Essays in Development and Social Economics

Wenhao Cheng

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Economics

University College London

Wednesday 5th June, 2024

I, Wenhao Cheng, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work. I collaborated with Jiahong Han and Jiacheng Xiao on Chapter 3; for all other parts, I am the sole author.

Abstract

This dissertation investigates the causes, evolution, and influence of social structures and norms within the domain of development economics.

Chapter 1 investigates the role of patrilineal kinship as a representative premodern social structure in facilitating male marriages in 19th-century China amid development spurred by a forced port opening. Males with higher centrality among other unmarried male relatives in their patrilineal family tree have an increased likelihood of securing a spouse after the port opening. A model of altruism within networks, supported by corresponding evidence, is developed to suggest that a larger surplus from development incentivizes unmarried males to uphold their connections with their kin groups, thereby enhancing their relevance in resource allocation.

Chapter 2 develop a model to discuss how social learning, specifically Naïve learning, helps coordinate product adoption in development programs. Individuals receive initial signals regarding the value of the product, communicate afterwards and make adoption decisions based on that. The model suggests that as beliefs converge, the result will converge to a unique cutoff equilibrium, as in a global game. It also shows that more adoption is expected with high inequality in network positions if the value of the product to be adopted is low and vice versa.

Chapter 3 investigates the historical origins of son preference and gender bias in China, examining the influence of rice and wheat production on parents' choice of offspring's sex ratio. It shows that provinces/prefectures with larger gaps in rice and wheat suitability exhibit higher male-to-female sex ratios at birth. Furthermore, individuals from regions with larger gaps in suitability tend to have more unequal gender norms.

Impact Statement

In this Impact Statement, I outline the potential impacts of my PhD thesis both within academia and beyond, on a chapter-by-chapter basis.

Chapter 1 contributes to the literature on how economic development affects social structures and institutions. Utilizing micro-level data and employing triple DID, my research provides evidence that development can enhance the influence of traditional institutions. It demonstrates that patrilineal kinship becomes more pronounced in supporting male marriages following the opening of a port. This finding remains relevant in contemporary times, as many traditional institutions are still ingrained in societies, even amid rapid economic development. This research underscores the importance of considering local social structures when devising and implementing development programs, such as cash transfer programs.

Chapter 2 connects the literature on social learning in networks with that on coordination games, demonstrating how naive learning can facilitate behavior coordination. It is motivated by the promotion of products, such as pesticides, in development programs, and proposes numerous policy implications in this context. For example, it suggests that when the intrinsic value of the product to be adopted is not very high, policymakers could empower influential individuals within the conversation network to play a more prominent role in promoting the product to increase adoption.

Chapter 3 contributes to the literature that links agricultural practices to gender norms by adding the cultivation of major crops as an important factor. Understanding the origins of gender inequality can help eliminate biases regarding gender.

Acknowledgements

I am greatly indebted to Antonio Cabrales and Aureo de Paula for guiding me through the PhD journey with their patience, generous support, and constructive feedback. I am also deeply grateful to Nikita Roketskiy, especially for his intensive feedback on Chapter 2 of the thesis. I would like to thank Duarte Gonçalves, Jonas Hjort, Philippe Jehiel, Hyejin Ku, Kalina Manova, Konrad Mierendorff, Franck Portier, Imran Rasul, Gabriel Ulyssea, Cui Wei, Ming Yang, and other faculty at UCL for helpful conversations. My thanks also go to the UCL Department of Economics for their financial assistance and multifaceted support, as well as to my PhD colleagues there for their unwavering support. I thank my coauthors, Jiahong Han and Jiacheng Xiao, without whom Chapter 3 of the thesis would have been impossible. Last but not least, I would like to thank my family, especially my parents, for always giving me the freedom, encouragement, and steadfast support to pursue my endeavors.

Contents

List of Figures

List of Tables

Introduction

This thesis comprises three distinct chapters, addressing topics related to the origins, dynamics, and effects of social structures and norms within the field of development economics. Each chapter features its own introduction, which summarizes the chapter's main points.

The first chapter of the thesis investigates the research question of how economic development mediates the influence of traditional institutions and explores the mechanisms behind this mediation. The study examines the impact of a forced port opening, an exogenous shock that spurs economic development, on the Chinese patrilineal kinship system, showcasing it as a representative traditional institution. Specifically, I concentrate on the support an unmarried male receives from his kin group for marriage. Using China multi-generatioanal panel dataset, I employ a triple DID approach to explore the impact of a forced port opening on the effect of patrilineal kinship ties, using marital status change in the subsequent period as the outcome. The findings indicate that the role of kinship ties becomes more pronounced after the port's opening, as males with a higher centrality among other unmarried male relatives in patrilineal family trees have increased chances of securing a spouse. If an unmarried male's centrality among unmarried males is at the median, the port opening boosts his probability of marrying within the next three years by 2.3%. If his centrality among unmarried males is at its maximum, this likelihood jumps to 18.6% and the average marginal response obtained by utilizing a logit model is 21.4%. A model of altruism in networks is then developed, accompanied by corresponding evidence, suggesting the mechanism that as economic development amplifies the surplus that can be obtained from one's kinship group, unmarried males exhibit a greater willingness to maintain their attachment to their kin groups, and their greater attachment makes the kin group power structure for resource allocation more relevant.

The second paper is motivated by the observation that development programs aimed at promoting products, such as pesticides or agricultural insurance, often encounter coordination problems and adoption rates are often influenced by people's beliefs regarding the usefulness of the product. This paper explores whether consensus formed through the exchange of beliefs can coordinate individuals' behavior. I develop a Naïve learning model in social networks. In the model, individuals in the initial period receive independently distributed signals regarding the intrinsic value of the product to form their initial beliefs. They communicate their initial beliefs with their friends in the network and update their beliefs according to the weights they assign to each friend, and choose to adopt or not after the communication. It shows that after enough rounds of updating, their beliefs converge to a consensus with mild conditions, and the outcome regarding adoption only depends on the value of the consensus. The distribution of the consensus value is influenced by both the intrinsic value and the network structure of information diffusion. When the value of the technology to be adopted is low and there is high inequality in network positions, characterized by the presence of a few opinion leaders, a higher level of adoption is expected. Conversely, when the intrinsic value is high, individuals do not rely as heavily on opinion leaders, resulting in greater adoption within a more equally distributed network structure.

Chapter 3 delves into the relationship between rice and wheat cultivation and the preference for sons in China. This investigation is spurred by the fact that rice cultivation historically required more kinship-level cooperation, with males playing a predominant role. Initially, we develop a theoretical model to outline the impact of rice and wheat cultivation on parental preferences for offspring sex ratios, where parents endogenously choose the within-family sex ratio, taking into account females' comparative advantage in wheat cultivation. The model indicates that the sex ratio is increasing in the disparity between rice and wheat productivity. Subsequent empirical analysis on the effect of rice and wheat production on the choice of offspring's sex ratio reveals that prefectures with significant gap in rice and wheat suitability exhibit higher male-to-female sex ratios at birth. One standard deviation increase in such a gap is associated with an increase in sex ratio by around 3.26%. Moreover, individuals from areas with pronounced disparities in crop suitability are more likely to adhere to unequal gender norms.

Chapter 1

Forced Opening and Reinforced Patrilineal Power: Theory and Evidence from Pre-modern China

1.1 Introduction

Traditional institutions persist in modern societies. For instance, there is mounting concern suggesting that patriarchal values and norms contribute to the recent rise in populism, autocracy, and global social divisions (Chenoweth and Marks, 2022; Sanders and Jenkins, 2022). This persistence is particularly noticeable during the early stages of economic development and industrialization. As an example, Goldin (1994) highlights a downward trend in female labour participation in the early development stages, which also exacerbates gender disparities in many other dimensions, intensifying overall economic inequality (Eastin and Prakash, 2013).

However, the mechanisms behind the persistence of traditional institutions can be complex. It could be a societal inclination to adhere to longstanding institutions, even if their removal might lead to a Pareto improvement. For example, Goldin (1994) posits that the decline in female labour participation might stem from patriarchal values that oppose women's work opportunities. Alternatively, it could be more about power structures: the benefits of economic development might empower those with advantageous positions in traditional institutions, further cementing these institutions' prominence.

In this study, I provide evidence supporting the latter mechanism by focusing on one of the most entrenched patriarchal institutions: the Chinese patrilineal kinship system. My paper consists of two parts. First, utilizing a forced port opening in the 19th century as a development-triggering shock, I provide causal evidence that economic take-offs amplify this institution's role in shaping kinsmen's individual outcomes. Second, I develop a model of altruism networks, providing evidence supporting its fitness, and detail the mechanism: as economic development amplifies the surplus, unmarried males exhibit a greater willingness to maintain their attachment to their kin groups to access more resources, and this increased attachment makes the kin group's power structure for resource allocation more relevant.

I focus on the Chinese patrilineal kinship system due to its representativeness of traditional institutions and the high relevance of its internal power structure to resource allocation. In 19th-century China, the kinship system played a pivotal role in determining an individual's economic fortunes. As Weber (1951) observed, kinship ties (which he refers to as sib relations) in China during this period garnered enormous support gained by patriarchal power, and based on that, ...(a sib organization) supported its members in need through mutual aid and free or cheap credit and economic organizations which went beyond the scope of the individual establishment rested almost wholly upon actual or imitated personal sib relationships.

I use the support provided by patrilineal kinship groups for males' marriages as a primary measure of the influence of these kin groups. What makes it an ideal proxy is the unique marriage pattern in 19th-century China. During that century, China experienced a high prevalence of celibacy and low nuptiality among males, which heightened the challenges men faced in finding spouses due to an abnormally high male-to-female sex ratio estimated to be around 1.08 to 1.33 during that century (Lee et al., 1994). Consequently, men faced significant challenges in finding spouses and often had to accumulate substantial amounts of money for marriage-related expenses, including search costs, bride price, and wedding costs, among others. They relied heavily on their kinship groups as a source of these resources, thereby highlighting the crucial role played by the patrilineal kinship system in resource allocation, where the resources available to a man, and consequently his success in the marriage market, depended significantly on the support from his patrilineal kin.

I study how this influence evolves in light of economic take-off, triggered by a forced port opening at that time. During the 19th century, certain regions in China were compelled by Western powers to open for global trade and foreign investment. These areas experienced accelerated development compared to non-treaty port regions, benefiting from international trade and foreign investment[1]. Within the scope of my data, I specifically examine the opening of *Niuzhuang* Port in Liaoning province. The opening of Niuzhuang Port began in 1858. Before this, it had a purely agricultural economy, but its subsequent prosperity has been extensively documented in historical texts. Leveraging this significant historical event and employing the triple Difference-in-Differences (TD) methodology, I explore the evolving influence of kinship networks given this major economic transformation.

This study relies on the China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN). It is an administrative dataset originally compiled by the imperial government for the purpose of governance. In addition to its rich demographic information, the data provides detailed information on kinship ties, allowing for the placement of individuals within their respective genealogies. To measure the support one receives from his kin group, this research adopts a social network analysis approach utilizing *decay centrality* within one's family tree among selected relatives. I primarily focus on two specific measures of decay centrality: centrality among married male relatives and centrality among unmarried male relatives, because they are expected to exert different impacts on an individual's likelihood of marriage.

In the baseline analysis, I divide the sample into two groups. The Near-Port Group consists of individuals from villages located closer to the port than the median distance among all individuals. The Far-Port Group comprises the remaining

[1]Jia (2014) provides evidence on the positive effects of forced treaty port openings on economic development.

individuals. I use the former as the treatment group and the latter as the control group. For further analysis, I also employ continuous distances, measured in units of 100 km.

The baseline results indicate that having a higher centrality among married male relatives positively impacts the likelihood of getting married. In contrast, the impact of higher centrality among unmarried male relatives is inconsistent, suggesting it may reflect a mixture of both support and competition. However, the effect of centrality among unmarried males has consistently increased due to the port opening. If an unmarried male's centrality among unmarried males is at the median, the port opening boosts his probability of marrying within the next three years by 2.3%. If his centrality among unmarried males is at its maximum, this likelihood jumps to 18.6% and the average marginal response obtained by utilizing a logit model is 21.4%. No similar effects are observed when considering centralities among female relatives.

I present evidence indicating that the increased welfare was related to surging agricultural product exportation and that the port opening increased disparity within kin groups. To ensure robustness, I investigate the presence of systematic missing observations and conduct an event study to address concerns about pre-trends, along with various other robustness checks and evidence against alternative channels, all provided in the appendix.

I develop a model based on altruism in patrilineal kinship family trees to explain the mechanism underlying the observed effects. In this model, unmarried males have the option to join a kinship fund, known as the marriage fund, which aims to help bachelors within the kin find a spouse. Participants contribute a fixed amount of wealth that is subsequently redistributed to maximize participants' total utility, taking into account a social utility component that reflects distance-based altruism. The likelihood of a male getting married is increasing in the wealth he has, which, in turn, is associated with the decay centrality within the group of participants if he is also in the marriage fund.

Married males, being more established and settled, are assumed to always participate in the marriage fund without requesting the return. Hence, as economic development leads to increased wealth, there is a larger surplus available for unmarried males who choose to participate in the marriage fund, resulting in a higher participation rate among them. Consequently, being centrally positioned among unmarried relatives becomes more important, as there are more unmarried males within the kinship fund group who can provide support and assistance to each other. In summary, economic development, resulting in higher average wealth, amplifies the surplus that unmarried males can obtain from their kinship groups. This provides an incentive for those unmarried males with a flexible level of attachment to their kin groups to reinforce these bonds, thereby elevating the significance of power structures embedded in patrilineal lineages.

Furthermore, I assess the degree of fitness between the model and the data. The model proficiently forecasts the probability of future marriages for unmarried males. Employing these predicted probabilities as the dependent variable yields outcomes that resemble the main results of the paper.

Related Literature The evolution of social norms or traditional institutions in economic development is a longstanding topic that garners significant interest from social science researchers. Some early thoughts include Rousseau (1775) suggesting that the culture and institutions of a society are determined by its modes of satisfying material needs, and Marx (1973) claiming that economic forces drive all social change. Recent research utilizing observational or experimental evidence aims to establish the causal relationship between certain dimensions of economic development and specific aspects of traditional institutions or norms. These aspects encompass political institutions (Robinson, 2006, economic organizations (Nabli et al., 1989; Desmet and Parente, 2014), religion and beliefs (Inglehart and Abramson, 1994; McCleary and Barro, 2006), and social structures (Heß et al., 2021; Banerjee et al., 2021). The findings are mixed: while some studies indicate that traditional institutions may be weakened or replaced as the economy develops, others fail to yield significant results in this regard. It is possible that the nature of this relationship is heavily influenced by the specific context being examined.

Particularly, when it comes to the early stages of development and patriarchal systems, some associated norms or institutions appear to remain quite persistent. Jayachandran (2015) notes that cultural institutions such as patrilocality and male-centered funeral rituals contribute to the persistence of gender inequality in developing countries, despite economic growth. Based on empirical findings, some researchers hypothesize that gender biases and patriarchal institutions are reinforced throughout the process of economic development (Forsythe et al., 2000; Eastin and Prakash, 2013). Szołtysek et al. (2017) propose a metric of patriarchy using historical data on familial behaviour and demographics, demonstrating that patriarchal institutions did not significantly decline in the 19th century in Europe. Consequently, it becomes necessary to closely investigate the influence and mechanisms of related institutions as well as their interaction with economic fundamentals. Yet, there remains a dearth of quantitative evidence that directly tackles this issue. My research seeks to address this gap by examining the evolving influence of patrilineal kinship on individual outcomes, with an exogenous shock to the agricultural economy.

This study also contributes to a distinct strand of literature focused on the effect of the Chinese kinship system. It has been well-understood that kinship has played a very significant role in Chinese history. During the late imperial China period, lineages or clans evolved into well-organized corporations characterized by joint households and shared ownership (Freedman, 2021; Watson, 1982), which distinguishes them from most Western countries and even other East Asian countries (Bengtsson et al., 2004). Due to its significance, there is ample evidence investigating the influence of kinship on individual outcomes in China, such as human capital (Shiue, 2017), economic status (Peng, 2004; Tang and Zhao, 2023) and demographics (Telford, 1995; Harrell and Pullum, 1995; Dong, 2018; Zhang, 2020). A particularly relevant study to mine is Campbell and Lee (2011), as they document positive correlations between how central a male is in his kin network (measured by the number of male relatives in various categories) and individuals' achievements in obtaining official positions, entering into first marriages, and engaging in reproduction in 19th-century northeast China. Although these studies effectively document the important effects of kinship, few of them consider the overall structure of the kinship network when assessing an individual's position within it. This oversight may lead to an underestimation of the impact of kinship, as the interests of individuals belonging to the same descent group are inherently interconnected[2]. Furthermore, most studies examine the influence of kinship in a static manner, overlooking its changing role. My study also aims to fill these gaps in the existing strand of literature.

The remainder of the paper is organized as follows. Section 2 provides the historical background. Section 3 presents the data and descriptive statistics. Section 4 discusses the empirical strategy. Section 5 presents the main empirical results. Section 6 includes robustness checks. Section 7 discusses the possible channels causing the empirical results. Section 8 presents a theoretical model. Section 9 evaluates the model's fit with the data. Section 10 concludes.

1.2 Historical Background

In this section, I present a brief introduction to the relevant historical background in 19th-century China. For a more detailed exploration of the historical context related to late imperial China's kinship, nuptiality, and the opening of Niuzhuang port, along with the supporting literature and historical evidence for the statements presented in this section, please refer to Appendix A.

The 19th century marked a period of multiple crises in China. First of all, it was regarded as the century of humiliation due to the imperial government's forced acceptance of a series of unequal treaties following military defeats against Western powers. As a consequence, the government was compelled to open several coastal cities, known as *Treaty Ports*, to global trade and foreign investment. Among these ports, Niuzhuang was established in 1858 following the signing of the *Tianjin Treaty* between imperial China and four Western countries: the UK, France, America, and

[2]For instance, concerning risk-sharing, an individual's consumption may depend on the structure of the entire kinship network, even if people only help out close relatives (Ambrus et al., 2014; Bourlès et al., 2021).

Russia. Despite its origins in military defeat, the opening of Niuzhuang Port brought about prosperity to the area. It rapidly became one of the busiest ports in imperial China, attributing numerous foreign corporations and contributing to the improvement of local residents' welfare.

The second significant crisis pertains to demographics. During the Qing dynasty, China experienced remarkable population growth, rising from approximately 150 million in the late 17th century to around 450 million in the late 18th century (Smil, 1993). This rapid population increase placed immense pressure on individuals' lives, particularly since China was still trapped in the Malthusian Trap during that time. Moreover, due to an unusually high male-to-female sex ratio, low nuptiality among males was prevalent in the Qing dynasty, particularly in the 19th century. Given the challenges of finding a spouse and the burdens of living stress, ordinary males often relied on their kinship groups for support in their marriages. Consequently, their probability of marriage, along with other individual outcomes, became heavily dependent on their position within the patrilineal kinship group (Lee et al., 1997; Bengtsson et al., 2004), which was widely recognized as a patriarchal institution in imperial China. With these phenomena, as well as the shocks caused by the forced opening of treaty ports[3], 19th-century China offers an ideal background for studying the evolving impact of patrilineal kinship during economic development.

1.3 Data

This paper utilizes data from the China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN). The whole dataset comprises information on over $260,000$ residents in present-day Liaoning province, China, spanning the years 1749 to 1909, with more than 1.5 million observations. Each unit of observation within the dataset includes a wide range of demographic characteristics, such as age, sex, number of sons and daughters, occupations, and more. This section offers an overview of the data utilized in the research and describes the related data processing methods. For

[3] Lee and Campbell (2005) documents some evidence indicating that the opening of the port had positive effects on the demographic outcomes of local individuals.

further discussion on the data limitations, please refer to Appendix B.4.

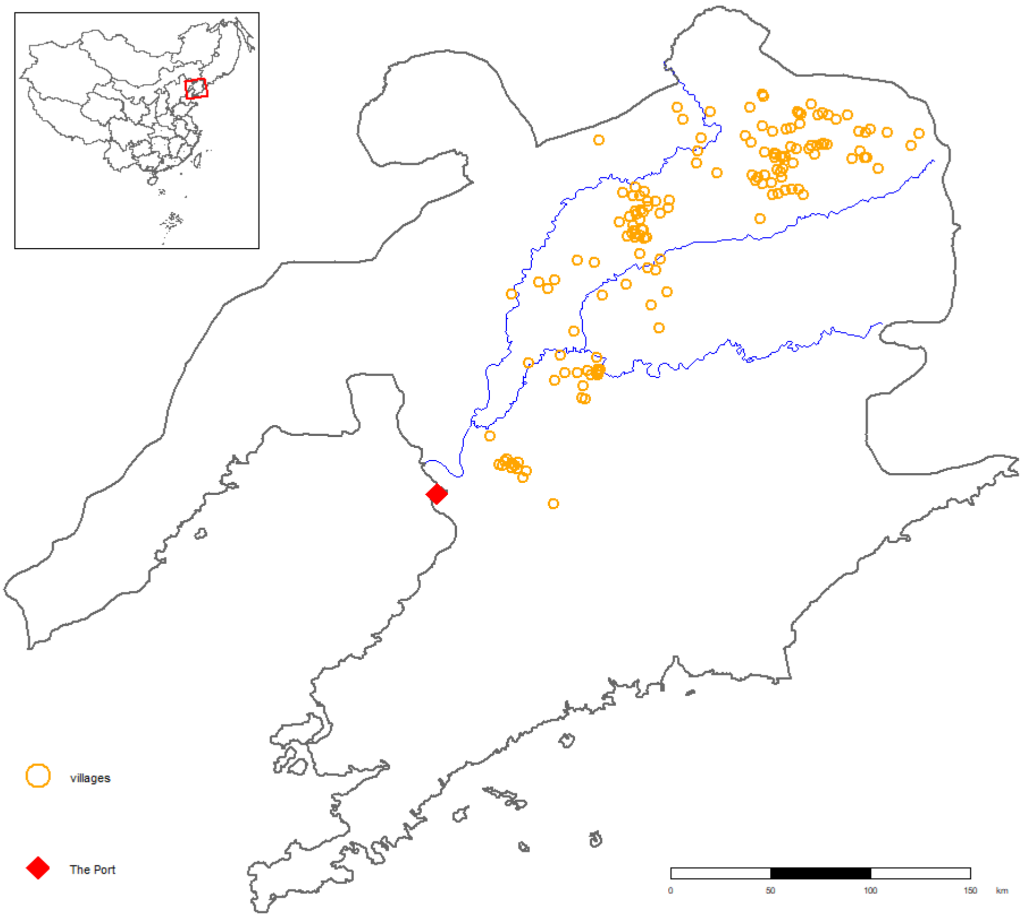
1.3.1 Individuals in the Sample

The dataset primarily derives its records from the population registers of the Liaodong Eight Banners. The Qing dynasty (1644-1912), which was the last dynasty of China, was an empire established by the Manchus, who originated from the Northeast of China, with Liaoning province, known as Fengtian Province at the time, being a central part of their domain. After the Manchus took over mainland of China, the Northeast was still governed as Manchurian provinces under the system of the Eight Banners, which divided its population into eight different administrative banners, including plain and bordered yellow, white, red, and blue, and kept them under paramilitary administration. The individuals in the CMGPD-LN dataset were from the Northeastern banners, and their records were maintained by the Shengjing Imperial Household Agency. The dataset predominantly consists of descendants of Han immigrants who migrated to Liaoning province during the late seventeenth and early eighteenth centuries (Ding et al., 2004) and were organized as hereditary labourers of the Imperial Household by the Qing government. Registration occurred every three years, making three years a unit of the period in this panel data.

The residents of the Eight Banners enjoy certain minor privileges, such as the opportunity for state employment. However, the majority of them still maintain a peasant lifestyle. At the same time, residents of the Eight Banners are restricted from freely immigrating, and those who fail to comply are documented as absconded. Additionally, they are prohibited from marrying individuals outside of the Eight Banner. Further details about the Eight Banners can be found in Appendices A.1 and B.1.

Geographically, the sample residents of CMGPD-LN are primarily located in the North, Central, South-Central, and South regions of Liaoning Province. In the baseline analysis, I intentionally excluded individuals residing in the mountainous areas southeastern to Niuzhuang District. This is due to the potential complexity of assessing the opening's impact on these areas, given their proximity to the port

Figure 1.1: Map of Liaoning (Fengtian) Province in 19th century



Note: The red diamond marks the location of the port, while the orange circles indicate the villages where the sample individuals reside.

yet mountainous terrain. By excluding these areas, a majority of the sample aligns within the Liaohe Plain and along the primary shipping rivers. Consequently, their distances to the port become comparable, establishing a reliable measure for treatment intensity.

Figure 1.1 illustrates a map displaying the sampled areas. It highlights Niuzhuang District with a red circular outline. The red triangle marks the location of the port. They originate from 697 villages and are affiliated with 13 districts. Despite being sourced from the same province, there is significant geographic variation within Liaoning, ranging from the Northernmost to the Southernmost areas.

1.3.2 Time Span

I choose the triennial periods from 1849 to 1885 as the periods for my empirical analysis. According to Lee et al. (2010), the dataset exhibits its highest quantity and quality during the mid to late nineteenth century, while earlier data suffers from a more significant issue of missing data. Furthermore, the 19th century in China was characterized by profound societal transformations and major historical events. For instance, the signing of the *Treaty of Nanking* in 1842 marked the first instance of China being compelled to open up, and the outbreak of the First Sino-Japanese War in 1894 turned Liaoning Province into a battlefield, among other significant events. Employing too many periods could expose the sample to a greater number of structural shocks, potentially undermining the validity of the findings.

Another rationale for excluding early years is the need to trace lineages when constructing kinship networks. Lineages cannot be accurately depicted for individuals in the early years of the data. For instance, if an individual's grandfather passed away before 1749 (the initial period of the data), we do not know who the individual's grandfather is. Consequently, if that individual and one of his or her cousins, who share the same grandfather, are included in the 1749 sample, their blood relationship would remain unknown. This situation introduces measurement errors, particularly given the significant issue of missing data in the 18th-century sample.

Thus, this study designates roughly ten years (three periods) before the opening as pre-treatment periods and about thirty years (ten periods) following the opening as post-treatment periods, covering the sample period from 1849 to 1885.

1.3.3 Kinship Network Construction

Each individual in the dataset is assigned a unique ID number for identification purposes. Also, the ID numbers of an individual's mother and father are recorded as two separate variables in the dataset, enabling the establishment of parental connections. By utilizing this information, it is possible to construct networks that represent kinship relationships by connecting individuals to their respective parents.

In the main empirical analysis, I establish family trees by linking each individual to their parents, or to their parents-in-law for females, and linking husbands to their wives, thereby creating kinship networks for analysis. The constructed family trees are traced back up to 30 years. This cutoff ensures that individuals who share a common ancestor who passed away 30 years ago are considered part of different kinship networks, and vice versa. This cutoff is subject to be varied in a robustness check.

Individuals belonging to the same kinship network and who are alive during a specific period are referred to as a *kinship group* at that particular period. Given a kinship network, I use *decay centrality* to measure how central an individual is in the network. For a given patrilineal kinship group indexed by k, denote C_{ik}^{S} as the patrilineal decay centrality of a member i that measures how central he is within a selected group of his relatives, N_k^S. C_{ik}^{S} is calculated as follows:

$$C_{ik}^{S} = \sum_{j \in N_k^S \setminus i} \alpha^{d(i,j)}$$

Here, $d(i,j)$ represents the distance (i.e., the length of the shortest path) between individuals i and j within the kinship network. The parameter α is a predefined value that determines the weight assigned to the decay factor. In the baseline analysis, α is set to 0.5. The expression abstracts the time period, as the kinship group can exist in any arbitrary period. It is important to note that throughout this paper, the term centrality specifically refers to the decay centrality defined in this manner.

The selected group N_k^S can vary to evaluate centrality across different subsets of relatives. For instance, if N_k^{Male} represents the living male relatives of individual i, then C_{ik}^{Male} would denote his decay centrality within this group, i.e. his patrilineal kin group. In this study, three primary measures are used: centrality among married male relatives, centrality among unmarried male relatives and centrality among married female relatives. Throughout this paper, they are referred to as *centrality among married males*, *centrality among unmarried males* and *centrality among married females*, respectively.

For a more detailed explanation and justification of the approach used to construct kinship networks in this paper, please refer to Appendix B.3.

1.3.4 Dependent Variable, Treatment and Covariates

The empirical analysis in this paper mainly involves comparing the effect of one's position in the network on their marriage before and after the forced opening of Niuzhuang Port. The primary measure for the outcome is a binary variable indicating whether an individual gets married within the next three years (in the subsequent register). Since only those who were unmarried before could potentially have a value of one for this dummy variable, I limit the sample in the empirical analysis to unmarried males only. Unmarried males for whom subsequent records and village locations are available account for 56,397 observations from 1849 to 1885.

Given that the representation of individuals residing in the Niuzhuang district is relatively small within the sample, it becomes challenging to establish whether the parallel trends assumption is met and whether the control group is a sufficiently accurate representation of what would have happened in the Niuzhuang District without the opening. Concerns regarding the spillover effect are also prominent, particularly considering that certain areas in other districts are even closer to the port area than the centre of the Niuzhuang District.

To address these concerns, I utilize the distance between an individual's village and the port to indicate treatment. Each individual in the CMGPD-LN dataset is associated with a unique village ID, indicating their place of residence. Additionally, a restricted file of CMGPD-LN contains information regarding the latitudes and longitudes of certain villages in the sample. By merging the dataset with the village information, I obtain the distance between each individual's village and the port area. I utilize the interaction between this distance and the dummy variable indicating the post or pre-treatment period as a measure of treatment intensity.

The control variables as individual outcomes consist of age-fixed effects, generation-fixed effects, birth order, estimated income (with non-zero values only for a few civil servants), whether his father is alive, whether his mother is alive, whether he is a clan chief, birth order and whether he is the oldest among all living siblings. I also control for the number of unmarried brothers to ensure that the main results are not influenced by variations occurring within each household. Additionally, the size of the kinship group and the size of unmarried males in the kinship group are considered kinship group characteristics. Appendix B.2 provides summary statistics for key variables.

1.4 Empirical Strategy

I employ a triple difference-in-difference approach to examine the effect of the treatment (the opening of the port) on the support provided by the patrilineal kinship network for men's marriage. The main specification is as follows:

$$\begin{aligned} Y_{ifvt} =& \alpha_f + \alpha_v + \alpha_t + \beta_1 D_v \times t + \beta_2 C^M_{ifvt} + \beta_3 D_v \times C^M_{ifvt} + \beta_4 Post_t \times C^M_{ifvt} \\ &+ \beta_5 D_v \times Post_t \times C^M_{ifvt} + \beta_6 C^U_{ifvt} + \beta_7 D_v \times C^U_{ifvt} + \beta_8 Post_t \times C^U_{ifvt} \\ &+ \beta_9 D_v \times Post_t \times C^U_{ifvt} + \beta_{10} D_v \times Post_t + \gamma' \mathbf{X}_{ifvt} + \varepsilon_{ifvt} \end{aligned} \tag{1.1}$$

where i, f, v and t are subscripts representing individuals, kin group founders, villages and record years, respectively. α_f, α_v and α_t represent the corresponding fixed effects. Y_{ifvt} is the outcome variable indicating whether that unmarried male get married within the next three years. C^M_{ifvt} and C^U_{ifvt} are the measures of decay centrality of that individual at that time among married and unmarried male relatives respectively, as previously defined. D_v indicates the proximity of village v to the port. In the baseline regression, it is a dummy variable set to 1 if the village is closer to the port than the median distance, i.e. belongs to the Near-port Group rather than the Far-Port Group. $Post_t$ is a dummy that indicates whether the year t is after 1858. $\mathbf{X}_{ifvt}$ represents the individual characteristics. ε_{ifvt} is an independently and identically distributed error term with an unconditional mean of zero.

This specification falls within the framework of the triple difference (TD) method but with C^M_{ifvt} and C^U_{ifvt} representing continuous treatment variables. To avoid confusion, in this paper, the term treatment exclusively refers to the port opening, despite the TD nature of the specification. For the purpose of discussing identification, I simplify equation (1.1) by combining C^M_{ifvt} and C^U_{ifvt} into a single variable, denoted as C_{ifvt}. The equation then becomes:

$$\begin{aligned} Y_{ifvt} = \alpha_f + \alpha_v + \alpha_t + \beta_1 D_v \times t + \beta_2 C_{ifvt} + \beta_3 D_v \times C_{ifvt} + \beta_4 Post_t \times C_{ifvt} \\ + \beta_5 D_v \times Post_t \times C_{ifvt} + \beta_6 D_v \times Post_t + \gamma' \mathbf{X}_{ifvt} + \varepsilon_{ifvt} \end{aligned} \tag{1.2}$$

The results regarding identification that hold for equation (1.2) also hold for equation (1.1). Let $\Upsilon_{ifvt} := \frac{\partial Y_{ifvt}}{\partial C_{ifvt}}$, which denotes the response of the probability of marriage for individual i from kin group k in district d at period t, with respect to changes in his centrality. Denote $\Upsilon^{(1)}_{ifvt}$ as this response when the individual is treated (due to the port opening), while $Y^{(0)}_{ifvt}$ represents the same response when the individual is untreated. The objective is to identify $E[\Upsilon^{(1)}_{ifvt} - \Upsilon^{(0)}_{ifvt} | D_v = 1, Post_t = 1]$, which captures the difference in centrality effects caused by treatment within the treated group[4]. I call this *average treatment effect on the treated about the response* (ATTR). Identifying ATTR does not rely on the standard OLS assumption of exogeneity (i.e., the error term having a conditional mean of zero) or parallel trends regarding the outcome. It is identified and equal to β_5 under two conditions: 1) parallel trends regarding $\frac{\partial \varepsilon_{ifvt}}{\partial C_{ifvt}}$ between the treated and control groups, and 2) parallel trends regarding $\Upsilon^{(1)}_{ifvt}$ between the treated and control groups. The proof and relevant discussion are available in Appendix D.

Hence, the effect of the port opening on the relationship between centrality and the outcome of interest can be estimated by examining the coefficient β_5 in equation (1.2), which corresponds to coefficient β_5 and β_9 in equation (1.1). Statistically significant β_5 or β_9 in equation (1.1) would suggest that the port opening changes the degree of how centralities influence the marriage rate.

If we replace the dummy D_v with a continuous D_v, namely, the distance between the village and the port, the interpretation of coefficient β_5 in equation (1.2) becomes complicated. Given treatment effect heterogeneity[5], it should not be interpreted as approximating any causal parameters (Callaway and Sant'Anna, 2021).

[4]In TD case, unconditional average treatment effect (ATE) is not identified. See Olden and Møen (2022).

[5]Treatment effect heterogeneity is expected in my case, as port opening preferentially benefits areas surrounding the port.

Furthermore, to the best of my knowledge, there have been very few findings so far concerning identification in triple DID scenarios with continuous treatment. Nonetheless, I include the results obtained using continuous D_v, as it is common practice in similar research.

1.5 Main Results

In this section, I present the main results based on specification (1) and its variations. Table 1.1 presents the main results. In Columns 1-5, the variable *Proximity* (D_v in equation (1.1)) serves as a dummy variable indicating whether an individual comes from a village located closer to the port than the median distance. Column 1 indicates a positive effect of port opening: being in the Near-Port Group after the opening on average increases the probability of marriage within the subsequent period (three years) by around 2%. In Columns 2 and 3, centrality among married males and centrality among unmarried males are introduced as the third interacting variables, individually and respectively. In Column 4, both are included simultaneously, as specified in equation (1.1). Column 4 suggests that port opening increases an individual's probability of marriage within the subsequent period by 9.4%, assuming his centrality among unmarried males is equal to 1, irrespective of other effects. This figure changes to 2.3% if his centrality among unmarried males is 0.25 (the median), and rises to 18.8% if it reaches 2 (the maximum). Column 5 reports the result of using a logit model instead of OLS. The numbers reported are average marginal responses. It implies that being in the Near-Port Group increases the probability of marriage within the next period by 21.4%, averaging with respect to centrality among unmarried males.

Columns 6-10 present the same estimations as 1-5, but with *Proximity* being measured as the geodesic distance in units of 100 km. Column 8 suggests that being 100 km closer to the port after the opening raises one's probability of marriage within the next three years by 6.3% if his centrality among unmarried males is 1. Unlike in Columns 1-5, the coefficients associated solely with centrality among unmarried males turn negative. This suggests strong competition among unmarried kinsmen around the port area prior to its opening.

To highlight the relevance of patrilineal kinship ties, Table 1.2 presents the effects of female relatives and the impact of port opening on these effects, compared to their male counterparts.

Column 1 reveals that both centrality among married and unmarried females positively influence marriage rates, while the effect from married female relatives is significant. However, Columns 2 and 4 indicate that the impact of port opening on these effects is not statistically significant. In Columns 3 and 5, when centralities among males and related interactions are included, the only significant triple interaction is the interaction involving port opening and centrality among unmarried males.

While having female relatives may enhance a male's likelihood of marriage through various means such as matchmaking and information diffusion, the results suggest that these factors remain unaffected by the port opening. In this paper, I argue that the port opening influences a male's chances of marriage through patrilineal kinship power resource allocation.[6]

[6] Regarding the marriage market, in Appendix D.2, I also argue that a greater number of women were married into the Near-Port area following the port's opening.

Table 1.1: Changing effects of centrality among married and unmarried males

	Dep. Var.: Marriage next period									
	Proximity: Whether belong to Near-Port Group					Proximity: Geodesic distance in 100km				
	OLS				Logit	OLS				Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Proximity×Post	0.020***	0.031*	0.001	0.023	0.073	-0.011	-0.015	-0.001	-0.010	-0.011
	(0.005)	(0.014)	(0.002)	(0.013)	(0.201)	(0.010)	(0.010)	(0.006)	(0.010)	(0.080)
Centrality among married males		0.053***		0.040**	0.129***		0.102**		0.109*	0.337***
		(0.012)		(0.012)	(0.113)		(0.040)		(0.053)	(0.398)
Centrality among unmarried males			0.090***	0.056**	0.141***			-0.020	-0.089	-0.341
			(0.022)	(0.015)	(0.209)			(0.056)	(0.049)	(0.917)
Proximity×Centrality among married males		0.026		0.044*	0.096*		-0.017		-0.025	-0.073
		(0.018)		(0.020)	(0.213)		(0.018)		(0.024)	(0.187)
Proximity×Centrality among unmarried males			-0.049*	-0.070**	-0.137***			0.043	0.055**	0.178**
			(0.020)	(0.019)	(0.213)			(0.022)	(0.020)	(0.367)
Post×Centrality among married males		-0.011		0.003	0.005		-0.027		-0.051	-0.129
		(0.019)		(0.021)	(0.271)		(0.049)		(0.052)	(0.549)
Post×Centrality among unmarried males			-0.049**	-0.053**	-0.109*			0.091*	0.119**	0.387**
			(0.018)	(0.021)	(0.250)			(0.046)	(0.040)	(0.768)
Proximity×Post ×Centrality among married males		**-0.015**		**-0.038**	**-0.089**		**0.005**		**0.018**	**0.046**
		(0.022)		(0.025)	(0.321)		(0.025)		(0.028)	(0.308)
Proximity×Post ×Centrality among unmarried males			**0.072*****	**0.094*****	**0.214*****			**-0.053***	**-0.063***	**-0.196***
			(0.018)	(0.021)	(0.283)			(0.023)	(0.024)	(0.404)
Village FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kin Founder FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Distance time trend	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kin group characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	55,937	55,937	55,937	55,937	55,937	55,937	55,937	55,937	55,937	55,937
Adjusted R^2	0.125	0.126	0.126	0.127	0.139	0.125	0.126	0.126	0.127	0.139

Note: Robust standard errors clustered at the district level in parentheses. Average marginal response and adjusted pseudo R^2 are reported for logit model.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.2: Changing effects of centrality among female and male relatives

		Proximity: Whether belong to Near-Port Group		Proximity: Geodesic distance in 100 km	
Dependent variable: Marriage next period	(1)	(2)	(3)	(4)	(5)
Centrality among married females	0.039***	0.050***	0.047***	0.102***	0.092**
	(0.006)	(0.003)	(0.012)	(0.022)	(0.037)
Centrality among unmarried females	0.174	0.187	0.195	0.011	0.047
	(0.106)	(0.137)	(0.136)	(0.082)	(0.103)
Centrality among married males			0.005		0.082
			(0.026)		(0.051)
Centrality among unmarried males			0.024		-0.177**
			(0.033)		(0.062)
Centrality among married females×Proximity × Post		-0.006	-0.018	0.008	0.010
		(0.014)	(0.037)	(0.008)	(0.019)
Centrality among unmarried females ×Proximity × Post		-0.029	-0.046	-0.040	-0.026
		(0.042)	(0.043)	(0.039)	(0.041)
Centrality among married males×Proximity × Post			-0.021		0.024
			(0.045)		(0.032)
Centrality among unmarried males×Proximity × Post			0.120***		-0.082***
			(0.015)		(0.019)
Other interactions	×	✓	✓	✓	✓
Village FE	✓	✓	✓	✓	✓
Village time trends	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Kin Founder FE	✓	✓	✓	✓	✓
Kin group characteristics	✓	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓	✓
Observations	55,937	55,937	55,937	55,937	55,937
Adjusted R^2	0.130	0.130	0.130	0.130	0.130

Note: Robust standard errors clustered at the district level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

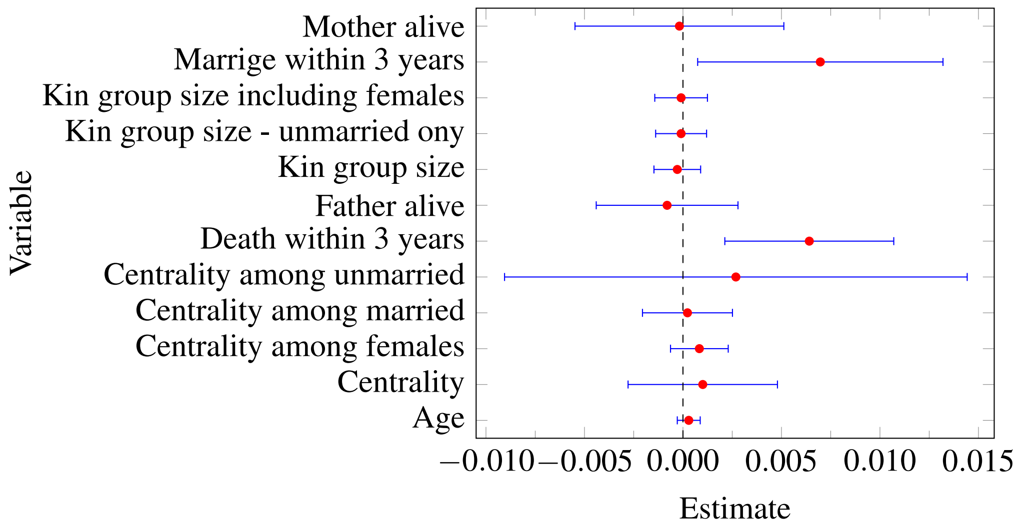
1.6 Robustness Checks

In this section, I conduct robustness checks to address the most significant concerns, which are systematic missing observations and pre-trends. Various other robustness checks are also conducted, which include altering age thresholds for kin group construction, varying the values of the altruism parameter, and excluding small kin groups. Detailed information about these additional robustness checks is provided in Appendix E.

1.6.1 Missing Observations

As stated in Appendix B.4, the most significant limitation of the data is the presence of many missing records. Systematic missing can introduce bias, given that only observations with available next records are included. To investigate whether systematic missing exists, I conduct regressions using all unmarried males between 1849 and 1885, including both those with available next registers and those without. The outcome variable is whether the next observation is available or not. Figure 1.2 presents the results of these regressions.

Figure 1.2: Testing if missing data is systematic



The obtained result supports the notion that the missing data is not systematic. It is reasonable to observe a correlation between marriage as well as death within the next 3 years and missing the next records. As long as marriage or death events are recorded in the next available record, the corresponding variables will take a value of one. For instance, if a man is recorded as unmarried in 1858 but married in 1864, with his record in 1861 missing, it indicates that he got married sometime between 1858 and 1864. Unsurprisingly, the likelihood of such an event occurring is higher than that of getting married within a mere 3-year span. This is precisely why observations lacking subsequent records should be excluded from the analysis.

