	3,124
R-squared	0.00	0.00	0.00	0.00	0.00
PANEL B: Full controls					
TOPSHARE	0.03	0.14***			
	(0.02)	(0.05)			
TOPSHARE^2		-0.17***			
		(0.06)			
HERF			0.01	0.16**	
			(0.02)	(0.07)	
HERF^2				-0.22***	
				(0.08)	
POLAR					0.05**
					(0.02)
Peak of Concavity		0.41		0.36	
Mean of Y	0.07	0.07	0.07	0.07	0.07
Observations	2,668	2,668	2,668	2,668	2,668
R-squared	0.16	0.16	0.16	0.16	0.16
# of townships	220	220	220	220	220

Note: Robust standard errors in parentheses. In Panel B, standard errors are cultured at township level. Dependent variable is a dummy variable which equals one if a village received an A cement grade, and zero otherwise. This measure captures the production of village public goods. An A cement grade is an indication of more production of public goods compared to a B or C grade. TOPSHARE is the household share of the largest clan in the village. TOPSHARE^2 is the squared term of TOPSHARE. HERF is the Herfindahl Index based on village clan structure. POLAR is the polarization index by García Montalvo and Reynal-Querol (2002). Panel A does not have control variables. Panel B includes various village characteristics including village size, demographic, proximity to urban infrastructure and geographical characteristics. Additionally, it includes township fixed effects and family clan identities. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Effects of TOPSHARE on public good production

	Dependent variable: (Prob. Getting an A grade)				
	(1)	(2)	(3)	(4)	(5)
TOPSHARE	0.02	0.13***	0.12***	0.13***	0.14***
	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)
TOPSHARE^2		-0.16***	-0.16***	-0.16***	-0.17***
		(0.05)	(0.06)	(0.06)	(0.06)
Basic control	N	N	Y	Y	Y
Urban proximity	N	N	N	Y	Y
Geography	N	N	N	N	Y
Township FE	N	N	Y	Y	Y
Mean of Y	0.07	0.07	0.07	0.07	0.07
# of townships		245	245	245	220
Observations	3,124	3,124	3,124	3,124	2,668

Note: Robust standard errors in parentheses. When township fixed effects are included, standard errors are cultured at the township leve. Dependent variable is a dummy variable which equals one if a village received an A cement grade, and zero otherwise. This measure captures production of village public goods. An A cement grade is an indication of more production of public goods compared to a B or C grade. TOPSHARE is the household share of the largest clan in the village. TOPSHARE^2 is the squared term of TOPSHARE.

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

Table 8: Effects of Family Name Heterogeneity on Size of Cultivated Land Donation per Households

Dependent variable: Size of Cultivated Land Donation per HH					
	(1)	(2)	(3)	(4)	(5)
PANEL A: No controls					
TOPSHARE	7.80*	37.83***			
	(4.06)	(11.41)			
$TOPSHARE^2$		-33.24***			
		(11.54)			
HERF			3.73	26.82**	
			(4.52)	(13.16)	
$HERF^2$				-29.87**	
				(13.40)	
Polarization					7.56
					(4.58)
Peak of Concavity		0.57		0.45	
Observations	207	207	207	207	207
R-squared	0.02	0.04	0.00	0.02	0.02
PANEL B: Full controls					
TOPSHARE	4.66	24.59*			
	(3.93)	(11.44)			
$TOPSHARE^2$		-22.09*			
		(10.60)			
HERF			1.10	7.05	
			(4.19)	(13.81)	
$HERF^2$				-7.68	
				(14.17)	
Polarization					2.07
					(5.03)
Peak of Concavity		0.56		0.46	
Observations	205	205	205	205	205
R-squared	0.17	0.19	0.17	0.17	0.17
# of Provinces	8	8	8	8	8

Note: Robust standard errors in parentheses. In Panel B, standard errors are cultured at the province level. Dependent variable is the size of donation of cultivated land per household for village projects in a village. This measure captures private contribution to public goods by village residents. TOPSHARE is the household share of the largest family name in the village. $TOPSHARE^2$ is the squared term of TOPSHARE. HERF is the Herfindahl Index based on family name distribution. POLAR is the polarization index by García Montalvo and Reynal-Querol (2002). Panel A does not have control variables. Panel B includes various village characteristics including village size, share of agricultural households in the village, irrigation rate, cultivated area per household, total area of the village. It also include province fixed effects. Top one percent of the outcome variable is dropped as outliers. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Effects of family clan heterogeneity on ownership of a wheeled agricultural machine

Dependent variable: Δ Number of Power Tillers per HH 1970-1980					
	(1)	(2)	(3)	(4)	(5)
PANEL A: No controls					
TOPSHARE	0.25***	0.61***			
	(0.06)	(0.18)			
$TOPSHARE^2$		-1.04**			
		(0.50)			
HERF			0.36***	1.11***	
			(0.11)	(0.30)	
$HERF^2$				-3.93***	
				(1.51)	
POLAR					0.21***
					(0.06)
Peak of Concavity		0.29		0.14	
Mean of Y	0.18	0.18	0.18	0.18	0.18
Observations	245	245	245	245	245
R-squared	0.07	0.09	0.05	0.08	0.05
PANEL B: Full controls					
TOPSHARE	0.11**	0.33**			
	(0.04)	(0.13)			
$TOPSHARE^2$		-0.63*			
		(0.36)			
HERF			0.11	0.47**	
			(0.07)	(0.21)	
$HERF^2$				-1.84*	
				(0.97)	
POLAR					0.09**
					(0.04)
Peak of Concavity		0.26		0.13	
Mean of Y	0.18	0.18	0.18	0.18	0.18
Observations	245	245	245	245	245
R-squared	0.60	0.61	0.60	0.60	0.60

Note: Robust standard errors in parentheses. Dependent variable is the changes in the number of household owned power tillers per household between 1970 and 1980. TOPSHARE is the weighted average of household share of the largest family clan in the village, aggregated at the township level with weight being the number of village households. $TOPSHARE^2$ is the squared term of TOPSHARE. HERF is the Herfindahl Index based on family clan distribution. POLAR is the polarization index by García Montalvo and Reynal-Querol (2002). Panel A does not have control variables. Panel B include various township level characteristics including township population size, size of cultivated land, literacy, irrigation rate, ownership of agricultural machines in 1970.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

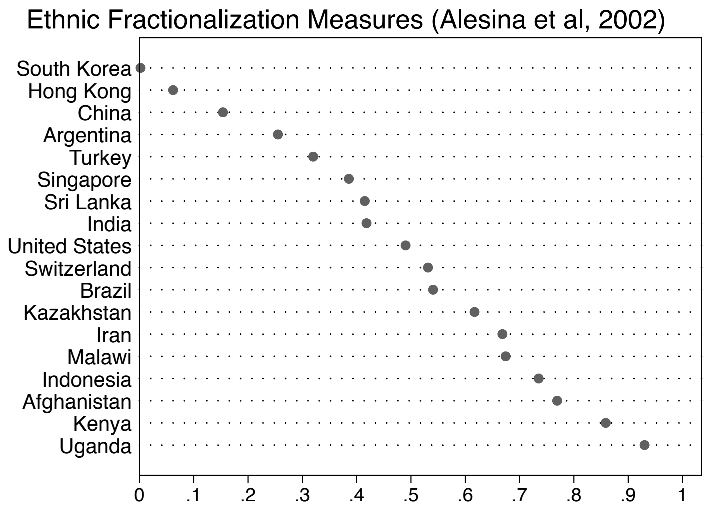


Figure 1: Comparison of ethnic fractionalization measures across countries
Source: Alesina et al. (2003)

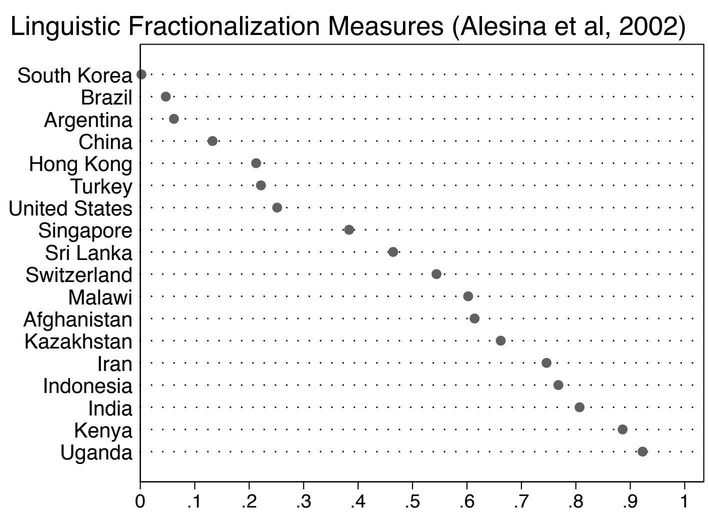


Figure 2: Comparison of linguistic fractionalization measures across countries
Source: Alesina et al. (2003)