1.6.2 Event Study

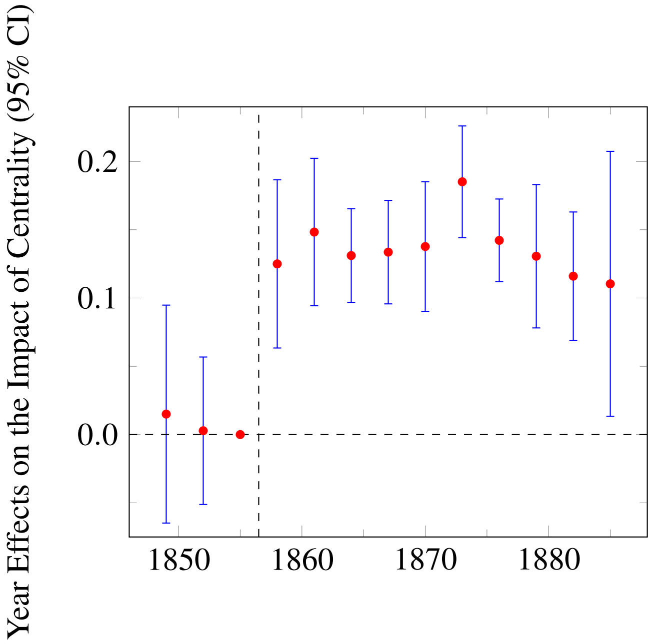
A common concern in research utilizing the DID approach is the potential presence of pre-trends. It is possible that the changing effects of centralities were already in place before the year 1858. An event study is the most common approach to address this concern. In this section, I present the findings obtained through an event study approach, based on specifications modified from equation (1.1). To capture the dynamic effects of centrality among unmarried males within the treated group, the following specification is estimated.

$$\begin{aligned} Y_{ifvt} =& \alpha_f + \alpha_v + \alpha_t + \beta_1 D_v \times t + \beta_2 C^M_{ifvt} + \beta_3 D_v \times C^M_{ifvt} + \beta_4 Post_t \times C^M_{ifvt} \\ & + \beta_5 D_v \times Post_t \times C^M_{ifvt} + \beta_6 C^U_{ifvt} + \beta_7 D_v \times C^U_{ifvt} + \beta_8 Post_t \times C^U_{ifvt} \\ & + \sum_{s=1;s\neq 3}^{13} \beta_{9,s} D_v \times C^U_{ifvt} \times \mathbf{I}\{t = 1846 + 3s\} + \beta_{10} D_v \times Post_t + \gamma' \mathbf{X}_{ifvt} + \varepsilon_{ifvt} \end{aligned} \tag{1.3}$$

where $\mathbf{I}\{t = 1846 + 3s\}$ is an indicator function that equates to 1 if $t = 1846 + 3s$. Hence, the variable $Post_t$ in the corresponding triple interaction term is replaced with year FE. The year 1855 serves as the baseline period. The coefficients $\beta_{9,s}$ then are plotted for the examination of the trends. The variable $Post_t$ is retained in the model because the focus of this paper is solely on investigating the dynamic impact of the port opening on the influence of kinship[7]. To capture the dynamic effect of the port opening on the influence of centrality among married males, the term $\beta_5 D_v \times Post_t \times C^M_{ifvt}$ will be modified accordingly.

[7] In studies where the interests lie in comparing across various combinations of dimensions, additional dynamic effects may be explored. For example, Alsan et al. (2020).

Figure 1.3: Event study focusing on the centrality among unmarried males, where D_v is binary (whether belong to Near-Port Group)



The results presented in Figures 1.3-1.6 indicate that there are no significant pre-trends in the effects of centralities, both among married and unmarried individuals, in the approximately ten years before the port opening. The opening of the port has a significantly positive impact on the effect of centrality among unmarried individuals, as evidenced by the clear decrease shown in Figures 1.3 and 1.5. Figures 1.4 and 1.6 demonstrate a negative shock of the opening on the effect of centrality among married individuals, although this effect is not as clear as the results in Figure 1.3 and 1.5.

Figure 1.4: Event study focusing on the centrality among married males, where D_v is binary (whether belong to Near-Port Group)

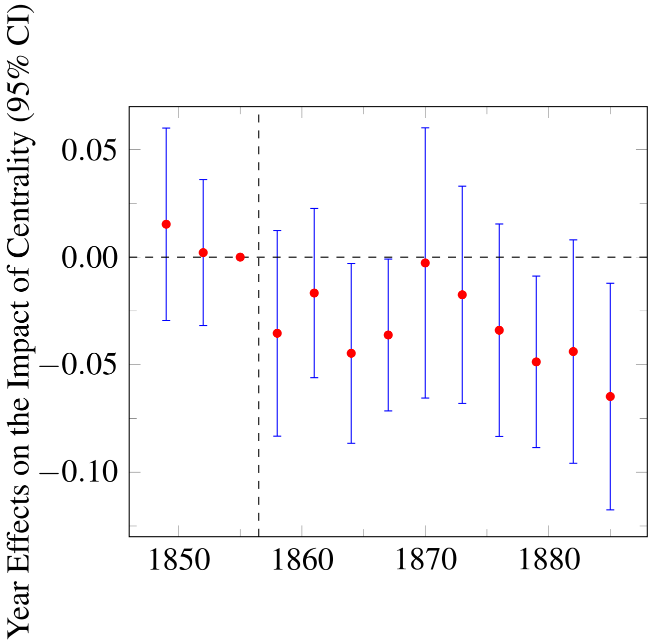


Figure 1.5: Event study focusing on the centrality among unmarried males, where D_v is continuous (geodesic distance to the port)

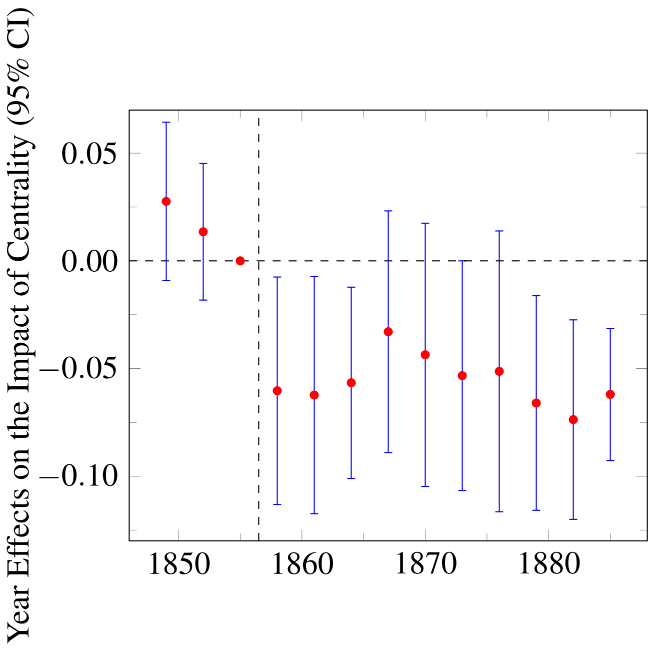
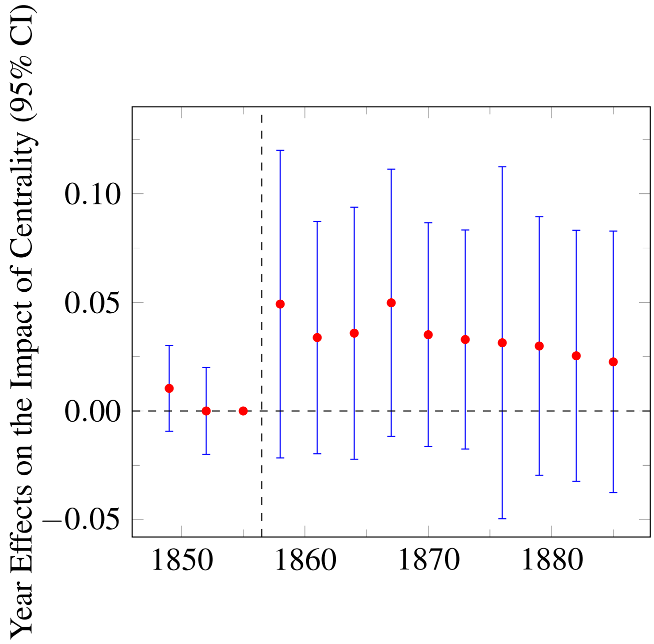


Figure 1.6: Event study focusing on the centrality among married males, where D_v is continuous (geodesic distance to the port)



1.7 Mechanism

I argue that post-opening shifts in centrality effects stem from the increased surplus and unequal resource distribution dictated by patrilineal kinship. In Section 7.1, I outline how the port opening boosts hereditary peasants' economic fortunes, even when migration is restricted, through soybean exports. In Section 7.2, I show increased within-kin-group inequality via the dispersion of couple age gaps. These findings can also be considered as motivational evidence for the model proposed in Section 8. Additionally, in Appendix F, I present evidence ruling out two alternative channels, namely through migration and changes in the sex ratio.

1.7.1 Soybean Exporting

After the opening of Niuzhuang Port, soybeans rapidly became a signature export not only for the province but also for many other parts of Northeastern China[8]. In 1865, three categories related to soybean—bean cakes, beans and peas, and bean oil—comprised more than 90% of Niuzhuang Port's exports. In contrast, rice only

[8] Kung and Li (2011) examine the impact of soybean exports on the economic prospects of immigrants in Northeast China, covering the years 1895 to 1934.

accounted for less than 5% (Shurtleff and Aoyagi, 2022).

The surge in soybean exports also led to the emergence of numerous river piers as trade hubs along three major rivers—Liao River, Hun River, and Taizi River—depicted in Figure 1.1 (Zhang, 2020). While the exact locations of these piers are difficult to pinpoint, areas closer to these rivers are generally expected to benefit more from soybean exports, all else being equal.

Based on the aforementioned historical facts, two hypotheses are posited:

Hypothesis I: Following the port opening, unmarried males living in villages more suitable for soybean cultivation will have a higher rise in the marriage probability.

Hypothesis II: This effect will be more evident in villages nearer to major rivers.

Due to limited historical crop composition data for that era and region, I make use of the Caloric Suitability Indices developed by Galor and Özak (2016a), which provide a measure of potential agricultural output in calories in history for various crops. These indices are available at a granular level, specifically for $5' \times 5'$ grid cells. I match these suitability indices to the villages in my sample and take the logarithm of these values for easier interpretation.

The findings lend support to both Hypotheses I and II. In Columns 1 and 2, the outcome variable is regressed on the logged soy suitability for samples before and after the port opening, respectively. The results indicate a dramatic increase in the impact of soy suitability post-opening. Column 3 presents the standard DID results, indicating that a 1% increase in soybean suitability results in an additional 0.596% increase in the probability of marriage, attributable to the port opening, which supports hypothesis I.

Columns 7 and 8 present the DID results, stratified by proximity to the nearest river to test hypothesis II. For the Near-River Group (distance below the median), the effect is larger, whereas in the Far-River Group (distance above the median), the effect is smaller and statistically insignificant.

Columns 4, 5, 6, 9, and 10 present analyses using rice suitability instead of

Table 1.3: Impact of crop suitability on marriage

Dep. var: Marriage next period	Before 1858	After 1858	Full sample	Before 1858	After 1858	Full sample	Dist. to river < median	Dist. to river ≥ median	Dist. to river < median	Dist. to river ≥ median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Soy suitability	-0.143	0.982**								
	(0.191)	(0.029)								
Rice suitability				-0.284	-0.166					
				(0.235)	(0.150)					
Soy suitability×Post			0.596***				0.642***	0.553		
			(0.115)				(0.132)	(0.419)		
Rice suitability×Post						0.348			0.277	0.600
						(0.252)			(0.430)	(0.311)
District FE	✓	✓	×	✓	✓	×	×	×	×	×
Village FE	×	×	✓	×	×	✓	✓	✓	✓	✓
Village time trends	×	×	✓	×	×	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kin Founder FE	×	×	✓	×	×	✓	✓	✓	✓	✓
Kin group characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	9,317	46,620	55,937	9,317	46,620	55,937	26,801	29,136	26,801	29,136
Adjusted R^2	0.112	0.096	0.127	0.102	0.095	0.127	0.121	0.133	0.121	0.127

Note: Robust standard errors clustered at the district level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

soy suitability. These columns yield no statistically significant results, except for an increased positive effect of rice suitability in Column 6. This may be due to rice being another export good, although its significance is much lower compared to soybeans.

1.7.2 Increased Within-kin-group Disparity

I present evidence showing that the port opening increased within-kin-group disparity in two aspects: the increased dispersion of the couple age gap within kin groups and the estimated income within kin groups.

The preference of male individuals in my sample for younger or older spouses is uncertain. While it might be assumed that males in patriarchal societies favour younger wives, historical evidence suggests that in 19th-century Northeast China, ordinary males often chose older wives due to their ability to contribute more effectively to family labour (Ding et al., 2004). Nevertheless, in either scenario, if inequality in males' competitiveness in the marriage market intensifies, a more pronounced age gap between couples is anticipated.

To delve into this matter, I perform a regression examining the dispersion in age difference between husbands and wives (calculated as the husband's age minus

Table 1.4: Port opening on within-kin-group disparity

	Dep. var: Within-kin-group SD of marriage age gap		Dep. var: Within-kin-group SD of estimated income	
Proximity:	Near-Port Group	Distance to port in 100 km	Near-Port Group	Distance to port in 100 km
	(1)	(2)	(3)	(4)
Proximity×Post	0.524*	-0.476***	1.575***	-1.448*
	(0.232)	(0.117)	(0.404)	(0.770)
Village FE	✓	✓	✓	✓
Year founder FE	✓	✓	✓	✓
Village FE	✓	✓	✓	✓
Observations	1,618	1,618	2,713	2,713
R^2	0.584	0.585	0.470	0.469

Note: Robust standard errors clustered at the district level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the wife's age) within individual kin groups. In this context, the unit of analysis is each time-specific kin group, and the outcome variable is the standard deviation of the couple age differences from marriages occurring in that period within the kin group.

Regarding the second aspect, the data reports estimated income in taels generated from an individual's state position, with a value equal to 0 if the individual holds no position. While the majority of the sample individuals are peasants, only 1.96% of sample individuals have a positive estimated income. This poses challenges for individual-level analysis, but it can still be utilized for analysis at the kin group level. For this analysis, I exclude kin groups with no kin member holding any state position and examine the impact of post-opening on within-kin-group estimated dispersion.

The proximity of a kin group to the port is based on the village housing the majority of that kin group's members. In most instances, members of a kin group are all based in the same village. I analyze only standard deviations calculated from observations with multiple at least two marriages for couple age gap, and with at least two kin group members for estimated income. The mean SD of the marriage age gap is 1.957, while the mean SD of estimated income is 6.564. Table 1.4 indicates increased disparities in male competitiveness in the marriage market as well as in estimated income after port opening.

1.8 The Model

In this section, I present a model that describes resource allocation within kinship networks for marriage, offering predictions that align with the empirical findings. Within the model, resources allocated for male marriage are distributed among unmarried kinsmen. Individuals[9] with greater centrality within their kinship group receive a larger share of the resources. While married clan members are assumed to always contribute to the allocation, an unmarried individual has the choice to participate or not. If he chooses to participate, he contributes his resources (endowment) to the kin group. What he receives then depends on his centrality within a subgraph of the complete patrilineal family tree, a subgraph that is formed by the participating individuals[10]. If an unmarried male decides not to participate in the allocation, he contributes nothing to the clan, receives nothing in return, and has no influence over the distribution of resources among others. However, he retains full control over his reserved resources. The likelihood of getting married, which determines an unmarried individual's private utility, depends only on the resources an unmarried individual owns in the end.

Within this framework, it is expected that individuals with greater centrality within their kinship networks are more likely to participate in the resource allocation process compared to peripheral individuals. With the increase in available resources resulting from economic development, unmarried individuals have a stronger incentive to join the fund because they stand to gain more from their married relatives. As more unmarried males participate and exert influence over the allocation, the importance of centrality among unmarried relatives, such as having a larger number of unmarried cousins, becomes more significant.

1.8.1 Model Premises

Kinship network. For simplicity, assume that there is only one kinship group in the economy (only consider male members, i.e. it is patrilineal). $N = \{1, ..., n\}$ is

[9]In this section, individuals refers to male kinsmen.

[10]Non-participation can be seen as renouncing the implicit self-insurance agreement within kin groups, as in contexts described in Ligon et al. (2000)

the set of all its kinship group members. So the size of the kinship group is $|N| = n$. M is the group of unmarried individuals among the kinship group and $|M| = m$. G is the *adjacency matrix* representing the family tree of the kinship group and g_{ij} is the matrix's element in ith row and jth column. It means that $g_{ij} = 1$ if i and j have a father-son relationship and 0 otherwise. A sequence of individuals $i_1, \ldots, i_k$ is a *path* in the kinship network if $i_j \neq i_{j+1}$ for all integer j from 1 to $k-1$, and the length of the path is $k-1$. Since G represents a family tree, it is a tree network and hence there is only one path between any two individuals. Let d_{ij} be the length of that path between i and j, and call d_{ij} the *distance* between i and j, and conventionally define $d_{ii} = 0$ for all i.

Kinship fund for supporting marriage. The model is static. Each individual is endowed with wealth $w > 0$, which can be interpreted as the average surplus a kinship group member is able to hand out for supporting another's marriage, reflecting the average level of wealth within the kinship group. The action of every unmarried individual i is to choose whether to join the fund for supporting marriage, i.e. choose $a_i \in \{0, 1\}$. Denote the set of fund participants by N_x. Among these participants, denote the set of unmarried participants by M_x. So $|M_x| = \sum_{i \in M} a_i$. Assume that a married individual is always in the marriage fund, i.e. if $i \in N \backslash M$, $i \in N_x$. Hence, $|N_x| - |M_x| = |N| - |M| = n - m$ which represents the number of married individuals in the entire kin group. Assume that both $n - m$ (the number of married) and m (the number of unmarried) are greater than 2. The model becomes trivial otherwise.

Marrige probability. After the formation of the marriage fund, wealth will be redistributed accordingly. Let w_i' denote the wealth of individual i after the allocation. If an unmarried individual i chooses not to participate in the marriage fund, his wealth will remain unchanged, i.e., $w_i' = w$. $\{w_i' : i \in M\}$ is called an *allocation*. The probability of individual i getting married is given by $1 - e^{-\lambda w_i'}$ if $w_i' \geq 0$ and 0 otherwise. λ is the parameter of the corresponding exponential distribution.

Although the model is static, $1-e^{-\lambda w'_i}$ can be interpreted as the probability of marriage within the next 3 years to better align with the data.

The parameter λ represents the general level of difficulty for a male to find a spouse. We can conceptualize this by considering individual i randomly drawing a potential spouse whose wealth follows an exponential distribution, and a successful match occurs if her wealth is smaller than that of individual i. Therefore, $\frac{1}{\lambda}$ represents the expected wealth of unmarried females in the marriage market or equivalently the wealth required for a male to marry such a woman. For further discussion on this setup regarding marriage probability, please refer to Appendix H.1.

Utility and resource allocation. Given w'_i, an unmarried individual i receives his private utility $u_i(w'_i) = (1-e^{-\lambda w'_i})v$, which is his probability of getting married multiplied by a constant payoff v. Only unmarried individuals are active and have private utility in the model. Besides private utility, an individual also has a social utility term from others' marriages. Let a given parameter $\alpha \in (0,1)$ measures the degree of altruism. One unit increase in $u_i(w'_i)$ will give $\alpha^{d_{ij}}$ units of social utility to individual j, so j is more altruistic to i if they are more closely related. There is a fund manager (FM)[11] who manages the allocation and maximizes the total utility given M_x. The FM's problem is

$$\max_{(w'_j: j\in M_x)} \sum_{i\in N_x} \sum_{j\in M_x} \alpha^{d_{ij}} u_i(w'_i) \tag{1.4}$$

$$s.t. \sum_{i\in M_x} w'_j = |N_x| w \tag{1.5}$$

$$w'_i \geq 0, \quad \forall i. \tag{1.6}$$

[11]One can think of this as the central authority of the kinship group e.g. the head or the elders of the kinship group.

and can be rewritten as

$$\max_{(w'_j : j \in M_x)} \sum_{j \in M_x} \delta_j u_j(w'_j) \tag{1.7}$$

$$s.t. \sum_{i \in M_x} w'_j = |N_x| w \tag{1.8}$$

$$w'_i \geq 0, \quad \forall i \tag{1.9}$$

where $\delta_i = \sum_{j \in N_x} \alpha^{d_{ij}}$ and call it *(decay) centrality within participants* of i. There is a unique solution to this problem due to the concavity. Given the FM's solution, define $w_i^{M_x\prime}$ as what the FM allocates to i if $i \in M_x$, and $w_i^{M_x\prime} = w$ if $i \in M$ but $i \notin M_x$.

Timing. To summarize, the timing of the game is as follows

1. Unmarried individuals simultaneously choose whether to participate in the safety N_x net or not, while all married individuals are always in N_x.
2. After N_x is formed, wealth is redistributed accordingly.
3. Unmarried individuals realize their w'_i as well as the probability of getting married. Everyone receives his payoff (private utility plus social utility). The game ends.

1.8.2 Analysis

Solution concept. Based on the model setup, an equilibrium must specify a decision (joining or leaving the kinship marriage fund) for every unmarried individual. While each individual chooses between joining or leaving, there are 2^m potential action profiles. It makes the analysis much harder, especially given one's decision affects the utilities of others.

A common way to handle the multiplicity is to impose additional assumptions for equilibrium selection. I impose a mild condition that allows an individual to switch from leaving to joining only when it is beneficial for himself and makes no one already in the marriage fund want to leave. The intuition will be provided in the following text. I then define my concept of equilibrium as follows.

Definition 1. *An* equilibrium $\xi \in \Xi$ *is defined as an action profile* $\boldsymbol{a}^{\xi} = (a_1^{\xi}, ..., a_m^{\xi})$ *which gives a set* M_x^{ξ} *and a vector* $(w_i'^{\xi} : i \in M)$ *such that*

1. $w_i'^{\xi} = w_i^{M_x^{\xi}\prime}$ *for each i,*

2. $\sum_{j \in M} \alpha^{d_{ij}} u_j(w_j'^{\xi}) \geq u_i(w) + \sum_{j \in M} \alpha^{d_{ij}} u_j(w_j^{M_x^{\xi} \setminus i\prime})$ *for each* $i \in M_x^{\xi}$*, and*

3. *There is no individual* $z \notin M_x^{\xi}$ *makes* $\sum_{j \in M} \alpha^{d_{ij}} u_j(w_j^{M_x^{\xi} \cup z\prime}) \geq u_i(w) + \sum_{j \in M} \alpha^{d_{ij}} u_j(w_j^{M_x^{\xi} \cup z\prime})$ *for every* $i \in M_x \cup z$*.*[12]

Basically, this solution concept consists of three conditions. First, given a marriage fund M_x^{ξ}, the allocation $\{w_i'^{\xi} : i \in M\}$ must solve the FM's problem so that the total utility is maximized. Second, no one in the marriage fund has an incentive to leave because his leaving gives a weakly lower utility (private plus social utility) than the status quo. Third, it is impossible to include any individuals in the marriage fund while making no one want to leave the new marriage fund. While the first and second conditions are standard, condition 3 needs more explanation. It says that one is able to deviate from leaving to joining to reach an equilibrium only when it makes no one already in the marriage fund want to leave. The intuition and further discussion are provided in Appendix H.3.

Since an individual cares about other relatives' private utilities in addition to their own, it is possible that individual i joins (leaves) the fund when it decreases (increases) their private utility but increases (decreases) their social utility. Assumption 1 below rules out this case.

Assumption 1. $\alpha \leq e^{-\lambda w}$.

Assumption 1 ensures that an individual's level of altruism isn't so high that they would act detrimentally to their own interests to benefit others. This assumption is both testable and mild. For example, the value of $1 - e^{-\lambda w}$ can be interpreted as the marriage probability of males without relatives, and it's estimated to be around 5%. Therefore, for this assumption to be valid, α only needs to be less than

[12] To clarify, w_i' is what individual i gets from an arbitrary allocation, $w_i^{M_x\prime}$ is what he gets from the allocation given by the FM with M_x, and $w_i'^{\xi}$ is what he gets in an equilibrium satisfying the conditions in definition 1.

95%.

Centrality and resource allocation. Given an equilibrium ξ, the set of individuals in the marriage fund can be expressed as $N_x^\xi := M_x^\xi \cup (N \backslash M)$, which is unmarried participants plus married males. This also gives a set of centralities within participants $\{\delta_i^\xi : i \in N\}$. δ_i^ξ can be separated into two parts, *centrality among married*, $\bar{\delta}_i := \sum_{j \in N \backslash M} \alpha^{d_{ij}}$, and *centrality among unmarried participants*, $\hat{\delta}_i^\xi := \sum_{j \in M_x^\xi \backslash i} \alpha^{d_{ij}}$. Thus, $\delta_i^\xi = \bar{\delta}_i + \hat{\delta}_i^\xi$. Notice that $\bar{\delta}_i$ does not depend on ξ, since married individuals are always in N_x. Because $\hat{\delta}_i^\xi$ is unobservable, any prediction based only on $\hat{\delta}_i^\xi$ will be untestable. What can be observed with the data is $\bar{\delta}_i$, and $\hat{\delta}_i := \sum_{j \in M \backslash i} \alpha^{d_{ij}}$, which denotes the observed centrality among unmarried. Denote $\hat{\delta}_i^\xi := \tau_i^\xi \hat{\delta}_i$ where τ_i^ξ is set to be equal to $\frac{\hat{\delta}_i^\xi}{\hat{\delta}_i}$ and hence on $[0,1]$ interval. The proposition below characterizes the equilibria.

Proposition 1. *Under Assumption* 1*, set* Ξ *is non-empty. So there exists some* $\xi \in \Xi$ *such that for any* $i \in M_x^\xi$*,*

$$w_i'^\xi = \frac{|N_x^\xi|}{|M_x^\xi|} w - \frac{1}{\lambda} \frac{1}{|M_x^\xi|} \sum_{j \in M_x^\xi} \ln \frac{\bar{\delta}_j + \tau_j^\xi \hat{\delta}_j}{\bar{\delta}_i + \tau_i^\xi \hat{\delta}_i}, \tag{1.10}$$

while $w_i'^\xi = w$ *when* $i \notin M_x^\xi$.

Several remarks can be given by the proposition. First of all, the benchmark of resources an unmarried gets from his clan is $\frac{|N_x^\xi|}{|M_x^\xi|} w$, which is the total resources to be allocated $|N_x^\xi| w$, divided by the number of unmarried who request resources, $|M_x^\xi|$. Obviously, this level will be higher if the average wealth w is higher. Second, centrality plays an important role in resource allocation, for those who join the marriage fund. The term $\sum_{j \in M_x^\xi} \ln \frac{\bar{\delta}_j + \tau_j^\xi \hat{\delta}_j}{\bar{\delta}_i + \tau_i^\xi \hat{\delta}_i}$ captures how central individual i is in the marriage fund, relative to other unmarried individuals who also joins. If every unmarried has the same centrality (which for example could be the case where G represents a nucleus family, with one father and several brothers), the resources will be distributed evenly among those in M_x^ξ. If j is more central than i in N_x, $\ln \frac{\bar{\delta}_j + \tau_j^\xi \hat{\delta}_j}{\bar{\delta}_i + \tau_i^\xi \hat{\delta}_i}$

is positive, and hence it can be interpreted as j 'takes away' $\frac{1}{\lambda}\frac{1}{|M_x^\xi|}\ln\frac{\bar{\delta}_j+\tau_j^\xi\hat{\delta}_j}{\bar{\delta}_i+\tau_i^\xi\hat{\delta}_i}$ units from i, which incurs a loss to i's 'fair share', $\frac{|N_x^\xi|}{|M_x^\xi|}w$. Suppose i is the most peripheral individual in N_x, then $\ln\frac{\bar{\delta}_j+\tau_j^\xi\hat{\delta}_j}{\bar{\delta}_i+\tau_i^\xi\hat{\delta}_i}$ is negative for all j and hence he has the least among M_x^ξ. However, he may still want to stay in the marriage fund, because either there are enough married individuals in the kinship group to support unmarried ones (so $\frac{|N_x^\xi|}{|M_x^\xi|}w$ is large enough because $|N_x^\xi|$ is large relative to $|M_x^\xi|$), or w is large enough (so the surplus married individuals contribute for unmarried to share is high enough).

Port opening. I adopt the approach of comparative statics and compare the results before and after the port opening. The port opening is indicated by an increase in wealth levels, both for the wealth of an unsupported average unmarried male and the expected wealth of an unmarried female. Suppose there are two states, denoted as $s \in l,h$, representing before and after the opening, respectively. Let w and λ before the opening be denoted as w_l and λ_l, respectively. After the opening, they become $w_h > w_l$ and $\lambda_h < \lambda_l$. We assume that $w_l\lambda_l \leq w_h\lambda_h$, i.e. a male with the lowest level of wealth (w_s) has been better off after the opening.

Comparative statics. The set of equilibria Ξ changes when w changes. Hence, denote Ξ_s as the set of equilibria in state $s \in \{l,h\}$. The aim is to compare some equilibria in Ξ_l with some in Ξ_h and see how things change. The question is, which pairs of equilibria should we use for such a comparison? Suppose that a kinship group reaches an equilibrium $\xi \in \Xi_l$, and then the opening of Niuzhuang suddenly raises w from w_l to w_h. Staying in the marriage fund becomes more beneficial if there is no newcomer joining because equation (1.10) shows that $w_i'^\xi$, as well as $w_i'^\xi - w$, is increasing in w, ceteris paribus. If no one outside of M_x^ξ wants to join, ξ is still an equilibrium. If some outsiders find joining beneficial without violating condition 3 in definition 1, let them join and then a new marriage fund is formed. This may attract more outsiders to join if they find themselves becoming quite central in the new marriage fund, but eventually, a new equilibrium will emerge and all individuals in M_x^{ξ} will still be in the new fund. This is formalized by lemma 1 below.

Lemma 1. *Under Assumption* 1*, if* $\xi_1 \in \Xi_l$*, there are* k $(k \geq 1)$ *equilibria in* Ξ_h *denoted by* $(\xi_{1,j} : j \in \{1,...,k\})$ *such that* $M_x^{\xi_1} \subseteq M_x^{\xi_{1,j}}$ *for every* j*.*

I compare each equilibrium ξ_i in Ξ_l with every its corresponding equilibrium, $\xi_{i,j}$ in Ξ_h. I show that for each such comparison and for each individual, being central within unmarried individuals becomes more important for one to get higher wealth. This is stated by proposition 2.

Proposition 2. *For every such* $\xi_{1,j}$*,* $\frac{\partial w_i'^{\xi_1}}{\partial \bar{\delta}_i} \geq \frac{\partial w_i'^{\xi_{1,j}}}{\partial \bar{\delta}_i}$ *for every* i*.*

It's valuable to delve deeper into how the opening influences the effects of centralities on marriage probability. This is detailed in Proposition 3.

Proposition 3. *For such a* $\xi_{1,j}$*,* $\sum_{i \in M} \frac{\partial u_i(w_i'^{\xi_1})}{\partial \hat{\delta}_i} \leq \sum_{i \in M} \frac{\partial u_i(w_i'^{\xi_{1,j}})}{\partial \hat{\delta}_i}$*, if*

1. $\tau_i^{\xi_{1,j}} \hat{\delta}_i \leq \bar{\delta}_i$ *for every* i*, and*
2. $\frac{\lambda_h}{\lambda_l} \leq \frac{(1+\frac{S}{|M_x^{\xi_1}|})w_l}{(1+\frac{S}{|N_x^{\xi_1}|})w_h}$ *where* $S = |M_x^{\xi_{1,j}}| - |M_x^{\xi_1}|$*.*

The two conditions in Proposition 3 aim to select the equilibria to be focused on that are reasonable in accordance with the data. In the condition $\tau_i^{\xi_{1,j}} \hat{\delta}_i \leq \bar{\delta}_i$, $\tau_i^{\xi_{1,j}}$ is determined endogenously. However, this condition is always less restrictive than $\hat{\delta}_i \leq \bar{\delta}_i$ since $\tau_i^{\xi_{1,j}}$ is set to be less than 1. Furthermore, $\hat{\delta}_i \leq \bar{\delta}_i$ itself is quite reasonable with the data. Evidence supporting this claim can be found in Appendix H.4. This ensures that for every individual in the post-wealth increase equilibrium, their centrality among unmarried participants does not exceed their centrality among married relatives. The underlying intuition of the second condition in Proposition 2 suggests that while the port opening benefits those without relatives or who do not participate in the fund, these benefits are likely to be quite limited. If the number of participants increases post-opening, the marriage market reacts in a way that disadvantages those who do not participate. This is because, within a given kinship group, a non-participant's wealth ranking would be comparatively lower if the number of participants remained unchanged.

Recall that u_i is defined as the probability of marriage multiplied by a constant. Thus, the expression $\sum_{i\in M} \frac{\partial u_i(w_i'^{\xi})}{\partial \hat{\delta}_i}$ corresponds to the average response of marriage probability with respect to centrality in that equilibrium. Therefore, Proposition 3 predicts the main empirical finding of this study as $\sum_{i\in M} \frac{\partial u_i(w_i')}{\partial \hat{\delta}_i} - \sum_{i\in M} \frac{\partial u_i(w_i'^{\xi_{1,j}})}{\partial \hat{\delta}_i}$ corresponds to the ACRR. Notably, drawing similar (or contrasting) propositions about centrality among married relatives is hard. One can show that centrality among married grows in importance for those who are already participants in the pre-wealth increase equilibrium. Yet, as more individuals join the kinship fund, the expected response might shift upward.

1.9 Model Fit

1.9.1 Predicted Marriage Probability

The model's results depend on multiple assumptions. While some assumptions can be tested, others cannot[13]. Also, the complexity of the kinship network complicates the derivation of an analytical solution for within-kin-group disparity. This section evaluates the model fit by using the model to estimate marriage probabilities for each male individual, comparing them with the observed marriage rates, and assessing whether they produce the same empirical results.

Using the model, one can estimate the probability of a male individual getting married within the next three years based on the available data. Premultiplying both sides of equation (1.10) by λ yields

$$\lambda w_i'^{\xi} = \frac{|N_x^{\xi}|}{|M_x^{\xi}|}\lambda w - \frac{1}{|M_x^{\xi}|}\sum_{j\in M_x^{\xi}} \ln \frac{\bar{\delta}_j + \tau_j^{\xi}\hat{\delta}_j}{\bar{\delta}_i + \tau_i^{\xi}\hat{\delta}_i} \tag{1.11}$$

$$= \frac{|N_x^{\xi}|}{|M_x^{\xi}|}\phi - \frac{1}{|M_x^{\xi}|}\sum_{j\in M_x^{\xi}} \ln \frac{\bar{\delta}_j + \tau_j^{\xi}\hat{\delta}_j}{\bar{\delta}_i + \tau_i^{\xi}\hat{\delta}_i} \tag{1.12}$$

[13]For example, it's impossible to determine if the second condition in Proposition 3 is met because the set of participants remains unobservable.

where $\phi = \lambda w$. The first step involves estimating ϕ. This can be accomplished by using all unmarried males who are the only unmarried males in their kin group. Suppose the number of married individuals in the kin group of such individual i is k_i (abstracting from the period for now). The probability of marriage predicted by the model for individual i is $1 - e^{-(k_i+1)\phi}$. Let Y_i denote whether individual i actually gets married or not. We can use Maximum Likelihood Estimation (MLE) and select ϕ to maximize the likelihood function:

$$\mathscr{L}(\phi) = \prod_{i \in N_O} [1 - e^{-(k_i+1)\phi}]^{Y_i} [e^{-(k_i+1)\phi}]^{1-Y_i} \tag{1.13}$$

as well as the corresponding log-likelihood function (LLF) to obtain the estimate for ϕ.

I segment the sample based on two dimensions: district and period. The estimation of ϕ is done individually for each of these cells, making the value of ϕ specific to a district-period cell. Consequently, the measure of exposure to the port opening cannot be more specific than this level. This approach differs from my primary results where I use village-level distance as a measure of exposure. The rationale behind this deviation is the small sample size of unmarried males within each village-period cell, which makes MLE become problematic in cases where either no one marries or everyone does within the next period for a particular cell.

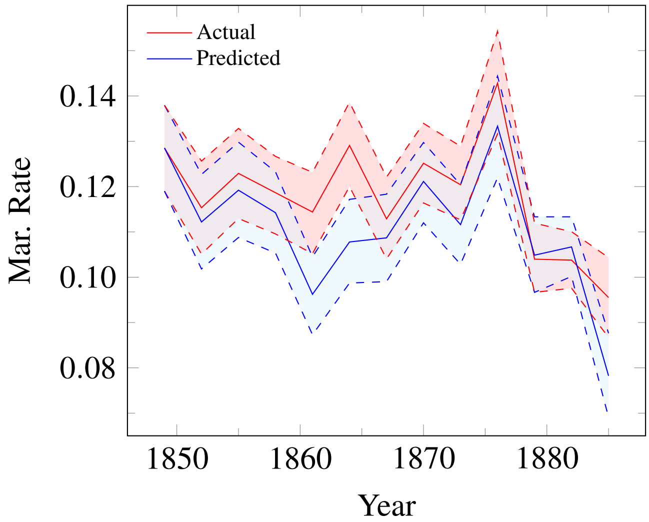
Given those values of ϕ, I then simulate the equilibrium for each kin group. The algorithm is as follows.

1. Start with an initial set of unmarried fund participants denoted as $M_x^{(1)}$, and let $M_x^{(1)} = M$, including all unmarried individuals as participants. Compute the value of $\lambda w_i'$ for each individual using equation (1.12).

2. Create a new participant group $M_x^{(2)}$ by including those individuals with $\lambda w_i' \geq \lambda w$. The remaining individuals exit the fund.

3. Repeat the process until convergence:

Figure 1.7: Comparing predicted mar. rate with actual mar. rate for participants



(a) Let the new participant group be denoted as $\bar{M}_x^{(1)}$. Select any individual $j \notin \bar{M}_x^{(1)} \cup P$, where $P = \emptyset$ initially, and add j to $\bar{M}_x^{(1)}$. Compute all λw_i based on the updated set.

(b) If all $\lambda w_i \geq \lambda w$, denote the new set as $\bar{M}_x^{(2)}$ and repeat from Step 3.a, with $\bar{M}_x^{(1)}$ replaced by $\bar{M}_x^{(2)}$.

(c) If $\lambda w_i < \lambda w$ for some i, repeat from Step 3.a, but add j to the set P.

4. The procedure continues until $\bar{M}_x^{(s)} = \bar{M}_x^{(s+1)}$. At this point, $\bar{M}_x^{(s)}$ constitutes an equilibrium and denote it by M_x^{ξ}.

Individuals in M_x^{ξ} are referred to as *participants*. Then, the estimated $\lambda w_i'^{\xi}$, and consequently, $1 - e^{-\lambda w_i'^{\xi}}$ can be obtained for each participant i. This quantity is referred to as the *predicted marriage probability*. In Figure 1.7, I compare the average predicted marriage probability with the average of the observed outcome, whether individuals get married within the next 3 years, categorized by districts and by periods, for participants.

1.9.2 Results Using Predicted Marriage Probability

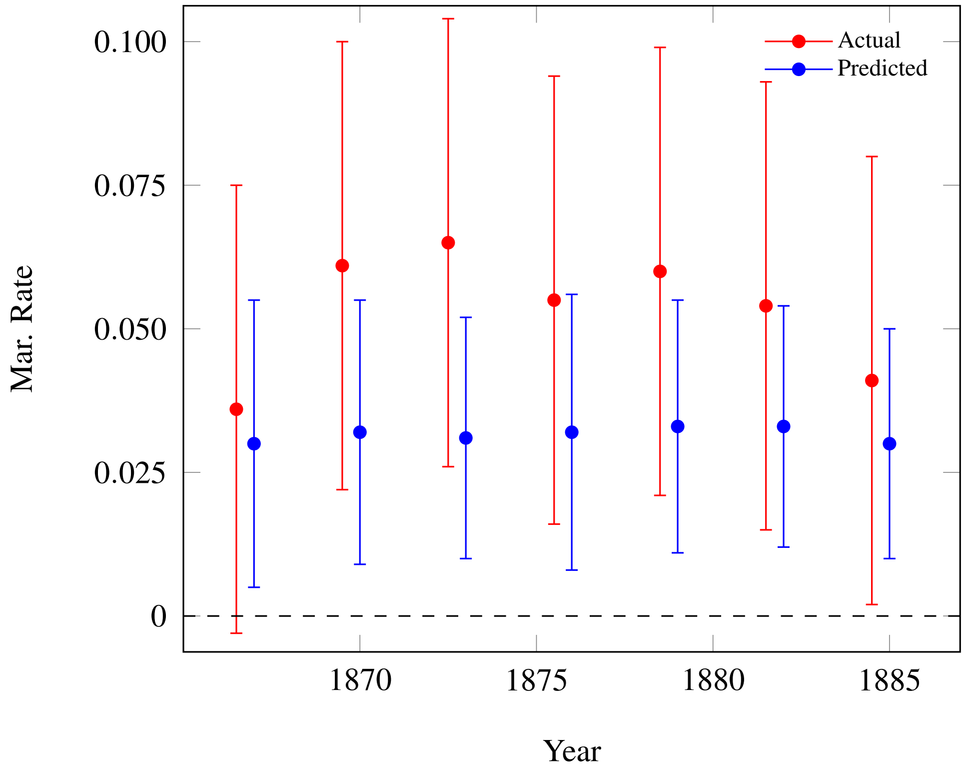
To assess whether the marriage probability predicted by the model yields similar results, I use the predicted marriage probability as the outcome variable in the regression of equation (1.1), comparing these outcomes with those obtained using the observed marriage outcomes. I vary the time span from ending at 1867 to ending at 1885 and plot the corresponding estimates for β_{10}.

Notably, utilizing village-level distances to the port as a measure of exposure to the port opening becomes irrelevant since no village-level information was employed in estimating $1-e^{-\lambda w_i}$. Consequently, I designate those in the same district as the port (Niuzhuang District) as the treatment group. For individuals who are not participants, I utilize $1-e^{-\phi}$ as their predicted marriage probability, where ϕ is specific to their respective district-period cell.

Figure 1.8 illustrates that the marriage probability generated by the model also indicates an increased effect of unmarried male relatives, albeit with a magnitude not as pronounced as when utilizing the observed marriage outcome, and it exhibits less variation across different time spans for the regression. This is understandable, considering that actual marriage outcomes are influenced by numerous additional factors, such as age, which are not accounted for in the model simulation. While I cannot discount the possibility of additional channels contributing to the main results, the simulation indicates that the proposed channel could indeed be one of them.

Furthermore, I show that the correlation between the predicted marriage probability and the observed future marital status remains significant even after controlling for all the variables used in generating that predicted probability. I also find that larger within-kin-group disparities, measured by the dispersion of resources $w_i'^{\xi}$, have emerged after the port opening. For more details, please refer to Appendix G.

Figure 1.8: Plotting estimated increased effect of unmarried male relatives by different year spans



1.10 Conclusion

While the persistence of traditional institutions during the early stages of economic development has been well-observed, there is a scarcity of evidence that examines the causal relationship between economic take-offs and the strength and effects of traditional institutions at the micro-level. In this paper, I utilize the forced opening of a treaty port in pre-modern China as a treatment that generates positive shocks to the agricultural economy and examine how this economic shock influences the role of the patrilineal kinship system in facilitating marriage among its male members over a span of approximately forty years (1849-1885).

My findings indicate that this patrilineal kinship institution become more pronounced, given men who hold central positions within their patrilineal networks, specifically among their unmarried male relatives, tend to marry earlier compared to those who are more peripheral in this context, following the port opening. This effect is not observed when centrality is measured based on the individual's female relatives.

I conduct several robustness checks and discuss several possible channels through which the port opening could impact marriage patterns. I also present evidence that the main effects are accompanied by increased within-kin-group disparity. Additionally, I develop a theoretical model to describe the mechanism by which the larger economic surplus resulting from the opening attracts more unmarried males to become more attached to their kinship networks and provide mutual support. Moreover, evidence supporting the fitness of this model is provided.

This research carries several important implications. Firstly, it reveals the persistence of traditional institutions during the initial stages of development by investigating specific social structures and providing the underlying mechanisms. This highlights the need for a nuanced understanding of the relationship between economic progress and cultural institutions and provides a starting point for future research. This also remains relevant in contemporary times, as institutions, norms, and cultural aspects related to redistribution often evolve at a slower pace. Many of them remain ingrained nowadays and can serve as a channel through which inequality is exacerbated.

Secondly, besides the commonly understood inequality between men and women, this study offers another dimension to the research on patriarchy, by emphasizing the inequality among men. By only focusing on gender inequality, it is hard to discern whether it is driven by patriarchal values or by power structures in resource allocation, given both amplifying the disparity. On the other hand, my research underscores the influence of the patriarchal power structure. It also demonstrates that challenging patriarchal institutions not only benefits women but also has advantages for men, or at least for those disadvantaged.

Lastly, this research underscores the importance of taking into account local cultures, norms, and social structures when devising and executing development programs, such as cash transfer programs. Neglecting these factors can result in suboptimal outcomes and exacerbate existing inequalities.

Chapter 2

Naïve Learning as a Coordination Device in Social Networks

2.1 Introduction

In many real-life scenarios, a common coordination problem arises when a group of individuals needs to make simultaneous adoption decisions for a technology or product that exhibits positive externality. This coordination problem stems from the fact that adoption may only be beneficial when a sufficient number of individuals adopt. Game theory literature has explored various approaches to address this coordination problem, including considering incomplete information (Carlsson and Van Damme, 1993; Morris and Shin, 2003, etc.), allowing individuals to signal their willingness to adopt (Farrell, 1987; Farrell and Saloner, 1988; Cooper et al., 1992, etc.), and using focal points as coordination devices (Schelling, 1960; Parravano and Poulsen, 2015, etc.). However, one aspect that has received limited attention in the literature is the influence of communication among individuals regarding the value of adoption before making their adoption decisions. In many cases, individuals engage in discussions or communicate with each other to assess the usefulness of a technology or product through their social networks. This communication process subsequently influences their adoption decisions. An illustrative example is the study by Cai et al. (2015), where a field experiment was conducted in rural China, offering weather insurance to local farmers. While the adoption of weather insurance has inherent value for the farmers, their decisions are also influenced by the adoption choices of others.[1] During each session, participants were introduced to the concept of insurance and then asked to simultaneously decide whether to opt for it or not. Prior to certain sessions, information about weather insurance had already been shared and discussed among local residents through their social networks.[2] In cases like this, policymakers are interested in examining whether coordination arises as individuals communicate their beliefs, and how the outcome is influenced by the structure of the information diffusion network.

This paper proposes a model of a coordination game with social learning in networks to characterize the above features. In the model, all individuals simultaneously decide whether to adopt a product or not while the adoption is strategically complementary. Before they make their adoption decisions, there is a process for individuals to form their beliefs over the underlying value of the product to be adopted. This belief-forming process is modeled as naïve learning in the framework of DeGroot (1974), where people weight the average of the beliefs they observe from others to form their own. The matrix consisting of these weights is referred to as the *listening matrix*, which represents the structure of the *conversation network* for such communication. With mild conditions, a consensus will be reached given infinite periods for communication, i.e., all individuals will have the same belief regarding the underlying value of the product.

The analysis reveals that, under mild assumptions, as individuals' beliefs converge to a consensus, a unique equilibrium emerges. In this equilibrium, each individual employs a cutoff strategy, meaning that an individual adopts the product if their belief surpasses their cutoff threshold at that time, and refrains from adopting otherwise. As beliefs converge, all individuals eventually use the same cutoff over infinite time. Consequently, given the consensus, the resulting outcome is either universal adoption or non-adoption. If the consensus surpasses a threshold determined by the underlying value, coordination is achieved, in the sense that given the

[1] The authors argue that this influence can be attributed to scale effects, the desire to imitate, or the presence of informal risk-sharing arrangements.

[2] They highlight the pivotal role of information diffusion through social networks in shaping the adoption of the insurance.

consensus universal adoption is attained when it is efficient. This outcome mirrors the outcome of a global game, where a unique equilibrium emerges when the difference between individuals' signals diminishes. The existence of this equilibrium shows that naïve learning can serve as a coordination mechanism, playing a similar role to that of a global game framework by providing an information structure that ensures coordination.

In addition, I explore the impact of conversation network structures on coordination. Specifically, I examine the effect of (eigenvector) centrality and its implications. I show that when the underlying value of the product is very low, a higher degree of inequality in centrality within the conversation network increases the likelihood of full adoption but results in lower social welfare. Conversely, when the underlying value of the product is very high, greater inequality in centrality reduces the probability of full adoption as well as overall social welfare. Intriguingly, there exists an intermediate range of the underlying value wherein a higher level of centrality inequality in the conversation network enhances both the probability of full adoption and social welfare. The underlying rationale is that when adoption holds substantial value, individual incentives are strong enough to facilitate coordination within an egalitarian conversation network. Conversely, if the adoption value is not significant, yet full adoption remains an efficient outcome, the presence of influential opinion leaders who possess greater influence over the consensus becomes instrumental in achieving a high consensus leading to a desirable outcome that would otherwise be less likely to occur.

I also extend the model to encompass more general forms of conversation networks. Specifically, I relax the assumptions made in the baseline model that information from one individual can directly or indirectly reach any other individual. By making these relaxations and a few additional assumptions, I demonstrate that a unique cutoff equilibrium still emerges as time progresses towards infinity, which is the same as in the baseline model. Additionally, the adoption continues to be influenced by the structures of the conversation network, in line with the underlying mechanisms observed in the baseline model. This extension renders the model more empirically relevant, as it no longer restricts the outcome to extreme configurations where either everyone adopts or no one adopts. Moreover, it generates more testable predictions.

The remainder of the paper is organized as follows. Section 2 provides an overview of the relevant literature. Section 3 introduces the baseline model. Section 4 presents the extension of the model. Section 5 concludes and offers a discussion. All proofs can be found in the appendix.

2.2 Related Literature

This section is situated within the context of existing literature on social learning, social networks, and coordination games. It offers a concise overview of the relevant literature and highlights key references that are particularly relevant to the research at hand.

The motivation of this paper lies in investigating technology product adoption within a social network for development programs, drawing upon previous studies in this field. Bandiera and Rasul (2006) provide evidence of the influence of farmers' network choices on their decisions to adopt a new crop. Adoption by neighbors in their study exhibits positive externalities with diminishing marginal returns, indicating that farmers are more likely to adopt when their neighbors do, but less likely when many others adopt. Conley and Udry (2010) use unique data on communication patterns among farmers to highlight dynamic learning from successful neighbors. Beaman et al. (2021) apply contagion diffusion models to social network data from Malawi, demonstrating that network-based targeting outperforms traditional approaches in promoting the adoption of productive agricultural technology. Additionally, there is a rich literature on selecting optimal seeds for adoption within social networks (Akbarpour et al., 2020; Jackson and Storms, 2017). However, these studies often overlook individuals' ex-ante belief formation regarding adoption payoffs, despite some consideration of learning from neighbors' experiences. In addition to the aforementioned literature, there are other studies that explore the impact of seeding new information within social networks on adoption decisions. Notable examples include the works of Banerjee (1992), Genius et al. (2014), Banerjee et al. (2019) and BenYishay and Mobarak (2019). These studies primarily focus on strategies to disseminate information effectively to maximize awareness and adoption of a particular technology or intervention. Their emphasis lies in understanding how to make information reach as many individuals as possible rather than examining the formation and influence of consensus on the perceived usefulness of the technology within the network structure. My research examines how adoption is influenced by the pattern of communication on beliefs about the usefulness of the technology within the social network framework prior to adoption.

To address this objective, the research focuses on belief formation and draws upon relevant literature on social learning theory in the context of social networks. In this paper, the learning process is assumed to follow the approach described byDeGroot (1974) as it aligns better with the context of promoting technology adoption in small, rural communities. This assumption is supported by the research conducted by Chandrasekhar et al. (2018), who conducted lab experiments in Indian villages and Mexican universities. Their findings revealed that approximately 10% of subjects in Indian villages exhibited Bayesian learning, while the percentage increased to 50% in universities. Besides, there is other evidence supporting that individuals do not process information as perfect Bayesians, although it does not directly compare Bayesian learning to naïve learning. Such experimental studies include Massey and Wu (2005), Kübler and Weizsäcker (2004), Çelen and Kariv (2004), etc.

There are many studies discussing how social learning affects agents' behavior, e.g., whether agents will take the same action based on the posterior of the value of such action they form in social learning (for example, Bikhchandani et al., 1992; Banerjee, 1992; Smith and Sørensen, 2000, etc.), and some of them assume that learning takes place in social networks. They do not, however, assume that agents' behavior has externalities such that one's payoff depends on others' actions, and hence coordination problems arise. While most of those papers show that agents' actions will converge as well as their beliefs, such results can fail to hold in the existence of a coordination failure. For example, the class model of Bikhchandani et al. (1992) assumes that agents receive a binary signal indicating that the value of adoption is either high or low and sequentially choose to adopt or not to adopt. The payoff is positive if and only if the true value is high and the agent adopts, and such a positive payoff is the same for each agent. So an agent does not worry that others will not adopt; it is beneficial for him or her to adopt as long as the value is high. However, things are different if the payoff of an agent also depends on other agents' actions. An agent may not adopt not only because he or she believes that the value is not high enough but because no one will adopt with him or her. Then, even if the value is very high, the adoption rate can be zero due to coordination failure.

The literature exploring the relationship between social learning and coordination in networks is relatively scarce in my knowledge, mainly due to the complexity of analysis involved. Among the existing works, the research by de Martí and Milán (2019) on global games of regime change is particularly relevant to my own study. In their model, individuals decide whether to initiate an attack on the current regime based on a threshold criterion. They communicate information about their threshold with their neighboring nodes in a social network. The authors identify a unique Bayesian Nash equilibrium characterized by cutoff strategies. Similar to my research, they find that the probability of regime change increases with inequality in the degree distribution when the threshold exceeds the mean, and vice versa. However, it is important to note that their model assumes one-time signal exchange exclusively among direct neighbors. While this assumption is reasonable in the context of regime change in a large society, it may be less applicable to the scenarios motivating my research. My study focuses on investigating how network structures influence the evolution of beliefs over time and the impact of inequality on individuals' ability to influence others' beliefs in the equilibrium. Therefore, my model addresses different aspects, emphasizing the dynamics of belief formation and the role of inequality in social influence on aggregated information, rather than focusing on the distribution of degrees.

2.3 Baseline Model

2.3.1 Premises of the model

I model a network with a finite number of individuals. The time periods are infinite, represented by $0,1,2,...,$ and there are n individuals who are potential adopters of a product. The individuals are indexed as $1,2,...,n$, and I denote the set of all individuals as $N := \{1,2,...,n\}$. Individual i refers to the ith individual in the set N, and it is assumed that $n \geq 2$.

The end of the game A time t^s is selected by nature at the beginning without being realized by any individual. Suppose it is drawn from a discrete probability distribution with $Pr(t^s = T) > 0$ for each integer $T > 0$, which is known by individuals. An individual at time t_s chooses whether to adopt the product. In the example illustrating the motivation, t_s is the moment when the weather insurance is sold to the farmers. [3] Let $a_i^t \in 0,1$ represent the adoption choice of individual i at time period t. Each individual must make an adoption decision for every period, but the decision can only be implemented when $t = t^s$. When $t = t^s$, the individual receives their payoff, which is defined as follows.

Payoff Payoffs from adopting the technology depend on the underlying fundamental state θ, continuously distributed over unbounded support $\mathbb{R}$. θ represents the underlying (intrinsic) value of adopting the product, which is not influenced by the adoption decisions of other individuals. Denote by $\mathbf{a}^t := (a_1^t,...,a_n^t)$ the action profile of all individuals at time t and $\mathbf{a}_{-i}^t := (a_1^t,...,a_{i-1}^t,a_{i+1}^t,...,a_n^t)$ denote the action profile of all individuals except i at time t. Each i obtains the following ex-post payoff given $\mathbf{a}^t$, if $t = t^s$.

$$u_i(a_i^t,\mathbf{a}_{-i}^t|\theta) = a_i^t(k\theta + \phi \sum_{j \in N \setminus i} a_j^t) \tag{2.1}$$

[3] Specifically, in this example, the farmers are unaware of when the second round sessions will occur, making t^s unknown to them.

where $\phi > 0$ and $k > 0$. ϕ is the network effect capturing the positive externality that i's neighbors' adoption imposes on i. The linear term $k\theta$ captures the payoff of adopting that is not related to the network effect, while k captures how the underlying value θ translates into an individual's payoff. In the example of weather insurance, θ can be understood as the general valuation of insurance for farmers across the entire society on average, while k incorporates other payoff-related factors such as the occurrence of extreme weather conditions or other relevant conditions specific to the community/village. Although individuals receive payoffs only at t^s, they do not know what t^s will be; hence, they must make a decision for each possible state regarding every time period t. Thus, the model is in nature static.

Given this payoff, it can be observed that as long as $\theta \geq 0$, it is profitable for everyone to adopt without any adopter in the economy. On the contrary, if $\theta \leq \frac{-\phi(n-1)}{k}$, it is unprofitable for any individual to adopt it even if everyone else does. θ can possibly take values higher than 0 or lower than $\frac{-\phi(n-1)}{k}$ since it is continuously distributed over the real number field and therefore can be arbitrarily large or small. This corresponds to the dominance regions assumption in global game literature, which supposes that the underlying state can be arbitrarily high or low so that every player makes the same decision regardless of others' decisions. It is trivial to formally make such an assumption in this paper since the linear payoff immediately implies that.

Information structure At time period 0, θ is observed with noise by all individuals, and individual i receives the signal which is named as *initial belief*: $b_i^0 = \theta + \varepsilon_i^0$, where $\varepsilon_i^0 \sim N(0, \sigma)$ and σ is finite. This generation process is common knowledge. $\forall i, j$, $\varepsilon_i^0 \perp \varepsilon_j^0$. Call b_i^t the *belief* of i at t, if $t \geq 1$. At $t \geq 1$, individual i assigns weight w_{ij} to every j's belief b_j^{t-1} to form his or her belief b_i^t. Therefore, all individuals' beliefs are updated in the *conversation network* represented by an $n \times n$ matrix W in which the $(i, j)^{th}$ element is denoted by w_{ij} and $w_{ij} \geq 0$ for all i, j. Call the matrix W *listening matrix*. Conventionally, assume that

$\sum_{j=1}^{n} w_{ij} = 1$. The updating rule of i's belief is[4]

$$b_i^t = \sum_{j=1}^{n} w_{ij} b_j^{t-1} = \theta + \sum_{j=1}^{n} w_{ij} \varepsilon_j^{t-1} = \theta + \varepsilon_i^t \tag{2.2}$$

where $\varepsilon_i^t = \sum_{j=1}^{n} w_{ij} \varepsilon_j^{t-1}$ for all i. It is easy to see that $b_i^t \sim N(\theta, \sigma_i^t)$ where $\sigma_i^t := Var(\varepsilon_i^t)$. σ_i^t is the variance of b_i^t. $\mathbf{b}^t = (b_1^t, ..., b_n^t)$ is the vector of beliefs at t of all individuals and $\sigma^t = (\sigma_1^t, ..., \sigma_n^t)$ is the vector of variances of these beliefs.

The spread of information is determined by the listening matrix W. If $w_{ij} > 0$, it indicates that the beliefs of individual j directly influence individual i. The matrix W is considered common knowledge, meaning that individual i is aware of all the w_{jz} for every j and z. It also implies that all variances and covariances relating to beliefs are also known by all individuals. The model aims to describe communication dynamics within small groups, such as a neighborhood in a small village or extended families living together in a rural area. In these settings, individuals regularly observe the interactions of others on a daily basis.

W can also be referred to as the *adjacency matrix* in the field of social network analysis. However, it differs from a standard adjacency matrix where each element takes binary values. Therefore, it is important to provide clear definitions for certain terms in order to avoid confusion.

Definition 1. A *path* is a sequence $w_{i_1 i_2}, ..., w_{i_{k-1} i_k}$ such that $w_{i_{l-1} i_l} > 0$ for every integer $1 < l \leq k$.

Definition 2. A listening matrix W is *strongly connected* if there exists a path from any individual to any other individual.

Definition 3. A listening matrix W is *strongly aperiodic* if $w_{ii} > 0$ for every i.[5]

Then the following assumption is imposed to control the pattern of information diffusion in the model.

[4]The evolution of beliefs can also be motivated by the setup by DeMarzo et al. (2003).

[5]Note that the standard definition for a network to be aperiodic is based on the requirement that the greatest common divisor of the lengths of all its cycles is one. In this study, I utilize the term *strong aperiodicity* which is a sufficient condition for aperiodicity. The reason for using strong aperiodicity is its ease of interpretation and understanding.

Assumption 1. *The listening matrix W is strongly connected and strongly aperiodic. $WW \neq W$.*

By assuming that the network is strongly connected, the information any individual gets will eventually get to affect anyone else. Moreover, by assuming strong aperiodicity, one must be the neighbor of himself or herself, which implies that one's new belief will be based on his or her previous beliefs. These assumptions guarantee that the listening matrix W of the corresponding Markov chain is irreducible and aperiodic, and therefore there is a stationary distribution of beliefs. Finally, the assumption $WW \neq W$ guarantees that the beliefs will evolve across periods. Referring to the example for motivation, although the entire village may not exhibit a strong sense of connectedness, it is assumed that it can be divided into distinct conversation networks that fulfill Assumption 1. Otherwise, the model would lack meaningful analysis. Note that throughout this paper, the conversation network W is taken as given, unless it discusses the relationship between network structures and adoption/social welfare. Furthermore, define individual j as a *neighbor* of individual i if the weight w_{ij} is greater than zero.

Lemma 1. *(DeGroot (1974)) Under Assumptions 1, W^t (the t-th power of listening matrix W) converges to a matrix which can be denoted by $W^* := \lim_{t\to\infty} W^t$. All elements in the jth column of W^* corresponding to individual j have a common value, denoted as p_j. The vector of beliefs $\boldsymbol{b}^t$ converges and $\boldsymbol{b}^* := \lim_{t\to\infty} \boldsymbol{b}^t$. We have $\boldsymbol{b}^* = (b^*, ..., b^*)$ and $b^* = \sum_{i\in N} p_i b_i^0$. $b^* \sim N(\theta, \sum_{i\in N} p_i^2 \sigma)$. Denote $\bar{\sigma} := \sum_{i\in N} p_i^2 \sigma$. So $b^* \sim N(\theta, \bar{\sigma})$.*

See DeGroot (1974) for a formal proof. In the long run, all beliefs converge to b^* and let us call b^* *the consensus*. The consensus is a weighted average of all individuals' initial beliefs, where the weight p_i is determined by one's centrality in the network. Call p_i the *social influence of i*. Given the conversation network structure, the consensus is affected by only the initial beliefs $\mathbf{b}^0$. Since b_i^0 is drawn from a normal distribution for each i, b^* is a linear combination of normally distributed variables, so it is also drawn from a normal distribution.

When making decisions at time t, it is important to clarify the information available to individuals. One option is to assume that individuals can use all past beliefs along with their current belief b_i^t. However, this complicates the model unnecessarily.

In this model, the expected adoption payoff consists of two components: the individual's expectation of the underlying value θ and the number of other adopters (externality). Within the framework of naive learning, it's intuitive that individuals should rely solely on their current belief b_i^t to estimate θ at time t. Past beliefs only matter if they influence the adoption decisions of others. For example, individual j may be more likely to adopt at time t if they have a higher b_j^{t-1}. However, b_j^{t-1} doesn't directly impact j's behavior at time t, given the realization of b_j^t. One justification for considering b_j^{t-1} is that it assists individual i in estimating b_j^t, which is not directly observable at time t. Nonetheless, this paper primarily focuses on outcomes as t approaches infinity. Lemma 1 guarantees belief convergence, rendering past beliefs of neighbors irrelevant for estimating their current beliefs in the long run. Hence, the paper assumes that individuals only use their current belief for making adoption decisions at the current time.

Assumption 2. *At time t, when an individual i decides whether to adopt or not, he or she only uses b_i^t in addition to the listening matrix W, i.e., an individual does not consider any beliefs observed at previous periods $t-1$ or earlier.*

The intuition is that if a piece of information does not impact anyone's estimation of θ, it will not affect the behavior of any individual. Intuitively, this assumption implies that an individual does not believe that the belief of one of their neighbors, observed yesterday, provides any additional information about whether that neighbor will adopt or not today. This is because the individual believes that their own current belief already incorporates all the information they have seen in the past, and given enough time for communication, all beliefs will eventually converge to close proximity to his or her own belief. Therefore, they do not consider their neighbor's belief from yesterday as relevant for predicting their neighbor's adoption decision today.

Given an individual i only observes b_i^t for his or her decision making, a strategy will be to assign an adoption decision a_i^t to all possible b_i^t. This paper uses Bayes Nash Equilibrium (BNE) as the solution concept. It means that an individual i has to have a posterior regarding all b_j^t for making his or her decision at time t. It is natural to assume that an individual i takes b_j^t as a normally distributed variable with b_i^t as the expectation and $\sigma_i^t + \sigma_j^t - 2\sum_{z\in N} w_{iz}^{(t)} w_{jz}^{(t)} \sigma$, which is denoted by $\sigma_{i,j}^t$, as the variance, where $w_{ij}^{(t)}$ denotes the corresponding element in matrix W^t. They are the conditional mean and variance of b_j^t given b_i^t, as it is easy to show that

$$E(b_j^t|b_i^t,W) = E(\theta + \varepsilon_j^t|b_i^t,W) = E(b_i^t - \varepsilon_i^t + \varepsilon_j^t|b_i^t,W) = b_i^t \tag{2.3}$$

and

$$Var(b_j^t|b_i^t,W) = Var(b_i^t - \varepsilon_i^t + \varepsilon_j^t|b_i^t,W) = Var(\varepsilon_i^t - \varepsilon_j^t|b_i^t,W) = \sigma_i^t + \sigma_j^t - 2\sum_{z\in N} w_{iz}^{(t)} w_{jz}^{(t)} \sigma \tag{2.4}$$

The nature of the model is static, as the multiple time periods serve solely for belief updating, and individuals receive payoffs only once when they have to decide whether to adopt or not, after which the game ends. Therefore, based on Assumption 2, an individual i's action is characterized solely by b_i^t given W.

Furthermore, the conversation network W and the variance of initial signals σ are both common knowledge and taken as given by individuals. As a result, the variances of beliefs, denoted as σ^t, are also common knowledge. Due to Assumption 2, σ^t is the only time-related factor that affects the strategy.[6] Therefore, in this context, a strategy is defined as a mapping of an individual's possible types (i.e., his or her current beliefs) to the two available actions: adopting and not adopting, given all variances at t (equivalently W).

Strategies A strategy of player i at time t is denoted as $s_i^{\sigma^t}$. Following standard approaches, it is a mapping that assigns for every possible b_i^t an adoption decision

[6]For this reason, I use σ^t instead of t as the superscript in the notation for strategies.

in $\{0,1\}$, i.e., $s_i^{\sigma^t}:\mathbb{R}\mapsto\{0,1\}$. $s_i^{\sigma^t}$ depends on time t only because σ^t change over time. Because the game is in nature static, I refrain from defining individual i's strategy as a vector containing $s_i^{\sigma^t}$ for all t. Also, denote $\mathbf{s}^{\sigma^t}:=(s_1^{\sigma^t},...,s_n^{\sigma^t})$. Given a strategy $s_i^{\sigma^t}$, there must be a set (possibly empty) containing all the values of beliefs with which i adopts using $s_i^{\sigma^t}$. Denote such a set by $B(s_i^{\sigma^t}):=\{b:s_i^{\sigma^t}(b)=1\}$. Denote by $F_{b_i^t,\sigma_{i,j}^t}$ the cumulative distribution function (CDF) of the random variable following the normal distribution $N(b_i^t,\sigma_{i,j}^t)$, which is the conditional posterior i has of b_j^t. Also, denote the corresponding probability density function (PDF) by $f_{b_i^t,\sigma_{i,j}^t}$. Given a strategy profile $\mathbf{s}^{\sigma^t}$ and one's own belief b_i^t, i's expected payoff if the games ends at t is

$$U_i(s_i^{\sigma^t},\mathbf{s}_{-i}^{\sigma^t}|b_i^t,W)=s_i^{\sigma^t}(b_i^t)[kb_i^t+\phi\sum_{j\in N\setminus i}\int_{x\in B(s_j^{\sigma^t})}f_{b_i^t,\sigma_{i,j}^t}(x)dx] \tag{2.5}$$

BNE can be hence defined.

Definition 4. A *Bayes Nash equilibrium (BNE) at t* is a strategy profile $s^{\sigma^t*}=(s_1^{\sigma^t*},...,s_n^{\sigma^t*})$ such that for every i, $s_i^{\sigma^t*}(b_i^t)=1$ iff $U_i(1,\mathbf{s}^{\sigma^t*}_{-i}|b_i^t,W)\geq 0$.

Since choosing not to adopt yields zero payoffs, one will adopt as long as his or her expected payoff is (weakly) positive. BNE happens when no one given any belief can switch from adopting to not adopting, or the other way around, to increase his or her expected payoffs. Throughout this paper, when it says equilibrium, it means BNE defined in Definition 4. Finally, the timing of the game is below.

Timing

1. Nature selects θ and the time t^s at which the game ends.
2. Every individual observes the beliefs his or her neighbors have at last time period if the time $t\geq 1$.
3. Every b_i^t is formed according to naïve learning if $t\geq 1$, or is drawn from the normal distribution $N(\theta,\sigma)$ if $t=0$.

4. If $t \geq 1$, Every individual chooses to adopt or not (but it will be realized only if $t = t^s$).

5. If it is t^s, every individual receives payoffs. The game ends. Otherwise, it goes to the next period.

2.3.2 Equilibrium

The game described earlier shares similarities with global games. A notable characteristic of global games is that in the limit as the precision of the signals players receive about the state of nature approaches infinity, and consequently, the signals become infinitely close to each other, a unique equilibrium emerges where players coordinate their actions if their signals exceed a certain cutoff threshold. In my model, a similar phenomenon occurs when the time for exchanging beliefs tends towards infinity. Lemma 1 can be utilized to demonstrate that

$$\lim_{t\to\infty} \sigma^t_{i,j} = \lim_{t\to\infty}(\sigma^t_i + \sigma^t_j - 2\sum_{z\in N} w^{(t)}_{iz} w^{(t)}_{jz} \sigma) \tag{2.6}$$

$$= 2\bar{\sigma} - 2\sum_{z\in N} p^2_z \sigma \tag{2.7}$$

$$= 2\bar{\sigma} - 2\bar{\sigma} \tag{2.8}$$

$$= 0 \tag{2.9}$$

Hence, the results derived from classical global game literature can be applied when the time approaches infinity and all beliefs converge to a consensus. However, even in finite time, an equilibrium still exists. The monotonicity and continuity of the expected utility function (equation (2.5)) ensure that a cutoff strategy, where an individual i adopts at time t if and only if their belief b^t_i is above a certain cutoff threshold, is always employed. Moreover, the fact that an individual i never adopts at time t if $b^t_i < -\frac{\phi(n-1)}{k}$ and always adopts if $b^t_i > 0$ guarantees that a profile of best responses constitutes a continuous function mapping a compact convex set of cutoff thresholds to itself. By applying the Brouwer fixed-point theorem, the existence of an equilibrium can be established. Proposition 1 presents this result.

Proposition 1. *Under Assumption* 2*, an equilibrium exists at each* $t > 0$*. Denote an equilibrium by* $\mathbf{s}^{\sigma^t *} := (s_1^{\sigma^t *}, s_n^{\sigma^t *})$*. Every equilibrium is a cutoff equilibrium, in the sense that given any* σ^t*, individual* i *adopts in the equilibrium if and only if* b_i^t *is higher than his or her cutoff, which is denoted by* $c_i^{\sigma^t}$*. Also denote* $\mathbf{c}^{\sigma^t} := (c_1^{\sigma^t}, \ldots, c_n^{\sigma^t})$ *and* $\mathbf{c}_{-i}^{\sigma^t} := (c_1^{\sigma^t}, \ldots, c_{i-1}^{\sigma^t}, c_{i+1}^{\sigma^t}, \ldots, c_n^{\sigma^t})$*.*

A cutoff in this context refers to a belief level at which an individual is indifferent between adopting and not adopting, taking into account the cutoff strategies of all other individuals. Thus, an equilibrium is characterized by a profile of cutoff strategies $\mathbf{c}^{\sigma^t}$ that satisfy the following condition

$$kc_i^{\sigma^t} + \phi \sum_{j \in N \setminus i} (1 - F_{c_i^{\sigma^t}, \sigma_{i,j}^t}(c_j^{\sigma^t})) = 0 \tag{2.10}$$

for every i. The system of equations is nonlinear, as $F_{c_i^{\sigma^t}, \sigma_{i,j}^t}(c_j^{\sigma^t})$ is nonlinear with respect to $c_j^{\sigma^t}$. Therefore, without additional assumptions, it is not possible to rule out the existence of multiple equilibria in finite time. One approach to ensure a unique equilibrium in finite time is to assume that at some time t, all $\sigma_{i,j}^t$ are sufficiently large so that other individuals' cutoff changes have a very small impact on an individual's decision-making. In such a case, a contraction mapping argument can be applied to prove the existence of a unique equilibrium at that time. However, it is mentioned that $\lim_{t \to \infty} \sigma_{i,j}^t = 0$, which means this assumption does not hold in the infinite time limit.

In the long run, multiplicity of equilibria does not exist due to the convergence of all posteriors. In the infinite time limit, every individual adopts the same cutoff strategy since they have the same posteriors regarding other individuals' beliefs. This results in equation 2.10 becoming linear, as $F_{c_i^{\sigma^t}, \sigma_{i,j}^t}(c_j^{\sigma^t}) = \frac{1}{2}$ for any variance $\sigma_{i,j}^t$ as long as $c_i^{\sigma^t} = c_j^{\sigma^t}$. Therefore, denoting the cutoff used by every individual in the limit as c, it can be shown that $c^* = -\frac{\phi(n-1)}{2k}$. This implies that if an individual's belief in the limit is equal to c^*, they will be indifferent between adopting and not adopting, assuming everyone else adopts with a probability of $\frac{1}{2}$. From other individuals' perspectives, the probability of adoption for an individual with belief

c^* is also equal to one-half.

This result is similar to the example presented by Morris and Shin (2003), where two players adopt (or invest) according to the same cutoff, and each player believes the other will adopt with a probability of one-half. Furthermore, a cutoff strategy $c^* > -\frac{\phi(n-1)}{2k}$ or $c^* < -\frac{\phi(n-1)}{2k}$ does not survive iterated deletion of strictly dominated strategies, which is a method used in global game literature to find a unique equilibrium. This is because $kc^* + \frac{1}{2}\phi(n-1) > (<) 0$, leading every individual to believe that everyone else will lower (raise) their cutoffs. Thus, the result stated below aligns with the results obtained in the global game literature.

Corollary 5. *Under Assumption* 1 *and* 2, $\lim_{t\to\infty} \boldsymbol{s}^{\sigma^t *} := \boldsymbol{s}^*$ *where* $\boldsymbol{s}^* := (s^*, \dots, s^*)$. *It equivalently implies that* $\lim_{t\to\infty} \boldsymbol{c}^{\sigma^t} := \boldsymbol{c}^*$ *where* $\boldsymbol{c}^* := (c^*, \dots, c^*)$, *i.e., in the unique equilibrium in infinite time, every individual will choose to adopt if and only if* $b^* \geq c^*$. *In short, every individual uses the same cutoff strategy in the unique equilibrium in the long run. Moreover,* $c^* = -\frac{\phi(n-1)}{2k}$.

In the context of the model, the unique equilibrium in the limit, denoted as $\boldsymbol{s}^*$ or $\boldsymbol{c}^*$, is referred to as the *limiting equilibrium* or *equilibrium in the limit* interchangeably. This research primarily focuses on the limiting equilibrium, driven by my main motivation of studying communication within small and rural communities, characterized by relatively intensive internal interactions. In such contexts, it is reasonable to believe that beliefs converge rapidly. In the limiting equilibrium, there are two possible cases if b^* is given. One case is when every individual adopts, which occurs when $b^* \geq -\frac{\phi(n-1)}{2k}$. The other case is when no one adopts, which occurs when $b^* < -\frac{\phi(n-1)}{2k}$.

The action that every individual takes in the limiting equilibrium, $s^*(b^*)$, is referred to as the *program outcome*, or *outcome* in short. While the program outcome of $s^*(b^*) = 1$ always leads to a higher number of adopters compared to the alternative case, it may not necessarily maximize the ex-post total utility, which depends on the underlying value θ. Therefore, when considering the perspective of a social planner or a policymaker, it becomes important to determine under what conditions adoption improves social welfare and how this depends on the structure of the conversation network W. These questions are discussed in Section 3.3 of the paper.

2.3.3 Coordination and network structure

In the framework of the baseline model, the program outcome can only fall into one of two cases: either universal adoption or no adoption. However, a social planner, without observing the true consensus belief b^*, can still infer the probabilities of these two cases based on his or her knowledge of the distribution of b^*. The main question then becomes how these probabilities depend on the structure of the conversation network W and the underlying value θ. In particular, the paper investigates how the disparity in social influences, which is determined by the structure of W, affects the probabilities of program outcomes and, consequently, social welfare.

To quantify the disparity among individuals in network positions, the paper employs a network centrality measure called *eigenvector centrality*. The left-hand unit eigenvector centralities of W are defined as an $n \times 1$ vector $\mathbf{e}$ satisfying the equation $\mathbf{e}'W = \mathbf{e}'$, where n is the number of individuals in the network. Appendix A provides further details on unit eigenvector centrality. Let $\bar{e}$ denote the average of the unit eigenvector centralities e_i for all individuals i in the network N. Additionally, let s_e represent the sample variance of unit eigenvector centrality, defined as $s_e = \frac{\sum_{i \in N}(e_i - \bar{e})^2}{n}$. The measure s_e provides a quantification of the disparity in unit eigenvector centralities across individuals in the network. By examining how this disparity, determined by the structure of W, influences the probabilities of program outcomes and social welfare, the paper sheds light on the relationship between conversation network structure and decision coordination.

Since every individual uses the same strategy s^* in the limiting equilibrium, the ex-post payoff for any individual in the limiting equilibrium depends only on $s^*(b^*)$ and θ. Denote the ex-post payoff by $u(s^*(b^*)|\theta) = s^*(b^*)[k\theta + \phi(n-1)s^*(b^*)]$. Recall that everyone adopts if and only if $b^* \geq -\frac{\phi(n-1)}{2k}$. Also, $b^* \sim N(\theta, \bar{\sigma})$. Then, the *social welfare*, or *expected total utility* interchangeably, is defined as

$$\bar{U} = n[k\theta + (n-1)\phi][1 - F_{\theta,\bar{\sigma}}(-\frac{\phi(n-1)}{2k})] \tag{2.11}$$

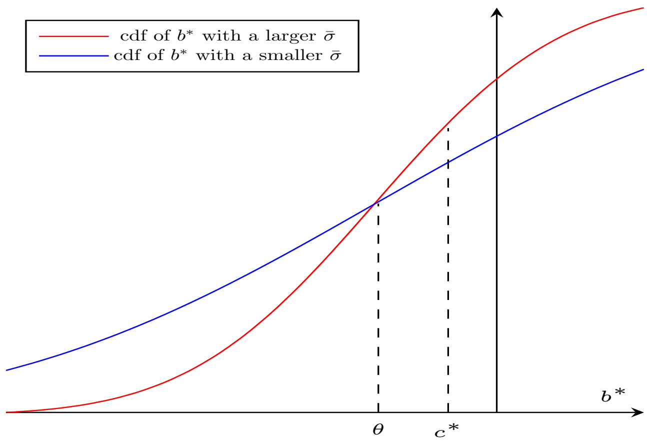


Figure 2.1: When $\theta < c^*$, a larger variance increases adoption.

where $1 - F_{\theta,\bar{\sigma}}(-\frac{\phi(n-1)}{2k})$ is the probability that $b^* \geq -\frac{\phi(n-1)}{2k}$. Imagine that a social planner who knows θ and W but not b^* wants to maximize the expected total utility given the distribution of b^*. Then the social welfare above is subject to be maximized by such a social planner. In addition to the total utility, it is also useful to define the *expected number of adopters*, denoted by $n(1 - F_{\theta,\bar{\sigma}}(-\frac{\phi(n-1)}{2k}))$, which is the number of all individuals times the probability that one adopts.

Proposition 2. *Under Assumption* 1 *and* 2*, the expected number of adopters is*

- *strictly increasing in* s_e *if* $\theta < -\frac{\phi(n-1)}{2k}$*, and*
- *strictly decreasing in* s_e *if* $\theta > -\frac{\phi(n-1)}{2k}$.

The social welfare $\bar{U}$ *on the other hand is*

- *strictly increasing in* s_e *if* $-\frac{\phi(n-1)}{k} < \theta < -\frac{\phi(n-1)}{2k}$*, and*
- *strictly decreasing in* s_e *if* $\theta > -\frac{\phi(n-1)}{2k}$ *or* $\theta < -\frac{\phi(n-1)}{k}$.

Since in the limit, the equilibrium is either that all adopt or that no one adopts, the expected number of adopters is increasing in the probability that the equilibrium in the limit is that everyone adopts. It depends on the distribution of b^*, with θ as its mean and $\bar{\sigma}$ as its variance. The variance is increasing in the inequality of network positions, s_e, as shown in the proof for Proposition 2. Then, if θ as the expectation of b^* is below the threshold $-\frac{\phi(n-1)}{2k}$, an increased variance will raise the probability that b^* is above that threshold. Otherwise, a larger variance will decrease that probability, as illustrated by Figure 2.1. Furthermore, although the threshold does not depend on the learning process, it does depend on the size of the network. A very large network (imagine n goes to infinity) will make the threshold $-\frac{\phi(n-1)}{2k}$ very small, so it will be easier to achieve adoption. This is not surprising. Because from individual i's perspective, there is a positive probability for another individual j to adopt, i.e., there is a positive probability that $b_j^t \geq 0$, and therefore j adopts even if he or she is the only adopter. Hence when n goes to infinity, the positive externality goes to infinity as well, which makes individual i always happy to adopt. This means that the probability that $b^* \geq c^*$ goes to 1 given θ and W.

The second part of Proposition 2 gives the conditions under which a higher inequality in network positions is good or bad for social welfare. If θ is too low (below $-\frac{\phi(n-1)}{k}$), a higher s_e increases adoption, but it does not worth it because individuals get negative ex-post payoffs even if everyone else adopts. On the other hand, if θ is higher than $-\frac{\phi(n-1)}{2k}$, a higher s_e decreases adoption while everyone adopting is an efficient outcome. It is notable that if θ is neither too low nor too high, i.e. between $-\frac{\phi(n-1)}{k}$ and $-\frac{\phi(n-1)}{2k}$, inequality in network positions is surprisingly good for social welfare, as it is still efficient if everyone adopts but the cutoff is higher than θ which makes full adoption less likely than no adoption.

Intuitively, when individuals are quite pessimistic about other individuals' beliefs, a more unequal network will make adoption more likely because they will lay their hopes on extreme beliefs possibly caused by few opinion leaders in their network. On the other hand, when individuals are optimistic about others' beliefs, any extreme opinion will only put adoption at stake. Loosely speaking, if centralities differ a lot across individuals in the network, from an individual's perspective it would be quite uncertain what this society is going to believe since ideas are greatly influenced by few people. If on average it is not beneficial for most people to adopt, a decentralized society is not good enough for achieving first-best, because the incentive to adopt is not strong enough at the individual level, hence an opinion leader

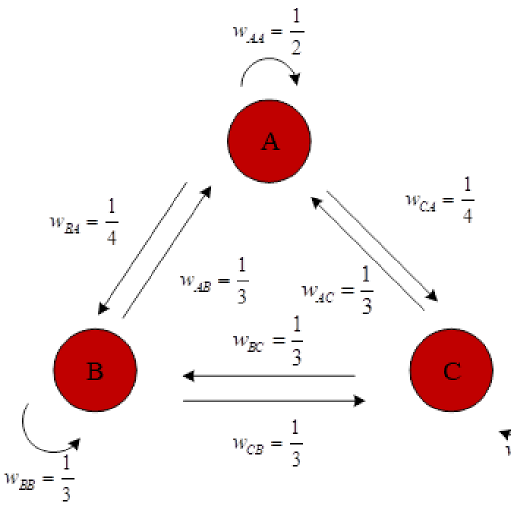


Figure 2.2: Conversation Network L in Example 1.

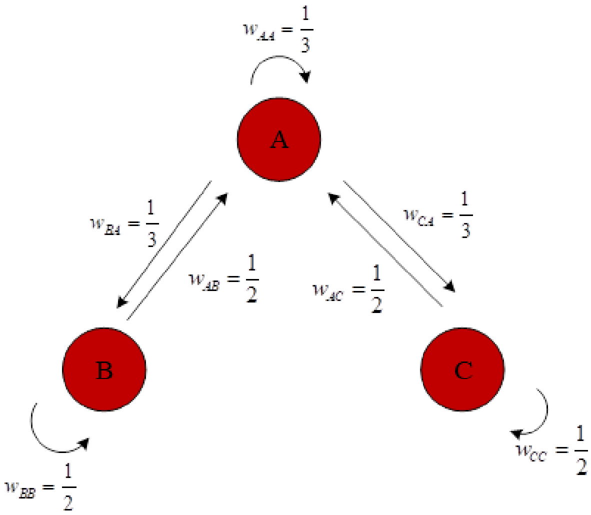


Figure 2.3: Conversation Network R in Example 1.

or dictator in the language of political economy is needed for an efficient outcome if there is still some value of adoption. If on average it is already beneficial for most people to adopt, then there is no reason for the society to call on a dictator. The following example is provided to illustrate Proposition 2.

Example 6. *Consider the two conversation networks in Figure 2.2 and 2.3. Call the conversation network in Figure 2.2 Network L and the conversation network in Figure 2.3 Network R. Both networks include three individuals, A, B and C. L and R appear in all relevant variables as superscripts representing that network. For example,* W^L *is the listening matrix of Network L. We have*

$$W^L := \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

This, L is a quite even conversation network, in which everyone listens to everyone. Individuals B and C equally put their weights on all individuals while A puts a half on himself or herself and then equally divides the other half between B and C. Making A put more weight on himself or herself is to guarantee that the rank of W^L *is larger than* 1*. We have* W^L *converges to*

$$W^{*L} := \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.4 & 0.3 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{pmatrix}$$

Therefore, individual A has higher social influence than B and C, because he or she listens less to others and more to himself or herself. Similarly, we have

$$W^{R} := \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

and

$$W^{*R} := \begin{pmatrix} 0.429 & 0.286 & 0.286 \\ 0.429 & 0.286 & 0.286 \\ 0.429 & 0.286 & 0.286 \end{pmatrix}$$

We can see that Network R is actually a star network in which individual A is connected to both B and C while the other two individuals are connected to only A. Intuitively, a star network should have a relatively larger inequality in social influence. Actually, the inequality in the social influence of Network L, s_e^L, is approximately 0.002 *while s_e^R is approximately* 0.004*. Assume that $\sigma = 0.5$, and we have $\bar{\sigma}^L \approx 0.170$ and $\bar{\sigma}^L \approx 0.173$.*

Suppose that we have $k = 10$, $\phi = 5$ and of course $n = 3$. Therefore, $-\frac{\phi(n-1)}{2k} = -0.5$ while $-\frac{\phi(n-1)}{k} = -1$. First, choose $\theta = -1.1$ which is below $-\frac{\phi(n-1)}{k}$. We can then compute $\bar{U}^L$ and $\bar{U}^R$, the total utility defined in equation 2.11 for Network L and R respectively. We have $\bar{U}^L \approx -0.0006$ while $\bar{U}^R \approx -0.0008$. So it is not beneficial for individuals to adopt, and higher inequality in social influence lower social welfare as it encourages adoption. If we instead have $\theta = -0.8$, which is betweeen $-\frac{\phi(n-1)}{k}$ and $-\frac{\phi(n-1)}{2k}$, then $\bar{U}^L \approx 0.232$ while $\bar{U}^R \approx 0.249$. Now it is beneficial to adopt, if everyone else does the same, and higher inequality in social influence improves social welfare as it encourages adoption. At last, set $\theta = -0.3$, which is higher than $-\frac{\phi(n-1)}{2k}$. Then $\bar{U}^L \approx 18.486$ while $\bar{U}^R \approx 18.399$. Now it is very beneficial for an individual to adopt if everyone else does, while a large disparity in network positions is harmful since it increases the probability of no adoption.