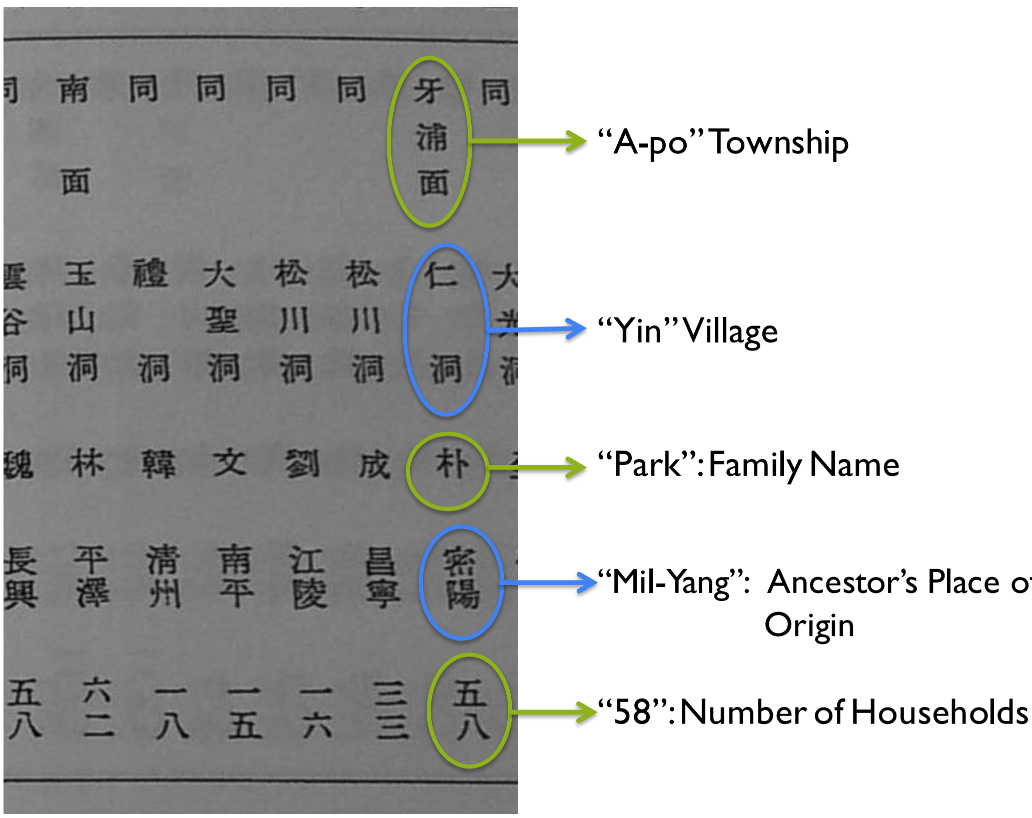


Figure 3: An example of family clan data from *Family Names in Chosun*

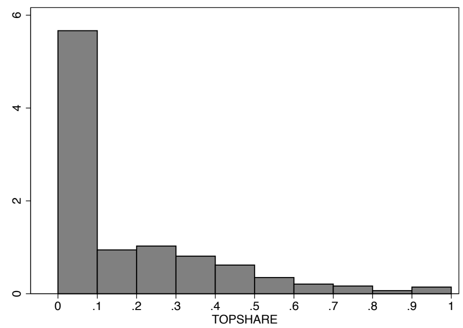


Figure 4: Histogram of TOPSHARE

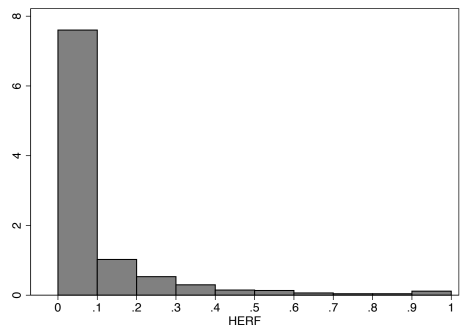


Figure 5: Histogram of HERFINDAHL

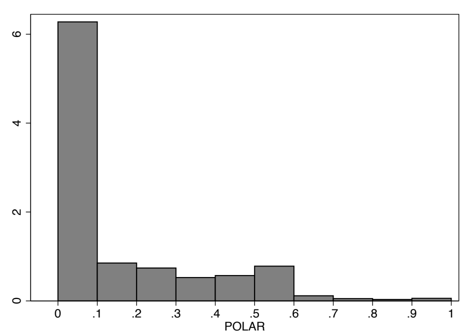


Figure 6: Histogram of POLARIZATION

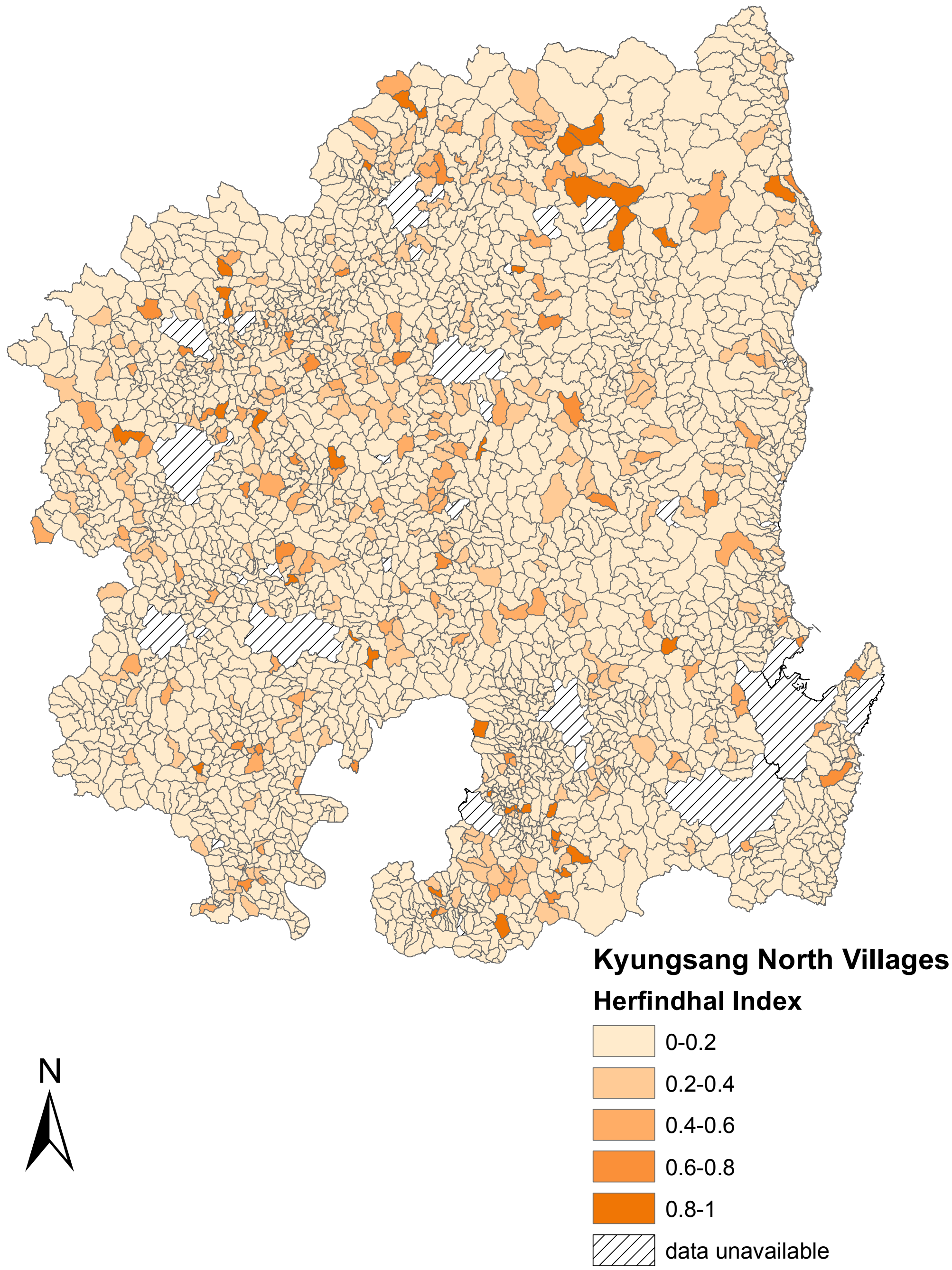


Figure 7: Spatial distribution of the Herfindahl Index for family clans

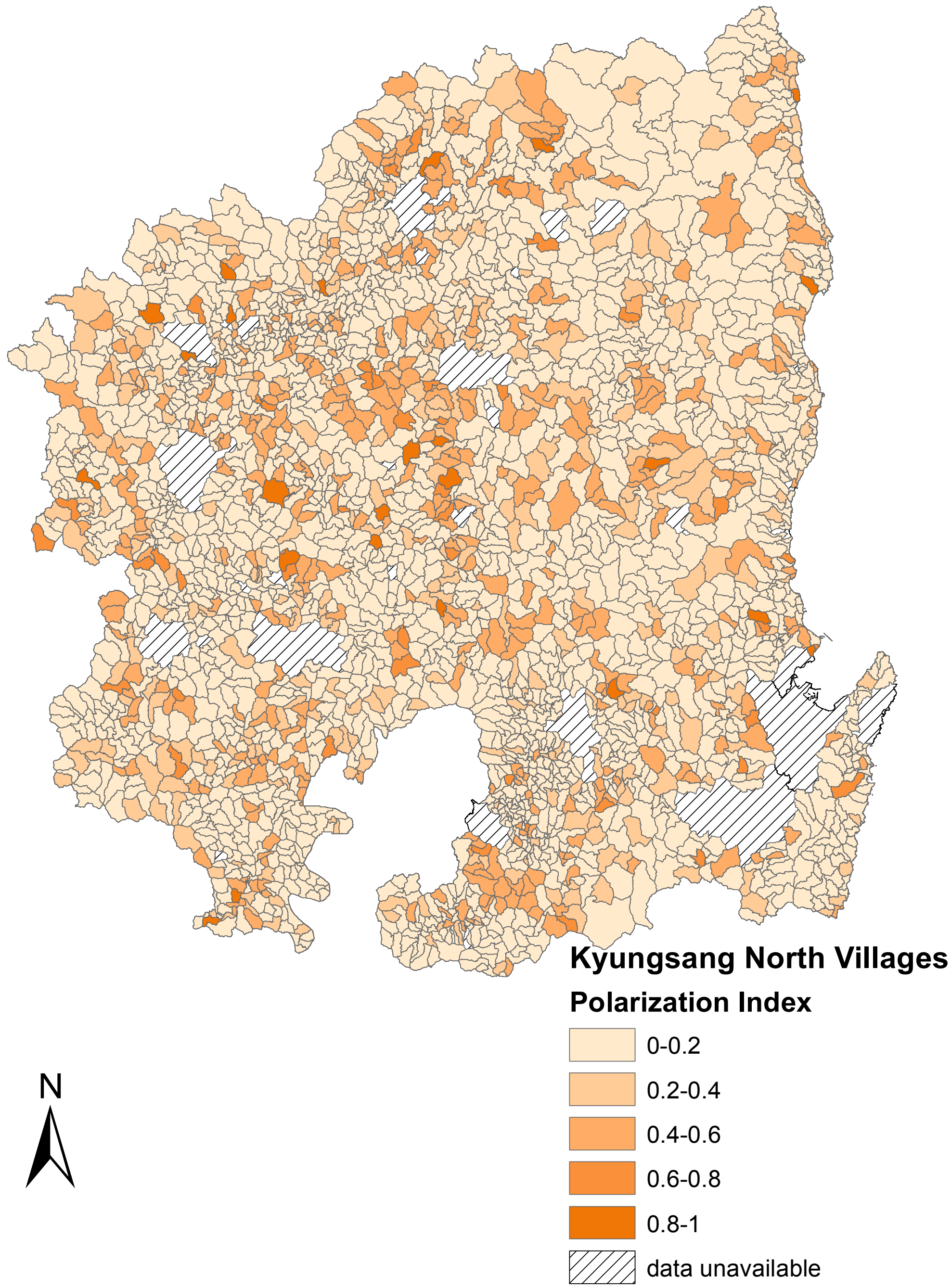
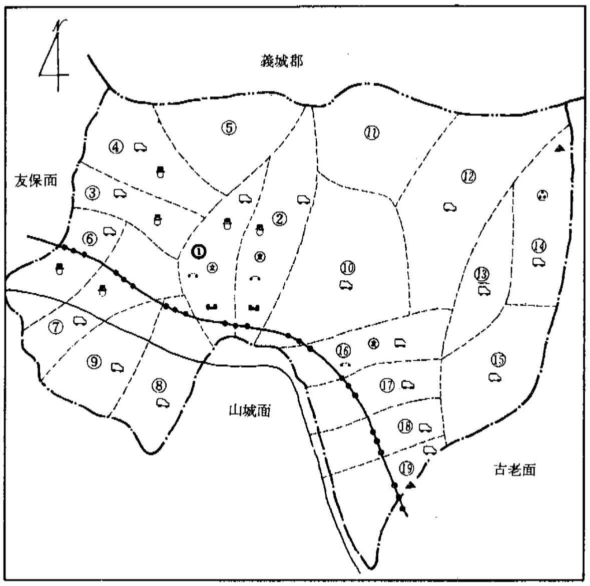


Figure 8: Spatial distribution of the polarization index for family clans

Code No. 9—7—6

軍 威 郡 義 興 面

1. 略　圖



2. 邑面—里洞距離表 (Km)

順位	里洞名	距離	順位	里洞名	距離
1	邑內 1	—	16	蓮桂 1	4.0
2	邑內 2	—	17	蓮桂 2	4.0
3	水北 1	1.0	18	梅城 1	6.0
4	水北 2	0.8	19	梅城 2	6.5
5	水北 3	2.0			
6	水西 1	1.6			
7	水西 2	1.4			
8	梨枝 1	1.6			
9	梨枝 2	1.8			
10	芭田洞	1.3			
11	新德 1	2.0			
12	新德 2	2.3			
13	芝湖 1	3.0			
14	芝湖 2	3.4			
15	芝湖 3	3.7			

3. 里洞別現況

Code No.	里洞名	새마을 區分	指導者 姓名	指導者 年齡	戶當耕地面積	家口 計	家口 農家	家口 非農家	人口 計	人口 14歲未滿	人口 14歲以上	地域類型	지붕改良 改良	지붕改良 未改良	自然部落數	結緣機關
332	邑內 1	자조	朴文達	32	0.3	223	104	119	1,010	411	599	1-1	125	39	1	
333	邑內 2	자조	金章煥	30	0.3	198	109	89	963	378	585	1-1	126	73	1	
334	水北 1	기초	朴正圓	45	1.0	71	62	9	354	127	227	1-1	70	25	1	軍威警察署
335	水北 2	자조	朴晩云	41	0.8	74	70	4	439	168	271	1-1	30	60	1	
336	水北 3	기초	朴守龍	33	0.7	29	29	—	174	73	101	1-1	5	30	1	
337	水西 1	기초	權五福	37	1.3	35	35	—	180	78	102	1-1	18	19	1	
338	水西 2	기초	崔五相	38	0.6	55	55	—	318	112	206	1-1	48	20	1	
339	梨枝 1	기초	崔龍煥	36	2.6	50	40	10	316	126	190	1-1	41	10	1	
340	梨枝 2	자조	都相武	36	1.1	96	96	—	550	221	329	1-1	98	34	1	
341	芭田洞	기초	崔成煥	33	0.8	86	84	2	498	194	304	1-1	76	34	3	
342	新德 1	기초	洪鍾建	29	0.8	45	40	5	227	84	143	1-1	30	14	1	
343	新德 2	자조	金東術	40	0.9	62	58	4	413	165	248	1-1	27	39	2	
344	芝湖 1	기초	李泰元	30	1.1	29	29	—	176	67	109	1-1	22	9	1	
345	芝湖 2	자조	全詳泰	35	0.8	57	56	1	298	128	170	1-1	18	48	1	
346	芝湖 3	자조	張壽根	25	0.7	50	47	3	241	89	152	1-1	19	35	1	
347	蓮桂 1	기초	金出煥	36	1.1	63	52	11	361	141	220	1-1	44	35	2	軍威農協
348	蓮桂 2	자립	孫遠植	35	0.9	76	73	3	449	173	276	1-1	76	20	1	
349	梅城 1	기초	朴正大	43	0.9	38	36	2	206	79	127	1-1	9	35	1	
350	梅城 2	자립	李洙鍾	33	0.9	60	54	6	333	133	200	1-1	23	49	1	
計	**19**				**0.9**	**1,397**	**1,129**	**268**	**7,506**	**2,947**	**4,559**		**905**	**628**	**23**	