There are valuable policy implications that Proposition 2 provides. It is possible for a policymaker to determine which condition regarding θ is satisfied. Although more detailed investigations would be necessary in real-world scenarios, a simplified example can illustrate this concept. The policymaker could initially allow a small group of individuals to adopt the product, thereby realizing the ex-post payoffs for this group. Then, the policymaker could inquire about the number of adopters from their social networks that would make adoption worthwhile for them. Suppose the average of these numbers is denoted as $\underline{n}$. It should hold that $k\theta + \phi\underline{n} = 0$, which implies $\underline{n} = -\frac{k\theta}{\phi}$. By estimating $-\frac{k\theta}{\phi}$ in this manner, it becomes possible to estimate $-\frac{\phi(n-1)}{2k\theta}$ and $-\frac{\phi(n-1)}{k\theta}$ since n is readily observable. Consequently, it can be determined whether θ is greater or smaller than $-\frac{\phi(n-1)}{k}$ and $-\frac{\phi(n-1)}{2k}$. Armed with this knowledge, the policymaker can determine whether a more equal conversation network increases the adoption rate, if the goal is to maximize the number of adopters, or whether it increases the expected total utility, if the aim is to maximize social welfare. Policies can then be formulated accordingly. For instance, if the policymaker discovers that θ is likely to be higher than $-\frac{\phi(n-1)}{2k}$, efforts to reduce inequality in social influence can be implemented. This could involve organizing social events where individuals have the opportunity to interact with new acquaintances, thereby diminishing disparities in social influence and increasing the likelihood of adoption. Conversely, if the findings suggest that θ is likely to fall between $-\frac{\phi(n-1)}{k}$ and $-\frac{\phi(n-1)}{2k}$, the policymaker could empower influential individuals within the conversation network to play a more prominent role in promoting the product[7]. Despite their personal beliefs, this strategy would, on average, lead to higher adoption rates and increased social welfare.

Although the baseline model offers some useful policy implications, it is not immune to criticism. The baseline model suggests that in any equilibrium, given the consensus b^*, the outcome is either full adoption or non-adoption. This is justified by considering that the entire community can be divided into small groups, where information diffusion and positive externalities occur within each group. However,

[7] Banerjee et al. (2019) provides an approach to identify those influential individuals.

while the baseline model may capture some features, it does not precisely reflect what happens in the real world. For instance, in the study on agricultural insurance adoption mentioned earlier, Cai et al. (2015) demonstrated that having a friend who has already adopted agricultural insurance increases the likelihood of an individual's own adoption. However, it is not necessary for both friends to adopt simultaneously; the influence can be asymmetrical, with one friend adopting while the other does not. To make the theory more empirically relevant, the next section extends the baseline model and considers relaxing assumption 1, which allows for the inclusion of conversation networks W which may not be strongly connected.

2.4 Beyond Strong Connectedness

In this section, I present an extended model that relaxes Assumption 2 to allow the conversation network to be not strongly connected. If the conversation network is not assumed to be strongly connected, the beliefs may not converge to the same consensus since the corresponding transition matrix of the Markov chains is no longer irreducible. In such cases, one approach to handle this type of listening matrix is to view it as a disjoint union of connected subnetworks called *closed communicating classes*, as defined by Golub and Sadler (2017)[8]. In this scenario, each closed communicating class will reach its own consensus, while the remaining nodes will take a weighted average of the consensus reached by the communicating classes they listen to. This allows for a more nuanced understanding of how beliefs evolve within different subgroups or communities which are weakly interconnected, as they can reach their own distinct consensus rather than a single global consensus, and hence allows for different outcomes across subgroups.

Although the extension relaxes Assumption 1 and introduces closed communicating classes, the assumption of strong aperiodicity for the whole conversation matrix W is still maintained for two reasons. First, this assumption is relatively weak, particularly when the weights w_{ii} can be arbitrarily small. It is reasonable to assume that individuals update their beliefs by at least a small amount based on

[8] For more comprehensive details, refer to Golub and Jackson (2010).

the beliefs they have already formed. Second, if W is not assumed to be strongly aperiodic, the remaining nodes that do not belong to any closed communicating class may experience cyclical changes in their beliefs. For instance, consider the case of two individuals i and j with $w_{ij} = w_{ji} = 1$, where they consistently use each other's previous period beliefs. In such a scenario, the outcomes of the game will not converge. Hence, Assumption 1 is replaced with the following assumption.

Assumption 3. *The listening matrix W is strongly aperiodic.*

To begin with, I define a subnetwork in the context of network analysis. A *conversation network* is represented by (N, ω) where N is the set of all individuals and ω is a set of weighted links defined as $\omega := \{w_{ij} : i \in N, j \in N\}$.

Definition 7. A *subnetwork l* of the conversation network (N, ω) is represented by (N_l, ω_l), satisfying the following conditions:

1. $N_l \subseteq N$
2. If $i, j \in N_l$ and $w_{ij} \in \omega$, then $w_{ij} \in \omega_l$.

The cardinality of N_l is denoted as n_l. The corresponding listening matrix of subnetwork l is represented by W_l, with its elements denoted as $w_{l,ij}$ corresponding to the weighted link w_{ij} in ω_l.

Note that a subnetwork can potentially consist of only one individual. Based on this definition, a closed communicating class can be defined as follows.

Definition 8. A subnetwork l of (N, ω) is a *closed communicating class* if it satisfies the following conditions:

1. $n_l \geq 2$.
2. W_l is strongly connected.
3. if $i \in N_l$ while $j \neq N_l$, $w_{ij} = 0$.

Intuitively, a closed communicating class can be understood as a conversation network for belief-exchanging that is internally cohesive, meaning its members interact closely (although maybe not directly) and a consensus will eventually emerge.

However, this closed communicating class does not listen to or receive input from individuals who are not part of the class. In other words, it forms a distinct subgroup or club that has its own internal dynamics of information diffusion but remains disconnected from other individuals outside the class. The entire conversation network in the baseline model can be regarded as a closed communicating class in the extended model. This means that the extended model demonstrates how a conversation network from the baseline model is embedded within a larger community consisting of multiple such components. Additionally, the extended model allows for the presence of individuals who do not belong to any closed communicating class. These individuals, referred to as *remaining nodes*, must also be taken into account as their decisions can impact those within closed communicating classes. It is worth noting that due to strong aperiodicity, it can be proven (as shown in Lemma 2 below) that every remaining node will eventually have their beliefs converge over an infinite time horizon. Therefore, it is valuable to expand the definition of closed communicating class and introduce a new concept called a *consensus class*, defined as follows.

The model becomes peculiar if belief convergence occurs within finite time. This is because W is assumed to be known by every individual. If $WW = W$, it implies that convergence occurs in one period, leading to a scenario where, for $t > 1$, individuals accurately know others' beliefs. This challenges the uniqueness of the equilibrium. To address this anomaly, an additional assumption below is introduced.

Assumption 4. *If $l \in M$ is a closed communicating class, $W_l W_l \neq W_l$.*

Definition 9. A subnetwork l is a *consensus class* if it is either a closed communicating class or a remaining node. Denote M as the set of all consensus classes in (N, ω). So $l \in M$ if a subnetwork l is a closed communicating class or a remaining node. Denote $M_c \subseteq M$ as the set of all closed communicating classes while $M_r \subseteq M$ as the set of all remaining nodes. So $M_c \cap M_r = \emptyset$ and $M_r \cup M_c = M$.

It can be demonstrated that for a remaining node, his or her consensus belief will be a weighted average of the consensus beliefs of certain consensus classes.

Moreover, if a consensus class is itself a closed communicating class, it means that its consensus belief is also formed by taking a weighted average of other consensus beliefs, with a weight of 1 assigned to itself. Therefore, in the limit, every consensus belief is formed by taking a weighted average of other consensus beliefs, potentially including itself. To simplify the notation, denote the weight that consensus class l assigns to consensus class m in forming its own consensus belief as w_{lm}.

Lemma 2. *Under Assumption* 3, $\lim_{t\to\infty} b_i^t = b_l^*$ *for all* $i \in N_l$ *and* $l \in M$*, where* b_l^* *is the consensus reached by consensus class* l*.* $b_l^* = \sum_{m\in M} w_{lm} b_m^*$*. Particularly, if* l *is a closed communicating class, its listening matrix converges:* $\lim_{t\to\infty} W_l^t = W_l^*$*. Denote* $p_{l,i}$ *as the social influence of* i *in such a closed communicating class* l*.*

Lemma 3. *Under Assumption* 3*, for all consensus class* $l \in M$*,* $b_l^* \sim N(\theta, \bar{\sigma}_l)$ *where*

1. $\bar{\sigma}_l = \sum_{i\in N_l} p_{l,i}^2 \sigma^2$ *if* l *is a closed communicating class.*
2. $\bar{\sigma}_l = \sum_{m\in M_c} w_{lm}^2 \bar{\sigma}_m + \sum_{g\in M_r} w_{lg}^2 \sigma$ *if* l *is a remaining node.*

Figure 2.4 provides an example of a conversation network that is not strongly connected (but is strongly aperiodic). This network is derived from the one shown in Figure 2.2 by adding two additional nodes, D and E. In this network, the red nodes A, B, and C form a closed communicating class, while the yellow nodes D and E represent the remaining nodes. Node D listens to node B, but when considering nodes A, B, C, and D together, they do not form a closed communicating class since there is no path from any of A, B, or C to D. In the long run, the belief of node D will converge to the consensus of the closed communicating class. Despite D assigning a weight of $\frac{1}{3}$ to his or her own belief, the influence of his or her own belief diminishes in the long run, as shown in the proof of Lemma 2. Consequently, the beliefs of node E also converge to the consensus of the closed communicating class. Now, let us consider a scenario where $w_{ED} = 0$, $w_{DE} = \frac{1}{6}$, $w_{DD} = \frac{1}{6}$, and $w_{EE} = 1$. In this case, E no longer listens to D and only relies on his or her own beliefs. D assigns equal weights to E as he or she does to him or herself. As a result, E only considers his or her initial beliefs, which remain unchanged over time. Since

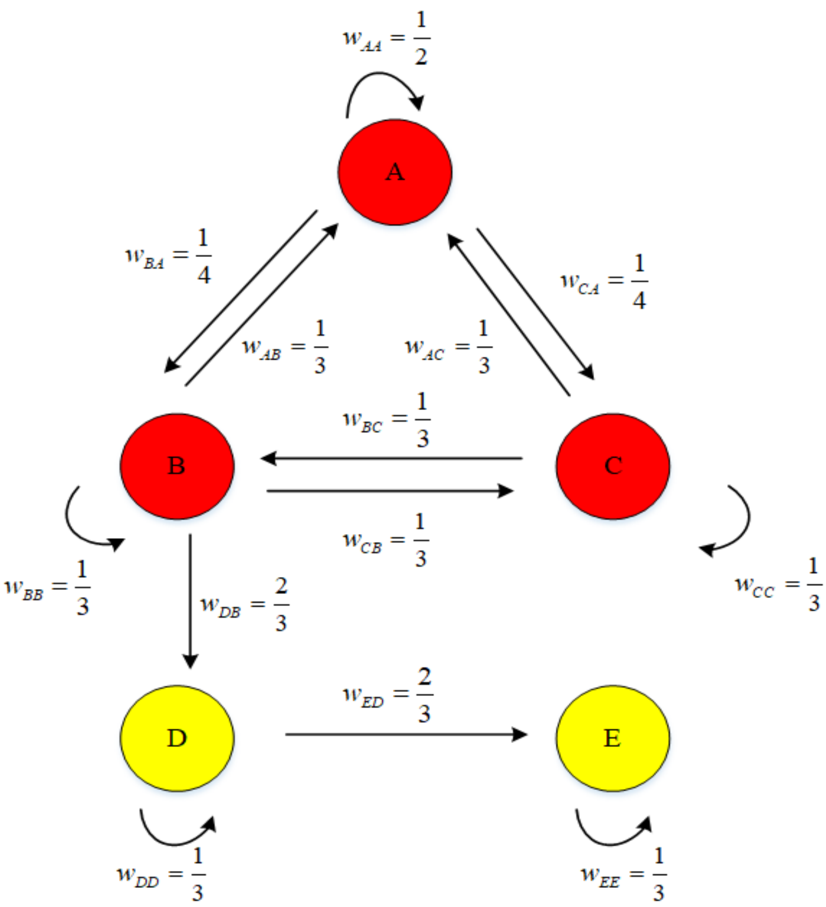


Figure 2.4: A conversation network that is not strongly connected.

D assigns a weight of $\frac{1}{6}$ to E, $\frac{1}{6}$ of D's belief always consists of E's initial belief, even in the long run.

The extended model retains all other setups, including timing, payoffs, strategies, and Assumption 2, from the baseline model. The only modification is the relaxation of Assumption 1 with the consideration of consensus classes. It is important to note that Proposition 1 does not rely on Assumption 1, indicating that an equilibrium still exists for all time periods. This is due to the preservation of the continuity of the expected utility function and the presence of upper and lower bounds regarding the cutoff an individual may choose. Therefore, even in the extended model, the existence of an equilibrium is guaranteed.

However, it is noticeable that in the baseline model, an individual only receives a positive externality from others who are connected to them in the conversation network. Adoption by individuals outside of their conversation network either has no impact on their payoff or is accounted for by the term $k\theta$. On the other hand, in the extended model, an individual can benefit from the adoption of another individual, even if they are not connected in the conversation network. In such cases, two individuals still need to form posteriors about each other's beliefs, even if they belong to different consensus classes. Their beliefs may not converge to the same consensus, leading to persistent uncertainty about each other's beliefs. Since individuals within the same closed consensus class eventually reach the same belief, the variance $\sigma^t_{i,j}$, which represents the variance of individual j's belief b^t_j as perceived by individual i based on i's own belief b^t_i, will converge to the same value for all individuals i belonging to one consensus class l and all individuals j belonging to another consensus class m. We can denote this convergence as $\lim_{t\to\infty}\sigma^t_{i,j} = \bar{\sigma}_{l,m}$, where $i \in N_l$, $j \in N_m$, and $l,m \in M$ with $l \neq m$. As a result, the process of iterated deletion of strictly dominated strategies, as in a global game, does not apply in the same way. To ensure the existence of a unique equilibrium in the limit, the following assumption is necessary.

Assumption 5. *$k > (n - n_l)\phi$ for every $l \in M$ and $\bar{\sigma}_{l,m} > \frac{1}{2\pi}$ for every $l,m \in M$.*

The assumption described serves to ensure that individuals or consensus classes do not overreact to changes in cutoff strategies of individuals from other consensus classes, thereby leading to a unique equilibrium. It has two parts, each with its own rationale. The first part of the assumption reflects the idea that the underlying value θ holds significant importance, exerting a larger influence on individuals' payoffs compared to the network effect. In other words, individuals are more sensitive to changes in the underlying value rather than the payoff derived from the network effect. This assumption is reasonable in many practical scenarios. For instance, when considering the adoption of agricultural insurance in a village, individuals are likely to first assess the benefits of the insurance itself before considering the number of others who participate. Their decision-making prioritizes the intrinsic value of the insurance rather than the influence of others. The second part of the assumption pertains to individuals' beliefs about other consensus classes. It posits that individuals do not hold extremely precise beliefs about the beliefs of individuals from other consensus classes. Consequently, no individual assumes that individuals from other consensus classes will possess beliefs that closely resemble their own. This aspect ensures that individuals are not highly sensitive to the behavior of other consensus classes in the long run. Importantly, the second part of the assumption is mild, as the variance of a normal distribution being represented by $\frac{1}{2\pi}$ is a very small value. Together, these assumptions contribute to the uniqueness of the equilibrium by preventing excessive sensitivity to changes in cutoff strategies from other consensus classes and emphasizing the importance of the underlying value θ in individuals' decision-making processes. Based on the described assumption, a unique equilibrium exists in the limit.

Proposition 3. *Under Assumptions* $2-5$, $\lim_{t\to\infty} c_i^{\sigma^t} = c_l^*$ *if* $i \in N_l$ *and* $l \in M$. *This implies that individuals belonging to the same consensus class will converge to the same cutoff strategy, denoted as* c_l^*, *in the long run. Furthermore, for any two closed communicating classes* l *and* m *in* M_c, *it holds that* $c_l^* \leq (\geq) c_m^*$ *if* $n_l > (<) n_m$, *and* $c_l^* = c_m^*$ *if* $n_l = n_m$.

It is not surprising that individuals within the same consensus class share the same long-run cutoff, as they base their decisions on identical information in the long run. The second part of Proposition 3 introduces a more intriguing aspect. It shows that the cutoff in the long run weakly decreases with an increasing closed communicating class size, implying that a larger group of individuals engaged in intensive internal communication is more prone to adoption. This is noteworthy because, based on the ex-post payoff function, an individual receives an equal benefit ϕ from the adoption of any other individual. Consequently, with regard to network externality, the adoption of any individual j has an equal impact on another individual i, irrespective of whether j belongs to the same closed communicating class as i or not. Nevertheless, this result demonstrates that individuals within a larger consensus class exhibit a greater propensity for adoption compared to those from a smaller class. The intuition is that as beliefs converge to the consensus, individuals in the same closed communicating class become increasingly confident about acting in a unified manner. Consequently, in larger closed communicating classes, their willingness to adopt is heightened, as they are aware that for all beliefs exceeding the cutoff, full adoption within the closed communicating class is certain. This result aligns with a significant body of empirical evidence, albeit some of which may be derived from slightly different contexts. For instance, Murendo et al. (2018) conducted a study on mobile money adoption in Uganda, which found that a household is more likely to adopt mobile money when it engages in daily communication about mobile money with a greater number of households. Another relevant example is the study conducted by Banerjee et al. (2015), which reveals a strong correlation between an individual's likelihood of adopting microfinance and his or her communication centrality, which can be closely associated with the size of the closed communicating class to which an individual belongs. It should be noted that this result is conditioned on the structure of the whole conversation network. It does not state that a closed communicating class will have a lower cutoff if we simply add more individuals to that class, as this will change the cutoff strategy used in other consensus classes. It instead says that given the same conversation network, if we pick two closed communicating classes from it, the one which has a larger size will have a lower cutoff and hence a higher propensity to adopt than the other. Due to this consideration, analyzing the impact of higher inequality in network positions within a closed communicating class on both adoption and the welfare of that class becomes challenging. Obtaining results like Proposition 2, which examines the relationship between inequality in network positions and adoption or total utility within a closed communicating class, is difficult due to the complex interactions. This is because larger inequality in network positions of a closed communicating class l can alter the cutoff strategies used by other classes through changes in $\bar{\sigma}_l$. Alternatively, I propose comparing two closed communicating classes of the same size, where one exhibits higher inequality in network positions while the other displays lower inequality. The results of this comparison are presented in Proposition 4; however, certain notations need to be clarified beforehand.

Recall that $p_{l,i}$ represents the social influence of individual i in closed communicating class l. Similarly, as defined in Section 3.3, let us denote the vector of left-hand unit eigenvector centralities of W_l as an $n \times 1$ vector $\mathbf{e}_l$, satisfying $\mathbf{e}_l' W_l = \mathbf{e}_l'$. The element in $\mathbf{e}_l$ corresponding to individual $i \in N_l$ is denoted as e_i. Now, consider $\bar{e}_l = \frac{\sum i \in N_l e_i}{n_l}$, which represents the average unit eigenvector centrality across all individuals in N_l. Additionally, let $s_e^l := \frac{\sum_{i \in N_l} (e_i - \bar{e}_l)^2}{n_l}$ denote the sample variance of unit eigenvector centrality among individuals in N_l. In order to analyze social welfare, it is necessary to define the expected total utility of a closed communicating class. However, this task is more intricate than the definition presented in Section 3.3. This is due to the fact that even if the social planner assumes that a closed communicating class will adopt, an individual's utility within that class still depends on the probabilities of adoption by other consensus classes. Let us denote the expected total utility of a closed communicating class as $\bar{U}_l$. To further elaborate, let d_l represent a partition of all consensus classes M, excluding class l, into two distinct sets: M_{d_l} and $\bar{M}_{d_l}$. Thus, we have $M_{d_l} \cup \bar{M}_{d_l} = M$ and $M_{d_l} \cap \bar{M}_{d_l} = \emptyset$. Let D_l denote the set of all such partitions. For simplicity, we denote the ex-ante probability of adoption by a consensus class l as α_l. Therefore, $\alpha_l = Pr(b_l^* \geq c_l^* | W, \theta)$, and $\alpha_l n_l$ corresponds to the expected number of adopters, as explained in Section 3.3.

Based on these notations, the expected total utility of a closed communicating class l can be defined as follows:

$$\bar{U}_l = \alpha_l n_l [k\theta + \phi(n_l - 1) + \phi \sum_{d_l \in D_l} \Pi_{m \in M_{d_l}} \alpha_m \Pi_{z \in \bar{M}_{d_l}} (1 - \alpha_z) \sum_{m \in M_{d_l}} n_m] \tag{2.12}$$

Although this expression may appear complex, it simply represents the total utility that the closed communicating class will obtain, expected over all possible scenarios regarding which consensus classes will adopt. Proposition 4 provides sufficient conditions to determine whether a larger inequality in network positions within a closed communicating class results in a lower or higher adoption rate, as well as the expected total utility, compared to another closed communicating class of the same size.

Proposition 4. *Under Assumptions $2-5$, consider two closed communicating classes l and m from the set M, where both classes have the same size, i.e., $n_l = n_m$.*

Suppose $s_e^l > s_e^m$, then the following conclusions can be drawn:

1. *If $\theta > -\frac{\phi(2n_l-1)}{2k}$, it holds that $\alpha_l < \alpha_m$ and $\bar{U}_l < \bar{U}_m$.*
2. *If $-\frac{\phi(n-1)}{k} < \theta < -\frac{\phi(n-n_l-\frac{1}{2})}{k}$, it holds that $\alpha_l > \alpha_m$.*
3. *If $\theta < -\frac{\phi(n-1)}{k}$, it holds that $\alpha_l > \alpha_m$ and $\bar{U}_l < \bar{U}_m$.*

Obtaining a threshold like Proposition 2, which determines when a more unequal conversation network increases or decreases social welfare, becomes challenging in this case. Even in the limit, there will still be uncertainty about other individuals' beliefs, making the threshold a function of the cutoffs of other consensus classes. Additionally, besides network structures, the size of a closed communicating class now affects its adoption probability, as Proposition 3 shows. Therefore, comparing a more equal closed communicating class with a less equal one is meaningless without keeping the size fixed. Nevertheless, it is still possible to derive sufficient conditions, and Proposition 4 follows the same rationale as Proposition 2. It indicates that if the underlying value of θ is very high (low), the adoption of a closed communicating class will be lower (higher) than a more unequal closed communicating class of the same size within the same conversation network. However, social welfare is higher in the more equal closed communicating class in both cases. Thus, similar to Proposition 2, an unequal conversation network is detrimental when the underlying value θ is either too high or too low, while increasing adoption if θ is relatively low.

This result, although not as strong as Proposition 2, holds value in terms of empirical relevance. It provides predictions that can be easily tested using appropriate data. For example, imagine conducting a program where adoption rates[9] and conversation networks are observed. As mentioned earlier, it is possible for the policymaker to assess whether the conditions regarding θ in the proposition are satisfied or not. Then these predictions can be tested by regressing adoption rates across closed communicating classes on the dispersion of unit eigenvector centrality, while

[9] It can be assumed that random shocks impact adoption decisions, leading to situations where full adoption may not be observed even if the cutoff of a closed communicating class is below its consensus.

controlling for the size of communicating classes and fixed effects of the whole conversation networks if applicable. In contrast, Proposition 2 is more challenging to test unless the program is conducted across multiple conversation networks that all satisfy Assumption 1. The extended model, being more empirically relevant, opens up avenues for counterfactual analysis by allowing for parameter estimation. Although obtaining a closed-form solution is not straightforward, numerical solutions for cutoffs in the limit can be obtained without much difficulty. This provides an opportunity for policymakers to determine an optimal subsidy (by increasing θ) that maximizes social welfare, considering the associated costs.

2.5 Conclusion and Discussion

This paper examines a coordination model in a network using naive learning over an infinite time horizon. Individuals update their beliefs regarding the value of adoption through the DeGroot learning process within a conversation network. Initially, each individual receives a private signal indicating the underlying value, and their beliefs are updated by taking the weighted average of their neighbors' beliefs. Individuals decide whether to adopt a product with strategic complementarity based on the beliefs they have formed through the naive learning process. The timing of adoption decisions is determined by nature.

The main contribution of this paper is the development of a framework that combines naive learning within social networks to analyze coordination problems. This framework contributes to the theoretical understanding of these fields and has potential applications in policy-making. The paper establishes the existence of an equilibrium in this framework, where individuals consistently use cutoff strategies that converge alongside their beliefs. This approach is similar to the framework of global games, a well-known model for solving coordination problems. While the standard global game model relies on the elimination of noise in information, the proposed model selects equilibrium based on the convergence of beliefs over an infinite time horizon.

In the baseline model, the convergence of beliefs leads to two outcomes: no adoption or universal adoption. The expected total utility (social welfare) is computed based on the consensus distribution. The paper explores how network structure affects social welfare and finds that a more unequal conversation network, characterized by inequality in unit eigenvector centrality, reduces social welfare. This holds when the state of nature is relatively high, encouraging coordination in an egalitarian and decentralized conversation network. However, when individual incentives are weak, individuals rely on extreme beliefs held by highly central individuals. Social welfare, measured as the long-term expectation of total utility over the consensus, consistently decreases with greater inequality in social influence. However, there is an interval of underlying value where higher centrality inequality improves adoption and social welfare. The extended model, which considers more general conversation network structures, produces similar results and provides greater empirical relevance.

This research can be further developed in several dimensions. First, while naive learning is supported by empirical evidence, it would be interesting to explore the combination of naive learning and Bayesian learning. It would be valuable to understand when individuals are likely to use Bayesian learning versus naive learning, possibly in short periods. For instance, in de Martí and Milán (2019), individuals communicate their signals only once, making it natural to assume that individuals can update their beliefs in a Bayesian way, given that updating is not complex for a single period.

Second, this model relies on the convergence of beliefs. However, if we consider adoption behavior within a finite time frame before consensus is reached, differences in beliefs will undoubtedly play a significant role. Unfortunately, analyzing this aspect becomes challenging as an individual's belief can vary greatly in different finite time periods. Determining someone's beliefs at any given time, even with a listening matrix influencing their social influence, becomes almost intractable. Moreover, complexity significantly increases in a coordination game.

However, there are still empirical implications that can be derived, leading to a potential third avenue for future development. While obtaining analytical solutions for finite time periods is challenging, it is relatively easy to compute numerical solutions. These heuristic results can be valuable, particularly for policy-making purposes. Although network data for development studies, especially in rural areas of developing countries, are typically expensive to collect, there has been a significant increase in the availability of such data.[10] Therefore, future research in this field is highly justified and warranted.

[10]For example, Alatas et al. (2016); Beaman et al. (2021); Atwell and Nathan (2022), etc.

Chapter 3

Son Preference and the Crops: Theory and Evidence from China

3.1 Introduction

Son preference is a prevalent phenomenon in many developing countries, and is especially an issue in South and East Asia. For example, it is documented that in China the male-to-female sex ratio at birth reached 1.2 in the first decade of 21st century (Li, 2007). Negative consequences directly from the distortion of sex ratio include infanticide (Hausfater, 1984; Langer, 1974), excess female child mortality which was considered the second source of missing women (Sen, 1992), and imbalances in marriage markets which not only increase the cost of a male to get married (Neelakantan and Tertilt, 2008; Hopkins et al., 2011), but also lead to behavioral changes such as rising criminality (Cameron et al., 2019).

Culture has been one of the main explanations for son preference in South and East Asia (Arnold and Zhaoxiang, 1992; Freedman et al., 1974; Williamson, 1976). In Chinese history, the dominance of traditional values promoted a universal preference for large clans, especially for many sons. Sons were considered as who would remain within the parental household to support the older generation after marriage, and who would continue the family line and tend the ancestral shrine (Croll, 1985; Yi et al., 1993). There has already been a vast literature that documents the importance of culture in causing son preference and the distortion of sex ratio, and some works also investigate the origins and persistence of such cultural phenomenon(Boserup, 2007; Alesina et al., 2013). However, the mechanism within this cultural transmission is still under-explored, and what features of some agricultural sectors will result in son preference remains a question.[1]

Our work sheds light on these questions by investigating the correlation between the difference in crops suitability and male-to-female sex ratio at birth, taking China as an example. China provides a prefect background for this study. Firstly, as we have already mentioned above, China has a severe problem of gender bias and very much distorted sex ratio at birth. Second, the son preference is very much rooted in its traditional values, which persists over centuries in Chinese history. The son preference in Chinese traditional values is well-noticed and proof can be easily found in numerous historical contexts. Last but not least, the agriculture in northern and southern China show a distinct pattern and is largely reflected by the geographical suitability of growing two main crops, wheat and rice. Geographically, South China is more suitable to grow rice, and thereby has a larger fraction of rice output than the North. In contrast, North China has a larger proportion of its agricultural output in wheat. Unlike wheat, growing rice demands a higher level of collaboration[2] within villages or clans, where patrilineal kinship systems, and consequently males, tend to play a more active role.

We first develop a theoretical model to characterize the effect of rice and wheat production on parents' choice over their off-springs' sex ratio. In our model, a household chooses the proportion of female and the proportion of male in its family planning, and then allocate its females and males between the two agricultural sectors, rice and wheat production. A household will have to pay a cost if it wants to have either the proportion of female or male greater than $\frac{1}{2}$, which is what we use as natural ratio. In other words, there is a cost in manipulating sex ratio at birth. We then assume that there are two types of households that differ in their cost of

[1] Most related literature adopts the assumption that women have comparative advantage over men in some sectors, see Qian (2008) and Xue (2018) etc., but rarely discusses how it is except men's physical strength.

[2] Talhelm et al. (2014) use a diverse and large set of cognitive tests to conclude that one reason that people in southern China show a greater collectivism than people in the North was societal patterns of farming rice versus wheat.

manipulating sex ratio, and a household with a lower cost is said to have greater son preference. We have wages all linear and increasing in a total production factor which is measured by crop suitability in our data. If a household puts some labor in rice production, it needs to pay a fixed cost, but its male labor will enjoy a benefit from collaboration which is increasing in the population working in rice production. This allows for the case where female has comparative advantage on wheat so that there is a specialization in production by gender. Our model suggests that the sex ratio is increasing in the gap between rice suitability and wheat suitability (the former minus the later). This prediction still holds when we extend our model to the dynamic case. Hence, if an area is geographically more suitable for growing rice than wheat (usually it also means that the area actually grows more rice than wheat), son preference will be more prevalent there and the sex ratio at birth will be distorted at a higher degree.

Empirically, we apply Geographic Information System (GIS) data on rice and wheat suitability and 1% population census data on male-to-female sex ratio at birth in China to test our theory. We show that, consistent with the predictions of our model, provinces/prefectures where the gap between rice and wheat suitability is lager also have a higher sex ratio at birth in both year 2000 and 2010. Our model thus provides a potential mechanism that could explain the regional variations in sex ratio that one observes in the data for China.

In addition, we apply 2012 wave Chinese General Social Survey (CGSS) at the individual level to see whether traditional agriculture is also correlated with gender norms. CGSS surveys individuals' views about gender difference in ability as well as gender roles. Using both OLS, probit and ordered probit estimator, we find positive correlations between the difference in rice and wheat suitability and attitudes reflecting gender inequality in today's society. In particular, individuals from provinces with lager gap between rice and wheat suitability also tend more to agree on that man's ability is greater than woman, and that being housewives gives the same sense of achievement as working for woman, as well as woman's role is to look after the family.

Our research contributes to several lines of literature. First of all, this paper adds to the literature about the origins of gender bias. One of the seminal works is Alesina et al. (2013, 2018). They examine the adoption of plough cultivation across countries as an important determinant of the historical gender division of labor and the evolution/persistence of gender norms as well as unbalanced sex ratio. However, while they focus on cross country comparisons, their framework fails to be applied in the case we are interested in, where the two major crops (rice and wheat) in China are both 'plough-positive', but there is still significant regional variations in sex ratio at birth. Moreover, while they emphasize on the empirical evidences, we also provide a theoretical model that can explain the empirical correlation between agricultural features and son preference.

Other related literature mainly examine the impact of existing economic structure of agriculture, or the change of it, on gender inequality. For instances, Qian (2008) studies economic reforms in China in the late 1970s and found that the development of tea-picking sector reduces gender inequality. Carranza (2014) finds that in parts of India with soil suitable for deep tillage, female will have a lower bargaining-power and it leads to a lower female labor force participation. Goldin (1995) and Mammen and Paxson (2000) discuss the relationship between economic development and gender bias. Our contribution lies in two aspects. Firstly, we develop a tractable model to describe the mechanism behind the empirical correlation between agricultural features and gender inequality, which to the best of our knowledge no research has done before. Second, compared to existing researches that study the impact of a specific industrial or agricultural sector on gender inequality, we look at two representative crops in a society experiencing a fairly long history of the dominance of agriculture, where the two crops clearly show distinct spatial distribution.

This paper is organized as follows. In section 2, we present our theoretical model and the results of it. In section 3, we describe the data we use for empirical investigations. Section 4 reports our empirical results on crops suitability and sex ratio, while section 5 presents our empirical results on crops suitability and gender norms. Section 6 concludes.

3.2 Model

3.2.1 Basic Setup

Consider a representative household (parents) is doing family planning by choosing the proportion of females and males in the next generation. The amount of both genders is normalized to measure 1 (maximal fertility). Denote the amount of males by L_m, and females by L_f we have: $L_m + L_f = 1$.

Off-springs will be working in the field for agricultural output. For the moment let us assume that there is only one output - wheat, and the production function is:

$$F^w(L_m, L_f) = A_w(\gamma L_m + L_f). \tag{3.1}$$

We assume $\gamma > 1$, which captures gender inequality in (pre-industrial) agricultural production (Alesina et al., 2013; Boserup, 2007). A_w in our model is TFP which mainly reflects how suitable the area is for growing wheat.

The natural male-to-female sex ratio is assumed to be 1. In other words, by nature we will have $L_m = L_f = \frac{1}{2}$. However, the parents can manipulate the sex ratio to maximize their payoff.[3] Nevertheless, by choosing L_f (or L_m) different from $\frac{1}{2}$, the parents need to pay a cost linear to the square of the degree of manipulation. For example, by having $L_f = a$ the cost is $c(\frac{1}{2} - a)^2$, where $c > 0$. Taking wages as given, the household's problem is given by:

$$\max_{L_f, L_m} u(L_f, L_m) = A_w \gamma L_m + A_w L_f - c(\frac{1}{2} - L_f)^2. \tag{3.2}$$

s.t.

$$L_f + L_m = 1, L_f \geq 0, L_m \geq 0.$$

[3] Sex manipulation in pre-modern China is a well-documented fact. The most observed ways include abandoning and infanticide. For example, see Lee and Feng (1999).

The interior solution follows:

$$L_f^* = \frac{A_w(1-\gamma)}{2c} + \frac{1}{2}, \tag{3.3}$$

$$L_m^* = \frac{1}{2} - \frac{A_w(1-\gamma)}{2c}. \tag{3.4}$$

Given $c > 0$ and $\gamma > 1$, apparently we have $L_f < \frac{1}{2}$. In other words, because man have advantages in doing physical works (such as using the plough), in a society that mainly produces wheat, at equilibrium the male-to-female sex ratio would be lager than 1. This result is consistent with Alesina et al. (2018), who find that countries that traditionally practiced plough agriculture which requires more physical strengthens (such as wheat and rice)-rather than shifting cultivation-have higher male-to-female sex ratio. In addition, the proportion of females is increasing in the cost of manipulating sex ratio. Moreover, higher A_w will increase male-to-female-sex ratio.

Now let us turn to rice production. We keep all the above settings but assume that

$$F^r(L_m, L_f) = A_r(\gamma L_m + L_f) + \bar{L}L_m. \tag{3.5}$$

where the subscript and superscript r mean rice. $\bar{L}$ is the measure of male labor force participating in rice production. Firstly, $\bar{L}$ can capture the fact that rice production involves cooperation at the village or clan level, such as building water conservancy facilities. In traditional societies, especially in traditional China, such cooperation was largely achieved through the operation and dominance of the patrilineal kinship system, as evidenced by extensive literature in economics, history, sociology, and anthropology[4]. Therefore, having male offspring also increases the potential for a family to maintain or take an active role in such cooperation.

Addtioanlly, in traditional Chinese culture, female will move out to live with her husband's family when married, which results in a 'lost' in labour for the fe-

[4]Related literature includes but is not limited to Alvard (2003); Greif and Tabellini (2010a); Xu and Yao (2015); Thomas et al. (2018); He et al. (2018) etc.

male's family. On the other hand, having a son in the family will bring labour outside (the wife) into the family through marriage, which is non-trivial for families that grow rice, as rice production demand a larger amount of labour (compared to wheat) due to the need for collaboration. Hence, for rice-producing families there is an additional value (that arises from marriage inequality) for having son. Although a concern is that those effects may be applied to both men and women, $\bar{L}L_m$ in equation (3.5) is not a crucial assumption for the main results of this paper. In appendix, we relax this assumption and our key predictions remain valid.

When the representative household chooses L_m and L_f, it takes $\bar{L}$ as given. However, in equilibrium we must have $\bar{L} = L_m$. In the case of representative household, it is without loss of generality to assume that $\bar{L}$ is a given constant. Under such setting, the interior solution becomes:

$$L_m = \frac{(\gamma - 1)A_r + \bar{L}}{2c} + \frac{1}{2}, \tag{3.6}$$

$$L_f = \frac{1}{2} - \frac{(\gamma - 1)A_r + \bar{L}}{2c}. \tag{3.7}$$

We again see that rice production itself will generate unbalanced sex ratio.

3.2.2 Full Model

Now, let us assume that there are two types of households $i \in \{h, l\}$, having different cost of manipulating the sex ratio c_i, where $c_h > c_l$ and each of measure $\frac{1}{2}$. We further assume that any household participates in rice production will have to pay a cost τ (e.g. the household need to spend time helping the others, etc). Suppose in total $\bar{L}$ males work on rice, the utility of type i household is:

$$u_i = A_w \gamma L^w_{mi} + A_w L^w_{fi} + (A_r \gamma + \bar{L}) L^r_{mi} + A_r L^r_{fi} - c_i \left(\frac{1}{2} - L^w_{fi} - L^r_{fi}\right)^2 - \tau 1\{L^r_{mi} + L^r_{fi} > 0\} \tag{3.8}$$

Hence the problem of a household i is:

$$\max_{L^w_{fi}, L^w_{mi}, L^r_{fi}, L^r_{mi}} u_i(L^w_{mi}, L^w_{fi}, L^r_{mi}, L^r_{fi}) \tag{3.9}$$

s.t.

$$L^w_{fi} + L^w_{mi} + L^r_{fi} + L^r_{mi} = 1,$$
$$L^w_{fi}, L^w_{mi}, L^r_{fi}, L^r_{mi} \geq 0.$$

where $\bar{L}$ is the benefits from collaboration works and τ is the cost of joining the collaboration, which will be incurred only if the household participates in rice production. We then have the following definitions:

Definition 1. An *equilibrium* is a set of choice variables $({L^w_{ml}}^*, {L^w_{fl}}^*, {L^w_{mh}}^*, {L^w_{fh}}^*, {L^r_{ml}}^*, {L^r_{fl}}^*$ $, {L^r_{mh}}^*, {L^r_{fh}}^*)$ that solves the household's problem for each types, and $\bar{L} = \frac{1}{2}({L^r_{ml}}^* + {L^r_{mh}}^*)$.

Definition 2. *Sex ratio* is $SR := \frac{1}{2}({L^r_{mh}}^* + {L^r_{ml}}^* + {L^w_{mh}}^* + {L^w_{ml}}^* - {L^r_{fh}}^* - {L^r_{fl}}^* - {L^w_{fh}}^* - {L^w_{fl}}^*)$.

Definition 3. A household i is said to be *specializing in wheat (rice)* production if ${L^w_{mi}}^* + {L^w_{fi}}^* = 0$ (${L^r_{mi}}^* + {L^r_{fi}}^* = 0$).

Assumption 1. $A_w^2 < 3A_r^2$.

2. $\tau > \gamma A_r - \frac{1}{2}(\gamma+1)A_w$ *and* $c_l \leq \min\left(\frac{1}{4}, \frac{\frac{(\gamma-1)^2}{4}A_w^2 - \frac{3(\gamma-1)^2}{4}A_r^2}{\tau - \gamma A_r + \frac{1}{2}(\gamma+1)A_w}\right)$.

For our model to behave well, we need the cost of participating in rice production high enough, and the cost of manipulating sex ratio for l type household low enough. This assumption guarantees that the l type household, conditional on being specializing, will always be better off specializing in rice production than specializing in wheat production. Given this assumption, we have the lemma below

Lemma 1. *In any equilibrium,* ${L^w_{ml}}^* = 0$.

All proofs are in the appendix. Lemma 4 indicates that in any equilibrium, l type households always allocate their male members to produce rice. In words, in the world we model, if a household is not so 'guilty' of being son-preferring, it will put all the resources in the sector favorable to males. Given this, we will have the following proposition

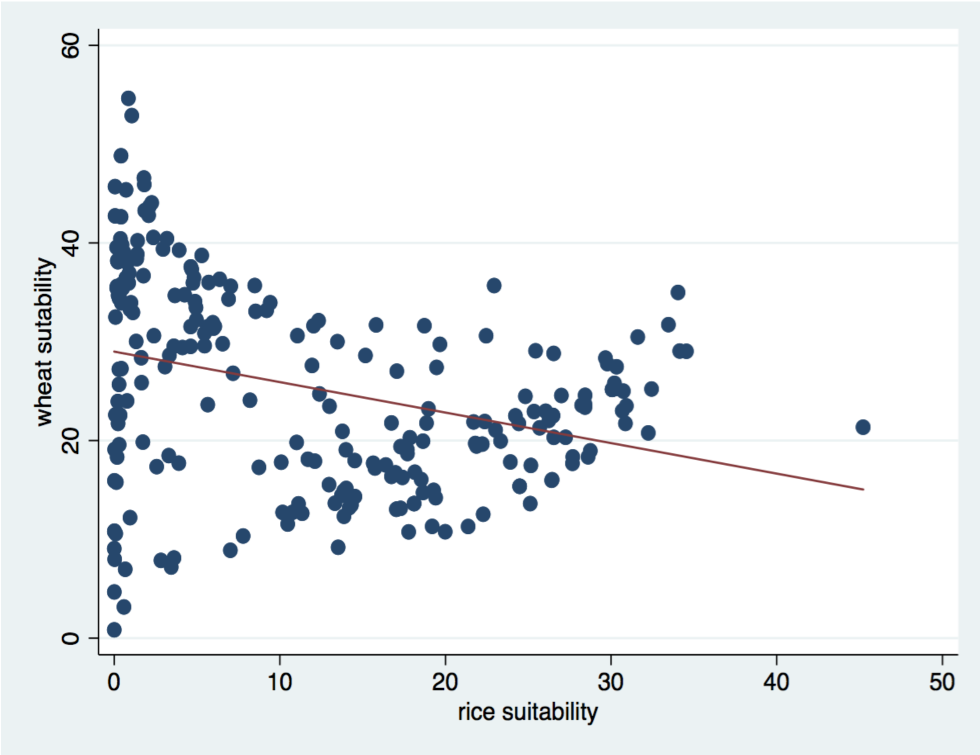


Figure 3.1: Correlation of Wheat and Rice Suitability

Proposition 1. *Suppose A_r changes to A'_r whereas A_w changes to A'_w and we have $A'_r - A'_w > A_r - A_W$. SR increases if $\frac{A'_w - A_w}{A'_r - A_r} < 1$.*

This proposition says that sex ratio is increasing in $A_r - A_w$, if A_r and A_w changes in the opposite direction, or if they change in the same direction, rice suitability changes more than wheat suitability. These conditions are supported by our empirical evidences. Firstly, in our data we observe a negative correlation between rice suitability and wheat suitability[5]. Secondly, if both rice suitability and wheat suitability increase across our sample, it must be the case that it goes from districts relatively not suitable for agricultural to more arable districts. Since in our sample nearly all districts are suitable for wheat whereas rice suitability varies more, it gives justification for the assumption that rice suitability changes more than wheat suitability when they both increase or decrease. As mentioned before, there could be a potential concern in regard of the assumption that only males generate positive externalities in producing rice. To address this concern, we relax the assumption

[5]The scatter plot is presented in Fig 3.4.

and assume that both males and females benefit from cooperation ($\bar{L}$). The results are given in appendix B.