1385

Figure 9: An example of village data from *New Village Comprehensive Survey*

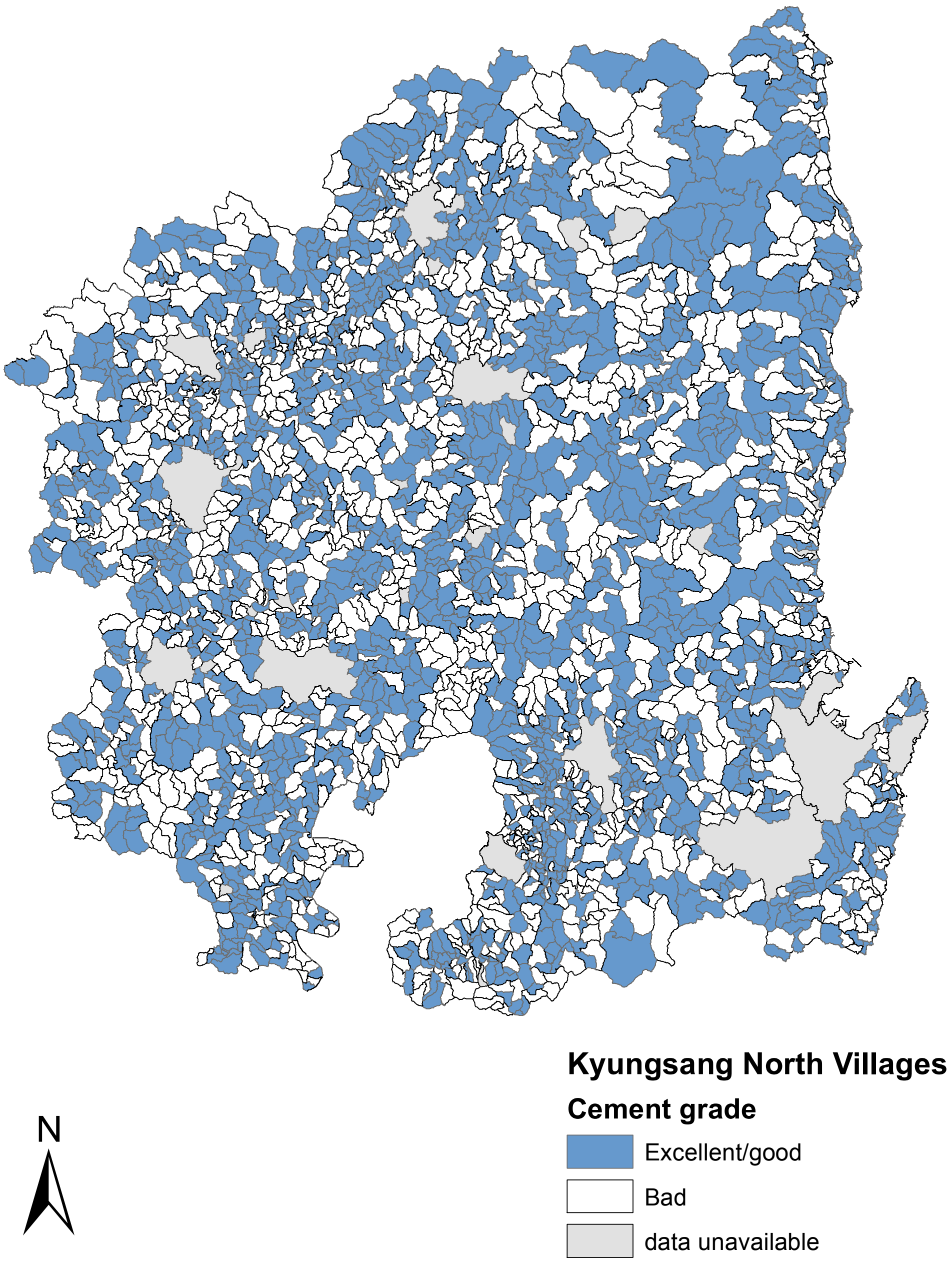


Figure 10: Spatial distribution of cement project grades

뒷산 2ha에 낙엽송을 조림하여 먼 훗날 후손에게 자랑스러운 마을을 물려주게 되었다.

○ 74년 1차 대통령특별지원 하사금으로 1,028평의 마을공동농장을 구입, 인삼재배를 실시하여 76년 늦여름 190만원의 순소득을 올려 호당 4만원씩 이윤배당을 하고 나머지 70만원은 마을금고에 적립하여 자녀교육이 어려운 가정에 대부하여 주었다.

○ 하수구정비 100m, 소하천 620m, 암거 2개소를 완성하여 보다 살기 좋은 마을을 만들었으며

○ 소득증대에도 게을리하지 않아 초지가 좋은 곳이라 새마을지도자는 낙농 꿈을 실현하기 위해 농협융자 400만원을 얻어 마을공동으로 유우 5두를 구입, 당년에 송아지 4두를 분만하여 개가를 올렸으며 아담한 공동축사 25평에 입식시켜 마을 뒷산 3ha를 개간하여 초지를 조성하여 80년에 낙농마을의 꿈을 키우고 있다.

○ 모든 기반을 확충하여 명실공히 부자마을을 만들었으며 74년 11월 8일 대통령각하로부터 새마을 훈장 협동장을 수여받았으며 57가구 357명의 주민은 지난날 끼니를 먹지 못하였던 날을 회고하고 있다.

5) 土地 및 現金喜捨狀況

○ 개미같이 땀흘려 일하는 이 마을 주민들에게 격려와 위로의 온정이 각계에서 보내져 왔고

○ 주민 또한 새마을사업을 위하여 문전옥답과 현금을 아까운 줄 모르고 내놓았다.

<土地喜捨狀況>

年 月 日	姓 名	地 目	地 積	當時換價額	備 考
1971. 3.20	朴氏宗中	山	250평	5,000	
〃	孫成模	畓	5〃	2,500	
〃	심재숙	田	10〃	3,000	
〃	장세환	畓	10〃	5,000	
〃	오수선	〃	20〃	10,000	
〃	洞소유	畓	20〃	10,000	
〃	朴氏宗中	〃	40〃	20,000	
〃	李병문	〃	10〃	5,000	
〃	金영환	〃	20〃	10,000	
〃	朴氏宗中	山	100〃	3,000	
〃	孫윤재	畓	40〃	20,000	
〃	朴氏宗中	〃	10〃	5,000	
〃	沈재숙	田	50〃	15,000	
1974. 2.10	沈재숙	畓	30〃	30,000	
〃	오철영	山	50〃	1,500	
〃	崔판술	畓	20〃	20,000	
〃	金영환	田	50〃	15,000	
〃	심원섭	畓	100〃	50,000	
〃	鄭氏宗中	山	100〃	3,00)	
〃	李창건	畓	100〃	50,000	
計			1,035〃	283,000	