We then model inter-generational transmission of values, which is defined by distinct costs of manipulating sex ratio[6]. Suppose we have discrete time periods $t \in \{0,1,2,...\}$. The measure of l type households at time t is p_t and type h is $1-p_t$. $p_0 > 0$ is a given constant. Denote the income of household i at time t by C_i^t. For example, recall that a household i specializing in wheat gets $F^{w\prime}_{L^w_{mi}} L^w_{mi} + F^{w\prime}_{L^w_{fi}} L^w_{fi}$ in the static model. If the household has the same labor input at a time period t in the dynamic model, we have $C_i^t = F^{w\prime}_{L^{wt}_{mi}} L^{wt}_{mi} + F^{wt\prime}_{L^{wt}_{fi}} L^{wt}_{fi}$. Assume for $t > 0$ we have $p_t = p_{t-1} + \frac{C_t^l - C_t^h}{C_t^l + C_t^h}(1-p_{t-1})$. So the richer type will attract the other type to join them in the next generation (e.g. through marriage). Let the sex ratio at time t denoted by SR_t. We have the following results:

Proposition 2. *Given any $p_0 > 0$, we have both SR_t and p_t weakly increase in t.*

Those households which are less morally constrained will be able to obtained a higher income, hence will thrive in the long run compared to the others. We then discuss how crops suitability will have an impact on this.

Proposition 3. *Given any values of c_l and c_h and $p_0 \geq \frac{1}{2}$, SR_t and p_t are both weakly increasing in $A_r - A_w$ for any t.*

The proof of this proposition is dropped, since it is only a copy of the proof from proposition 6. From proposition 6 we know that p_t is increasing in t, and does not converge to any value less than 1, there should be a time period t where $p_t > \frac{1}{2}$. If we can treat this time period as period 0, then the assumption that $p_0 \geq \frac{1}{2}$ will not affect the generality of our results.

The above results paves the way to link cultural values to the underlying economic variables. In a society where the method of production is in favor of males, those with son preference (or other reasons lowering costs of manipulating sex ratio) will eventually crowd out those without it, making such feature a ubiquitous

[6]One way to think of this is that the 'cost' of manipulating sex ratio could be the 'psychological cost' or the 'moral cost'. Those have low cost (e.g. stronger son preference and lower morality) feel less 'painful' when having to infanticide.

cultural phenomenon. A concern is that our main model of agriculture may fail to be applied to the demographics of a industrialized economy. However, according to above results, the distortion of sex ratio will persist to some extend. Suppose the agricultural sectors in our model is suddenly replaced by industrial sectors at some t, as long as there is a (minor) difference between female and male's productivity, probably caused by discrimination at work place or in education, the values of son preference should still matter and the sex ratio will still be distorted. Most importantly, our model provides a possible mechanism that can explain the regional variations in son preference that we observe in the data for China, as it happens that traditionally the major crops in south China (where son preference is stronger) is rice, while the major crops in north china (where male-to-female sex ratio is relatively lower) is wheat. In the next part of this paper, we link our model with the data through empirical analysis.

3.3 Estimation Specification and Data Construction

3.3.1 Estimation Method

To empirically test the effect of rice and wheat production on male-to-female ratio at birth, we estimate the following specification:

$$SR_i = \alpha + \beta(rice - wheat)_i + \lambda X_i + \varepsilon_i \tag{3.10}$$

where SR_i is the male to female ratio at birth in prefecture i from 2010 Population Census Data. $(rice - wheat)_i$ is the difference between rice and wheat suitability for cultivation in prefecture i, which are extracted from Caloric Suitability Indices (CSI) developed by Galor and Özak (2016b). X_i is a vector of geographical variables. The empirical analysis comes through cross sectional comparison. In particular, we would like to see whether there is significant positive correlation between male-to-female sex ratio at birth and the differences of the suitability in growing rice versus growing wheat, across prefectures in mainland China. In our baseline regressions, we use prefecture-level data and consequently we have 238 observations in total.

We also construct province-level data aggregated from prefecture data and end up with 27 observations. The reason that we have prefecture level data but still want to analyse results at the province level is because later on when we move on to the gender norm analysis at the individual level, the data that we use are unfortunately only representative at the province level.

3.3.2 Crops Suitability Data

For Geographic Information System (GIS) data on rice and wheat suitability, we apply Caloric Suitability Indices (CSI) developed by Galor and Özak (2016b)[7] which is available at the prefecture level. The suitability captures crops' potential output based on the land which is suitable for rice/ wheat cultivation in the post-1500 era, and we exploit the estimates based on agro-climatic conditions that are unaffected by human intervention. We simply choose prefecture as our basis unit of analysis as there could be more variations that exists within single province.

We also test the effect of rice and wheat production on male-to-female ratio at birth in province level. It enhances our baseline results using prefecture level data. In this case,we aggregate the rice and wheat suitability into province level. There are two calculation methods for rice and wheat suitability at the province level. One is that we take the average across prefectures for each provinces. The other is that we use rice and wheat suitability of the provincial capital as proxy for the overall suitability of the provinces.

Figure 3.1 and Figure 3.2 show the provincial suitability of growing rice and wheat using the first method which is used for constructing provincial level crops suitability, respectively [8]. As we can see, south/southeast China is relatively more suitable of growing rice (e.g. the blue area in Figure 3.1), while north China is more

[7]CSI estimates potential yields of crops suitable for cultivation in post 1500 era for each cell of size 5× 5. Full descriptions of the index can be found at: https://ozak.github.io/Caloric-Suitability-Index/

[8]Notice that we exclude 5 provinces in mainland China: Xinjiang, Xizang, Neimenggu, Jilin and Heilongjiang. There are two justifications. Firstly, historically most of these regions are not considered as the main grain-producing-area, especially before modern China. Second, for regions such as Xinjiang, Xizang and Neimenggu, they are municipality that minorities live. For minorities, one-child policy does not apply (e.g. the majority of them can have two children), which could bias our results.

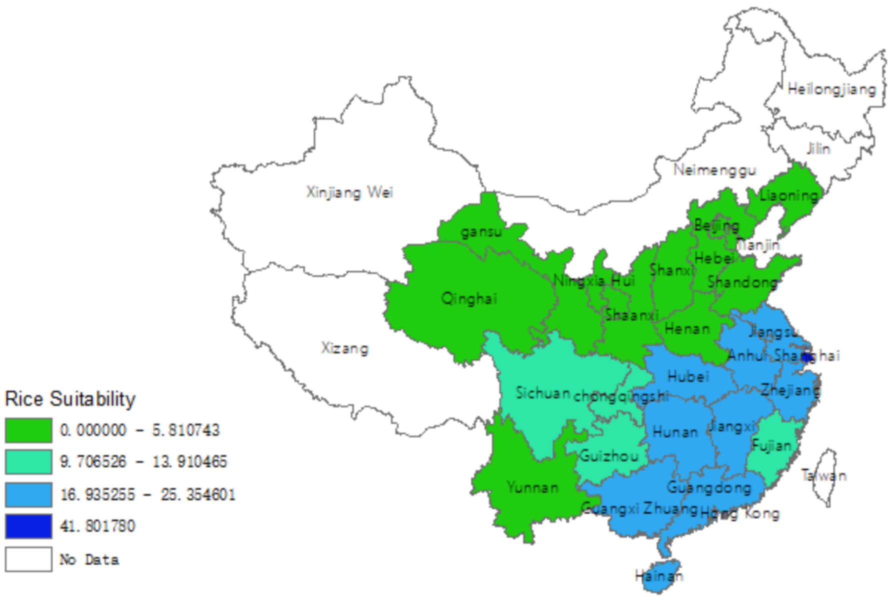


Figure 3.2: Suitability of Growing Rice

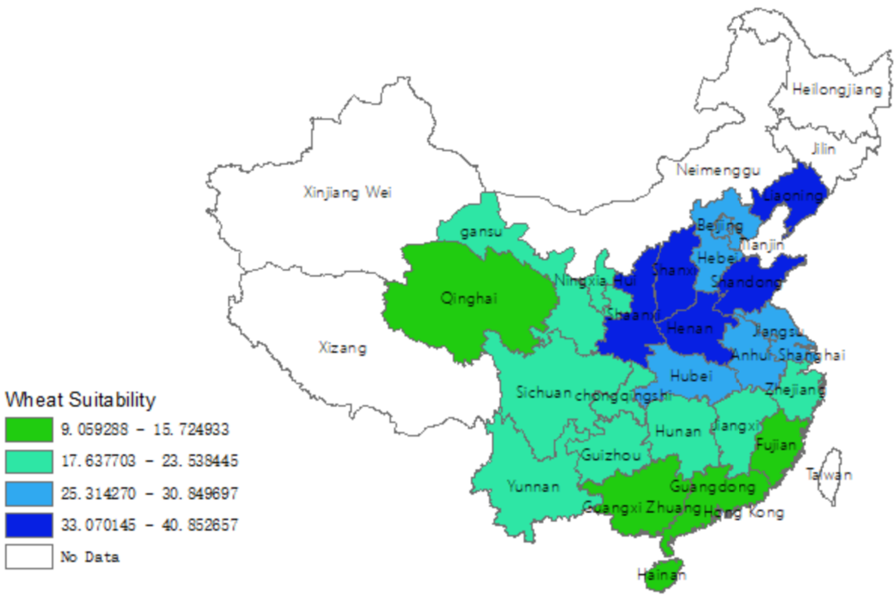


Figure 3.3: Suitability of Growing Wheat

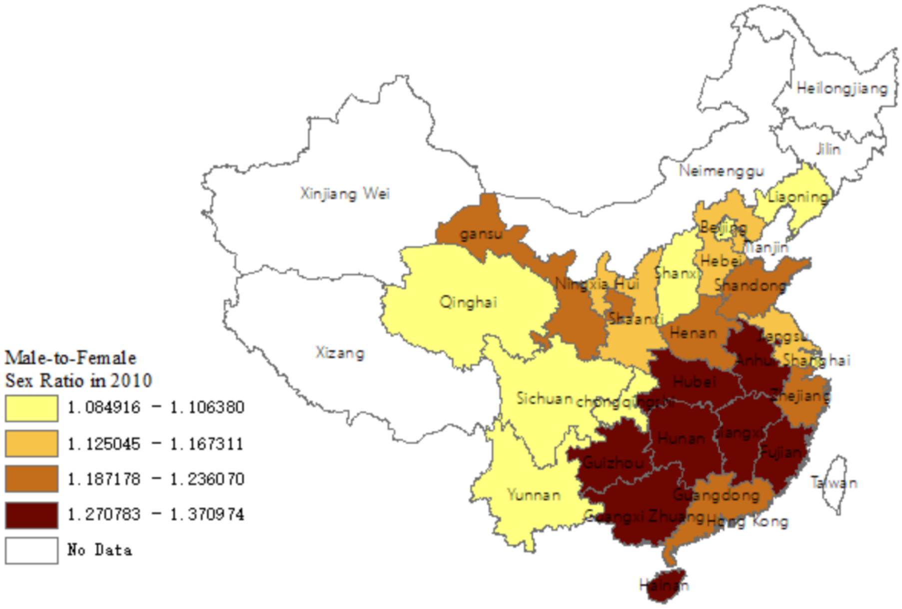


Figure 3.4: Male-to-Female Sex Ratio at Birth in 2010 1% Population Census

suitable of growing wheat (e.g. the dark blue area in Figure 3.2). These figures are consistent with the fact that traditionally the major crops in south China is wet rice, while people in the north usually grow wheat[9]. We only include the area which is regarded as 'China proper' and exclude peripheral areas. In history people in those peripheral areas mostly lived a nomadic life, so agricultural suitability should have a limited impact on them. Also, the problem of missing data for some control variables is very severe in some of those areas.

3.3.3 Population Census Data

We use 2010 Population Census Data to construct male-to-female ratio at birth. The census data provides the number of registered new-born male infants as well as the number of new-born female infants, both at province level and prefecture level, which allows us to calculate the male-to-female ratio[10] at birth. As shown in Figure 3.3, in 2010, male-to-female sex ratio (at birth) are much higher in south China than

[9]This is even reflected in traditional food culture. In north China, people prefer noodles (which is usually made from wheat) over rice, while it is the opposite in south China.

[10]Sex ratio equals the number of registered new-born male infants over the number of registered new-born female infants.

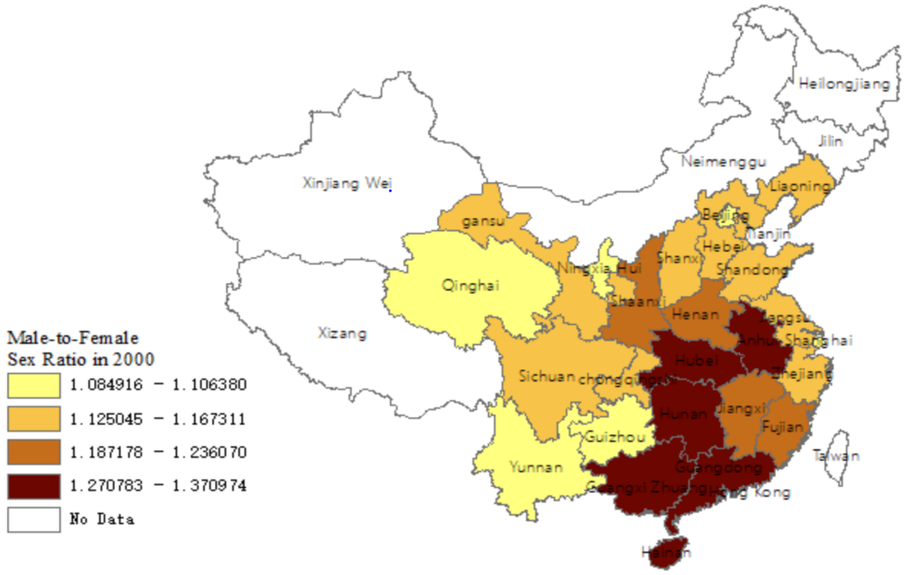


Figure 3.5: Male-to-Female Sex Ratio at Birth in 2000 Population Census

the north (e.g. the dark red area). Together with the figures that we show above, one can see a clear pattern that correlates rice suitability and sex ratio. Moreover, for regions that are more suitable of growing wheat than growing rice, the ratio is still larger than one, indicating a general son preference across China, which is also consistent with the prediction of our model as well as the literature (Alesina et al., 2013). In case we may have a bias sample in one particular year, we also apply 2000 population census data as robustness check. The sex ratio is calculated as the same way as 2010. The data is displayed in Figure 3.4. The pattern is almost identical to the one that uses 2010 census data, suggesting that it is a general phenomenon that son preference is more severe in the south than in the north in China.

So far it seems to be that rice suitability is positively correlated with son preference in China. Nevertheless, recall that wheat agriculture itself would also lead to unbalanced sex ratio (due to the fact that it also requires physical work which men have absolute advantages). Therefore, when we try to explain the regional differences in son preference, it is important to not just look at rice/wheat suitability alone. Instead, according to our model, what we should look at is the difference between rice suitability and wheat suitability ($A_r - A_w$). In addition, although the suitability of growing rice/wheat is a 'natural variable' and thereby should be very unlikely to suffer from endogeneity problem, some other geographical variables may still lead to omitted variable bias if we do not control for them. For example, it is possible that rugged areas is more suitable for planting one crop than the other, and rugged areas may be less-developed so have a higher sex ratio. Hence, in order to have more robust results, in section 4 we regress sex ratio on $A_r - A_w$, controlling for the following geographical variables, which are provided in Chen et al. (2020) at the prefecture level:

Log Distance to Coast This variable is measured as the (log) distance between a prefecture's centroid to the closest point on the coast. Throughout the province level analysis, we simply use the provincial capital's distance to coast as proxy for the overall distance between the provinces and the coast. This variable could potentially affects the suitability of growing rice/wheat, while also being correlated with son preference. Distance to coast may affect the crops cultivation as well as the irrigation system. Also, prefectures nearby coast are likely to be beneficiaries of western knowledge and cultural beliefs, which affects the son preference.

Terrain Ruggedness This index is constructed by calculating the difference in elevation between adjacent cell grids[11]. At the province level, we follow the same methods as when we calculate the provincial rice/wheat suitability. That is, we use both method (1) and (2) to calculate this index at the province level.

Overall Agricultural Suitability This variable measures the overall suitability of land for agriculture, and again could be correlated with rice/wheat suitability and son preference simultaneously. Both method (1) and (2) are applied when aggregating this variable (from prefecture level) to province level. Descriptive statistics of these variables are reported in Table 3.1. We present the regression results in section 4. In the next subsection 3.4, we move on to talk about the data that we use for gender norm at the individual level.

[11] As explained in Chen et al. (2020), "*The Digital Elevation Model (DEM) is typically spaced at the 90 square-metre cell grids across the entire surface of the earth on a geographically projected map.* "

Table 3.1: Descriptive Statistics

Variables	Obs	Mean	Std.Dev.	Min	Max
sex ratio 2000	238	1.19	0.12	1.02	1.7
sex ratio 2010	239	1.18	0.08	1.02	1.4
rice-wheat suitability difference	239	-13.14	17.16	-53.79	23.86
log distance to coast	239	12.51	1.14	9.73	14.25
Terrain Ruggedness	239	0.18	0.15	0.005	0.80
Agriculture Suitability	239	3.04	0.72	0.55	4.84

Note: The descriptive statistics is summarised based on prefecture-level variables.

3.3.4 Gender Norm Measurement

In the second part of our empirical analysis, we further explore whether traditional agricultural activities shape gender norm today. For measurements of gender norm at the individual level, we apply data from 2012 wave Chinese General Social Survey (CGSS). CGSS is the first national continuous social survey project in China implemented by academic institutions in 2005, which could be considered as the Chinese counterpart of the General Social Survey (GSS) in the U.S. It aims to monitor systematically the changing relationship between social structure and quality of life in both urban and rural China. To do so, CGSS collected information in various areas of individual's demographic background, attitudes and behaviours, and socioeconomic status, among others. The sample covers both rural and urban areas from 28 provinces. In its 2012 wave, respondents are asked to choose from options "Strongly disagree", "Disagree", "Agree", "Strongly agree", and "Neither agree nor disaggree" for the following statements: (1) 'Man has greater ability than woman' (2) 'The role of man is to earn money while the role of woman is to take care of the family' (3) "For woman, being housewives gives the same sense of achievement as being in the labour force". Based on these information, we construct our measures of gender norm following exactly the same procedure as in Alesina et al. (2013). To be more precise, we first omit observations in which the respondents answer 'I do not know' or 'Neither agree nor disaggree'. We then code 0 if respondents answer 'Strongly disagree' or 'Disagree', and 1 for 'Agree' or 'Strongly agree'. We then regress this binary variable on $A_r - A_w$ based on the province that the individual comes from, using both OLS and probit estimator. Other geographical variables mentioned in subsection 3.3 are also included as controls.

Alternatively, we can rank the answers from 'Strongly Disagree' = 1 to 'Strongly agree' = 4 and apply ordered probit estimator, which is a generalization of the probit model to the case of more than two outcomes of an ordinal dependent variable. Results from using this method are also presented as robustness check.

3.4 Results: Sex Ratio

3.4.1 Prefecture Level

In this section, we regress sex ratio on $A_r - A_w$ at the prefecture level. The results are shown in Table 3.2. In our sample we have in total 238 prefectures, which in principle is large enough to provide more consistent estimations than the province level regressions. The correlations between sex ratio and $A_r - A_w$ remain significantly positive.The baseline regression shows that one standard deviation increase in $A_r - A_W$ is associated with an increase in sex ratio by around 3.26%. In other words, at the prefecture level we observe that son preference is stronger in places that the differences between rice and wheat suitability are larger.

We also use A_r/A_w instead of $A_r - A_w$ as our independent variable for robustness check at the prefecture level. Results are reported in table 3.3. From the table all coefficients remain significantly positive. That is, son preference is positively correlated with the ratio between rice and wheat suitability.

3.4.2 Province Level

We then analyze results at the province level. Recall that our data on rice and wheat suitability (as well as control variables such as terrain ruggedness and overall agricultural suitability) are collected at the prefecture level. We thereby need to aggregate the data to province level through the two methods that are described in subsection 3.2.

Table 3.4 presents results using method (1), in which we take the average rice and wheat suitability (as well as terrain ruggedness and overall agricultural suitability) across prefectures for each provinces. The key variable of interest is the

Table 3.2: Sex Ratio and the Ratio Between Rice and Wheat Suitability at Prefecture Level

	(1) sexratio2000	(2) sexratio2000	(3) sexratio2010	(4) sexratio2010
A_r/A_w	0.0687***	0.0668***	0.0382***	0.0347***
	(0.014)	(0.016)	(0.009)	(0.010)
Distance to coast		0.0226***		0.0108*
		(0.008)		(0.006)
Ruggedness		-0.2092***		-0.1428***
		(0.048)		(0.039)
Agriculture suitability		0.00194		0.0044
		(0.014)		(0.009)
Observations	238	238	238	238
Adjusted R^2	0.096	0.143	0.061	0.093

Standard errors in parentheses

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

Note: Dependent variable is the number of registered new-born male infants over the number of registered new-born female infants in population census data. A_r/A_w is the ratio between the suitability in growing rice versus growing wheat, measured at prefecture level. coast_dist, tr_index, ari_suitability are controls of log distance to coast, terrain ruggedness and overall agricultural suitability respectively. Robust standard errors at province level are in parentheses.

difference between rice suitability and wheat suitability ($A_r - A_w$). Column (1) and (2) regress male-to-female sex ratio in 2000 1% population census on $A_r - A_w$ without and with controls respectively, whereas column (3) and (4) use the sex ratio from 2010 1% population census. As we can see, in all columns we have significantly positive correlation between $A_r - A_w$ and sex ratio, which is consistent with our theory that sex ratio would be more unbalanced in regions where the differences between rice and wheat suitability are larger.

Nevertheless, it is worth to mention that although the coefficients that we get throughout this paper are to some extent the 'casual effects' because our independent variable is constructed by meteorological and geographic factors and hence unlikely to suffer from endogeneity problem, the numbers themselves (e.g. 0.0023 in column (1)) does not have any economic meaning. Therefore, we prioritize our analysis on the sign of the correlation.

Table 3.3: Sex Ratio and the Ratio Between Rice and Wheat Suitability at Prefecture Level

	(1) sexratio2000	(2) sexratio2000	(3) sexratio2010	(4) sexratio2010
A_r/A_w	0.0687***	0.0668***	0.0382***	0.0347***
	(0.014)	(0.016)	(0.009)	(0.010)
Distance to coast		0.0226***		0.0108*
		(0.008)		(0.006)
Ruggedness		-0.2092***		-0.1428***
		(0.048)		(0.039)
Agriculture suitability		0.00194		0.0044
		(0.014)		(0.009)
Observations	238	238	238	238
Adjusted R^2	0.096	0.143	0.061	0.093

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Dependent variable is the number of registered new-born male infants over the number of registered new-born female infants in population census data. A_r/A_w is the ratio between the suitability in growing rice versus growing wheat, measured at prefecture level. coast_dist, tr_index, ari_suitability are controls of log distance to coast, terrain ruggedness and overall agricultural suitability respectively. Robust standard errors at province level are in parentheses.

In table 3.5, we use method (2) to obtain $A_r - A_w$ at the province level. That is, we use data from provincial capital as proxy for each provinces. The results are almost identical to the ones in table 2, suggesting that the significant positive correlation between $A_r - A_w$ and sex ratio is robust to the method that we use to aggregate data from prefecture to province level. In addition, although our model suggests us to look at the correlation between $A_r - A_w$ and sex ratio, we also consider regressing sex ratio on the ratio between rice and wheat suitability (A_r/A_w) as robustness check. Table 7 reports regression results using A_r/A_w, which are calculated through method (1), whereas Table 8 presents results using provincial capital's A_r/A_w. We again get significant positive correlations, which leads to the same conclusion that sex ratio would be more unbalanced in regions that are more suitable of growing rice relatively to growing wheat.

To summarize, in support to our theory, we empirically show that in China

Table 3.4: Sex Ratio and the Difference in Rice and Wheat Suitability Using Average across Prefectures for Each Provinces

	(1) sexratio2000	(2) sexratio2000	(3) sexratio2010	(4) sexratio2010
$A_r - A_w$	0.0023**	0.0025**	0.0018**	0.0020***
	(0.001)	(0.001)	(0.001)	(0.001)
Distance to coast		0.0138		0.0101
		(0.018)		(0.015)
Ruggedness		-0.1933*		-0.1142
		(0.109)		(0.080)
Agriculture suitability		0.0001		0.0073
		(0.034)		(0.019)
Observations	27	27	27	27
Adjusted R^2	0.177	0.157	0.220	0.194

Standard errors in parentheses

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

Note: Dependent variable is the number of registered new-born male infants over the number of registered new-born female infants in 1% population census data. $A_r - A_w$ is the difference of the suitability in growing rice versus growing wheat, which is aggregated from prefecture level to province level using the average across prefectures for each provinces. coast_dist, tr_index, ari_suitability are controls of log distance to coast, terrain ruggedness and overall agricultural suitability respectively. Conventional standard errors are in parentheses.

the regional variation of son preference is significantly positively correlated with the difference of the region's suitability of growing rice versus growing wheat. In other words, our evidence suggests that traditional agriculture practices affect the economic variables today. In the next section, we extend our analysis to individual level investigating the effects of agricultural practices on beliefs.

Table 3.5: Sex Ratio and the Difference in Rice and Wheat Suitability Using Provincial Capital as Proxy for Each Provinces

	(1)	(2)	(3)	(4)
	sexratio2000	sexratio2000	sexratio2010	sexratio2010
$A_r - A_w$	0.0023**	0.0019*	0.0020***	0.0018***
	(0.001)	(0.001)	(0.001)	(0.001)
Distance to coast		0.0061		0.0062
		(0.018)		(0.014)
Ruggedness		-0.1425		-0.0805
		(0.094)		(0.079)
Agriculture suitability		0.0080		0.0072
		(0.027)		(0.013)
Observations	27	27	27	27
Adjusted R^2	0.191	0.147	0.307	0.259

Standard errors in parentheses

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

Note: Dependent variable is the number of registered new-born male infants over the number of registered new-born female infants in 1% population census data. $A_r - A_w$ is the difference of the suitability in growing rice versus growing wheat, which uses provincial capital as proxy for each provinces. coast_dist, tr_index, ari_suitability are controls of log distance to coast, terrain ruggedness and overall agricultural suitability respectively. Conventional standard errors are in parentheses.

3.5 Results: Gender Norm

In this section, we check whether son preference translates into gender norms that discriminate females, which is most relevant to what we have been looking at in this paper. Specifically, we want to see whether individuals from provinces that $A_r - A_w$ are higher also tend more to have biased gender view on average.

We consider three dimensions of gender norms. Firstly, we look at general gender discrimination against females, which is reflected in statement (1) from subsection 3.4: 'Man has greater ability than woman'. Panel A in Table 3.6 describes the regression results. In column (1) and (2), the dependent variable equals to 0 if the respondents answer 'Strongly disagree' or 'Disagree' towards statement (1), and 1 for 'Agree' or 'Strongly agree'. Column (1) estimates results using OLS estimator while column (2) applies probit estimator. In column (3), ordered probit

Table 3.6: Gender Norm and the Difference in Rice and Wheat Suitability

	(1) OLS	(2) Probit	(3) Ordered Probit
Panel A: Gender Discrimination			
$A_r - A_w$	0.0012	0.0029	0.0024
	(0.001)	(0.002)	(0.002)
Observations	2,589	2,589	2,589
Panel B: Woman's Role in Family			
$A_r - A_w$	0.0029	0.0053	0.0051*
	(0.002)	(0.003)	(0.003)
Observations	1,171	1,171	1,171
Panel C: Labour Force Participation			
$A_r - A_w$	0.0040***	0.0110***	0.0103***
	(0.001)	(0.002)	(0.002)
Observations	1,185	1,185	1,185

Cluster robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Panel A, B and C capture different dimensions of gender norms, and are based on statement (1), (2), and (3) described in subsection 3.4 respectively. Dependent variable in column (1) and (2) are binary variable which equals to 0 if the respondents answer "Strongly disagree" or "Disagree" towards the respective statement, and 1 for "Agree" or "Strongly agree". Dependent variable in column (3) ranges from "Strongly Disagree" = 1 to "Strongly agree" = 4. $A_r - A_w$ is the difference of the suitability in growing rice versus growing wheat, which is aggregated from prefecture level to province level using the average across prefectures for each provinces. All columns include controls of log distance to coast, terrain ruggedness and overall agricultural suitability. All columns use cluster robust standard errors at the province level.

estimator is applied, in which the dependent variable ranks from 1= 'Strongly Disagree' to 4='Strongly Agree'. All estimations use cluster robust standard errors at the province level, since the exogenous variations in these regressions are mainly from the provincial level (rather than the individuals). Results suggest that individuals from provinces that $A_r - A_w$ are higher are on average more likely to agree that man has greater ability than woman, although the positive correlation is statistically insignificant, probably due to the fact that while the exogenous variations come from the province level, we only have 27 provinces in the sample.

Secondly, we look at individuals' view towards woman's role in a family, which is indicated by statement (2) from subsection 3.4: 'The role of man is to earn money while the role of woman is to take care of the family'. This is an important aspect to look at as it could directly affect female labour force participation. Results are showed in Panel B of Table 3.6. Both OLS, probit and ordered probit estimators suggest that biased gender view is more likely to exist among individuals who are from provinces with higher $A_r - A_w$, but the coefficients are barely significant.

Lastly, we look at individuals' attitude towards female labour force participation, which is proxy by statement (3) from subsection 3.4: 'For woman, being housewives gives the same sense of achievement as being in the labour force'. Panel C of Table 3.6 reports the results. The coefficients turn out to be positive and are statistically significant at 1% across three different estimators. That is, individuals who are from higher $A_r - A_w$ provinces are on average more likely to agree that for female, it takes no difference between being housewives and working.

The results seem suggest that rice-planting leads to a discriminating view on female labor participation, while does not cause a general discrimination on women's ability. This is an interesting fact and an explanation can be that since women have comparative advantage on wheat rather than rice, they will be excluded more or less from labor participation in regions suitable for planting rice rather than wheat. So in areas with high $A_w - A_r$, women worked less in history when agriculture dominates and this would become one of the social norms shaping people's ideas today. However, this is only caused by division of labor in agriculture while other aspects of women's ability may still be regarded leading to the insignificant coefficients in panel A. This is only one possible explanation and of course we cannot rule out the possibility that beliefs of people today include many factors besides agricultural practices in history. Overall, our results suggest that gender bias is more severe in regions that the differences between rice and wheat suitability are larger. Traditional agriculture practices could develop the gender norms in a society.

3.6 Discussion of Channels

Throughout this paper we are arguing that one of the channels that the suitability gap between planting rice and wheat will lead to an unbalanced sex ratio is through the higher level of cooperation required by rice production which increases the comparative advantage of males. In this section we argue that it is a reasonable and convincing channel among all possible channels, and rule out some other possibilities that arouse most concerns. For the purpose of justifying our channel through cooperation, two questions should be answered. First, is it true that the a higher rice suitability as well as a lower wheat suitability is associated with a higher level of cooperation? Second, is it true that such cooperation leads to son preference?

Regarding the first question, there is sufficient evidence which shows that there is a correlation between rice production and collectivism. Talhelm et al. (2014) find that people in China who are from regions more favourable to rice than wheat tend to be more collectivism than individualism. They argue that that is very much due to the feature of rice production - rice farmers often shared labor and coordinated irrigation in the planting, which is not the case for wheat. Fan et al. (2021) use this rice theory and show that founders from regions with stronger collectivist cultures (rice production) have more family members as their managers, retain more firm ownership within the family and share the controlling ownership with more family members. We also examine the rice theory using our data by regressing the log of number of genealogy per capita, which is a common-used proxy for the strength of cooperation within kinship network, on the gap between rice and wheat suitability, with the same controls in the baseline. The coefficient is significantly positive (0.007 with sd equal to 0.003) which supports the rice theory.

The second question raises more concerns, since women's participation in agriculture rose greatly in China in recent decades, a phenomenon called feminization of agriculture (De Brauw et al., 2008; De Brauw et al., 2013). Therefore it is natural to question the argument that men benefit more than women from agricultural collaboration. However, in traditional China, agriculture, especially planting crops, was largely regarded as men's job. For females, their tasks were 'inside' the house, including taking care of the children, the household chores etc. Old Chinese sayings such as *nanzhuwai nüzhunei* (men's work centres around outside of the home while women's work centres around inside of the home) reflect this gender division of labour (Meng, 2014). Moreover, the construction of agricultural infrastructure such as irrigation system was a representative public good provided by informal institutions of kinship network as a coordination device in late imperial China (Fei, 1946), in which males played a dominant role (Stone and King, 2018). For these reasons, it is not that natural to exclude collaboration as a channel linking rice production and son preference.

We conduct a mediation test as in Baron and Kenny (1986) to further support our argument and to rule out two other channels that raise most concerns. The first is the possibility that the correlation between rice-wheat suitability and sex ratio is through economic development. For example, regions more suitable for planting rice than wheat may be less developed today, because of relatively more rooted agriculture, so lead to a more unbalanced sex ratio. Call this economic performance channel. The second possibility is that regions having more rice production were more wealthy in ancient time, given rice provides more calorie than wheat, therefore more people there were able to participate in *keju*, the traditional Chinese exam for selecting civil servants, which women were not eligible to take, and son preference could be caused by the benefit of it. Call this human capital channel. For economic performance channel, we use the log of gdp per capital as the mediator while for human capital channel we use the log of the number of jinshi (who were successful in keju exam). Finding a mediator for collaboration is more challenging, since to our knowledge there is no available proxy directly measuring the collaboration in agriculture in traditional China. Instead, we use three proxies for social capital to approximately measure the level of social collaboration, which are the number of genealogy per capita, the number of charitable organisations in 1840 and the number of social organizations including farmers' associations etc. in 2008 following Chen et al. (2020) in which the details of these variables can be found. We conduct principal component analysis (PCA) to generate a single factor mostly preserving the information of the three variables, and use this PCA index as a mediator for collaboration. The PCA factor is associated with an eigenvalue equal to 1.53878, which gives the validity of using it. The results are given in table 3.7.

Table 3.7: Mediation test

	Collaboration	GDP per capita (logged)	Jinshi density (logged)
Effects of $A_r - A_w$	0.00642*	0.00216	0.00701***
on mediators	(0.0037)	(0.0023)	(0.0025)
Indirect effects of $A_r - A_w$	0.00041*	-0.00011	-0.00014*
on sex ratio through mediators	(0.0002)	(0.0001)	(0.0001)
Observations	237	239	239

Note: Dependent variable is sex ratio at birth in 2010. Robust errors at province level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: Descriptive Statistics

Variables	Obs	Mean	Std.Dev.	Min	Max
sex ratio 2000	238	1.19	0.12	1.02	1.7
sex ratio 2010	239	1.18	0.08	1.02	1.4
rice-wheat suitability difference	239	-13.14	17.16	-53.79	23.86
log distance to coast	239	12.51	1.14	9.73	14.25
Terrain Ruggedness	239	0.18	0.15	0.005	0.80
Agriculture Suitability	239	3.04	0.72	0.55	4.84

Note: The descriptive statistics is summarised based on prefecture-level variables.

The results show that the indirect effect through our mediator of collaboration as well as the effect of $A_r - A_w$ on collaboration is positive and significant, which supports our prediction. The sign of the effect through economic performance is negative, which is against the economic performance channel that rice production increases sex ratio through rice. This coefficient is also insignificant and the effect of $A_r - A_w$ on economic performance is positive and insignificant. Lastly, rice production leads to a higher level of human capital in ancient China. However, this indirect effect of $A_r - A_w$ on sex ratio through Jinshi density is significantly negative, which is opposite to the prediction of human capital channel, and more importantly, it implies that if we exclude the effect of this channel, the correlation between $A_r - A_w$ and sex ratio will be more positively significant.

We must admit that the above results by no means show that collaboration is the only way to son preference. There are many other possible channels. For example, it can be the case that rice production is simply more labor-intensive than wheat, so sons were preferred due to physical strength. We cannot rule out all the possible channels, but the purpose of this paper is to show that the collaboration is one of the reasonable explanations, and to provide a theoretical foundation of it.

3.7 Conclusion

The origins of gender bias has been the interest of scholars in social sciences for quite a long time. We explore the composition of the agriculture as one aspect of it, by focusing on the case where some crops are favorable to men especially through collaboration. Our theoretical model suggests that an area will have a higher male-to-female sex ratio whenever the natural environment in history is relatively suitable for planting the crop favorable to men compared to the other and hence more likely to specialize in producing such a crop. Also, we have showed that the existence of such a crop will make families with strong son preference better off than the others, hence more families will become such a type in the long run, causing the persistence of the culture of son preference.

We formally test our predictions by utilizing GIS and population census data. Consistent with our theoretical model, we find a robust positive correlation between the gap of a region's suitability of growing rice versus growing wheat and the region's male-to-female sex ratio at birth. We further look at gender norms at the individual level and find that individuals from provinces with lager gap between rice and wheat suitability also tend more to agree on that man's ability is greater than woman, and that being housewives give the same sense of achievement as working for woman, as well as woman's role is to look after the family. In other words, traditional crops (agriculture) shapes the differences in gender views today. We also perform mediation test as a supporting evidence for the collaboration channel.

Appendix to Chapter 1

A Expanded Historical Background

A.1 Chinese Kinship Systems: A General Overview and Specifics of the Sample Population

The Chinese kinship system is recognized as one of the most intricate and extensive kinship systems worldwide. It is characterized by a hierarchical, patrilocal, and agnatic structure, and its long-term impact on development remains a subject of debate. Weber (1951) suggested that Chinese kinship served as a social safety net, assisting individuals in ancient China during economic hardships. However, this support system also had the unintended consequence of dampening individual motivation for upward mobility, thereby impeding the development of an urban status class. Furthermore, the functions of Chinese kinship, such as serving as informal courts and providing shelter, undermined the establishment of formal institutions.

On the other hand, numerous studies in recent decades have underscored the crucial role of kinship in China's industrialization. The presence of unconditional trust within kinship groups has facilitated a distinct form of capitalization, characterized by the prevalence of family firms (Redding, 2013; Greif and Tabellini, 2010b). Additionally, abundant evidence reveals that kinship organizations in China, particularly in the South, actively engage in collective activities, effectively addressing coordination problems, providing public goods, and distributing resources among clan members. This phenomenon holds true even in Northeast China, an area historically considered less influenced by lineages or clans than other parts of China (Cohen, 1990; Zheng, 2001; Szonyi, 2002; Campbell and Lee, 2008).

The significance of kinship or clans in resource distribution is particularly evident in the context of male clan members' marriage. Since securing a spouse for male clan members necessitated resource allocation due to the highly skewed sex ratio, decisions regarding marriage priorities were likely made collectively within kinship groups (Chen et al., 2014). Furthermore, when parents with daughters assessed prospective husbands, they considered not only the individual but also their position within the clan, as it determined the extent of resources they could access from their clan (Campbell and Lee, 2011).

The concept of kinship in China is far from being immutable and has undergone significant transformations over time. The traditional Chinese kinship system, as we currently understand it (as a corporate entity bound by a shared ancestor), did not fully emerge until the early 11th century during the Song dynasty (Johnson, 1977). During the late Qing dynasty, which corresponds to the time span covered by my data, patriarchal kinship reached its zenith, serving as the building block of society. This period witnessed a remarkable increase in the number of known clan genealogies, which served as symbols of clan/lineage unity. While there were only 485 compiled genealogies during the preceding Ming dynasty, there were $12,726$ compiled genealogies during the Qing dynasty (Greif and Tabellini, 2010b). This substantial gap cannot be solely attributed to the Qing dynasty's proximity to the present era. Also, as population growth in the Qing dynasty intensified economic stress on ordinary people, kinship became increasingly vital. An individual's position within their patriarchal kinship group came to significantly determine their fate (Lee et al., 1997).

The sample population in my data comes from Liaoning (formerly known as Fengtian) Province in Northeast China. In North China, institutional kinship systems are believed to play a less significant role than they did in the South. However, kinship groups in the North were still important in the Qing dynasty. As Cohen (1990) documents, kinship groups in the North were organized on a smaller scale than their Southern counterparts, being commonly referred to as lineages instead of clans as in the South; however, these Northern kin groups also held common proper-ties and served as essential social platforms for various forms of mutual assistance. His research also highlighted that in the South, an individual's power within his kin group was significantly influenced by his property, wealth, social standing, and political status. In contrast, in the North, the disparities within kin groups were more dictated by established genealogical hierarchies.

The individuals in my dataset come from the *Baqi*, also known as *The Eight Banners*, social group. As mentioned in Section 3.1, the sampled individuals primarily consisted of hereditary peasants and descendants of Han immigrants from North China Plain, even though the popular perception often associates members of the Eight Banners with elite Manchurians. The immigrants maintained numerous rituals related to kinship organizations. Their lineages exerted significant influence on the daily affairs of the people, though they were less independent from the government compared to their Southern counterparts (Ding et al., 2004). Several studies (Bengtsson et al., 2004; Lee et al., 1997; Campbell and Lee, 2008; Campbell and Lee, 2011) have empirically documented this influence within this group of people.

A.2 Low Male Nuptiality

As one of the most influential social theorists in history, Malthus (1986) posited that ancient China was a society predominantly governed by positive checks and experienced a prevalence of universal and early marriage. According to his assertions, the extraordinary encouragements that have been given to marriage, which have caused the immense production of the country to be divided into very small shares and have consequently rendered China more populous, in proportion to its means of subsistence, than perhaps any other country in the world. However, subsequent studies, such as the work by Lee and Feng (1999), have found this argument to be incomplete. They highlight the presence of excess female infant and adult mortality, as well as disparities in the customary age of marriage for males and females, which have resulted in an imbalanced marriage market along gender lines. Consequently, ordinary Chinese males, at least in Qing dynasty, have faced significant difficulties in finding suitable marriage partners. As Ownby (2002) states, poorer men had to delay their marriages by six years in comparison with richer men, and that twenty-five percent of men were unable to marry at all.

The low nuptiality among males also led to social security issues and even posed a threat to the stability of the empire. Poor and unmarried males served as a major constituency for violent revolt in a number of major uprisings. As a notable example, historians attribute the *Nian Rebellion* of 1851-1868, which took place in the North China Plain with an overall male-to-female ratio of 129 men for every 100 women, resulting in casualties exceeding $100,000$, alongside numerous other instances of violence, to low nuptiality rates among males in that region (Hudson and Den Boer, 2002).

Due to the surplus of males and scarcity of females, even for those who were fortunate enough to get married, the process of marriage entailed a significant financial burden, as grooms were required to pay a substantial bride price. This economic strain particularly affected peasants who often faced difficulties in meeting their financial demands. The ability of a man to secure a wife relied heavily on the support provided by his family and clan, as poverty served as a primary obstacle to marriage (Jiang et al., 2015). The marriage market was full of friction, and the process of searching and matching suitable partners was heavily dependent on one's family. Individuals had limited agency in choosing their own marriage partners, as parental or familial decisions held a greater significance. A fundamental rule in searching for a suitable match was to align with another household of similar social and economic status, known as ”men dang hu DUI.” Households typically engaged their social networks and employed the assistance of matchmakers to conduct the search for a suitable spouse, especially ensuring that the bride's father's status matched the groom's father's status [12].

Due to the unique feature of the marriage market in 19th century China and the significance of kinship within it, the marriage of males serves as a particularly suitable subject for studying the influence of patrilineal kinship when compared to other available variables. This is because male marriage is directly linked to resource allocation only during the years when a male is actively participating in the

[12] See Baber (1934) and Hamon and Ingoldsby (2003).

marriage market. In contrast, other variables have their limitations. For instance, mortality rates may reflect the care provided by a mother or other female relatives, and this influence cannot be fully isolated, even when controlling for the number of female relatives during the observation period, as an individual's mortality can be influenced by the care they received during childhood.

A.3 Port Opening in Niuzhuang

Following the First Opium War in 1842, the imperial government entered into a series of unequal treaties with the major industrialized countries of the Western world. These treaties compelled the government to open designated coastal cities, known as Treaty Ports, to global trade and foreign investment. This dramatic shift from the previous closed-door policies led to significant shocks to the local agrarian economy, particularly in the regions surrounding the treaty ports, where international trade and foreign investment had been virtually nonexistent until then.

In 1858, due to their military defeat in the Second Opium War, the imperial government was compelled to sign the *Treaty of Tientsin* with four Western countries: the Russian Empire, the Second French Empire, the United Kingdom, and the United States. This treaty included a provision for the opening of *Niuzhuang* in Liaoning for international business. However, it was *Yingkou*, which is also located in the same district but approximately 50 km away from the city of Niuzhuang, that effectively served this purpose. The preparations for the port opening were not carried out until 1860[13]. In fact, the early preparations, including strengthening market regulation, approving the admission of missionaries, granting the construction and occupation of a new dock to the West, and more, were carried out by local governors without the official agreement of the involved foreign governments. The treaty does not explicitly state which location, Niuzhuang City or Yingkou Port, was supposed to be opened. However, in 1861, Thomas T. Meadows, the first British consul of Niuzhuang, conducted an on-the-spot investigation in the Niuzhuang district. He suggested that the British consulate should be located in Yingkou, thus resolving

[13] Since my data is at a three-year interval, the 1858 sample covers the years 1858-1861, so in my sample, treatment began in 1858.

the issues surrounding the opening of Yingkou with the empire's local governor[14].

Although Yingkou was the port that was opened, it was often referred to as Niuzhuang Port in many texts, possibly because it held significant importance within the Niuzhuang district during that period. In order to avoid confusion and maintain consistency with existing references, this paper will refer to the opened port as Niuzhuang Port rather than Yingkou Port.

The opening of the port quickly brought prosperity to the port area. Many multinational companies, including Swire, Jardine Matheson, Mobil, Arnhold, and others, established their branches in Niuzhuang Port. In 1861, immediately after the opening, a total of 34 international merchant ships berthed in Niuzhuang Port. This number increased to 271 just four years later (He, 1989). The opening also had a profound impact on the well-being of the residents in the port area and nearby districts. As stated by Agassiz (1894), poverty, in the sense of actual want, is hardly known, except in time of famine or flood; and the fact the great numbers of workmen resort to Newchwang[15]...shows that there is no lack of employment for those who look for it. Undoubtedly, the port opening event constituted a significant shock to the local economy.

B Additional Data Description

B.1 Original Source

The CMGPD-LN data originates from the Qing government's population registers focused on the Eight Banner population. These foundational records are housed in the Liaoning Provincial Archives. The compilation and organization of the data was undertaken by a research team from the Hong Kong University of Science and Technology, led by James Lee and Cameron Campbell.

The imperial government rigorously registered the population under the Eight Banner system in Northeast China, seeing them as closely attached to the royal family and ruling elite, although this perceived connection was more theoretical

[14] These historical events are well documented in Zhang (2020).
[15] Newchwang refers to Niuzhuang under the Wade–Giles romanization system for Mandarin.

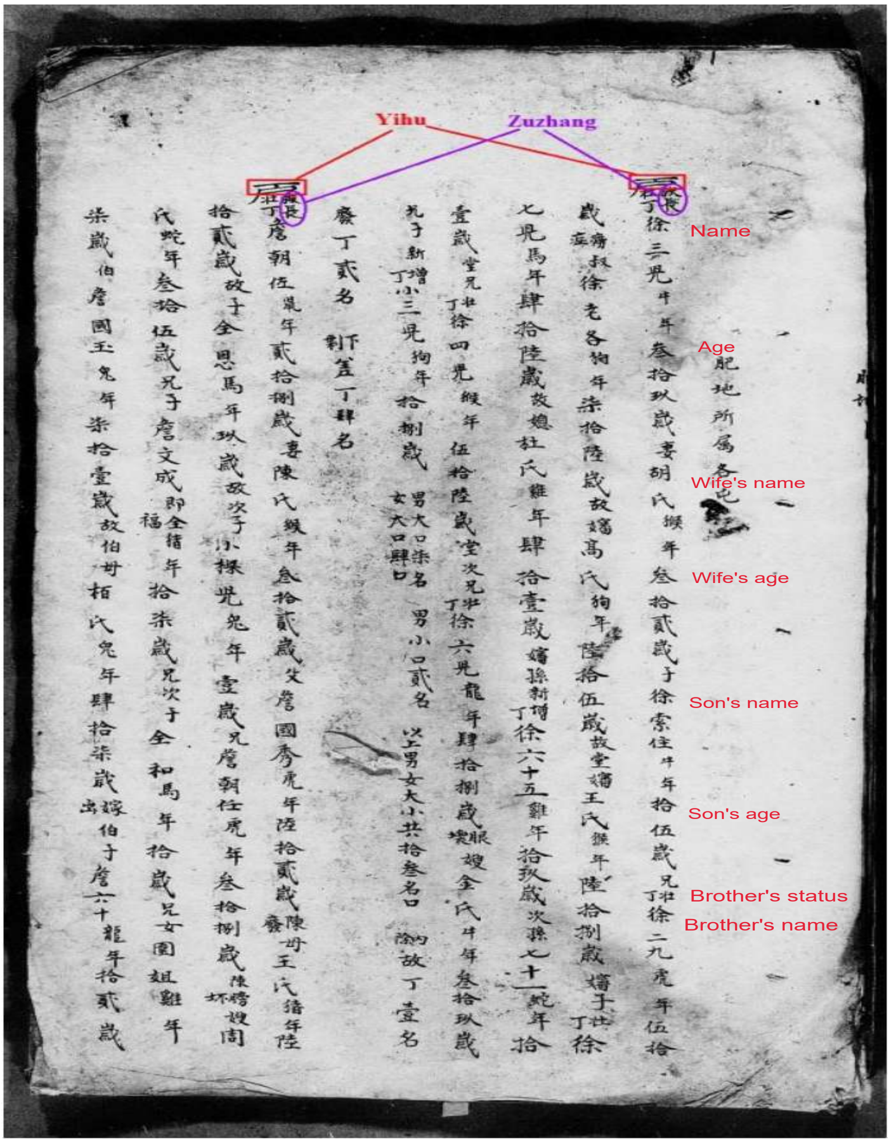


Figure B.1: A sample page of a register. Source: Lee et al. (2010)

than real, with many in the sample having just an ordinary social status.

The basic unit in most registrations is either a household group or a residential household, and these registers adhere to a mostly consistent format. Figure B.1 shows a sample page from a 1783 population register. The basic unit shown on this page is a household group (Yihu), and it shows two such groups, though the second is only partially covered. Each household group's record begins with details about the household head (Zuzhang) and then proceeds with data on the remaining household members. In the figure, I annotate the first column's details in English[16],

[16] Ages are in Chinese reckoning system (*sui*).

with the rest of the page following a comparable structure.

B.2 Descriptive Statistics

Table B.1 provides summary statistics for key variables. Observations of unmarried males take up around 43% of all the observations of males, and not surprisingly, they are younger and less central compared to the average of all. An unmarried male on average has centrality (among male relatives) equal to around 1. An example of such a case is having his father and four uncles alive. As discussed in Appendix B.4, this is subject to be underestimated due to missing data.

The average size of kin groups appears to be relatively small, approximately around 10 when females are included and 4 if not. This can be attributed to the existence of numerous small-sized kin groups, in addition to the impact of missing data. Remarkably, the maximum size reaches 114 when females are included. As a robustness check, Appendix E.3 presents results obtained by excluding individuals from kin groups with a low number of unmarried males.

B.3 Additional Details of Kinship Network Construction

In this section, I provide an explanation of the definitions used for kinship networks and (decay) centrality as adopted in this paper.

As mentioned earlier, the dataset assigns a unique ID number to each individual and records the ID numbers of their close relatives, including father, mother, wife, husband, and grandfather. This information enables the construction of kinship networks by establishing connections between individuals and their respective relatives.

The empirical analysis in this research focuses on examining the impact of patrilineal kinship. Therefore, I begin by constructing patrilineal kinship family trees, which solely consist of male members. By utilizing all father-son relationships within a lineage, it is possible to reconstruct the complete patrilineal descent. In each time period, I consider all recorded father-son relationships within that period and also within the preceding ten periods (equivalent to a span of 30 years). For instance, to construct the patrilineal kinship network of a male individual in 1849,

Table B.1: Summary Statistics

Panel A	(1)	(2)	(3)	(4)
Individual characteristics (all males)	Mean	SD	Min	Max
Age	30.413	19.126	1	87
Next die	0.050	0.217	0	1
Centrality among males	1.173	0.770	0	5.500
Centrality among married males	0.734	0.517	0	3.969
Centrality among unmarried males	0.438	0.431	0	2.000
Centrality among married females	1.005	0.615	0	4.422
Distance to port	1.911	0.678	0.115	3.107
Distance to port plus closest river	2.124	0.746	0.363	3.351
n	131,302			
N	34,419			
Panel B	(7)	(8)	(9)	(10)
Individual characteristics (unmarried males)	Mean	SD	Min	Max
Age	16.153	13.648	1	86
Next marry	0.112	0.315	0	1
Centrality among males	1.006	0.568	0	5.550
Centrality among married males	0.719	0.402	0	3.969
Centrality among unmarried males	0.287	0.277	0	2.000
Centrality among married females	0.750	0.427	0	4.422
Distance to port	1.891	0.687	0.115	3.107
Distance to port plus closest river	2.102	0.745	3.351	0.363
n	56,397			
N	21,369			
Panel C	(13)	(14)	(15)	(16)
Kin group characteristics	Mean	SD	Min	Max
Size	9.611	10.719	1	114
Size without females	4.367	10.719	1	1 52
No. of unmarried males	2.384	2.092	1	26
No. of kin groups	30,808			

Note: Panels A and B include only observations with the next registers and village location available. Panel B additionally excludes individuals who are recorded as dead in the next register. Panel C includes only kin groups that have at least one male with the next register available. Kin groups are year-specific; relatives appearing in two different years are considered as belonging to two different kin groups.

I utilize all his male relatives present in the data between 1819 and 1849. For these males, I establish links between each individual and his father, resulting in the formation of a patrilineal family tree for that kinship group. This patrilineal family tree serves as the basis for defining the patrilineal kinship network throughout this study.

There are two important considerations to note in the construction of these kinship networks for the baseline analysis. First, all males, including children, are incorporated into the kinship networks due to the prevalence of child marriage in pre-modern China, commonly enacted through ”minor marriages of child brides”[17]. Such marriages occur when families seek to secure brides for their sons at a young age. The data shows that males are married as young as 2 years old, and approximately 1.5% of individuals marry before the age of 12. While this could potentially be attributed to measurement error, the possibility of minor marriages cannot be ruled out. If minor marriages are present in the sample, there are two compelling reasons to include child-married individuals. First, minor marriages aptly reflect kinship support and patriarchal influence, as children would require assistance from their male relatives to enter these marriages. Second, minor marriages typically demand similar, if not greater, resources as regular marriages, given the costs associated with adopting a girl. Therefore, if children in the sample could potentially be child grooms, they are relevant for resource competition among unmarried males and should be included in measures of centrality among this group.

Secondly, under this procedure, two individuals are considered part of the same kinship group only if they share the same ancestors no earlier than 30 years prior. The purpose of establishing kinship groups in the analysis is to capture the support that an individual receives from such a group. It is unlikely that individuals who are very distantly related will provide each other with daily support. Therefore, the 30-year threshold ensures a meaningful and relevant connection within kinship groups.

Given that the vertices in the constructed patrilineal kinship networks represent

---

[17] See Eastman et al. (1989) and Telford (1995).

male individuals and the edges represent father-son relationships, I employ *decay centrality* as a measure of a male individual's centrality within such a network. Decay centrality is a network analysis measure that considers the distance between a selected vertex and every other vertex, weighted by a decay parameter. Suppose the decay parameter is denoted as α. In this context, a male individual receives a centrality value of α from his father, α^2 from his grandfather, α^3 from one of his paternal uncles, and so on. For a given patrilineal kinship group indexed by k and denoted by the set N_k, denote the decay centrality of a member i as C_{ik}. As presented in Section 3.3, C_{ik} is calculated as follows:

$$C_{ik} = \sum_{j \in N_k \setminus i} \alpha^{d(i,j)}$$

where $d(i,j)$ represents the distance (i.e., the length of the shortest path) between individuals i and j within the patrilineal kinship network and the parameter α is a predefined value that determines the weight assigned to the decay factor.

Decay centrality is considered more suitable for reflecting altruistic support within kinship networks compared to other centrality measures such as eigenvector centrality or Katz centrality, which are defined in a recursive manner. The rationale behind this lies in the nature of altruistic support and resource allocation within kinship networks. Decay centrality captures the notion of altruistic support by assigning weights based on the distance between individuals in the kinship network. As shown in Section 9, an individual's decay centrality corresponds to their weight in the total utility a social planner maximizes, when the utility function includes a term that captures altruism (social utility). On the other hand, centrality measures like eigenvector centrality or Katz centrality assume that a vertex is more central if any of the edges connected to it are more central. However, it is unlikely that, for example, an uncle would be more inclined to allocate resources to his nephew simply because other relatives are more willing to allocate resources to the uncle. Indeed, if we view the allocation process as a form of gift-giving, it is possible that an uncle might be more inclined to allocate resources to his nephew if he receives more resources from other relatives. However, this research focuses on the collective decisions made by the kinship network regarding the distribution of common resources to support the unmarried individuals, while personal gift-giving is not the concern. Even if personal gift-giving were considered, it is important to note that the uncle receiving more gifts would still allocate those gifts based on blood distance, giving more to his son and relatively less to his nephew. In the end, the resources an individual receives can still be reflected by their distance-based decay centrality within the kinship network.

Figure B.2(a-c) presents three illustrative examples of the construction of patrilineal kinship networks. Figure B.2(a) represents a simple patrilineal network comprising the grandfather, father, uncle, and son. A cousin sister is not included for constructing this patrilineal kinship network. The son's decay centrality in this network is calculated as $0.5 + 0.5^2 + 0.5^3 = 0.875$. In Figure 1(b), the network is larger than in Figure 1(a) as it traces back to the great-grandfather. As long as the great-grandfather (or any other) passed away no earlier than 30 years prior, these individuals remain connected. However, when computing an individual's centrality, only their living relatives contribute to their centrality score. Figure 1(c) depicts a scenario where the great-grandfather passed away more than 30 years ago. In this case, the son does not receive credit from the great-grandfather or the granduncle in the computation of his centrality.

In addition to the decay centrality derived from distances to everyone else in the network, I also calculate two other decay centrality measures: *centrality among married* and *centrality among unmarried*. These measures are based on the distances to married male individuals and unmarried male individuals, respectively. Continuing with the example in Figure 1(a), assume that both the father and grandfather are married, while the uncle remains unmarried. In this case, the centrality among married males for the son would be calculated as $0.5 + 0.5^2 = 0.75$, representing the centrality he has within the network of married individuals. On the other hand, the centrality among unmarried males for the son would be $0.5^3 = 0.125$, which he gets from the uncle.

Figure B.2: Examples illustrating patrilineal kinship network construction

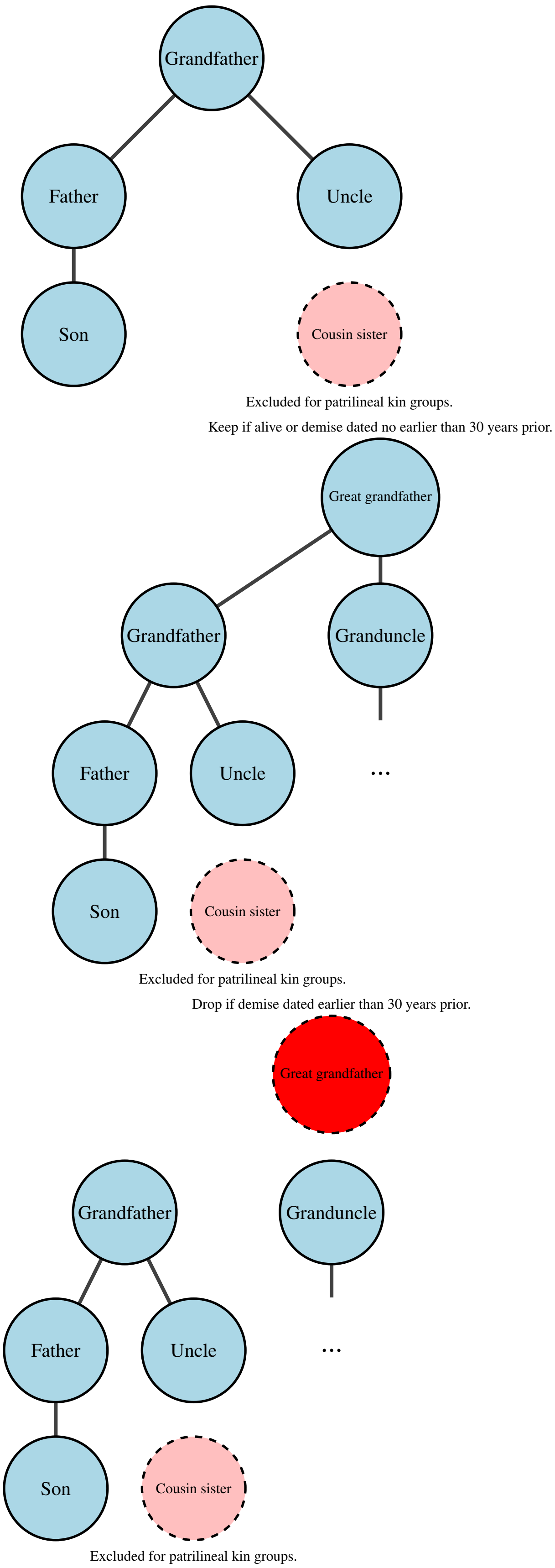
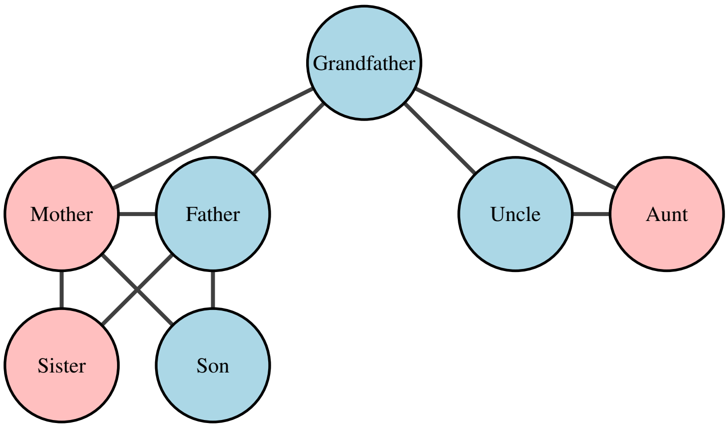


Figure B.3: A Kinship network with female relatives



Constructing a kinship networks including females involves not only father-son links but also mother-son and wife-husband links. It's important to note that in this dataset, married women's parents are considered to be their parents-in-law rather than their biological parents. The centrality computed by considering an individual's distances to female relatives is referred to as *centrality among females*. Figure B.3 provides an example that extends from Figure 1(a) by adding the mother, sister, and aunt of the son. Although there are multiple paths between two individuals, such as between the son and the mother, the computation of centrality considers only the shortest path. Therefore, the distances between the son and the mother, sister, and aunt are 1, 2, and 3, respectively, resulting in his centrality among females being equal to 0.875 (calculated as $0.5+0.5^2+0.5^3$).

The parameter of decay centrality, the threshold of age for being included in the kinship group, and the time span for tracing lineages are subject to variation in the robustness checks conducted in this study.

B.4 Data Limitations

Although the data is comprehensive and detailed, it does have several limitations. First of all, it has a significant number of missing observations due to destroyed, misplaced, damaged, or otherwise unavailable registers. For example, an individual may have records for 1858 and 1864 but no available record for 1861. More than 60% individuals have at least one period missing in between. This leads to missing links in the constructed networks and, as a result, generates measurement errors in the decay centrality. To address this issue, as part of a robustness check, I impute individuals into the periods in which they are absent but present in periods before and after, then constructed kinship networks based on the imputed data. In addition to networks, this problem also affects the outcome variables. For instance, if an unmarried individual is not observed in the next register but is observed as married in some periods afterwards, the data assumes that this individual gets married in the next register. Due to this issue, I restrict the sample to only include observations where the next register is available.

Another limitation of the data is the scarcity of observations for unmarried females. Furthermore, for females in the dataset, the ID numbers of their parents correspond to their biological parents if they have never been married. However, if a female is already married or is expected to marry in the future, her parents' ID numbers represent those of her mother/father-in-law. It makes analyzing larger kinship networks that take intermarriage into account considerably difficult. However, as previously mentioned, kinship in imperial China is considered to be patrilocal and it makes the support from married-in female relatives' natal families quite limited.

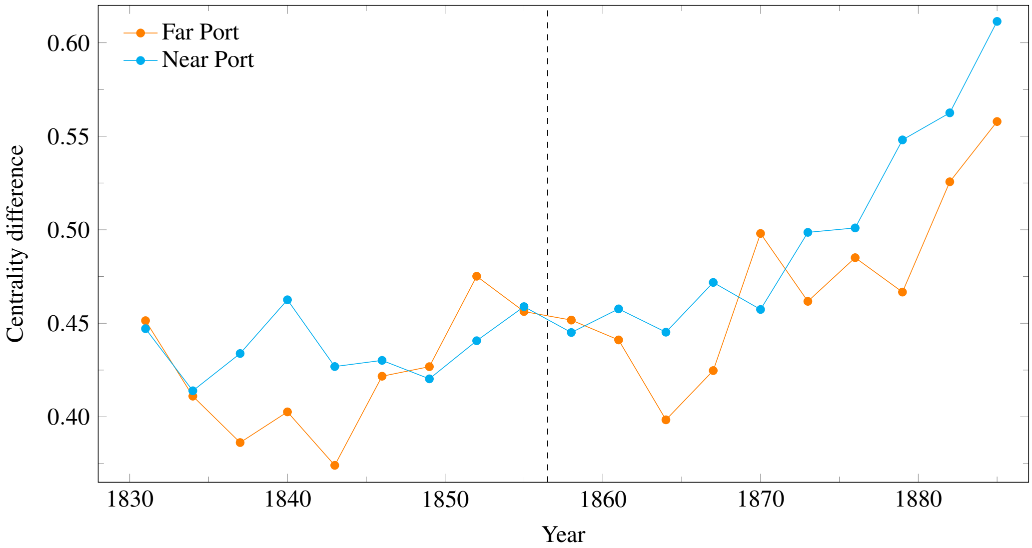
C Motivational Evidence

In this section, I present some motivational evidence showing the changing effects of kinship ties on one's marriage status.

Figure C.1 illustrates the mean differences in centrality among male relatives, between married and unmarried males, spanning periods from 1831 to 1885. The variable on the y-axis represents the mean of centrality among male relatives for all married males, minus that for all unmarried males. This is computed separately for the Near-Port Group (those residing closer to the port, below the median distance) and the Far-Port Group (those farther from the port, above the median distance). The blue dashed line indicates the time point when the port opened.

Both the Near-Port and Far-Port groups exhibit a rise in such mean difference in centrality over the time span, reflecting the challenging circumstances males faced in the marriage market, as evidenced by many historical accounts. However,

Figure C.1: Mean difference in centrality among males between married and unmarried males



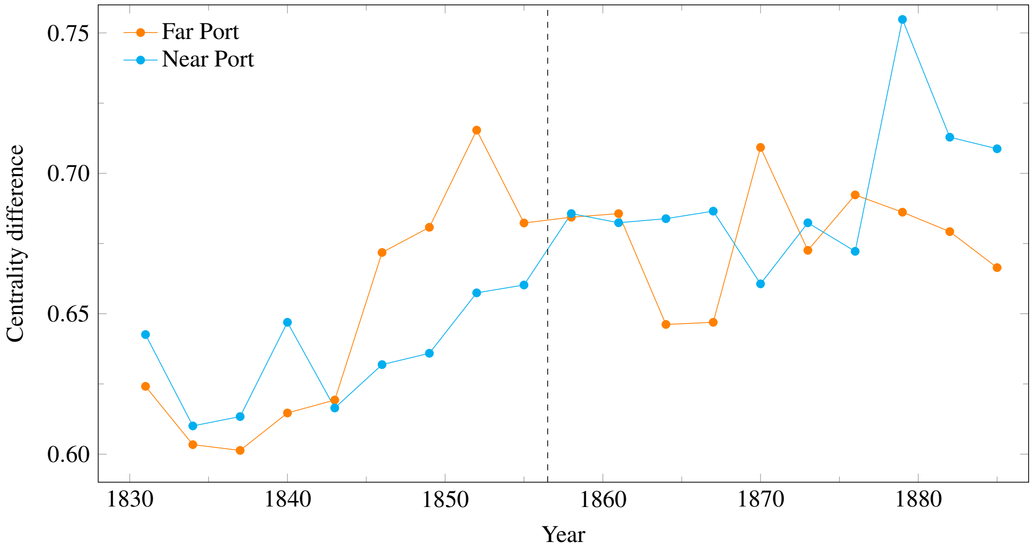
Note: The dashed line indicates the port opening.

before the port opening, the Far-Port Group only experienced a moderate increase in centrality difference, while it remained essentially stagnant for the Near-Port Group. Subsequent to the opening, the Near-Port Group experienced a pronounced escalation in centrality difference, despite the Far-Port Group also seeing a simultaneous rise, albeit less quickly. This pattern suggests that the port opening makes patrilineal kinship network positions more strongly associated with one's marital status.

Figure C.2 is configured in a manner analogous to Figure C.1 but delineates the dynamics of mean differences in centrality among females instead of males. As an ever-married female's natal family is unobservable, centrality among unmarried females is computed using 0.5^2 multiplied by the number of unmarried sisters. Despite the correlation between centrality among females and males, Figure C.2 does not exhibit a pattern as pronounced as that observed in Figure C.1.

Interestingly, Figure C.2 illustrates that the centrality difference steadily increases for both the Near-Port and Far-Port Groups about 20 years prior to the port opening, but stabilizes following it. While this paper does not provide a certain

Figure C.2: Mean difference in centrality among females between married and unmarried males



Note: Centrality among unmarried females is computed as 0.5^2 times the number of unmarried sisters, due to the unobservable natal families of ever married females. The blue dashed line indicates the port opening.

answer, the reason might be that the increasing difficulty for males in the marriage market before the port opening heightened the importance of information during times of ubiquitous resource scarcity, wherein female relatives could offer more assistance. On the other hand, the patriarchal system, crucial in resource allocation, gained prominence due to an influx of resources following the port opening.

D Identification

D.1 Identifying ATTR

This section demonstrates the identification of $E[\Upsilon^{(1)}_{ifvt} - \Upsilon^{(0)}_{ifvt} | D_v = 1, Post_t = 1]$, referred to as the ATTR in this paper. The discussion is based on the specification outlined in equation (1.2). I establish that the ATTR is equivalent to β_5. While the core elements of the proof are based on Olden and Møen (2022), I adapt it to suit the specifics of this paper.

Assumption D.1. $E[\frac{\partial \varepsilon_{ifvt}}{\partial C_{ifvt}} | D_v = 1, Post_t = 1, f, v, t, X_{ifvt}] - E[\frac{\partial \varepsilon_{ifvt}}{\partial C_{ifvt}} | D_v = 1, Post_t = 0, f, v, t, X_{ifvt}] = E[\frac{\partial \varepsilon_{ifvt}}{\partial C_{ifvt}} | D_v = 0, Post_t = 1, f, v, t, X_{ifvt}] - E[\frac{\partial \varepsilon_{ifvt}}{\partial C_{ifvt}} | D_v = 0, Post_t = 0, f, v, t, X_{ifvt}]$.

This assumption is non-standard; however, it is evidently weaker than the standard assumption of exogeneity, i.e., $E[\varepsilon_{ifvt}|D_v, Post_t, f, v, t, X_{ifvt}] = 0$. Moreover, in the context where Υ_{ifvt} is considered as the outcome for analysis, and DID rather than TD is employed, the standard assumption of exogeneity requires $E[\frac{\partial \varepsilon_{ifvt}}{\partial C_{ifvt}}|D_v, Post_t, f, v, t, X_{ifvt}] = 0$, which is also stronger than Assumption D.1, as Assumption D.1 allows for the existence of unobservable factors that influence both the effects of centrality and some independent variables, as long as these confounding factors remain stable across treated and control groups or, if not, to be cancelled out by using difference-in-differences. Another reason for adopting Assumption D.1, as opposed to the standard exogeneity assumption, is the predominantly demographic nature of the variables in the dataset. This characteristic raises concerns about omitted variable bias. Further discussion on this assumption can be found in Appendix D.2.

Assumption D.2. $E[\Upsilon^{(0)}_{ifvt}|D_v = 1, Post_t = 1, f, v, t, X_{ifvt}] - E[\Upsilon^{(0)}_{ifvt}|D_v = 1, Post_t = 0, f, v, t, X_{ifvt}] = E[\Upsilon^{(0)}_{ifvt}|D_v = 0, Post_t = 1, f, v, t, X_{ifvt}] - E[\Upsilon^{(0)}_{ifvt}|D_v = 0, Post_t = 0, f, v, t, X_{ifvt}]$.

If we treats Υ_{ifvt} as the outcome to be analyzed, then Assumption C.2 is the standard parallel trends assumption in DID approach.

Proposition D.1. *Suppose the model represented by equation (1.2) is correctly specified. Under Assumption C.1 and C.2, $E[\Upsilon^{(1)}_{ifvt} - \Upsilon^{(0)}_{ifvt}|D_v = 1, Post_t = 1]$ is identified and equal to β_5.*

Proof. Given equation (1.2), we have

$$E[\Upsilon_{ifvt}|D_v = 1, Post_t = 1, f, v, t, X_{ifvt}] - E[\Upsilon_{ifvt}|D_v = 1, Post_t = 0, f, v, t, X_{ifvt}]$$
$$= \beta_4 + \beta_5 + E[\frac{\partial \varepsilon_{ifvt}}{\partial C_{ifvt}}|D_v = 1, Post_t = 1, f, v, t, X_{ifvt}] - E[\frac{\partial \varepsilon_{ifvt}}{\partial C_{ifvt}}|D_v = 1, Post_t = 0, f, v, t, X_{ifvt}] \tag{D.1}$$

and

$$E[\Upsilon_{ifvt}|D_v=0,Post_t=1,f,v,t,X_{ifvt}]-E[\Upsilon_{ifvt}|D_v=0,Post_t=0,f,v,t,X_{ifvt}]$$
$$=\beta_4+E[\frac{\partial\varepsilon_{ifvt}}{\partial C_{ifvt}}|D_v=0,Post_t=1,f,v,t,X_{ifvt}]-E[\frac{\partial\varepsilon_{ifvt}}{\partial C_{ifvt}}|D_v=0,Post_t=0,f,v,t,X_{ifvt}]. \tag{D.2}$$

Subtract equation (C.1) by equation (C.2) and use Assumption C.1, we have

$$\begin{aligned}\beta_5=&(E[\Upsilon_{ifvt}|D_v=1,Post_t=1,f,v,t,X_{ifvt}]-E[\Upsilon_{ifvt}|D_v=1,Post_t=0,f,v,t,X_{ifvt}])\\&-(E[\Upsilon_{ifvt}|D_v=0,Post_t=1,f,v,t,X_{ifvt}]-E[\Upsilon_{ifvt}|D_v=0,Post_t=0,f,v,t,X_{ifvt}])\\=&(E[\Upsilon^{(1)}_{ifvt}|D_v=1,Post_t=1,f,v,t,X_{ifvt}]-E[\Upsilon^{(0)}_{ifvt}|D_v=1,Post_t=0,f,v,t,X_{ifvt}])\\&-(E[\Upsilon^{(0)}_{ifvt}|D_v=0,Post_t=1,f,v,t,X_{ifvt}]-E[\Upsilon^{(0)}_{ifvt}|D_v=0,Post_t=0,f,v,t,X_{ifvt}])\end{aligned} \tag{D.3}$$

Substitute Assumption C.2 into equation (C.3), we have

$$\begin{aligned}\beta_4=&E[\Upsilon^{(1)}_{ifvt}|D_v=1,Post_t=1,f,v,t,X_{ifvt}]-E[\Upsilon^{(0)}_{ifvt}|D_v=1,Post_t=1,f,v,t,X_{ifvt}]\\&+E[\Upsilon^{(0)}_{ifvt}|D_v=0,Post_t=1,f,v,t,X_{ifvt}]-E[\Upsilon^{(0)}_{ifvt}|D_v=0,Post_t=0,f,v,t,X_{ifvt}]\\&-E[\Upsilon^{(0)}_{ifvt}|D_v=0,Post_t=1,f,v,t,X_{ifvt}]+E[\Upsilon^{(0)}_{ifvt}|D_v=0,Post_t=0,f,v,t,X_{ifvt}]\\=&E[\Upsilon^{(1)}_{ifvt}-\Upsilon^{(0)}_{ifvt}|Port_d=1,Post_t=1,f,v,t,X_{ifvt}]\\=&E[\Upsilon^{(1)}_{ifvt}-\Upsilon^{(0)}_{ifvt}|D_v=1,Post_t=1]\end{aligned} \tag{D.4}$$

The last equality comes from the fact that β_5 is a constant and hence remains the same if we take the expectation of it with respect to f,v,t and X_{ifvt} given $Post_t=1$ and $D_v=1$. It completes the proof. □

D.2 Possible Concerns

Omitted-Variable Bias. The existence of omitted variables could still be a concern, however, it is less likely than in other counterpart empirical analyses.

First, an omitted variable of concern can only be more specific than kin group founder level, as kin group founder FE has been included in the specification. But as mentioned in Section 3.1, the majority of the individuals in the sample are hereditary peasants and have quite similar social-economic backgrounds, which makes omitted variables less of a concern.

Second, an omitted variable is a concern only when it influences the effects of centrality differently between the treated and control groups. To see this, suppose the omitted variable is Z_{ifvt} and $\varepsilon_{ifvt} = \gamma_1 Z_{ifvt} \times C_{ifvt} + \gamma_2 Post_t \times Z_{ifvt} \times C_{ifvt} + \gamma_3 D_v \times Z_{ifvt} \times C_{ifvt} + \gamma_4 D_v \times Post_t \times Z_{ifvt} \times C_{ifvt}$. Then Assumption C.1 is violated if $\gamma_4 \neq 0$.

Although it is hard to come up with such a confounding factor Z_{ifvt} as a real example, one possible scenario could be that centrality also assists unmarried males in their search, such as by introducing them to suitable unmarried females. Then Z_{ifvt} can represent one's searching cost, e.g. opportunity cost of spending time on searching for a spouse instead of working, so one's central position will be more helpful if his search cost is lower. However, it is hard to justify that given one's own search cost, his centrality becomes more important after the opening. Furthermore, even it does, ones' female relatives should play the same role as males, if not a more significant role, in introducing suitable spouses for an unmarried male. However, no similar effects have been found regarding centrality among females, as present in Section 5.

Migration. Another concern regarding endogeneity is related to migration. For example, it is possible that individuals who have a central position in their kinship network might choose to migrate and work in the port area since they are more supported and hence have more freedom in choosing their careers. This could result in increased wealth and a higher likelihood of marriage.

However, it is important to note that free immigration is officially prohibited for individuals in the sample, and any violation of this rule is documented as absconded. The dataset includes information on whether an individual absconds or not within the next three years. In Section E.2, I examine the impact of the port opening on the probability of an individual's absconding and find no evidence showing that the opening of the port leads to increased immigration. It is also possible that there were immigrants outside of the Eight Banner system who could have had an impact on the local marriage market if there was an influx of female or male immigrants into the port area. However, as previously mentioned, individuals governed by the Eight Banners system were generally prohibited from marrying outside of their own group. Although this rule may not have been strictly enforced in the late 19th century, people belonging to the Eight Banners displayed a strong reluctance to marry individuals from outside their own group. Even in the 1910s, following the collapse of the Qing dynasty, marriages within the former Eight Banners population were estimated to account for a significant majority (Ding et al., 2004). Furthermore, the hypothesis that central individuals are more likely to migrate contradicts related empirical findings (Munshi and Rosenzweig, 2016).

Marriage Market. Another concern, albeit not directly related to the specification, is that the effect of port opening might be constrained by the supply in the marriage market if the market is highly localized. For instance, if men only married women within their own village, there would be no effect if all male villagers became wealthier, assuming they were not significantly differently affected by the port opening.

However, historical records indicate that in Liaoning Province during the late Qing dynasty, women typically married men who lived quite far away (Ding et al., 2004). It is also well-established that wealth is one of the most crucial factors in determining a man's ability to find a spouse (Shiue and Keller, 2022). Consequently, it would be natural to conjecture that following the port opening, women living far from the port would marry men living closer to it.

Despite the significant missing observations of females, descriptive evidence supports this notion. Figures D.1 and D.2 display the number of married females observed in the data, in the Near-Port area and Far-Port area respectively, with the dashed line marking the port's opening. Despite the volatility, it is observable that

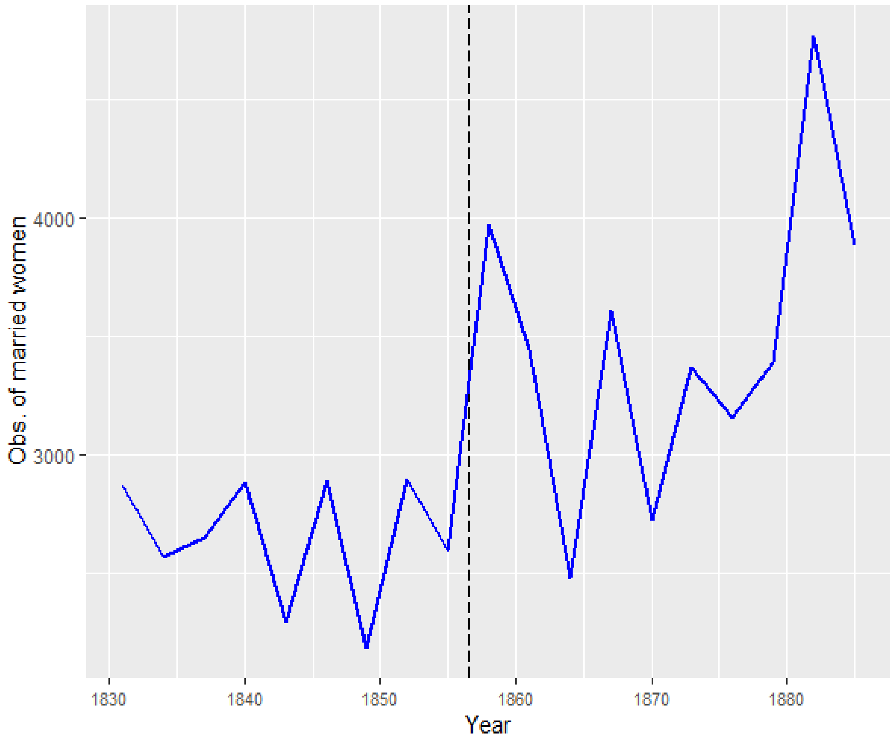


Figure D.1: Number of married female observations across years in Near-Port area

the number of married females in the Near-Port area exhibits an increasing trend after the port's opening, whereas in the Far-Port area, the trend appears more flat.

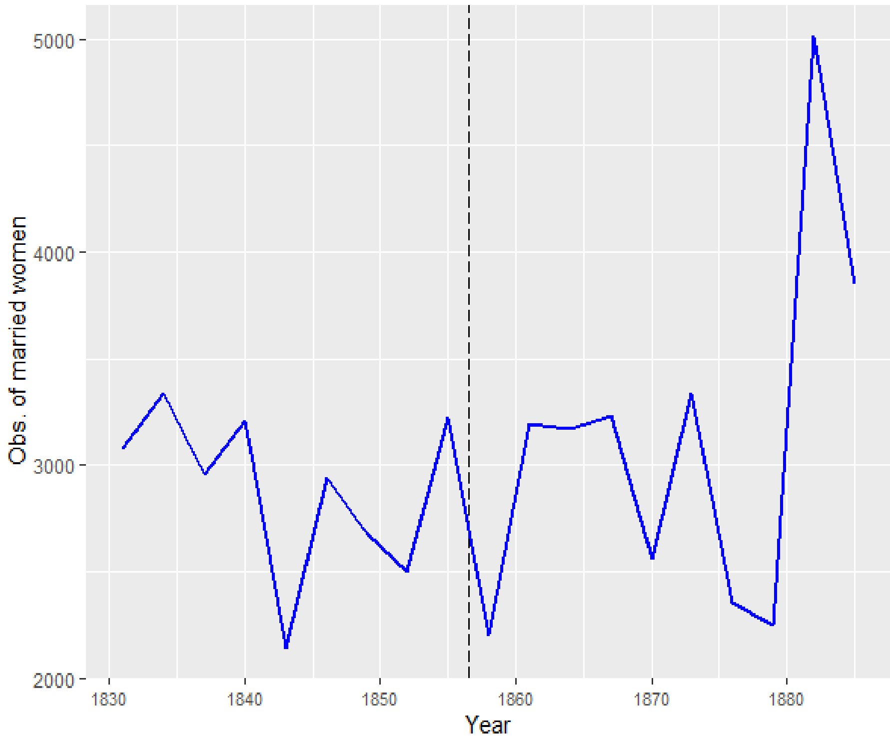


Figure D.2: Number of married female observations across years in Far-Port area

E Other Robustness Checks

E.1 Age Thresholds for Kin Groups

In the baseline analysis, I include all linked relatives regardless of age, since I want to consider the effect of child marriage. However, assuming that one's child relatives have an influence on his marriage is a strong assumption. In this section, I vary the age thresholds for constructing kin groups as an additional robustness check.

The results from Table E.1 indicate that the heightened support from unmarried male relatives post-port opening remains consistent when varying the age threshold for constructing kin groups. Notably, as seen in Column (5), the coefficient of Post×Centrality among unmarried males becomes significantly positive when the age threshold is set to 15. This suggests that the support from non-child unmarried male relatives post-port opening increased across all areas covered in the sample, not just those near the port.

Another concern is that the main results in the paper could be influenced by increased fertility. For instance, families benefiting more from the port opening could have both higher birth rates and enhanced resources for marriage, resulting in a misleading correlation between port opening and the influence of unmarried male relatives. To address this, I adjust the time span in Columns (3), (4) and (6). Take Column (2) as an example: its time span is 1849-1867 while the age threshold is 9. This ensures that children born after the port opening (1858) are not considered in kin group constructions, highlighting the effect purely from centrality among those unmarried males born prior to the port opening. Generally, the outcomes remain consistent, as evidenced by Columns (3), (4) and (6).

E.2 Varying Altruism Parameters

This section provides evidence demonstrating the robustness of the main results to variations in the decay centrality parameter, denoted as α in the expression for C_{ik}^{S} as presented in Section 3.3.

As shown in Table E.2, the main findings indicating that port opening enhances support from unmarried male relatives remain consistent, albeit with predictably smaller coefficient magnitudes as α increases. Notably, the adjusted R^2 value is higher when $\alpha = 0.5$ and $\alpha = 0.6$, as in Table 1, compared to other values of α. This observation suggests that values of α both higher and lower than these may not accurately represent the true underlying parameter.

E.3 Excluding Kin Groups with Few Unmarried Males

The issue of missing links results in a relatively small average kin group size, with on average only around two unmarried males in each kin group. The mechanism I suggest becomes implausible if the main results are driven by those from kin groups with a single unmarried male, since there is no resource competition in these instances. In this section, I demonstrate that the findings hold true even when excluding kin groups with a limited number of unmarried males.