Figure 11: An example of a land donation list from *Glorious Footsteps*

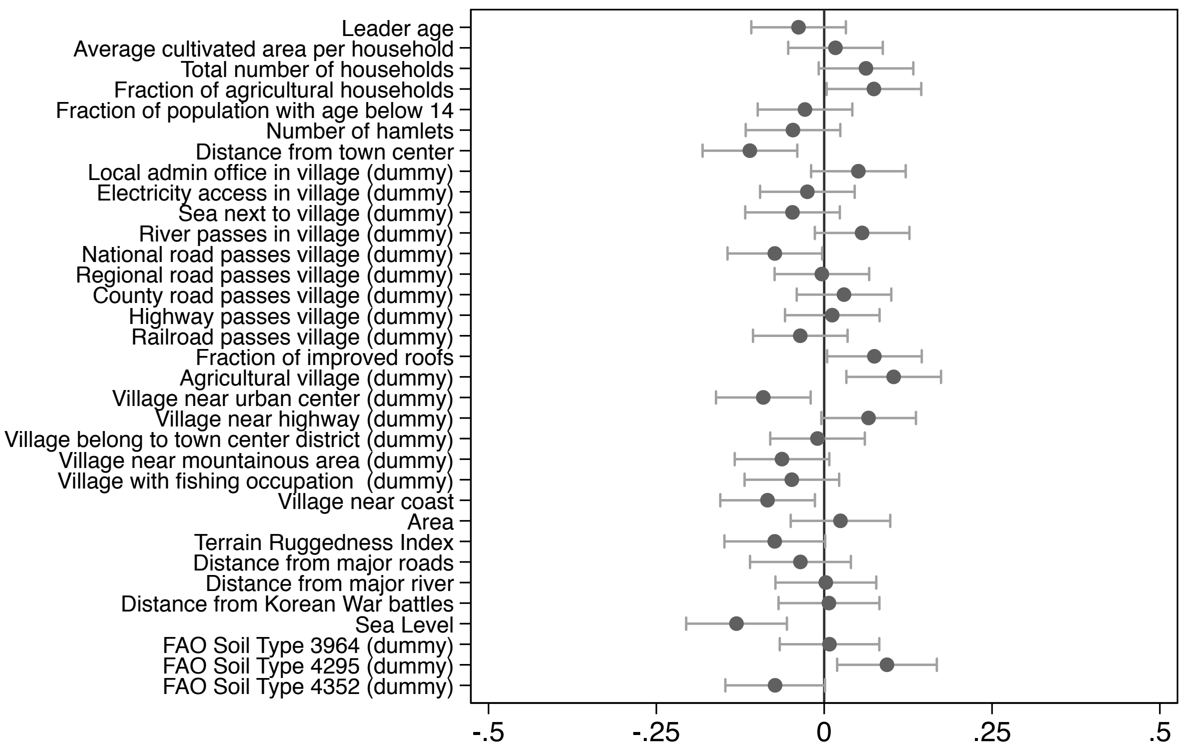
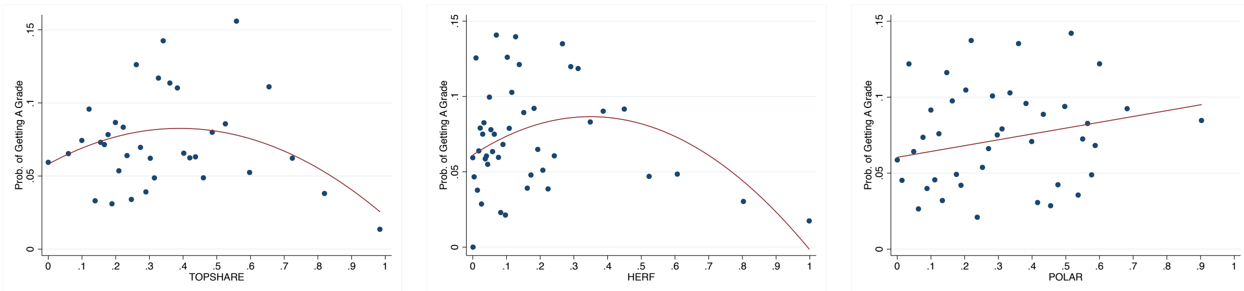
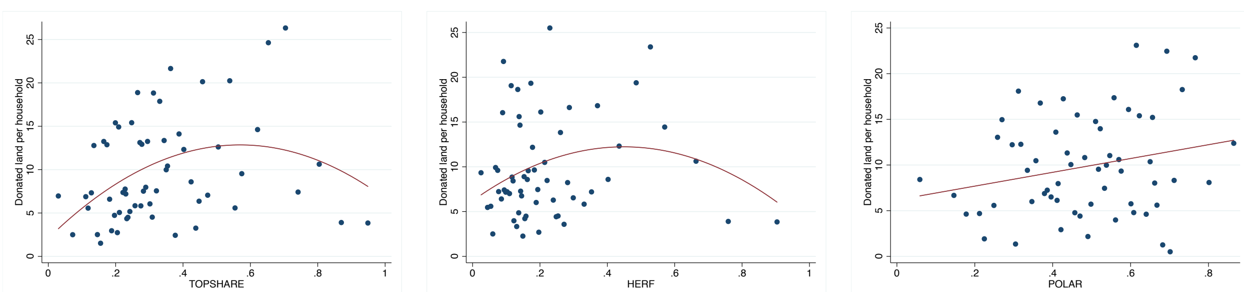


Figure 12: Standardized mean differences of village characteristics between villages with TOPSHARE above (group 1) and below the median (group 0). The dots represent the standardized mean differences of each variable. The lines represent the 95% confidence intervals. The median value of TOPSHARE is zero. The values are calculated by $(\mu_1 - \mu_0)/\sigma_0$, following Kline (2013), where μ_1 and μ_0 are the means of a variable for group 1 and group 0, respectively, and σ_0 is the standard deviation of group 0.



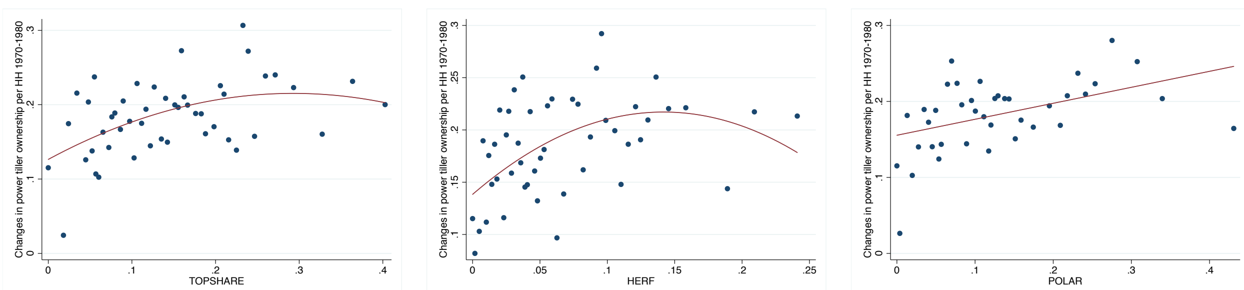
(a) TOPSHARE (b) HERFINDAHL (c) POLARIZATION

Figure 13: Scatter plot showing relationship between probability of getting an A grade (y axis) and group heterogeneity measures (each sub-figure) in the raw data. Each point in the graphs represents the average value of the probability of getting an A grade in a bin with approximately 80 villages in each bin. There are 40 bins. The lines represent the quadratic fit of the data for (a) and (b), and the linear fit for (c).



(a) TOPSHARE (b) HERFINDAHL (c) POLARIZATION

Figure 14: Scatter plot showing relationship between the amount of cultivated land donated per household (y axis) and group heterogeneity measures (each sub-figure) in the raw data. Each point in the graphs represents the average value of the probability of getting an A grade in a bin with approximately 4 villages in each bin. There are 60 bins. The lines represent the quadratic fit of the data for (a) and (b), and the linear fit for (c).



(a) TOPSHARE (b) HERFINDAHL (c) POLARIZATION

Figure 15: Scatter plot showing relationship between the change in number of power tillers owned per household between 1970 and 1980 (y axis) and group heterogeneity measures (each sub-figure) in the raw data. Each point in the graphs represents the average value of the probability of getting an A grade in a bin with approximately 5 villages in each bin. There are 50 bins. The lines represent the quadratic fit of the data for (a) and (b), and the linear fit for (c).

Appendix

A Appendix Tables

Table A1: Timeline of New Village Beautification Project

1969	Excess supply of cement by cement industry
1970	Distribution of cement bags to villages
1970-1971	Production of village public goods
1971	Evaluation by the government (A,B,C grades)
1971	Decision for rewarding villages

Table A2: Top priority village projects identified from a government survey

Order	Description of Project
1	Village access roads to be straightened and widened
2	Old bridges over streams to be reconstructed
3	Village roads to be widened and straightened
4	Sewage systems in village area to be improved
5	Thatched roofs to be replaced by cement made tiles
6	Old fences of farm houses to be repaired
7	Traditional wells for drinking water must be improved
8	Village hall to be constructed
9	River banks to be repaired
10	Feeder roads to fields to be developed
11	Rural electrification to be speeded up
12	Village owned telephones to be installed
13	Village owned bathhouse to be built
14	Children playground to be constructed
15	Laundry place in riverside to be improved
16	Trees and flowers to be planted for beautification

Source: Park (1998)

Table A3: Effects of Family Clan Heterogeneity on Public Good Improvement for non-split villages

	Dependent variable: (Prob. Getting an A grade)				
	(1)	(2)	(3)	(4)	(5)
PANEL A: No controls					
TOPSHARE	0.03	0.23***			
	(0.02)	(0.07)			
TOPSHARE2		-0.28***			
		(0.08)			
HERF			0.01	0.25***	
			(0.03)	(0.09)	
HERF2				-0.32***	
				(0.10)	
POLAR					0.09**
					(0.04)
Observations	1,298	1,298	1,298	1,298	1,298
R-squared	0.00	0.01	0.00	0.01	0.01
PANEL B: Full controls					
TOPSHARE	0.04	0.18**			
	(0.04)	(0.09)			
TOPSHARE2		-0.20**			
		(0.09)			
HERF			0.02	0.14	
			(0.04)	(0.13)	
HERF2				-0.16	
				(0.13)	
POLAR					0.06
					(0.05)
Observations	1,087	1,087	1,087	1,087	1,087
R-squared	0.17	0.17	0.17	0.17	0.17
# of townships	202	202	202	202	202

Note: Robust standard errors in parentheses. Samples are restricted to villages that did not split between 1930 and 1970. In Panel B, standard errors are cultured at the township level. Dependent variable is a dummy variable which equals one if a village received an A cement grade, and zero otherwise. This measure captures production of village public goods. An A cement grade is an indication of more production of public goods compared to a B or C grade. TOPSHARE is the household share of the largest clan in the village. TOPSHARE^2 is the squared term of TOPSHARE. HERF is the Herfindahl Index based on village clan structure. POLAR is the polarization index by García Montalvo and Reynal-Querol (2002). Panel A does not have control variables. Panel B includes various village characteristics including village size, demographic, proximity to urban infrastructure and geographical characteristics. It includes township fixed effects and clan identity dummies.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Effects of clan heterogeneity on the amount of cultivated land per household

Dependent variable: Amount of Cultivated Land per HH					
	(1)	(2)	(3)	(4)	(5)
PANEL A: No controls					
TOPSHARE	-0.12	1.07*			
	(0.23)	(0.61)			
TOPSHARE^2		-1.31**			
		(0.61)			
HERF			-0.26	0.45	
			(0.26)	(0.85)	
HERF^2				-0.93	
				(0.90)	
Polarization					0.09
					(0.31)
Observations	205	205	205	205	205
R-squared	0.00	0.01	0.00	0.00	0.00
PANEL B: Full controls					
TOPSHARE	-0.10	-0.22			
	(0.10)	(0.27)			
TOPSHARE^2		0.13			
		(0.25)			
HERF			-0.12	-0.40	
			(0.12)	(0.38)	
HERF^2				0.36	
				(0.37)	
Polarization					-0.15
					(0.13)
Observations	205	205	205	205	205
R-squared	0.89	0.89	0.89	0.89	0.89
# of Provinces	8	8	8	8	8

Note: Robust standard errors in parentheses. In Panel B, standard errors are cultured by province. Dependent variable is the amount of cultivated land per household in a village. The unit of the dependent variable is hectares. TOPSHARE is the household share of the largest family name in the village. TOPSHARE^2 is the squared term of TOPSHARE. HERF is the Herfindahl Index based on family name distribution. POLAR is the polarization index by García Montalvo and Reynal-Querol (2002). Panel A does not have control variables. Panel B includes various village characteristics including village size, share of agricultural households in the village, irrigation rate, cultivated area per household, total area of the village. It includes province fixed effects. Top one percent of the outcome variable is dropped as outliers.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Effects of family clan heterogeneity on ownership of a wheeled agricultural machine

Dependent variable: Δ Number of Power Tillers per HH 1970-1980	(1)	(2)	(3)	(4)	(5)
PANEL A: Full controls					
TOPSHARE	0.11**	0.33**			
	(0.04)	(0.13)			
$TOPSHARE^2$		-0.63*			
		(0.36)			
HERF			0.11	0.47**	
			(0.07)	(0.21)	
$HERF^2$				-1.84*	
				(0.97)	
POLAR					0.09**
					(0.04)
Peak of Concavity		0.29		0.14	
Mean of Y	0.18	0.18	0.18	0.18	0.18
Observations	245	245	245	245	245
R-squared	0.60	0.61	0.60	0.60	0.60
PANEL B: Full controls with the average share of A cement villages					
Share of A grade villages	0.08***	0.08**	0.09***	0.08***	0.09***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
TOPSHARE	0.11**	0.31**			
	(0.04)	(0.12)			
$TOPSHARE^2$		-0.59*			
		(0.35)			
FRAC			-0.12	2.96*	
			(0.07)	(1.70)	
$FRAC^2$				-1.70*	
				(0.94)	
POLAR					0.09**
					(0.04)
Observations	245	245	245	245	245
R-squared	0.61	0.62	0.61	0.61	0.61

Note: Robust standard errors in parentheses. Dependent variable is the changes in the number of household owned power tillers per household between 1970 and 1980. TOPSHARE is the weighted average of household share of the largest family clan in the village, aggregated at the township level with weight being the number of village households. $TOPSHARE^2$ is the squared term of TOPSHARE. HERF is the Herfindahl Index based on family clan distribution. POLAR is the polarization index by García Montalvo and Reynal-Querol (2002). Panel A does not have control variables. Panel B includes various township level characteristics including township population size, size of cultivated land, literacy, irrigation rate, ownership of agricultural machines in 1970.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Mean comparisons of village characteristics for non-split villages

	TOPSHARE above median	TOPSHARE below median	Diff.	t-stat
Outcome				
Cement Grade (=1 if Excellent/Good)	0.60	0.53	0.07**	2.55
Heterogeneity Measures				
Herfindahl Index	0.26	0.00	0.26***	23.16
Polarization Index	0.37	0.00	0.37***	39.47
Village Characteristics				
Leader age	39.8	39.7	-0.07	0.18
Average cultivated area per household	0.97	0.97	-0.00	-0.13
Total number of households	122	111	11.27***	2.81
Fraction of agricultural households	0.88	0.88	0.01	0.91
Fraction of population with age$<$14	0.39	0.39	-0.00	-1.38
Number of of sub-villages	2.36	2.40	-0.04	-0.55
Distance from town center	4.74	5.41	-0.66***	-3.19
Local admin office in village (dummy)	0.03	0.03	-0.00	-0.02
Electricity access in village (dummy)	0.31	0.29	0.02	0.89
Sea next to village (dummy)	0.01	0.02	-0.01	-1.18
River pass in village (dummy)	0.44	0.38	0.05*	1.9
National road passes village (dummy)	0.13	0.16	-0.03	-1.39
Regional road pases village (dummy)	0.17	0.16	0.01	0.39
County road passes village (dummy)	0.17	0.15	0.01	0.63
Highway passes village (dummy)	0.02	0.02	-0.00	-0.14
Railroad passes village (dummy)	0.07	0.05	0.01	0.86
Fraction of improved roofs	0.38	0.37	0.02	1.05
Agricultural village (dummy)	0.81	0.77	0.04	1.53
Near urban village (dummy)	0.06	0.08	-0.02	-1.17
Near highway village (dummy)	0.06	0.04	0.01	1.10
Village belong to town district (dummy)	0.03	0.03	-0.01	-0.78
Village near mountainous area (dummy)	0.03	0.04	-0.01	-0.71
Village with more than 50% of people with fishing occupation (dummy)	0.01	0.01	-0.00	-0.26
Village located less than four kilometers from coast (dummy)	0.01	0.02	-0.02**	-2.54
Sample Size	485	813		

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A7: Mean comparisons of village characteristics of split v.s. non-split villages

	non-split villages	split villages	Diff.	t-stat
Outcome				
Cement Grade (=1 if Excellent/Good)	0.55	0.51	0.05**	3.06
Heterogeneity Measures				
Clan Share	0.18	0.19	-0.01	-1.28
Herfindahl Index	0.10	0.07	0.02***	3.69
Polarization Index	0.14	0.15	-0.01	-1.15
Clan dummy	0.37	0.51	-0.14***	-7.65
Village Characteristics				
Leader age	39.7	39.6	0.12	0.52
Average cultivated area per households	0.97	0.92	0.05**	2.14
Total number of households	115	226	-111***	-18.30
Fraction of agricultural households	0.88	0.84	0.04***	6.40
Fraction of population with age$<$14	0.39	0.39	0.00	0.10
Number of of sub-villages	2.39	1.88	0.51***	12.96
Distance from town center (km)	5.16	4.22	0.94***	7.18
Local admin office in village (dummy)	0.03	0.05	-0.02***	-3.32
Electricity access in village (dummy)	0.30	0.31	-0.01	-0.73
Sea next to village (dummy)	0.02	0.04	-0.02***	-2.78
River pass in village (dummy)	0.40	0.35	0.05***	3.28
National road passes village (dummy)	0.15	0.15	0.00	0.16
Regional road pases village (dummy)	0.16	0.21	-0.05***	-3.54
County road passes village (dummy)	0.16	0.16	-0.00	-0.34
Highway passes village (dummy)	0.02	0.02	0.00	0.23
Railroad passes village (dummy)	0.06	0.07	-0.01	-1.19
Fraction of improved roofs	0.37	0.42	-0.05***	-5.68
Agricultural village (dummy)	0.79	0.78	0.01	0.61
Near urban village (dummy)	0.07	0.07	0.00	0.47
Near highway village (dummy)	0.05	0.04	0.00	0.42
Village belong to town district (dummy)	0.03	0.06	-0.02***	-3.46
Village near mountainous area (dummy)	0.04	0.02	0.02***	3.68
Village with more than 50% of people with fishing occupation (dummy)	0.01	0.02	-0.01***	-3.39
Village with more than 50% of people with aquaculture occupation (dummy)	0.00	0.00	-0.00*	-1.66
Village located less than four kilometers from coast (dummy)	0.02	0.02	-0.00	-0.21
Sample Size	1298	1826		