Table E.3 shows that the main finding – the port opening bolstered support from unmarried male relatives – remains consistent even when excluding kin groups with few unmarried males. Notably, both such effect and the adjusted R squared value

Table E.1: Changing effects of centrality among married and unmarried males

	Dep. var: Marriage next period					
Age Thresholds for Kin Groups:	9 years old		12 years old		15 years old	
	Full sample	1849-1867	Full sample	1849-1870	Full sample	1849-1873
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity×Post	0.016**	0.016	0.017**	0.011	0.018**	0.006
	(0.005)	(0.017)	(0.005)	(0.010)	(0.007)	(0.013)
Centrality among married males	0.023	0.011	0.027**	0.017	0.023**	0.019
	(0.007)	(0.025)	(0.008)	(0.024)	(0.007)	(0.020)
Centrality among unmarried males	0.015	0.052**	-0.067**	-0.031	-0.107***	-0.082***
	(0.015)	(0.016)	(0.019)	(0.027)	(0.010)	(0.004)
Proximity×Centrality among married males	0.047***	0.047**	0.038***	0.037**	0.036***	0.036**
	(0.007)	(0.017)	(0.006)	(0.013)	(0.006)	(0.013)
Proximity×Centrality among unmarried males	-0.110***	-0.102***	-0.080**	-0.063**	-0.091**	-0.066**
	(0.022)	(0.016)	(0.027)	(0.023)	(0.033)	(0.020)
Post×Centrality among married males	-0.006	-0.016	-0.009	-0.012	-0.008	-0.005
	(0.013)	(0.014)	(0.013)	(0.017)	(0.012)	(0.012)
Post×Centrality among unmarried males	-0.020***	-0.021	0.026	0.030	0.062***	0.068***
	(0.004)	(0.012)	(0.014)	(0.016)	(0.010)	(0.008)
Proximity×Post	-0.031*	-0.021	-0.022	-0.026	-0.020	-0.026*
×Centrality among married males	(0.013)	(0.015)	(0.014)	(0.016)	(0.013)	(0.013)
Proximity×Post	0.106***	0.093**	0.079***	0.078*	0.068**	0.068*
×Centrality among unmarried males	(0.015)	(0.025)	(0.021)	(0.034)	(0.021)	(0.028)
Village FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Kin Founder FE	✓	✓	✓	✓	✓	✓
Distance time trend	✓	✓	✓	✓	✓	✓
Kin group characteristics	✓	✓	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓	✓	✓
Observations	55,937	26,762	55,937	30,774	55,937	35,610
R^2	0.130	0.147	0.130	0.146	0.128	0.135

Note: Proximity indicates whether the individual belongs to the Near-Port Group. Robust standard errors clustered at the district level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.2: Changing effects of centrality among married and unmarried males

Dep. var: Marriage next period	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.7$
	(1)	(2)	(3)	(4)
Proximity×Post	0.021	0.024*	0018***	0.030***
	(0.012)	(0.036)	(0.004)	(0.006)
Centrality among married males	0.146***	0.080***	0.023***	0.012***
	(0.026)	(0.009)	(0.002)	(0.001)
Centrality among unmarried males	0.187	0.084	0.026	0.015
	(0.103)	(0.051)	(0.018)	(0.011)
Proximity×Centrality among married males	0.119***	0.086***	0.049***	0.031***
	(0.019)	(0.010)	(0.004)	(0.002)
Proximity×Centrality among unmarried males	-0.321***	-0.170***	-0.013***	-0.037***
	(0.050)	(0.028)	(0.008)	(0.008)
Post×Centrality among married males	0.013	0.007	0.003	0.001
	(0..043)	(0.027)	(0.008)	(0.006)
Post×Centrality among unmarried males	-0.174**	-0.095**	-0.038***	-0.020**
	(0.057)	(0.031)	(0.003)	(0.008)
Proximity×Post ×Centrality among married males	-0.086	-0.063*	-0.034***	-0.026**
	(0.054)	(0.033)	(0.008)	(0.007)
Proximity×Post ×Centrality among unmarried males	0.336***	0.178***	0.080***	0.037***
	(0.066)	(0.036)	(0.007)	(0.009)
Village FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Kin Founder FE	✓	✓	✓	✓
Distance time trend	✓	✓	✓	✓
Kin group characteristics	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓
Observations	55,937	55,937	55,937	55,937
Adjusted R^2	0.126	0.126	0.127	0.126

Note: Proximity indicates whether the individual belongs to the Near-Port Group. Robust standard errors clustered at the district level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

tend to roughly rise as the threshold for unmarried males increases. This suggests that the main result is largely driven by individuals from kin groups with a higher number of unmarried males.

The coefficient of Proximity×Port leans towards being negative when kin groups with fewer than 3 unmarried males are omitted. However, its interpretation is challenging since this coefficient implies the effect of port opening on the marriage likelihood of those without relatives – an implausible scenario for individuals in these samples. This actually indicates that, in these samples, after the port opening, marriage probability rises steeply with centrality among unmarried males. As a result, the conjecture is that the port opening negatively impacts those with a centrality among unmarried males of zero.

Table E.3: Changing effects of centrality among married and unmarried males

Dep. var: Marriage next period	Unmarried males>1	Unmarried males>2	Unmarried males>3	Unmarried males>4	Unmarried males>5
	(1)	(2)	(3)	(4)	(5)
Proximity×Post	0.028	0.023	-0.012	-0.056**	-0.046
	(0.029)	(0.024)	(0.032)	(0.020)	(0.041)
Centrality among married males	0.044*	0.047**	0.069***	0.073***	0.099***
	(0.019)	(0.013)	(0.013)	(0.016)	(0.014))
Centrality among unmarried males	0.047**	0.052**	0.050*	0.019	0.009
	(0.015)	(0.020)	(0.024)	(0.024)	(0.037)
Proximity×Centrality among married males	0.049	0.062**	0.052**	0.023	0.007
	(0.028)	(0.020)	(0.015)	(0.022)	(0.020)
Proximity×Centrality among unmarried males	-0.069***	-0.066**	-0.096**	-0.057**	-0.054
	(0.015)	(0.020)	(0.028)	(0.016)	(0.050)
Post×Centrality among married males	0.009	0.007	-0.010	-0.015	-0.039***
	(0.032)	(0.017)	(0.019)	(0.010)	(0.003)
Post×Centrality among unmarried males	-0.045***	-0.050***	-0.058***	-0.043***	-0.030
	(0.012)	(0.006)	(0.008)	(0.011)	(0.026)
Proximity×Post ×Centrality among married males	-0.049	-0.059**	-0.040	-0.004	-0.007
	(0.033)	(0.021)	(0.021)	(0.014)	(0.014)
Proximity×Post ×Centrality among unmarried males	0.089***	0.103***	0.129***	0.120***	0.111*
	(0.019)	(0.025)	(0.031)	(0.020)	(0.050)
Village FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Kin Founder FE	✓	✓	✓	✓	✓
Distance time trend	✓	✓	✓	✓	✓
Kin group characteristics	✓	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓	✓
Observations	45,347	34,664	25,880	18,999	14,046
Adjusted R^2	0.129	0.135	0.140	0.144	0.145

Note: Proximity indicates whether the individual belongs to the Near-Port Group. Robust standard errors clustered at the district level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F Mechanism

F.1 Altered Sex Ratios

The first potential channel I explore is the influence of altered sex ratios. For instance, there could be a possibility that the sex ratio around the port decreased after the port opening, potentially due to shifting social norms regarding son preference. If this was the case, finding a spouse would have become easier for unmarried males around Niuzhuang, potentially transforming competitive unmarried relatives into supporters. To explore this hypothesis, I investigate whether the years following the port opening saw a reduction in the sex ratio at birth in the Near-Port areas. I calculate the sex ratios at birth as the ratio of all girls reported as newborns to all boys reported as newborns each year at the village level, adding one to the denominator to avoid infinity.

Table F.1 displays the results from regressing the village-level sex ratio at birth on the port opening. Across all four columns, the port opening does not significantly affect the sex ratio at birth, and the coefficients' magnitude is minimal. These results do not corroborate the hypothesis that kinship's increased effect on marriage resulted from alterations in the sex ratio impacting the marriage market. While some might contend that population changes outside the Eight Banners, not accounted for in the sample, could still influence the marriage market, Section 3.1 and Appendix D.2 clarify that individuals under the Eight Banners administration typically do not intermarry with those outside.

Nevertheless, I exercise caution in interpreting this result. The reported number of girls is evidently underestimated, with approximately 80% of observations indicating no girls born at all. Even if this underestimation is uncorrelated with the port opening, it constrains the validity of the analysis.

F.2 Immigration

Another potential factor of concern is the possibility that after the port opening, some males may have left their villages to seek economic opportunities in the port area. As also discussed in Appendix D.2, individuals who have a central position

Table F.1: Opening on village level sex ratio

Dep. var: Sex ratio at birth	(1)	(2)	(3)	(4)
Proximity×Post	-0.003	-0.003	-0.005	-0.005
	(0.010)	(0.011)	(0.012)	(0.011)
Year FE	✓	✓	✓	✓
Village FE	×	×	✓	✓
Observations	2,013	2,013	2,013	2,013
Adjusted R^2	0.014	0.014	0.260	0.260

Note: Proximity indicates whether the individual belongs to the Near-Port Group. Robust standard errors clustered at district level in parentheses for column (1) and (3), and at village level for column (2) and (4). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

might choose to work in the port area since they are more supported and therefore can immigrate more easily. However, as previously mentioned, free migration was not permitted for individuals governed by the Eight Banners. These individuals were not allowed to leave their settled villages, and anyone who did leave would be marked as absconded in the data. Although the strict enforcement of this policy varied, it was consistent throughout the Qing dynasty.

To examine this channel, we can investigate whether there were more absconded unmarried males in Niuzhuang after the port opening, under the assumption that the policy was similarly stringent in Niuzhuang compared to other districts. The results presented in Table F.2 suggest that the port opening has a minimal and statistically insignificant effect on the likelihood of absconding, for both all males and unmarried males. While the coefficients are not significant, Columns 2 and 4 hint that a central position among unmarried males is associated with a reduced tendency to abscond. Column 4 indicates that after the port opening, those who are central among unmarried males are less inclined to abscond. This observation aligns with the mechanism I propose and does not support the immigration channel, although it is not statistically significant.

Table F.2: Opening on absconding

Dep. var: Abscond next period	All males		Unmarried males	
	(1)	(2)	(3)	(4)
Proximity×Post	<0.001	<0.001	-0.001	-0.001
	(<0.001)	(0.001)	(0.001)	(0.001)
Centrality among unmarried males		-0.001		-0.004
		(0.002)		(0.003)
Proximity × Centrality among unmarried males		<-0.001		0.003
		(0.002)		(0.002)
Post × Centrality among unmarried males		<0.001		0.002
		(0.001)		(0.002)
Proximity×Post × Centrality among unmarried males		<0.001		-0.002
		(0.001)		(0.002)
Year FE	✓	✓	✓	✓
Kin founder FE	✓	✓	✓	✓
Village FE	✓	✓	✓	✓
Kin group characteristics	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓
Observations	118,192	118,192	55,938	55,938
Adjusted R^2	0.133	0.133	0.251	0.251

Note: Proximity indicates whether the individual belongs to the Near-Port Group. Robust standard errors clustered at district level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

G Further on Model Fitness

A further concern of the model fitness is whether the results, derived from using the predicted marriage probability, are merely driven by district-period factors and changes in centralities. Specifically, does the predicted marriage probability maintain a correlation with the actual probability even when district-period factors and centralities — elements used in model simulation — are controlled?

Table G.1 explores this dimension of the analysis. Across all regression models presented, the predicted marriage probability maintains a significant correlation with the observed marriage outcome, even when accounting for all pertinent fac-

Table G.1: Predicted marriage probability on observed marriage outcome

	Dep. var: Marriage next period			
	(1)	(2)	(3)	(4)
Predicted marriage prob.	0.305***	0.289***	0.145***	0.153***
	(0.047)	(0.034)	(0.010)	(0.018)
Centrality among married males		0.006	0.015	-0.026*
		(0.009)	(0.009)	(0.014)
Centrality among unmarried males		0.006	0.003	0.029**
		(0.008)	(0.008)	(0.013)
District×Year FE	×	×	✓	✓
Kin Founder FE	×	×	×	✓
Observations	86,785	86,785	86,785	86,785
Adjusted R^2	0.003	0.003	0.006	0.023

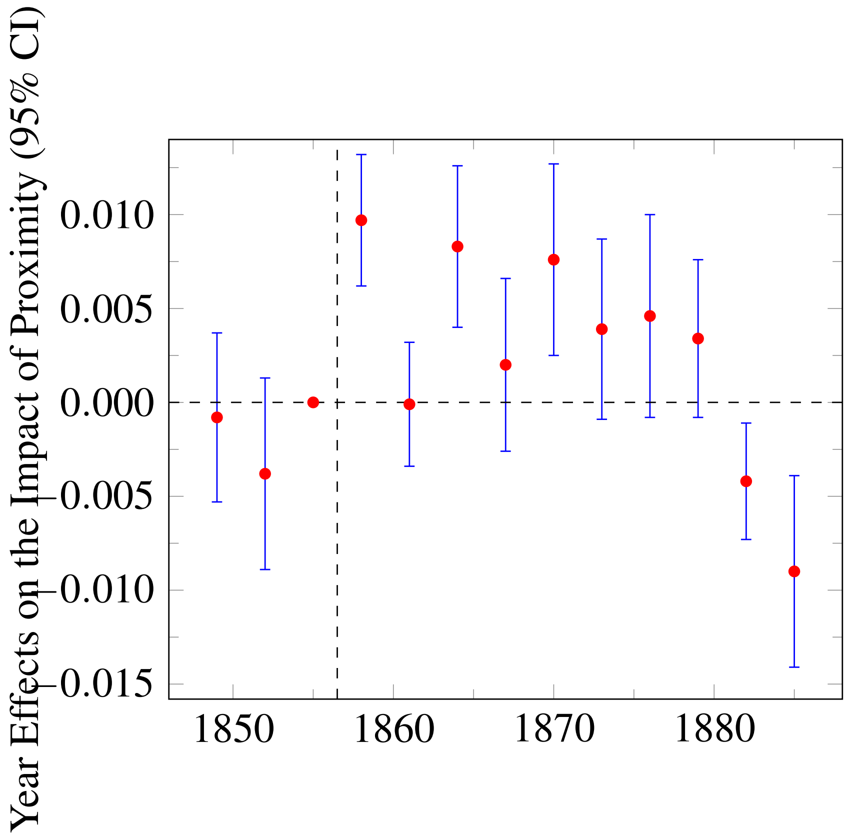
Note: Robust standard errors clustered at the district level in parentheses. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

tors. Notably, the coefficient for the predicted marriage probability in Column 4 surpasses that in Column 3, which does not control for the kin founder fixed effect. This lends credibility to the model, given that it is constructed and simulated on a kin-group-specific basis. Furthermore, the sign of the coefficient of centrality among married males becomes negative, which may occur because its positive effect has been fully accounted for by the predicted marriage probability.

Another question to consider is whether the model produces a larger within-kin-group disparity in the allocation of resources, which could resonate with the observed increase in within-kin-group disparity discussed in Section 7.2 and underscore the influence of patrilineal power structures.

To check this, I examine whether the dispersion of resources, $w_i'^{\xi}$, has been increased by the port opening. Analyzing this dispersion at the year-specific kin group level aforementioned poses challenges. Since the simulated resources are only available for unmarried males and many such kin groups contain only a few such individuals, as illustrated in Table B.1, discussing dispersion in this context may be meaningless. For the purposes of this analysis, I define a kin group as a group of individuals in a given year who share a common kin group founder, while centralities still trace kinship ties back only to 30 years ago, as illustrated in

Figure G.1: Event study on within-kin-group resource dispersion



Appendix B.3. The location of such a kin group to the port is based on the village housing the majority of that kin group's members in that year. I exclude all kin groups that include only one unmarried male.

Such a kin group is my unit of analysis. I use the standard deviation of simulated resources within the kin groups as my outcome variable. I regress this outcome on the interactions between year dummies and the kin groups' proximity to the port (Near-Port group or not), while controlling for year fixed-effects, village fixed-effects, and founder fixed-effects. Figure G.1 plots the results of this event study.

Clearly, there is a discernible pattern of increased dispersion in the Near-Port group following the port opening, although this increase is not significantly positive for each individual year. These effects diminish over time. The negative effects observed in the last two periods could be attributed to dramatic shocks in the political and economic environments. For instance, since 1882, China engaged in several conflicts in Korea, adjacent to Liaoning (Fengtian) Province. Additionally, in 1886, a significant flood in Liaoning's major rivers disrupted the transportation system. These events could potentially reduce the ϕ estimated for near-port districts, thereby driving males out of the marriage fund and consequently decreasing the dispersion of resources.

H Discussion on Model Assumptions

H.1 Exponential Marriage Probability

The probability of individual i getting married is given by $1-e^{-\lambda w_i'}$ if $w_i' \geq 0$ and 0 otherwise. λ is the parameter of the corresponding exponential distribution. The intuition is as follows. It at first assumes that the wealth distribution of unmarried females in the marriage market follows an exponential distribution with λ as the parameter[18] and w is the average wealth of every unmarried male without any support from his kinship group. Then it assumes that after the allocation, individual i will meet an unmarried female in the marriage market, and the match will be successful if the wealth of the woman is below the wealth of i, w_i'.[19] The model assumes that w_i' is the wealth one gets for marriage purposes only and is mainly determined by one's position in his kinship network, however, it also depends on and reflects the total wealth he has, which as mentioned above, also largely hinges on one's position in the kinship. Women also need to raise money (dowry) for marriage in imperial China. In the Qing dynasty, elites tend to prepare dowry higher than the bride price when they marry daughters, for their daughters' family status after marriage. However, in the lower class, which most individuals in my sample belong to, parents ask for a dowry higher than the bride price when marry their daughters, treating this as an opportunity to ease their financial difficulties (Mao, 2007). It should be noted that the model emphasizes the importance of wealth in determining competitiveness

[18] Exponential distribution is regarded as proper to describe the distribution of wealth for the vast majority and when money can be transferred. For example, see Drăgulescu and Yakovenko (2001).

[19] Even in modern society when a woman within a marriage market outearns a man, the marriage rate declines (Bertrand et al. (2015)). This was even more the case in pre-modern China, with a culture so-called *Men Dang Hu Dui*.

in the marriage market while abstracting from all other factors[20].

H.2 Social Utility and the Parameter α

There are several ways to interpret α and the altruistic allocation for marriage. For example, one can interpret v as the expected number of children one is going to have after marriage since in pre-modern ages love-based marriage is rare and one of the main purposes of marriage is to have offspring and increase one's family labour force, which is especially true in the lower classes where the data is from.[21] Then the rule of allocation could mean that the FM maximizes the expected number of children, and the expected number of children of individual i is weighted by how closely related his children will be to the rest of the kinship group, captured by α. That is because the kinship group as a lineage is expected to evolve to maximize its inclusive fitness and its genes passed on (Dawkins, 2016), and one can think of the FM as the head of the kinship group who acts for the continuation of his lineage. Another explanation could be that the central authority of the kinship group wants to allocate resources to an individual who has a more central position, knowing that in the future his children will favour more people in the kinship group since they will have more closely related relatives.

H.3 Solution Concept

Based on the model setup, without further assumptions, the existence of multiple equilibria can be ruled out only in very small or trivial kinship networks. For example, if there is only one unmarried individual, he will join as long as there are some other married kinship group members. However, when there are two unmarried individuals, there can be two equilibria if the resources are only enough to make $w_i' > w$ for one of them. It becomes nearly impossible to list all the possible equilibria when there are more individuals, not only because of the externality but also due to the presence of both strategic complements and substitutes. For instance, one additional unmarried individual joining the marriage fund means more resources to

[20] Shiue and Keller (2022) documents the importance of wealth in the marriage market in ancient China.

[21] It is true that traditional marriage is also about getting good in-laws for political or economic benefits, but it is much less so in the lower classes. (Coontz, 2006)

split among them, but it also means other unmarried individuals get more support, especially those who are closely related to the new participant.

Selecting a unique equilibrium becomes possible by imposing strong assumptions, such as only selecting the equilibrium that maximizes the total utility among all equilibria. However, the primary purpose of the model is to yield predictions to be tested with the data. Imposing overly strong assumptions may weaken the credibility of such tests. Thus, I refrain from finding and focusing on a unique equilibrium. Instead, I demonstrate that the desired results hold for each solution or equilibrium interchangeably, selected by the mild conditions I impose.

The intuition for the third condition in Definition 1 is as follows. One can think of the marriage fund as a coalition in cooperative games literature. The coalition of the status quo is M_x^ξ, and therefore if anyone wants to form a new coalition unilaterally, he has to join the current one. The third condition guarantees that given the equilibrium, if someone deviates to join, it must make some individuals want to leave, and therefore the new coalition will not deliver an allocation that blocks[22] the allocation given by the old one (some individuals in the new coalition will be unhappy and leave and hence the new coalition will not be formed).

One may propose a standard solution concept that no one can join to make himself better off. This concept is considered more standard but in fact, represents a stronger equilibrium. Any equilibrium meeting this criterion must also satisfy the third condition in Definition 1 since if no individual can join to make himself better off, it becomes impossible for someone's joining to benefit all participants. However, an equilibrium satisfying the third condition in Definition 1 does not necessarily adhere to this alternative solution concept, as some individuals may join and improve their own situation while adversely affecting other participants.

Apart from being a stronger concept, this alternative solution approach seems counter-intuitive as it fails to rule out scenarios where the joining of certain individuals from M_x^ξ leads to a desire for others to leave. Since an individual's payoff relies on their position in the marriage fund, some individuals leaving may jeopardize the

[22]If an allocation blocks another, it means given the former everyone in the coalition is weakly better than given the latter. For definitions in cooperative games, see Driessen (2013).

stability of the entire marriage fund. As a result, those individuals may end up with lower utility than before. In essence, I assume that an individual will abstain from joining the marriage fund if they anticipate that their participation will render the new marriage fund unstable.

H.4 Discussion of Condition 1 in Proposition 3

The first condition in Proposition 3, $\tau_i^{\xi_{1,j}}\hat{\delta}_i \leq \bar{\delta}_i$ for every i, is not standard due to the endogeneity of $\tau_i^{\xi_{1,j}}$. However, this condition is less restrictive than $\hat{\delta}_i \leq \bar{\delta}_i$ as $\tau_i^{\xi_{1,j}}$ is set to be smaller than 1. Should there be evidence supporting the plausibility of $\hat{\delta}_i \leq \bar{\delta}_i$, then it further validates the condition $\tau_i^{\xi_{1,j}}\hat{\delta}_i \leq \bar{\delta}_i$.

However, to adequately test with the data whether the main results remain consistent under the condition $\hat{\delta}_i \leq \bar{\delta}_i$, it is insufficient to simply exclude observations that exhibit higher centrality among unmarried males than married males. Given that the model predicts equilibria for specific kin groups, $\tau_i^{\xi_{1,j}}\hat{\delta}_i \leq \bar{\delta}_i$ must be valid for every individual within a kin group included in the sample. If an individual does not have higher centrality among unmarried males compared to married males, but some of his relatives do, then Proposition 3 does not apply and hence he should be excluded from the data, along with his relatives.

Furthermore, as demonstrated in Appendix I.5, the condition $\tau_i^{\xi_{1,j}}\hat{\delta}_i \leq \bar{\delta}_i$ aims to ensure that $\frac{\partial w_i'^{\xi}}{\partial \hat{\delta}_i}$ increases as $\tau_i^{\xi_1}$ approaches to $\tau_i^{\xi_{1,j}}$. This expected increase hinges on the assumption that $\tau_i^{\xi}\hat{\delta}_i \leq \bar{\delta}_i$ for every i across equilibria, both before and after the event. However, the condition $\tau_i^{\xi_1}\hat{\delta}_i \leq \bar{\delta}_i$ is encompassed by $\tau_i^{\xi_{1,j}}\hat{\delta}_i \leq \bar{\delta}_i$ because, given the model we have $\tau_i^{\xi_{1,j}} \geq \tau_i^{\xi_1}$. Nonetheless, for the sake of plausibility, I am hesitant to assume that $\tau_i^{\xi_{1,j}}\hat{\delta}_i \leq \bar{\delta}_i$ also occurs in my data, particularly because the third condition of the solution concept deviates from standard models.

In light of the two aforementioned points, to validate the condition $\tau_i^{\xi_{1,j}}\hat{\delta}_i \leq \bar{\delta}_i$, I choose to conduct regressions by excluding all kin groups containing at least one individual whose centrality among unmarried males exceeded that among married males during any period. Should the main findings remain robust under this increased level of strictness, it would be reasonable to conclude that the predictions of Proposition 3 are indeed valid.

The results are given in Table H.1. Compared to the baseline resultsin Table 1, there is even a larger effect of port opening on the influence of centrality among unmarried males. The average marginal response in Column 5 is astonishling 0.691, which is more than triple of the result in Table 1. Noticeably, the sample used for regressions in this section has a higher kin group sizes than the baseline sample, whose average is raised from 8.57 to 9.51. It could because small kin groups are more likely to have more unmarried males than married males, such as a father with three sons. The fact further validates the argument that the enhanced effect of unmarried male relatives are not explained by the existence of many small kin groups.

In Columns 3-5, the coefficient of the treatment (Proximity×Post) shifts to an insignificant negative value. This shift could be attributed to the notably smaller proportion of observations with zero centrality among unmarried males in this subset, which stands at only 3.37%, compared to 19.4% in the baseline sample. The reduced fraction is likely a result of excluding smaller kin groups. In these columns, the coefficient of the treatment can be interpreted as the impact of the port opening on individuals with no unmarried male relatives (Column 3) or no male relatives at all (Columns 4-5). The scarcity of such observations could lead to an underestimation of effects. Also, considering the very small number of those with no male relatives, the negative trend may reflect a lower marriage probability for individuals with only a few male relatives, who are likely negatively affected as more people join the fund to share resources, but they have insufficient support to compete effectively.

Table H.1: Changing effects of centrality among married and unmarried males: excluding kin groups with any individual whose centrality among unmarried males ever surpassed that among married males

	Dep. Var.: Marriage next period				
	OLS				Logit
	(1)	(2)	(3)	(4)	(5)
Proximity×Post	0.003	0.037	-0.059	-0.013	-0.052
	(0.008)	(0.043)	(0.032)	(0.058)	(0.085)
Centrality among married males		-0.003		-0.020	-0.065
		(0.063)		(0.051)	(0.593)
Centrality among unmarried males			0.095	0.102	0.275
			(0.105)	(0.091)	(1.592)
Proximity×Centrality among married males		0.047		0.079	0.263**
		(0.050)		(0.050)	(0.531)
Proximity×Centrality among unmarried males			-0.099	-0.129*	-0.544**
			(0.067)	(0.054)	(1.036)
Post×Centrality among married males		0.010		0.032	0.085
		(0.071)		(0.068)	(0.842)
Post×Centrality among unmarried males			-0.072	-0.083	-0.247
			(0.061)	(0.034)	(0.677)
Proximity×Post ×Centrality among married males		-0.049		-0.093	-0.293*
		(0.060)		(0.057)	(0.629)
Proximity×Post ×Centrality among unmarried males			0.165**	0.201***	0.691***
			(0.067)	(0.052)	(0.915)
Village FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Kin Founder FE	✓	✓	✓	✓	✓
Distance time trend	✓	✓	✓	✓	✓
Kin group characteristics	✓	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓	✓
Observations	15,281	15,281	15,281	15,281	12,360
Adjusted R^2	0.098	0.102	0.099	0.099	0.089

Note: Proximity indicates whether the individual belongs to the Near-Port Group. Robust standard errors clustered at the district level in parentheses. Average marginal response and adjusted pseudo R^2 are reported for logit model.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I Model Proposition Proofs

I.1 Lemma I.1

The proof of the main results relies on the following lemma. The statement asserts that an individual will leave the fund if his wealth upon joining is lower than w, and conversely, will choose to join the fund if doing so results in wealth higher than w.

Lemma 1. *Under Assumption 1, M_x does not constitute an equilibrium if*

1. $w_i^{M_x\prime} < w$ *for some* $i \in M_x$*, or*
2. $w_i^{M_x \cup i\prime} \geq w$ *for some* $i \notin M_x$*.*

Proof. At first, I prove that a M_x does not constitute an equilibrium if $w_i^{M_x\prime} < w$ for some $i \in M_x$. Suppose there is such an individual $i \in M_x$ and denote $\Delta w_i = w - w_i^{M_x\prime}$. Then his leaving increases his private utility by $u(w) - u(w - \Delta w_i)$. So he will stay only when his leaving decreases his social utility by more than $u(w) - u(w - \Delta w_i)$. If he leaves, Δw_i will be taken from the fund, which may lower other participants' private utility and hence individual i's social utility. However, the loss of individual i's social utility cannot exceed $\alpha(|M_x| - 1)u(\frac{\Delta w_i}{|M_x|-1})$, which is i's loss in social utility, assuming every other participant are all i's closest relatives, all have wealth allocated equal to $\frac{\Delta w_i}{|M_x|-1}$ before i's leaving, and all have 0 wealth after i's leaving. Individual i will have no higher social utility loss than this, due to the concavity of the utility function.

Hence, i will leave as long as

$$u(w) - u(w - \Delta w_i) \geq \alpha(|M_x| - 1)u(\frac{\Delta w_i}{|M_x| - 1}) \tag{I.1}$$

$$\Rightarrow u(w - \Delta w_i) + \alpha(|M_x| - 1)u(\frac{\Delta w_i}{|M_x| - 1}) \leq u(w) \tag{I.2}$$

Take the derivative of the left-hand side of the inequality (I.2) with respect to Δw_i

to obtain

$$\alpha\lambda e^{-\lambda\frac{\Delta w_i}{|M_x|-1}} - \lambda e^{-\lambda(w-\Delta w_i)} \tag{I.3}$$

$$\leq \lambda e^{-\lambda(w+\frac{\Delta w_i}{|M_x|-1})} - \lambda e^{-\lambda(w-\Delta w_i)} \tag{I.4}$$

$$\leq 0 \quad . \tag{I.5}$$

From I.3 to I.4, Assumption 1 is used. Hence, the LHS of inequality I.2 is decreasing in Δw_i, but when Δw_i reaches its lowest value, 0, it becomes $u(w)$ which is equal to the RHS. Hence, the inequality (I.2) always holds. Then all M_x that yield some $w_i^{M_x\prime} < w$ will not constitute any equilibrium. Therefore, from this point on, all M_x will be assumed to satisfy $w_i^{M_x\prime} \geq w$ for every i. Then proving the second point of Lemma I.1 is equivalent to proving that an individual $i \in M_x$ has no incentive to leave as long as $w_i^{M_x\prime} \geq w$. Denote $\Delta w_i = w_i^{M_x\prime} - w$. We have $\Delta w_i \leq (n-m)w$. This is because the largest surplus i can grab is the sum of the married individual's wealth, as every unmarried participant at least gets w. If he leaves, he losses $u(w+\Delta w_i) - u(w)$ on his private utility. Δw_i might be redistributed among other participants to increase his social utility, however, his social utility will not increase by more than $\alpha(|M_x|-1)(u(w+\frac{\Delta w_i}{|M_x|-1}) - u(w))$, assuming every other participant are all i's closest relatives, all have wealth equal to w before i's leaving, and all obtain $\frac{\Delta w_i}{|M_x|-1}$ by i's leaving. For i to be happy to stay, we need

$$u(w+\Delta w_i) - u(w) \geq \alpha(|M_x|-1)(u(w+\frac{\Delta w_i}{|M_x|-1}) - u(w)) \tag{I.6}$$

$$\Rightarrow u(w+\Delta w_i) - \alpha(|M_x|-1)u(w+\frac{\Delta w_i}{|M_x|-1}) \geq [1-\alpha(|M_x|-1)]u(w) \tag{I.7}$$

Suppose $\alpha < \frac{1}{|M_x|-1}$, i.e., $1-\alpha(|M_x|-1) > 0$, we have

$$u(w+\Delta w_i) - \alpha(|M_x|-1)u(w+\frac{\Delta w_i}{|M_x|-1}) \tag{I.8}$$

$$\geq u(w+\frac{\Delta w_i}{|M_x|-1}) - \alpha(|M_x|-1)u(w+\frac{\Delta w_i}{|M_x|-1}) \tag{I.9}$$

$$\geq [1-\alpha(|M_x|-1)]u(w+\frac{\Delta w_i}{|M_x|-1}) \tag{I.10}$$

$$\geq [1-\alpha(|M_x|-1)]u(w) \tag{I.11}$$

Hence, inequality (I.7) holds. On the other hand, suppose $\alpha \geq \frac{1}{|M_x|-1}$. Take the derivative of the LHS of inequality I.7 with respect to Δw_i, it gives

$$\lambda e^{-\lambda(w+\Delta w_i)} - \lambda\alpha e^{-\lambda(w+\frac{\Delta w_i}{|M_x|-1})} \tag{I.12}$$

$$\geq \lambda e^{-\lambda(w+\Delta w_i)} - \lambda e^{-\lambda w} e^{-\lambda(w+\frac{\Delta w_i}{|M_x|-1})} \tag{I.13}$$

$$\geq \lambda e^{-\lambda(\frac{|M_x|-2}{|M_x|-1}\Delta w_i - w)} \tag{I.14}$$

$$> \lambda e^{-\lambda(\Delta w_i - w)} \tag{I.15}$$

$$\geq \lambda e^{-\lambda(n-m-1)w} \tag{I.16}$$

$$\geq 0 \tag{I.17}$$

From I.13 to I.14, Assumption 1 is used. From I.15 to I.16, the fact that $\frac{|M_x|-2}{|M_x|-1} < 1$ is used. From I.16 to I.17, the fact that $\Delta w_i \leq (n-m)w$ is used. Hence, the LHS of inequality I.7 is increasing in Δw_i, but when Δw_i reaches its lowest value, 0, it becomes the same as the RHS of I.7. Hence, the inequality I.7 always holds. It completes the proof. □

I.2 Proof of Proposition 1

Proof. At first, I prove that equation (1.10) must hold for any participant i. The first-order condition of the FM's problem gives

$$\delta_i e^{-\lambda w_i'} = \delta_j e^{-w_j'} \tag{I.18}$$

Take logs to obtain

$$\ln \delta_i - \lambda w_i' = \ln \delta_j - \lambda w_j', \tag{I.19}$$

and sum over all $i \in M_x$ gives

$$|M_x| \ln \delta_i - \lambda |M_x| w_i' = \sum_{j \in M_x} \ln \delta_j - \lambda \sum_{j \in M_x} w_j' \tag{I.20}$$

$$= \sum_{j \in M_x} \ln \delta_j - \lambda |N_x| w \tag{I.21}$$

Rearrange the equation gives

$$w_i' = \frac{|N_x|}{|M_x|} w - \frac{1}{\lambda |M_x|} \sum_{j \in M_x} \ln \frac{\delta_j}{\delta_i}. \tag{I.22}$$

Rewriting δ_i as $\bar{\delta}_i + \tau_i \hat{\delta}_i$ and δ_j as $\bar{\delta}_j + \tau_i \hat{\delta}_j$ proves equation (1.10).

Lemma I.1 ensures that an individual will join the fund if doing so results in his wealth greater than w and will leave the fund if staying in the fund gives him less than w. The algorithm presented in Section 9.1 guarantees the existence of an equilibrium since the resulting configuration ξ satisfies Definition 1, and it must emerge given that the network is finite. It completes the proof. □

I.3 Proof of Lemma 1

Proof. Suppose w has been raised from w_l to w_h, while the set of unmarried participants remains $M_x^{\xi_1}$. Lemma I.1 implies that regardless of the value of λ, a participant i will not leave the fund as long as $w_i' \geq w$. We have

$$w_i' - w = (\frac{|N_x^\xi|}{|M_x^\xi|} - 1)w - \frac{1}{\lambda}\frac{1}{|M_x^\xi|}\sum_{j\in M_x^\xi} \ln \frac{\bar{\delta}_j + \tau_j^\xi \hat{\delta}_j}{\bar{\delta}_i + \tau_i^\xi \hat{\delta}_i}$$

which increases in w since it is assumed that $|N_x^\xi| > |M_x^\xi|$[23]. So those who are in $M_x^{\xi_1}$ will stay in the fund if the set of participants stay the same.

Hence, if the inclusion of any other individual $j \notin M_x^{\xi_1}$ prompts someone to leave, $M_x^{\xi_1}$ remains an equilibrium. Otherwise, allow those individuals j to join and repeat this process until either no one wants to join or their joining makes someone in the fund want to leave. This will result in a new equilibrium. In either case, the new equilibrium will provide a set of unmarried participants that weakly includes the previous set. This completes the proof. □

I.4 Proof of Proposition 2

Proof. Recall that given an arbitrary equilibrium ξ, $w_i'^\xi$ is given by equation (1.10). Take the derivative of $w_i^{\xi\,\prime}$ with respect to $\hat{\delta}_i$, we get

$$\frac{\partial w_i'^\xi}{\partial \hat{\delta}_i} = \frac{1}{\lambda |M_\xi|}|M_\xi|\frac{1}{\frac{\bar{\delta}_i}{\tau_i^\xi} + \hat{\delta}_i} \tag{I.23}$$

$$= \frac{1}{\lambda}\frac{1}{\frac{\bar{\delta}_i}{\tau_i^\xi} + \hat{\delta}_i} \tag{I.24}$$

Now let us compare this derivative, between $\xi_1 \in \Xi_l$ and $\xi_{1,j} \in \Xi_h$. We have

$$\frac{\partial w_i'^{\xi_1}}{\partial \hat{\delta}_i} = \frac{1}{\lambda_l}\frac{1}{\frac{\bar{\delta}_i}{\tau_i^{\xi_1}} + \hat{\delta}_i} \tag{I.25}$$

[23] We have $n - m > 2$ and married individuals are always participants. So besides unmarried participants, there are at least 2 married participants.

and

$$\frac{\partial w_i'^{\xi_{1,j}}}{\partial \hat{\delta}_i} = \frac{1}{\lambda_h} \frac{1}{\frac{\bar{\delta}_i}{\tau_i^{\xi_{1,j}}} + \hat{\delta}_i} \tag{I.26}$$

By definition, $M_{\xi_1} \subseteq M_{\xi_{1,j}}$. Hence, $\tau_i^{\xi_1} \leq \tau_i^{\xi_{1,j}}$ for every i. Also, $\lambda_h < \lambda_l$ by assumption. It then follows that $\frac{\partial w_i'^{\xi_1}}{\partial \hat{\delta}_i} \leq \frac{\partial w_i'^{\xi_{1,j}}}{\partial \hat{\delta}_i}$, which completes the proof. □

I.5 Proof of Proposition 3

Proof. Taking the derivative of $u(w_i'^{\xi})$ with respect to $w_i'^{\xi}$, we obtain

$$\begin{aligned}
\frac{\partial u(w_i'^{\xi})}{\partial w_i'^{\xi}} &= \lambda e^{-\lambda w_i'^{\xi}} \frac{\partial w_i'^{\xi}}{\partial \hat{\delta}_i} && \text{(I.27)}\\
&= \lambda e^{-\lambda\left(\frac{|N_x^{\xi}|}{|M_x^{\xi}|} w - \frac{1}{\lambda}\frac{1}{|M_x^{\xi}|}\sum_{j\in M_x^{\xi}} \ln \frac{\bar{\delta}_j + \tau_j^{\xi}\hat{\delta}_j}{\bar{\delta}_i + \tau_i^{\xi}\hat{\delta}_i}\right)} \frac{\partial w_i'^{\xi}}{\partial \hat{\delta}_i} && \text{(I.28)}\\
&= \lambda e^{-\lambda\left(\frac{|N_x^{\xi}|}{|M_x^{\xi}|} w - \frac{1}{\lambda}\frac{1}{|M_x^{\xi}|}\sum_{j\in M_x^{\xi}} \ln \frac{\bar{\delta}_j + \tau_j^{\xi}\hat{\delta}_j}{\bar{\delta}_i + \tau_i^{\xi}\hat{\delta}_i}\right)} \frac{1}{\lambda} \frac{1}{\frac{\bar{\delta}_i}{\tau_i^{\xi}} + \hat{\delta}_i} && \text{(I.29)}\\
&= e^{-\lambda \frac{|N_x^{\xi}|}{|M_x^{\xi}|} w}\, e^{\sum_{j\in M_x^{\xi}} \ln \left(\frac{\bar{\delta}_j + \tau_j^{\xi}\hat{\delta}_j}{\bar{\delta}_i + \tau_i^{\xi}\hat{\delta}_i}\right)^{\frac{1}{|M_x^{\xi}|}}} \frac{1}{\frac{\bar{\delta}_i}{\tau_i^{\xi}} + \hat{\delta}_i} && \text{(I.30)}\\
&= \frac{\bigcap_{j\in|M_x^{\xi}|} (\bar{\delta}_j + \tau_j^{\xi}\hat{\delta}_j)}{\frac{\bar{\delta}_i^2}{\tau_i^{\xi}} + 2\hat{\delta}_i\bar{\delta}_i + \tau_i^{\xi}\bar{\delta}_i^2} e^{-\lambda \frac{|N_x^{\xi}|}{|M_x^{\xi}|} w}, && \text{(I.31)}
\end{aligned}$$

where equation (1.10) is used to obtain I.28, and equation I.25 is used to obtain I.29. Recall that when the equilibrium switches from ξ_1 to $\xi_{1,j}$, $\tau_i^{\xi_1} \leq \tau_i^{\xi_{1,j}}$ for every i. Taking the derivative of $\frac{\bar{\delta}_i^2}{\tau_i^{\xi}} + 2\hat{\delta}_i\bar{\delta}_i + \tau_i^{\xi}\bar{\delta}_i^2$ with respect to τ_i^{ξ}, we obtain $-\bar{\delta}_i^2 \tau_i^{\xi\,-2} + \hat{\delta}_i^2$, which is negative given the first condition in Proposition 3. Also, $\bigcap_{j\in|M_x^{\xi}|} (\bar{\delta}_j + \tau_j^{\xi}\hat{\delta}_j)$ increases when the equilibrium switches from ξ_1 to $\xi_{1,j}$, since $\tau_i^{\xi_1} \leq \tau_i^{\xi_{1,j}}$ for every i and $|M_x^{\xi_1}| \leq |M_x^{\xi_{1,j}}|$. So the term $\frac{\bigcap_{j\in|M_x^{\xi}|} (\bar{\delta}_j + \tau_j^{\xi}\hat{\delta}_j)}{\frac{\bar{\delta}_i^2}{\tau_i^{\xi}} + 2\hat{\delta}_i\bar{\delta}_i + \tau_i^{\xi}\bar{\delta}_i^2}$ becomes larger after the opening.

The last thing is to prove that $e^{-\lambda_l \frac{|N_x^{\xi_1}|}{|M_x^{\xi_1}|} w_l} \leq e^{-\lambda_h \frac{|N_x^{\xi_1,j}|}{|M_x^{\xi_1,j}|} w_h} \iff \frac{|N_x^{\xi_1}|}{|M_x^{\xi_1}|} \lambda_l w_l \geq \frac{|N_x^{\xi_1,j}|}{|M_x^{\xi_1,j}|} \lambda_h w_h$. Denote $S = |M_x^{\xi_1,j}| - |M_x^{\xi_1}|$, the inequality becomes

$$\frac{|N_x^{\xi_1}|}{|M_x^{\xi_1}|} \lambda_l w_l \geq \frac{|N_x^{\xi_1}| + S}{|M_x^{\xi_1}| + S} \lambda_h w_h. \tag{I.32}$$

Rearrange it to obtain

$$\frac{\lambda_h}{\lambda_l} \leq \frac{(1 + \frac{S}{|M_x^{\xi_1}|}) w_l}{(1 + \frac{S}{|N_x^{\xi_1}|}) w_h}, \tag{I.33}$$

which is given by the second condition in Proposition 3. So far it is assumed that i is in the fund both before and after the opening. If i joins the fund only after the opening, $\frac{\partial w_i'^{\xi}}{\partial \hat{\delta}_i}$ only increases for such an individual since $\frac{\partial w_i'^{\xi}}{\partial \hat{\delta}_i} = 0$ if $i \notin M_x^{\xi}$, but $\frac{\partial w_i'^{\xi}}{\partial \hat{\delta}_i} \geq 0$ if $i \in M_x^{\xi}$. This completes the proof. □

Appendix to Chapter 2

A A brief introduction to eigenvector centrality and Katz centrality

Consider n individuals (or nodes/vertices in the language of social network analysis) form a network and suppose the network is represented by the $n \times n$ adjacency matrix G such that g_{ij} is the element at ith row and jth column in G. Assume that $g_{ij} \geq 0$ for any i and j but do not impose any other restrictions on g_{ij}. So G can be a listening matrix W as long as $g_{ij} = w_{ij}$ for every i and j.