* $p<0.10$, ** $p<0.05$, *** $p<0.01$

B Appendix Figures

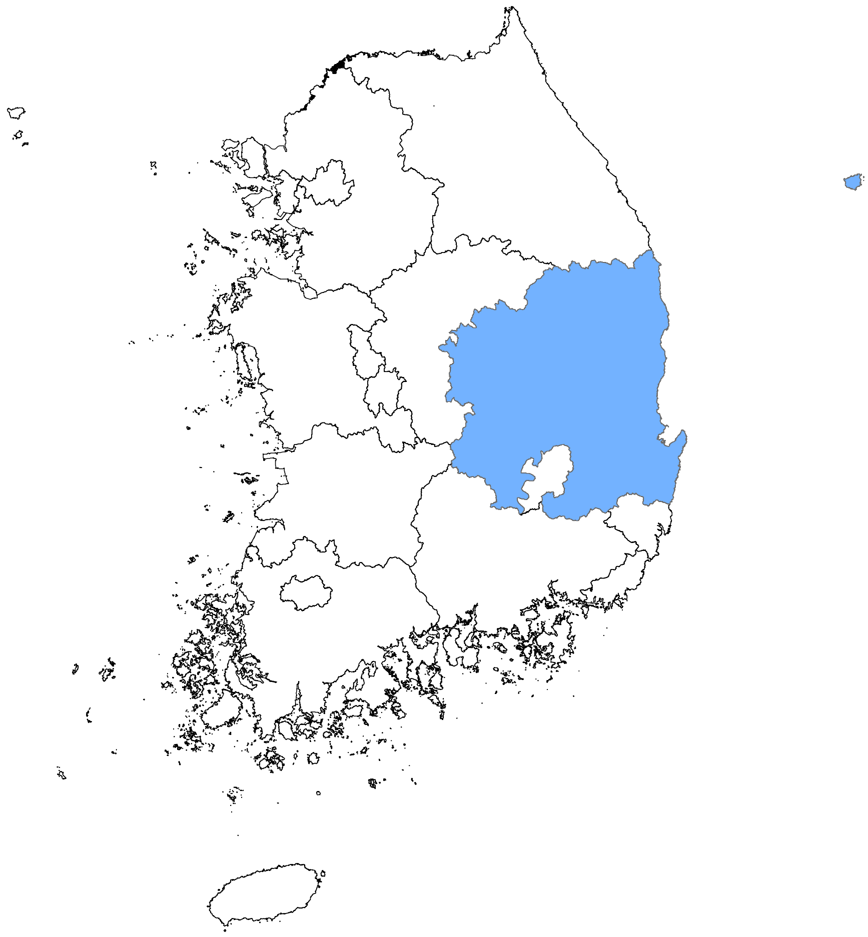


Figure A1: A map of South Korea with provincial boundaries. Kyungsang North Province is highlighted.

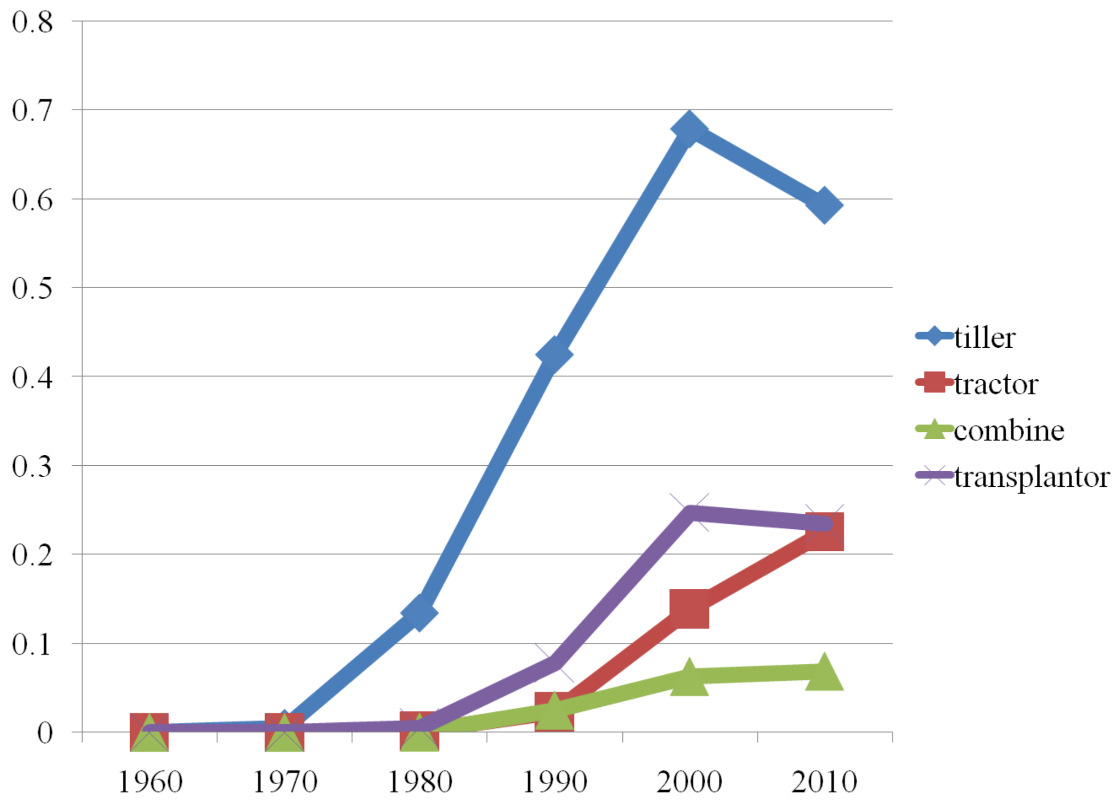


Figure A2: The average number of agricultural machines per agricultural household
(Source: agricultural census)

Figure A3: Power tillers (left) and traditional technologies for land tilling (middle) and transportation (right)

(Source: http://cfile6.uf.tistory.com/image/160FB14B4F9D6EED296D56,
http://www.zipul.com/coding/sub4/sub5.asp?bseq=9&mode=view&cat=-1&aseq=448&page=6&sk=&sv=,
http://blog.joins.com/media/folderlistslide.asp?uid=silhouette7&folder=12&list_id=6922086)

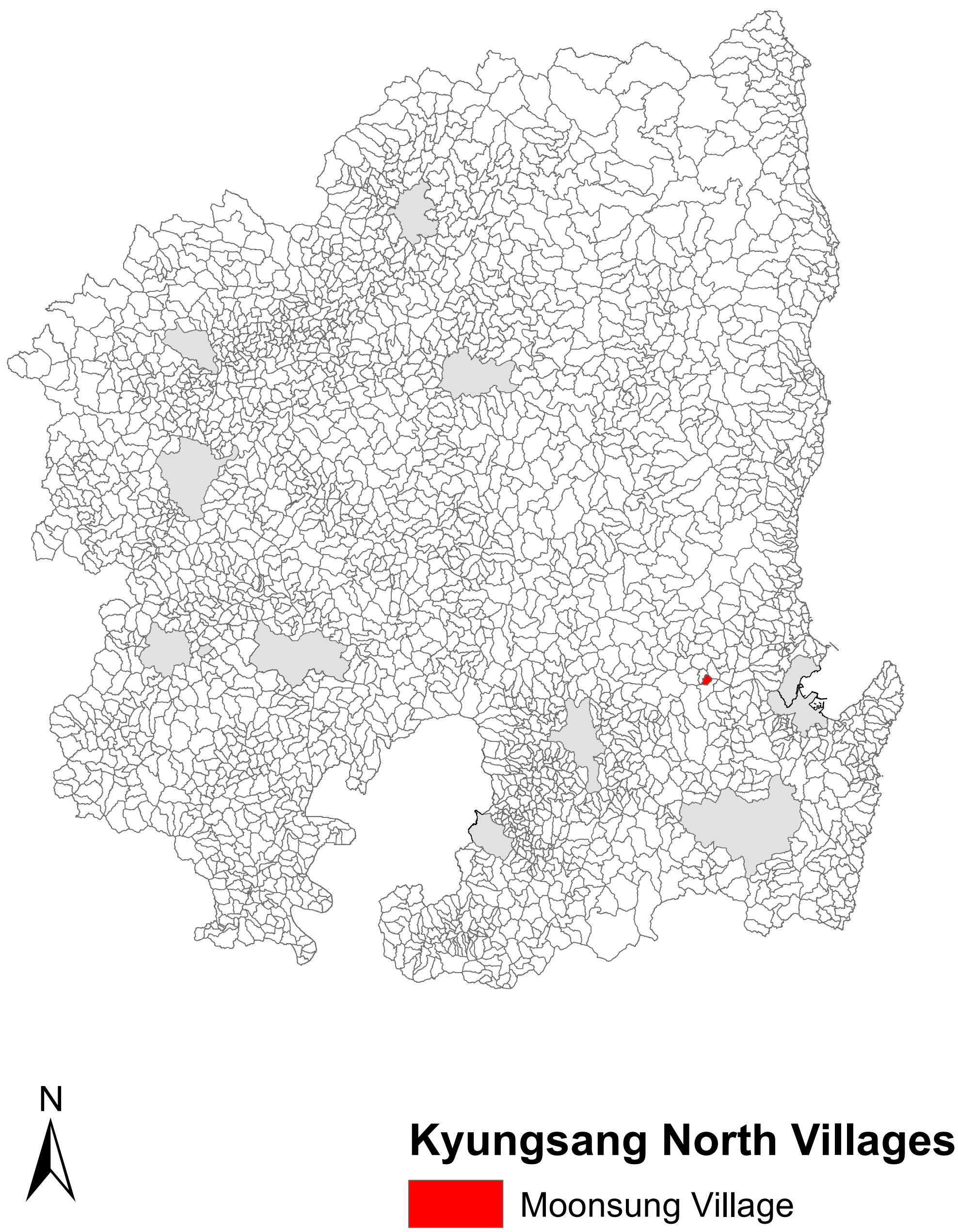


Figure A4: Location of Moonsung village in the case study

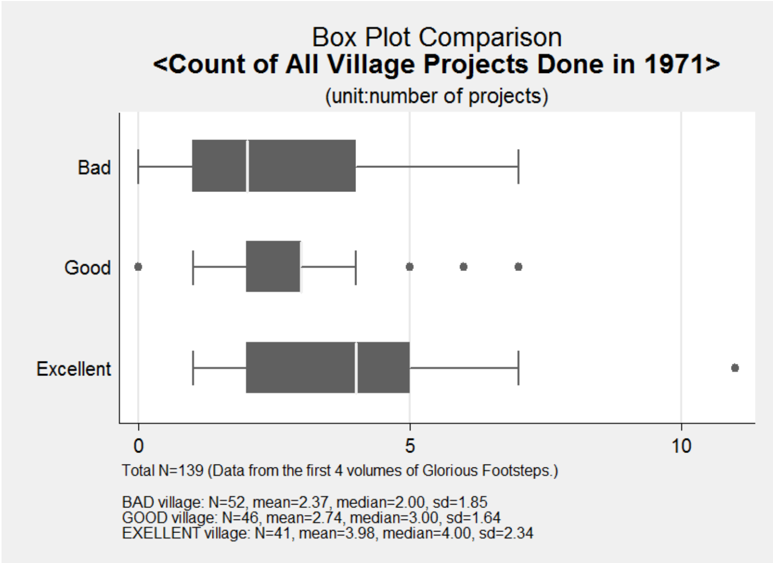


Figure A5: Number of Village Projects done in 1971 by Village Grades

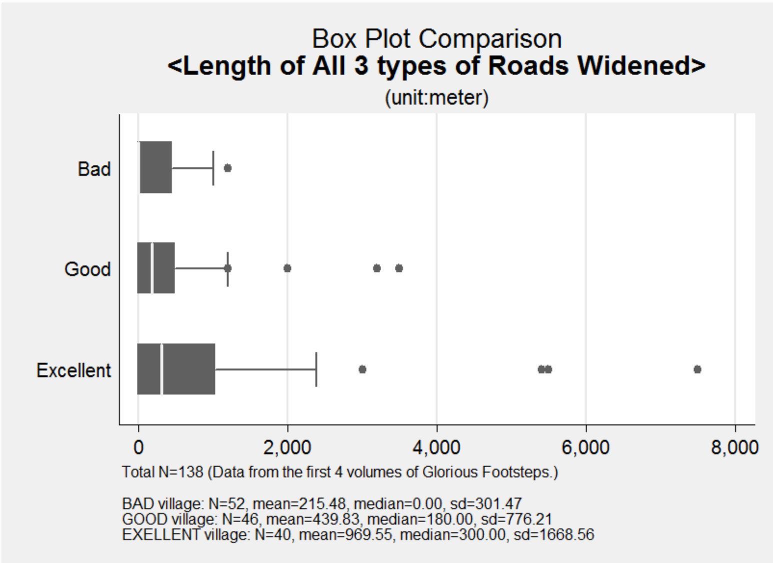


Figure A6: Length of Roads Widened in 1971 by Village Grades

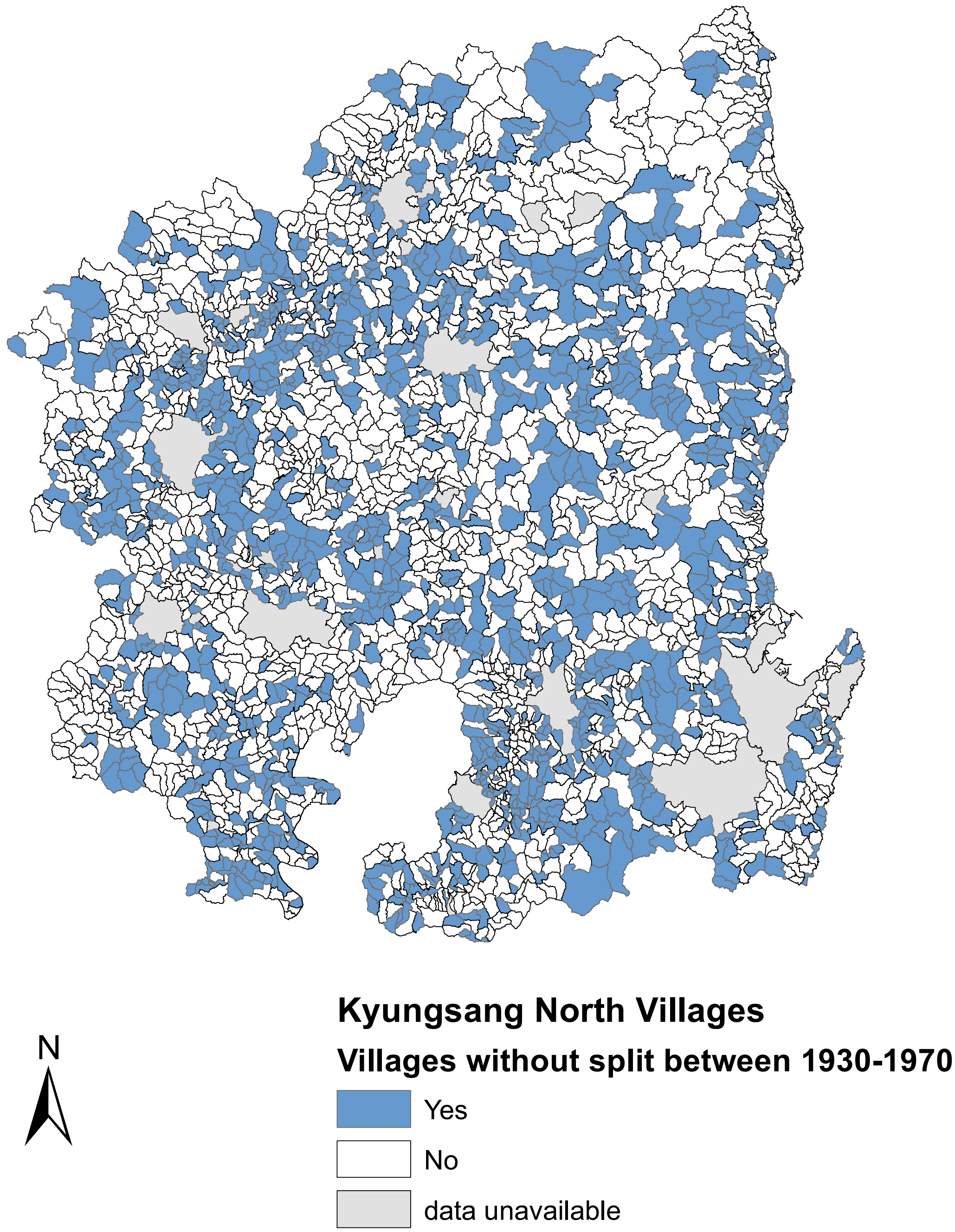


Figure A7: Location of villages that did not experience geographical split between 1930-1970