Eigenvector centrality is one of the measures that characterize how central an individual is in the whole network. The idea is that an individual will be considered more central if one of his neighbours, one of his neighbours' neighbours or so on becomes more central. Eigenvector centrality is defined in a self-referential way. Specifically, let e_i be the eigenvector centrality of individual i and the $n \times 1$ vector $\mathbf{e}$ is the eigenvector centrality associated with the network G. We have $e_i = \sum_{j=1}^{n} \frac{g_{ij}}{\lambda} e_j$ where λ is an eigenvalue of the matrix G. Writing the equation in matrix notation gives

$$\lambda \mathbf{e} = G\mathbf{e} \tag{A.1}$$

If $\lambda = 1$ given that G has an unit eigenvalue, then $\mathbf{e}$ is given by

$$\mathbf{e} = G\mathbf{e} \tag{A.2}$$

and we call such $\mathbf{e}$ *unit eigenvector centrality* of network G. Sometimes one does not become more central if the individuals to whom he is linked to are more central, but because the individuals who are linked to him are more central. These two aspects can be different if $g_{ij} \neq g_{ji}$ for some i and j. In such a case, we may want to define the eigenvector centrality as

$$\lambda \mathbf{e}' = \mathbf{e}'G \tag{A.3}$$

The $\mathbf{e}$ satisfying the equation is called *left-hand eigenvector centrality* of network G. Similarly, the $\mathbf{e}$ satisfying

$$\mathbf{e}' = \mathbf{e}'G \tag{A.4}$$

is called *left-hand unit eigenvector centrality* of network G. In the model, I use the left-hand unit eigenvector centrality of the conversation network W to represent one's influence in shaping the consensus, because at first W is the listening matrix and w_{ij} is the influence j has on i i.e. how much i listens to j. So it makes sense to use $e_i = \sum_{j=1}^{n} \frac{g_{ji}}{\lambda} e_j$ as the influence i has on others, directly or indirectly. Moreover, unit eigenvector centrality is used because of the fact that there is a unique eigenvector of W that has nonnegative values if W is strongly connected, aperiodic and row stochastic i.e. $\sum_{j=1}^{n} w_{ij} = 1$.[24]

B Proof of Proposition 1

Proof. At first, I prove that in any equilibrium of any time period, if it exists, an individual uses a cutoff strategy. The expected payoff as in equation (2.5) shows that an individual i will never adopt if $b_i^t < -\frac{\phi(n-1)}{k}$ and will always adopt if $b_i^t > 0$. Denote $-\frac{\phi(n-1)}{k}$ by $\underline{b}$. Then I show that there are no such points denoted by e_1 and e_2 such that

1. $(e_1, e_2) \in [\underline{b}, 0]$ and
2. an individual i will adopt if $b_i^t \in (\underline{b}, e_1)$ but not if $b_i^t \in (e_1, e_2)$.

[24] see section 8.3.5 of Jackson (2010).

To start, it is important to observe that the above argument, combined with the continuity of expected utility, implies that an individual i will receive zero payoffs if b_i^t equals e_1, or if b_i^t equals $\underline{b}$, i.e.,

$$k\underline{b} + \phi \sum_{j \in N \setminus i} \int_{x \in B(s_j^{\sigma^t})} f_{\underline{b},\sigma_{i,j}^t}(x)dx] = 0 \tag{B.1}$$

and

$$ke_1 + \phi \sum_{j \in N \setminus i} \int_{x \in B(s_j^{\sigma^t})} f_{e_1,\sigma_{i,j}^t}(x)dx] = 0 \tag{B.2}$$

Since $k\underline{b} < ke_1$, combining equation (B.1) and (B.2) gives

$$\phi \sum_{j \in N \setminus i} \int_{x \in B(s_j^{\sigma^t})} f_{e_1,\sigma_{i,j}^t}(x)dx] < \phi \sum_{j \in N \setminus i} \int_{x \in B(s_j^{\sigma^t})} f_{\underline{b},\sigma_{i,j}^t}(x)dx] \tag{B.3}$$

There must be some j such that

$$\int_{x \in B(s_j^{\sigma^t})} f_{e_1,\sigma_{i,j}^t}(x)dx] < \int_{x \in B(s_j^{\sigma^t})} f_{\underline{b},\sigma_{i,j}^t}(x)dx] \tag{B.4}$$

If the above inequality holds, the probability measure of $B(s_j^{\sigma^t})$ will have to be larger with a normal distribution represented by $f_{e_1,\sigma_{i,j}^t}$ than with a normal distribution represented by $f_{\underline{b},\sigma_{i,j}^t}$. Now divide $B(s_j^{\sigma^t})$ into two parts which are denoted as $B_1(s_j^{\sigma^t}) := \{x : x \in B(s_j^{\sigma^t}) \wedge x \in (\underline{b}, e_1)\}$ and $B_2(s_j^{\sigma^t}) := \{x : x \in B(s_j^{\sigma^t}) \wedge x \in (e_2, 0)\}$ respectively. Recall that $F_{\underline{b},\sigma_{i,j}^t}$ and $F_{e_1,\sigma_{i,j}^t}$ are the cdfs corresponding to $f_{\underline{b},\sigma_{i,j}^t}$ and $f_{e_1,\sigma_{i,j}^t}$ respectively. Since they correspond to normal distributions with mean $\underline{b}$ and e_1 respectively, we have $F_{e_1,\sigma_{i,j}^t}(e_1) - F_{e_1,\sigma_{i,j}^t}(\underline{b}) = F_{\underline{b},\sigma_{i,j}^t}(e_1) - F_{\underline{b},\sigma_{i,j}^t}(\underline{b})$, since they have the same variance. So from individual i's perspective, $Pr(b_j^t \in B_1(s_j^{\sigma^t}) | b_i^t = \underline{b}) = Pr(b_j^t \in B_1(s_j^{\sigma^t}) | b_i^t = e_1)$. Regarding $B_2(s_j^{\sigma^t})$, from individual i's perspective, $Pr(b_j^t \in B_2(s_j^{\sigma^t}) | b_i^t = \underline{b}) < Pr(b_j^t \in B_2(s_j^{\sigma^t}) | b_i^t = e_1)$, since all points in $B_2(s_j^{\sigma^t})$ are closer to e_1 than to $\underline{b}$ and $e_1 > \underline{b}$. Then we have $\int_{x \in B(s_j^{\sigma^t})} f_{e_1,\sigma_{i,j}^t}(x)dx] > \int_{x \in B(s_j^{\sigma^t})} f_{\underline{b},\sigma_{i,j}^t}(x)dx]$. Hence, there is a contradiction and the inequality $B.4$ does not hold. It is still possible that individual i does not adopt if $b_i^t < e_2$, but adopts if $b_i^t \in (e_2, e_3)$ and does not adopt if $b_i^t \in (e_3, e_4)$ where $e_4 \leq 0$. However, if so, we can use e_2 as $\underline{b}$ and repeat the above steps to show that this is never the case. Moreover, the continuity of the expected utility guarantee that there will be no belief such that individual i will adopt if b_i^t is in an interval containing such a belief but not adopt if b_i^t is equal to that belief, or the other way round. Hence, cutoff strategies will be used. Denoted by c_i^t the cutoff strategy individual i uses at time t.

As mentioned earlier, c_i^t must be between $-\frac{\phi(n-1)}{k}$ and 0 as an individual will adopt for sure if his or her belief is above 0 and not adopt for sure if his or her belief is below $-\frac{-\phi(n-1)}{k}$. Hence, $(c_1^t, ..., c_n^t) \in [-\frac{\phi(n-1)}{k}, 0]^n$. Therefore, a profile of best responses is a continuous function mapping a compact convex set of cutoffs to itself. Then applying Brouwer fixed-point theorem completes the proof.

□

C Proof of Corollary 1

Proof. By definition, a cutoff equilibrium $\mathbf{c}^{\sigma^t} := (c_1^{\sigma^t}, ..., c_n^{\sigma^t})$ solves

$$kc_i^{\sigma^t} + \phi \sum_{j \in N \setminus i} (1 - F_{c_i^{\sigma^t}, \sigma_{i,j}^t}(c_j^{\sigma^t})) = 0 \tag{C.1}$$

for all i. Notice that $c_i^{\sigma^t}$ has superscript σ^t only because it belongs to a solution to the above system of equations which includes σ^t, and this is the only reason why cutoffs depend on t. So as t changes, $c_i^{\sigma^t}$ changes only because σ^t changes. In order to avoid confusion, I drop the superscript and denote a solution to the above system of equations (hence a cutoff equilibrium at that time t) by $(c_1, ..., c_n)$. I tend to show that as $t \to \infty$, it is impossible to find any i and j such that $c_i > c_j$. To begin with, without loss of generality suppose $c_1 > c_2$. Rewrite equation (C.1) for individual 1 and 2 as

$$kc_1 + \phi(1 - F_{c_1, \sigma_{1,2}^t}(c_2)) + \phi \sum_{j \in N \setminus (1 \cup 2)} (1 - F_{c_1, \sigma_{1,j}^t}(c_j)) = 0 \tag{C.2}$$

and

$$kc_2 + \phi(1 - F_{c_2,\sigma^t_{2,1}}(c_1)) + \phi \sum_{j\in N\backslash(1\cup 2)} (1 - F_{c_2,\sigma^t_{2,j}}(c_j)) = 0 \tag{C.3}$$

Since $\lim_{t\to\infty}\sigma^t_{i,j} = \lim_{t\to\infty}\sigma^t_{k,z}$ for any i,j k and z, $\lim_{t\to\infty}[\phi\sum_{j\in N\backslash(1\cup 2)}(1 - F_{c_1,\sigma^t_{1,j}}(c_j)) - \phi\sum_{j\in N\backslash(1\cup 2)}(1 - F_{c_2,\sigma^t_{2,j}}(c_j))] > 0$ because $c_1 > c_2$. Due to the same reason, $\lim_{t\to\infty}[\phi(1 - F_{c_1,\sigma^t_{1,2}}(c_2)) - \phi(1 - F_{c_2,\sigma^t_{2,1}}(c_1))] > 0$. At last, $kc_1 > kc_2$. Hence, the RHS of equation (C.2) is strictly larger than the RHS of equation (C.3) as $t\to\infty$, which leads to a contradiction since they are supposed to be equal if equation (C.2) and (C.3) are satisfied. Therefore, in the limiting equilibrium, every individual must use the same cutoff. Suppose the cutoff in the limit used by every individual is denoted by c^*. c^* must satisfy

$$\lim_{t\to\infty}[kc^* + \phi \sum_{j\in N\backslash i} (1 - F_{c^*,\sigma^t_{i,j}}(c^*))] = 0 \tag{C.4}$$

for every i. Obviously, $F_{c^*,\sigma^t_{i,j}}(c^*)) = \frac{1}{2}$ regardless of the value of $\sigma^t_{i,j}$. Hence, equation (C.4) can be rewritten as

$$kc^* + \phi\frac{n-1}{2} = 0 \tag{C.5}$$

Solving equation (C.5) gives $c^* = -\frac{\phi(n-1)}{2k}$. □

D Proof of Proposition 2

Proof. Recall that we denote $\sum_{i\in N} p_i^2\sigma$ by $\bar{\sigma}$. Define $\bar{p} := \frac{\sum_{i\in N} p_i}{n}$ and $s_p = \frac{\sum_{i\in N}(p_i-\bar{p})^2}{n}$. The first step is to prove that $\bar{\sigma}$ is increasing in s_p. By definition $\bar{\sigma} = (p_1^2 + ... + p_n^2)\sigma$. $s_p = \frac{p_1^2+...+p_n^2-2(p_1+...+p_n)+n\bar{p}^2}{n}$. Since $p_1 + ... + p_n = 1$, $s_p = \frac{p_1^2+...+p_n^2-2+\frac{1}{n}}{n}$. Hence $\bar{\sigma} = (n * s_p - \frac{1}{n} + 2)\sigma$ which is increasing in s_p.

It then follows that s_e is increasing in $\bar{\sigma}$, because s_e is just s_p, sine $\mathbf{p}$ is just the unit eigenvector centrality of W. Notice that we must have $\mathbf{p}'W\mathbf{b}^* = \mathbf{p}'\mathbf{b}^*$ for any $\mathbf{b}^*$. To see this, notice that $W\mathbf{b}^* = W^*\mathbf{b}^*$, so $\mathbf{p}'W\mathbf{b}^* = \mathbf{p}'W^*\mathbf{b}^*$. $\mathbf{p}'$ is any one of rows of W^* (each row of W^* is the same), and all p_i are sumed up to 1, so

$\mathbf{p}'W\mathbf{b}^* = \mathbf{p}'W^*\mathbf{b}^* = \mathbf{p}'\mathbf{b}^*$ and so $\mathbf{p}'W = \mathbf{p}'$. It then proves that $\bar{\sigma}$ is strictly increasing in s_e.

Recall that we have $b^* \sim N(\theta, \bar{\sigma})$. Hence, if $\theta > -\frac{\phi(n-1)}{2k}$, $Pr(b^* > -\frac{\phi(n-1)}{2k}|W, \theta) := 1 - F_{\theta,\bar{\sigma}}(-\frac{\phi(n-1)}{2k})$ is decreasing in $\bar{\sigma}$, and so decreasing in s_e, and if $\theta < -\frac{\phi(n-1)}{2k}$, this probability is increasing in $\bar{\sigma}$, and so increasing in s_e. Recall that the expected number of adopters is $n[1 - F_{\theta,\bar{\sigma}}(-\frac{\phi(n-1)}{2k})]$. So the expected number of adopters is increasing in s_e if $\theta < -\frac{\phi(n-1)}{2k}$ and decreasing in s_e if $\theta > -\frac{\phi(n-1)}{2k}$. According to equation (2.11), we know that $\bar{U}$ is negative when $\theta < -\frac{\phi(n-1)}{k}$ and positive when $\theta > -\frac{\phi(n-1)}{k}$. If $\bar{U}$ is positive, it is increasing in $1 - F_{\theta,\bar{\sigma}}(-\frac{\phi(n-1)}{k})$, which increases in $\bar{\sigma}$ as well as s_e if and only if $\theta < -\frac{\phi(n-1)}{2k}$. If $\bar{U}$ is negative, it is decreasing in $1 - F_{\theta,\bar{\sigma}}(-\frac{\phi(n-1)}{k})$ and the relationship between $1 - F_{\theta,\bar{\sigma}}(-\frac{\phi(n-1)}{k})$ and s_e is the same. It completes the proof.

□

E Proof of Lemma 2

Proof. Lemma 1 makes it clear that this result holds for any closed communicating class since a closed communicating class itself, if considered as a conversation network, satisfies Assumption 1. Then the task is to show that for a remaining node i, $\lim_{t\to\infty} b_i^t = \sum_{m\in M} w_{lm} b_m^*$, i.e. the beliefs of a remaining node also converge and it becomes a weighted average of other consensuses. By definition, a remaining node i's belief at time t is $b_i^t = \sum_{j\in N} w_{ij} b_j^{t-1}$. $\sum_{j\in N} w_{ij} b_j^{t-1}$ can be divided into two parts: the part from closed communicating classes and the part from other remaining nodes respectively. The former can be denoted by $\sum_{m\in M_c}\sum_{j\in N_m} w_{ij} b_j^{t-1}$. Obviously,

$$\lim_{t\to\infty} \sum_{m\in M_c}\sum_{j\in N_m} w_{ij} b_j^{t-1} \tag{E.1}$$

$$= \sum_{m\in M_c}\sum_{j\in N_m} w_{ij} b_m^* \tag{E.2}$$

$$= \sum_{m\in M_c} w_{im} b_m^* \tag{E.3}$$

where $w_{im} = \sum_{j \in N_m} w_{ij}$. On the other hand, the fraction of b_i^t from other remaining nodes can be denoted by $b_i^t = \sum_{j \in M_r} w_{ij} b_j^{t-1}$, recall that M_r is the set of all remaining nodes. It is possible for a remaining node j to receive information from other remaining nodes as well. I aim to show that b_i^t is a combination of either past beliefs from closed communicating classes or initial beliefs from other remaining nodes. The following algorithm provides a clear explanation of this concept.

1. Consider consider time period t and a specific remaining node i. Select a consensus class denoted as l_1, which individual i listens to. In other words, if l_1 represents another remaining node, then $w_{il_1} > 0$. Alternatively, if l_1 represents a closed communicating class, then $\sum_{j \in N_{l_1}} w_{ij} > 0$. It is also possible for l_1 to refer to node i itself.

2. Repeat Step 1 to generate a sequence $l_1, \ldots, l_K$, stopping when either l_K corresponds to a closed communicating class or when $t = K$.

   It is worth noting that if, for some $k < K$, l_k is a remaining node who only listens to him or herself, i.e., $w_{l_k j} = 0$ for all $j \neq l_k$, then all consensus classes in the sequence following l_k must also be the node itself, and we have $K = t$. A sequence generated by the aforementioned algorithm illustrates the transmission of information from various "original sources" to the remaining node i. All such sequences for node i describe all the sources from which the belief b_i^t originates, as t approaches infinity. Considering that the conversation network is assumed to be finite, with a finite number of nodes denoted by n, the total count of sequences that can be generated for node i is also finite. Denote the set of all such sequences for node i as S. Since the node i can be arbitrary, I do not include it as a superscript or subscript in S. An arbitrary sequence for node i is $s \in S$. We can rewrite the sequence s as $l_1^s, ..., l_K^s$. The contribution from l_K^s to b_i^t can be expressed as $w_{il_1^s} \Pi_{k=2}^{K-2} w_{l_k^s l_{k+1}^s} \sum_{j \in N_{l_K^s}} w_{l_{K-1}^s j} b_j^{t-K}$. Please note that a closed communicating class does not listen to other consensus classes. Therefore, all l_k in the sequence will correspond to remaining nodes if $k < K$. The weight $w_{l_{K-1}^s j}$ represents the influence that j in the consensus class l_K^s exerts on the remaining node l_{K-1}. If l_K^s is a closed communicating class, the belief b_k^{t-K} converges to $b_{l_K^s}^*$. On the other hand, if l_K^s is a remaining node, according to the algorithm, it must be the case that $K = t$. Consequently, $b_{l_K^s}^{t-K}$ is the initial belief $b_{l_K^s}^0$. So it is either

$$lim_{t\to\infty} w_{il_1^s} \Pi_{k=2}^{K-2} w_{l_k^s l_{k+1}^s} \sum_{j\in N_{l_K^s}} w_{l_{K-1}^s j} b_j^{t-K} = w_{il_K^s}^s b_{l_K^s}^* \tag{E.4}$$

or

$$\lim_{t\to\infty} w_{il_1^s} \Pi_{k=2}^{K-2} w_{l_k^s l_{k+1}^s} \sum_{j\in N_{l_K^s}} w_{l_{K-1}^s j} b_j^{t-K} = w_{il_K^s}^s b_{l_K^s}^0 \tag{E.5}$$

where $w_{il_K^s}^s = w_{il_1^s} \Pi_{k=2}^{K-2} w_{l_k^s l_{k+1}} \sum_{j\in N_{l_K^s}} w_{l_{K-1}^s j}$. In particular, $E.5$ is positive only when all $w_{l_k^s l_{k+1}}$ are equal to 1 beyond a certain cutoff for k. Otherwise, as t approaches infinity, the term $\Pi_{k=2}^{K-2} w_{l_k^s l_{k+1}^s}$ converges to 0, since in this case $t = K$, so K approaches infinity as well. The belief b_i^* then can be computed by summing up $w_{il_K^s}^s b_{l_K^s}^*$ or $w_{il_K^s}^s b_{l_K^s}^0$ over all sequences $s \in S$. This completes the proof. □

F Proof of Lemma 3

Proof. The first part of the lemma can be easily derived from Lemma 1. As for the second part, if l is a remaining node, we have $b_l^* = \sum_{m\in M_c} w_{lm} b_m^* + \sum_{g\in M_r} w_{lg} b_g^0$, as proven in the proof of Lemma 3. It is then evident that the variance of b_l^* aligns with what Lemma 3 states. □

G Proof of Proposition 3

Proof. Proposition 1 is still applicable in the extended model as it relies solely on Assumption 2. Therefore, an equilibrium exists for any given time period t. First, I intended to prove that individuals within the same consensus class adopt the same cutoff strategy in the long run, i.e., $c_i^* = c_j^*$ if $i, j \in N_l$ and $l \in M$.

This proof is only necessary for closed communicating classes, as a remaining node is the only individual in his or her consensus class. Still, a cutoff equilibrium

$\mathbf{c}^{\sigma^t} := (c_1^{\sigma^t}, ..., c_n^{\sigma^t})$ solves

$$kc_i^{\sigma^t} + \phi \sum_{j \in N \setminus i} (1 - F_{c_i^{\sigma^t}, \sigma_{i,j}^t}(c_j^{\sigma^t})) = 0 \tag{G.1}$$

For simplicity, denote a solution to the system of equations described above, which represents a cutoff equilibrium at time t, by $(c_1, c_2, ..., c_n)$, as shown in the proof of Corollary 1. Consider the scenario where $c_1 > c_2$ and individuals 1 and 2 belong to the same closed communicating class l. In this case, it must hold true that

$$kc_1 + \phi(1 - F_{c_1, \sigma_{1,2}^t}(c_2)) + \phi \sum_{j \in N_l \setminus (1 \cup 2)} (1 - F_{c_1, \sigma_{1,j}^t}(c_j)) + \phi \sum_{z \notin N_l} (1 - F_{c_1, \sigma_{1,z}^t}(c_z)) = 0 \tag{G.2}$$

and

$$kc_2 + \phi(1 - F_{c_2, \sigma_{2,1}^t}(c_1)) + \phi \sum_{j \in N_l \setminus (1 \cup 2)} (1 - F_{c_2, \sigma_{2,j}^t}(c_j)) + \phi \sum_{z \notin N_l} (1 - F_{c_2, \sigma_{2,z}^t}(c_z)) = 0 \tag{G.3}$$

The difference between this proof and the proof for Corollary 1 is that we have two additional terms, namely $\phi \sum_{z \notin N_l} (1 - F_{c_1, \sigma_{1,z}^t}(c_z))$ and $\phi \sum_{z \notin N_l} (1 - F_{c_2, \sigma_{2,z}^t}(c_z))$, in equation (G.2) and (G.3), respectively, as compared to equation (C.2) and (C.3). However, it can be shown that in any equilibrium, $\lim_{t \to \infty} [\phi \sum_{z \notin N_l} (1 - F_{c_1, \sigma_{1,z}^t}(c_z)) - \phi \sum_{z \notin N_l} (1 - F_{c_2, \sigma_{2,z}^t}(c_z))] > 0$. This is due to the fact that $c_1 > c_2$ and $\sigma_{1,j}^t$ converges to the same value as $\sigma_{2,j}^t$ as $t \to \infty$, for every j. By following the same steps as in the proof for Corollary 1, we can conclude that two individuals in the same consensus class must have the same cutoff in any limiting equilibrium. To establish the uniqueness of the equilibrium in the limit, I proceed with the following proof. Given that each consensus class has only one cutoff strategy in the limit, we can represent the cutoff equilibrium as $(c_l^* : l \in M)$, which denotes the profile of cutoff strategies in the limit for all consensus classes. It must hold that

$$kc_l^* + \phi\frac{(n_l-1)}{2} + \phi\sum_{m\in M\setminus l} n_m(1-F_{c_l^*,\bar{\sigma}_{l,m}}(c_m^*)) = 0 \tag{G.4}$$

is satisfied for all l, where $1-F_{c_l^*,\bar{\sigma}_{l,m}}(c_m)$ is multiplied by n_m because every individual in m adopts with the same probability from the perspective of an individual in l. Now, the question arises: What will be the best response of consensus class l if another consensus class m has changed its cutoff strategy c_m^*? Implicit differentiation gives

$$\begin{aligned}
\frac{\partial c_l^*}{\partial c_m^*} &= \frac{\phi n_m f_{c_l^*,\bar{\sigma}_{l,m}}(c_m^*)}{k-\phi\sum_{z\in M\setminus l} n_z\frac{\partial F_{c_l^*,\bar{\sigma}_{l,z}}(c_z^*))}{\partial c_l^*}} && \text{(G.5)}\\
&< \frac{\phi n_m f_{c_l^*,\bar{\sigma}_{l,m}}(c_m^*)}{k} && \text{(G.6)}\\
&= \frac{\phi n_m}{k\sqrt{2\pi\bar{\sigma}_{l,m}}} e^{-\frac{(c_m^*-c_l^*)^2}{2\bar{\sigma}_{l,m}}} && \text{(G.7)}\\
&\le \frac{\phi n_m}{k\sqrt{2\pi\bar{\sigma}_{l,m}}} && \text{(G.8)}\\
&< \frac{n_m}{n-n_l} && \text{(G.9)}
\end{aligned}$$

The equation (G.6) is derived from the observation that $F_{c_l^*,\bar{\sigma}_{l,z}}(c_z^*)$ always decreases in c_l^*, given c_m^*. Equation (G.8) follows from the fact that $e^{-\frac{(c_m^*-c_l^*)^2}{2\bar{\sigma}_{l,m}}} \le 1$. Equation (G.9) is obtained by applying Assumption 5. Based on these observations, we can conclude that the total differentiation of c_l^* with respect to the cutoff strategies of other consensus classes is smaller than 1. By applying the Banach fixed point theorem, we can establish the existence of a unique limiting equilibrium.

To complete the argument, we still need to show that if a closed communicating class is larger than another, the larger class will have a smaller cutoff. Assume that

$n_m > n_l$ and $c_m^* > c_l^*$. It must hold that:

$$kc_l^* + \phi\frac{(n_l-1)}{2} + \phi n_m(1 - F_{c_l^*,\bar{\sigma}_m}(c_m^*)) + \phi \sum_{z\in M\setminus(l\cup m)} n_z(1 - F_{c_l^*,\bar{\sigma}_z}(c_z^*)) = 0 \tag{G.10}$$

and

$$kc_m^* + \phi\frac{(n_m-1)}{2} + \phi n_l(1 - F_{c_m^*,\bar{\sigma}_l}(c_l^*)) + \phi \sum_{z\in M\setminus(l\cup m)} n_z(1 - F_{c_z^*,\bar{\sigma}_z}(c_z^*)) = 0 \tag{G.11}$$

Note that the beliefs between two different closed consensus classes are independent. Therefore, we have $\bar{\sigma}_{l,m} = \bar{\sigma}_m$ for any closed communicating classes l and m. By subtracting the left-hand side of equation (G.11) from the left-hand side of equation (G.10), we obtain the following expression:

$$k(c_m^* - c_l^*) + \frac{\phi}{2}(n_m - n_l) + \phi[(1 - F_{c_m^*,\bar{\sigma}_l}(c_l^*))n_l - (1 - F_{c_l^*,\bar{\sigma}_m}(c_m^*))n_m] \tag{G.12}$$

$$> k(c_m^* - c_l^*) + \frac{\phi}{2}(n_m - n_l) + \frac{\phi}{2}(n_l - n_m) \tag{G.13}$$

$$= k(c_m^* - c_l^*) \tag{G.14}$$

$$> 0 \tag{G.15}$$

$(G.13)$ arises from the observation that when $c_m^* > c_l^*$, we have $F_{c_m^*,\bar{\sigma}_l}(c_l^*) < \frac{1}{2}$ and $F_{c_l^*,\bar{\sigma}_m}(c_m^*) > \frac{1}{2}$. Consequently, it is impossible for $c_m^* > c_l^*$ to hold when $n_m > n_l$, as $(G.12)$ would not be zero while $(G.10)$ and $(G.11)$ are satisfied. This leads to a contradiction. Now, we need to show that if $n_m = n_l$, it must be the case that $c_m^* = c_l^*$. Suppose that $n_m = n_l$ but $c_m^* > c_l^*$. In such a scenario, $(G.12)$ would become:

$$k(c_m^* - c_l^*) + \phi n_l[F_{c_l^*,\bar{\sigma}_m}(c_m^*) - F_{c_m^*,\bar{\sigma}_l}(c_l^*)] \tag{G.16}$$

$k(c_m^* - c_l^*)$ is positive because $c_m^* > c_l^*$. $F_{c_l^*,\bar{\sigma}_m}(c_m^*) - F_{c_m^*,\bar{\sigma}_l}(c_l^*)$ is positive because $F_{c_l^*,\bar{\sigma}_m}(c_m^*) > \frac{1}{2}$ while $F_{c_m^*,\bar{\sigma}_l}(c_l^*) < \frac{1}{2}$, given $c_m^* > c_l^*$. Hence, $(G.16)$ is positive. Therefore, if $n_m = n_l$, it follows that $c_m^* = c_l^*$; otherwise, it would lead to a contradiction. This completes the proof.

□

H Proof of Proposition 4

Proof. According to Proposition 3, when $n_l = n_m$, it must hold that $c_l^* = c_m^*$. Given the profile of cutoff strategies for all consensus classes in a limiting equilibrium, the following equation must be satisfied

$$kc_l^* + \phi(2n_l - 1) + \phi \sum_{z \in M \setminus (l \cup m)} n_z(1 - F_{c_l^*, \bar{\sigma}_z}(c_z^*)) = 0 \tag{H.1}$$

Here, I am using the fact that $\bar{\sigma}_{l,m} = \bar{\sigma}_m$ if l is a closed communicating class, and $\phi n_m(1 - F_{c_l^*, \bar{\sigma}_m}(c_m^*)) = \frac{1}{2}\phi n_l$ when $n_l = n_m$ and $c_l^* = c_m^*$. As a result, the solution for c_l^* becomes dependent on all other c_m^* values, making it difficult to obtain a closed-form solution for c_l^*. However, since the left-hand side of $(H.1)$ is continuous and strictly increasing in c_l^*, it must hold that $c_l^* < (>) \theta$ if

$$k\theta + \phi \frac{(2n_l - 1)}{2} + \phi \sum_{m \in M \setminus (l \cup m)} n_m(1 - F_{\theta, \bar{\sigma}_m}(c_m^*)) > (<) 0 \tag{H.2}$$

. It can be shown that

$$k\theta + \phi \frac{(2n_l - 1)}{2} + \phi \sum_{m \in M \setminus (l \cup m)} n_m(1 - F_{\theta, \bar{\sigma}_m}(c_m^*)) \tag{H.3}$$

$$> k\theta + \phi \frac{(2n_l - 1)}{2} \tag{H.4}$$

which is larger than 0 if $\theta > -\frac{\phi(2n_l - 1)}{2k}$. Hence, if $\theta > -\frac{\phi(2n_l - 1)}{2k}$, the cutoff strategy of both consensus classes l and m in the limit is smaller than θ. This implies that $\alpha_l < \alpha_m$ since $s_e^l > s_e^m$, as Proposition 2 establishes. Additionally, because $(H.4)$ is greater than zero, we have $\bar{U}_l < \bar{U}_m$. This is because we must have $0 < k\theta + \frac{\phi(2n_l - 1)}{2} < k\theta + \phi(n_l - 1) + \phi n_l \alpha_l$, since $\alpha_l = Pr(b_l^* \geq c_l^* | \theta, W) > \frac{1}{2}$ given $c_l^* < b_l^*$. The term $\phi n_l \alpha_l$ represents the ex-post payoff an individual in N_l will receive from the adoption of individuals in m. (note that $n_m = n_l$ and $\alpha_m = \alpha_m$) Therefore, when only considering the adoption of individuals within consensus classes l and m, the expected ex-post payoff of adopting for an individual in consensus class l is already positive. The positivity of the expected ex-post payoff ensures that the expected total utility of consensus class l increases with α_l. This proves the first part of Proposition 4. Similarly, we can show that

$$k\theta + \phi\frac{(2n_l - 1)}{2} + \phi \sum_{m \in M \setminus (l \cup m)} n_m(1 - F_{\theta,\bar{\sigma}_m}(c_m^*)) \tag{H.5}$$

$$< k\theta + \phi\frac{(2n_l - 1)}{2} + \phi(n - 2n_l) \tag{H.6}$$

$$< k\theta + \phi(n - n_l - \frac{1}{2}) \tag{H.7}$$

which is negative when $\theta < -\frac{\phi(n-n_l-\frac{1}{2})}{k}$. Hence, if $\theta < -\frac{\phi(n-n_l-\frac{1}{2})}{k}$, we have $c_l^* > \theta$. This verifies the second part of Proposition 4. However, it is important to note that the negativity of $(H.5)$ here does not guarantee the negativity of the expected ex-post payoff, as the latter depends on the cutoffs of other consensus classes. However, we can establish that the ex-post payoff is always negative if $k\theta + \phi(n-1) < 0$, which holds true when $\theta < -\frac{\phi(n-1)}{k}$. In other words, adoption is not beneficial even if everyone else does. Consequently, the expected total utility of consensus class l is negative if $\theta < -\frac{\phi(n-1)}{k}$, indicating a decreasing trend with respect to α_l. This concludes the third part of Proposition 4. □

Appendix to Chapter 3

A Proof of Lemma 2

Proof. At first we prove that with assumption 1, if the l type household is going to specialize in one production sector, it will always choose rice rather than wheat. Suppose the household l specializes in rice, its problem reduces to:

$$\max_{L_{fl}^r, L_{ml}^r} (A_r\gamma+\bar{L})L_{ml}^r + A_r L_{fl}^r - c_i(\frac{1}{2}-L_{fl}^r)^2 - \tau \tag{A.1}$$

s.t.

$$L_{fl}^r + L_{ml}^r = 1 \tag{A.2}$$

Solve this problem we have:

$$L_{ml}^{r\,*} = \frac{(\gamma-1)A_r+\bar{L}}{2c_l} + \frac{1}{2} \tag{A.3}$$

$$L_{fl}^{r\,*} = \frac{1}{2} - \frac{(\gamma-1)A_r+\bar{L}}{2c_l} \tag{A.4}$$

Denote the utility of l household with this solution by u_l^r, we have:

$$u_l^r = (A_r\gamma+\bar{L})(\frac{(\gamma-1)A_r+\bar{L}}{2c_l}+\frac{1}{2}) + A_r(\frac{1}{2}-\frac{(\gamma-1)A_r+\bar{L}}{2c_l}) - c_l(\frac{(\gamma-1)A_r+\bar{L}}{2c_l})^2 - \tau \tag{A.5}$$

Similarly, we can find the utility of l household specializing in wheat is:

$$u_l^w = A_w\gamma(\frac{(\gamma-1)A_w}{2c_l}+\frac{1}{2})+A_w(\frac{1}{2}-\frac{(\gamma-1)A_w}{2c_l})-c_l(\frac{(\gamma-1)A_w}{2c_l})^2 \tag{A.6}$$

We need to find a sufficient condition for $u_l^r - u_l^w > 0$. It is easy to show that $\frac{\partial u_l^r}{\bar{L}} > 0$, so if $u_l^r - u_l^w > 0$ for some low $\bar{L}$, it must also be the case when $\bar{L}$ increases. By denifition, we have $\bar{L} \geq \frac{(\gamma-1)A_r+\bar{L}}{4c_l}+\frac{1}{4}$, so $\bar{L} > \frac{(\gamma-1)A_r}{4c_l}$. With assumption 1 we have $\bar{L} > (\gamma-1)A_r$. So we have

$$u_l^r - u_l^w > (2\gamma-1)\frac{(\gamma-1)A_r^2}{2c}+\frac{1}{2}(2\gamma-1)A_r+\frac{1}{2}A_r-\frac{(\gamma-1)A_r^2}{2c}-c(\frac{(\gamma-1)A_r}{2c})^2 \tag{A.7}$$

$$-\tau-(\frac{\gamma(\gamma-1)A_w^2}{2c}+\frac{1}{2}A_w\gamma+\frac{1}{2}A_w-\frac{(\gamma-1)A_w^2}{2c}-c(\frac{(\gamma-1)A_w}{2c})^2) \tag{A.8}$$

$$=(\frac{1}{2}(2\gamma-1)A_r+\frac{1}{2}A_r-\frac{1}{2}A_w\gamma-\frac{1}{2}A_w-\tau)c+\frac{3(\gamma-1)^2}{4}A_r^2-\frac{(\gamma-1)^2}{4}A_w^2 \tag{A.9}$$

For l specializes in rice, we require

$$u_l^r - u_l^w > 0 \tag{A.10}$$

$$\Rightarrow (\gamma A_r - \frac{1}{2}(\gamma+1)A_w-\tau)c > \frac{(\gamma-1)^2}{4}A_w^2-\frac{3(\gamma-1)^2}{4}A_r^2 \tag{A.11}$$

The RHS is always negative with the first part of assumption 1. So if $(\gamma A_r - \frac{1}{2}(\gamma+1)A_w-\tau)$ is negative, it gives a reasonable condition for c. Otherwise it always hold since $c > 0$. So as long as assumption 1 holds, if the l household is going to specialize, it will always specialize in rice rather than wheat.

It is trivial to show that l will not have males on wheat and females on rice, since it makes l household pay τ but receive a lower return on rice. Putting tighter, we have that l household will always have its males on rice. □

B Proof of Proposition 2

Proof. Firstly, suppose $A_w > A_r$. Given the households' problem, it is obvious to see that a household will allocate all females on producing wheat if $A_w > A_r$, since if the household participates in rice production, its females will not enjoy the benefit from collaboration $\bar{L}$, however get a lower wage than A_w. Since we know that l type household always have males working on rice, there are only two possible cases: 1. Both types of households have their males on rice and females on wheat. 2. All on wheat except l type's males.

Case 1 emerges when $\bar{L}$ is sufficiently high so both types will put male labor into rice production. In such case, we have $SR = \frac{\gamma A_r - A_w + \bar{L}}{2c_h} + \frac{\gamma A_r - A_w + \bar{L}}{2c_l}$. We then have $SR'_{A_r} = \frac{\gamma}{2c_h} + \frac{\gamma}{2c_l}$ and $SR'_{A_w} = -\frac{1}{2c_h} - \frac{1}{2c_l}$. If the increase of $A_r - A_w$ is caused by both A_r and A_w rising, the marginal rate of change is $SR'_{A_r} + SR'_{A_w} = \frac{\gamma}{2c_h} + \frac{\gamma}{2c_l} - \frac{1}{2c_h} - \frac{1}{2c_l} > 0$. If the increase of $A_r - A_w$ is caused by A_r rising and A_w falling, it is obvious to see that SR will rise as well. Our assumption rules out the case where $A_r - A_w$ increases because both A_r and A_w decrease but A_w decreases more than A_r.

Case 2 happens when $\bar{L}$ is quite low so that only males from l type households want to stay in rice production. In this case, $SR = \frac{(\gamma-1)A_w}{2c_h} + \frac{\gamma A_r - A_w + \bar{L}}{2c_l}$. If the increase of $A_r - A_w$ is caused by A_r rising and A_w falling, the marginal rate of change is $SR'_{A_r} - SR'_{A_w} = \frac{\gamma}{2c_l} - \frac{\gamma-1}{2c_h} + \frac{1}{2c_l} > 0$. If the increase of $A_r - A_w$ is caused by both A_r and A_w rising, it is obvious to see that SR will rise as well. Again, our assumption rules out the case where $A_r - A_w$ increases because both A_r and A_w decrease but A_w decreases more than A_r.

Case 2 has smaller $A_r - A_w$ (lower A_r) than case 1, given all other parameters the same. So when $A_r - A_w$ increases, we may have SR jump from case 2 to case 1. However, subtracting the SR in case 1 by SR in case 2 we have $\frac{\gamma(A_r - A_w) + \bar{L}}{2c_h}$. Suppose the males working on rice from h type in case 1 is L_1, and males working on wheat from household h in case 2 is L_2. We know that males from h type households will only switch from wheat to rice (and hence we jump from case 2 to case 1) if $\gamma A_r + \bar{L} > \gamma A_w$ (otherwise there is no switch, given there is still a tax τ to pay), and thereby we have $\frac{\gamma(A_r - A_w) + \bar{L}}{2c_h} > 0$ at the jump, so SR will only be increased by the jump.

Now suppose $A_w < A_r$, so all members in l types households will produce rice. There are two cases as well. The first case is everyone produces rice, let us call it case 3, while the second case is all members in l types households produce rice while all members in h types households produce wheat, and we call it case 4. In case 3, $SR = \frac{(\gamma-1)A_r+\bar{L}}{2c_l} + \frac{(\gamma-1)A_r+\bar{L}}{2c_h}$. Apparently SR is increasing in A_r, and our assumption makes sure that when A_r increases $A_r - A_W$ always increases. So SR is increasing in $A_r - A_w$ in case 3. In case 4, $SR = \frac{(\gamma-1)A_r+\bar{L}}{2c_l} + \frac{(\gamma-1)A_w}{2c_h}$. We have $SR'_{A_r} - SR'_{A_w} = \frac{\gamma-1}{2c_l} - \frac{\gamma-1}{2c_h} > 0$, so SR increases when A_r increases but A_w decreases. If A_r and A_w both increase (so is SR), our assumption guarantees that $A_r - A_w$ always increases in the case.

Finally, When $A_r - A_w$ increases from negative to positive, we may have the following two possible situations:

Situation 1 We start with case 2, then jump to case 1, and finally jump to case 3

Situation 2 We start with case 2, then jump to case 4, and finally jump to case 3

We will never have the situation that we move from case 1 to case 4, because if h type males are already working on rice when A_r is relatively low (such as in case 1), increasing A_r (such that $A_r - A_w$ turns to positive) will never put them back to work on wheat. Meanwhile, we could have the case that we jump directly from case 2 to case 4, this happens when for any $A_r < A_w$, it is never optimal for h type males to produce rice.

To close the proofs, we need to show that SR in case 3 is larger than SR in case 1. This is straightforward: when $A_r = A_w$, we have SR in case 1 equals to SR in case 3, hence SR will only be increased by the jump. Similarly, in situation 2 when we jump from case 2 to case 4, SR will only be increased at the jump. Finally, subtracting the SR in case 3 by SR in case 4 we have $\frac{(\gamma-1)A_r+\bar{L}}{2c_h} - \frac{(\gamma-1)A_w}{2c_h} > 0$, so SR will only be increased by the jump from case 4 to case 3. Overall, no matter we are in situation 1 or 2, we always have SR increasing in $A_r - A_w$. □

C Proof of Proposition 3

Proof. It is sufficient to show that for any p_0, $C_l^0 > C_h^0$, and C_l^t is increasing in p_t for any t.

At first we show that $C_l^0 > C_h^0$. From the proof of proposition 6, we know there are four cases to be discussed: 1. Both types have males working on rice and females working on wheat. 2. Everyone works on wheat except low type's males. 3. Everyone produces rice. 4. h type households produce wheat while l type households produce works rice.

It is trivial to show that in case 1 and 3, l type households should be 'richer' than h type households, since the two types allocate their labors in the same way, but l type faces a lower cost of increasing the amount of males. So l type will have a larger amount of males. Since the wages are linear to both types, it means l type will receive a higher income. Also, from the proof of lemma 2, we know in case 4, l type should again be richer, because l type is always better in specializing in rice than specializing in wheat, while the income from specializing in wheat for l type is higher than the income of specializing in wheat for h type. So it follows naturally that in case 4 l type households will have a higher income than h type households.

The only case left to be discussed is case 2. In case 2, Given p_0, $C_l^0 = (A_r\gamma + \bar{L})\frac{\gamma A_r - A_w + \bar{L}}{2c_l} + \frac{1}{2}) + A_w(\frac{1}{2} - \frac{\gamma A_r - A_w + \bar{L}}{2c_l})$ where $\bar{L}$ is a function of p_0 but positive for any $p_0 > 0$. We also know that if case 2 happens, we must have $\gamma A_r + \bar{L} > \gamma A_w$ and $A_w > A_r$. Thereby we have $C_l^0 = (A_r\gamma + \bar{L})\frac{\gamma A_r - A_w + \bar{L}}{2c_l} + \frac{1}{2}) + A_w(\frac{1}{2} - \frac{\gamma A_r - A_w + \bar{L}}{2c_l}) > (A_r\gamma + \bar{L})\frac{\gamma A_w - A_w}{2c_l} + \frac{1}{2}) + A_w(\frac{1}{2} - \frac{\gamma A_w - A_w + \bar{L}}{2c_l}) > A_w\gamma(\frac{(\gamma-1)A_w}{2c_l} + \frac{1}{2}) + A_w(\frac{1}{2} - \frac{(\gamma-1)A_w}{2c_l}) = K$. While $C_h^0 = A_w\gamma(\frac{(\gamma-1)A_w}{2c_h} + \frac{1}{2}) + A_w(\frac{1}{2} - \frac{(\gamma-1)A_w}{2c_h})$, the only difference between K and C_h^0 is the value of cost. Take derivative of K with respect to c_l, we have $\frac{\partial K}{\partial c_l} = -\frac{(\gamma-1)A_w^2\gamma}{2}c^{-2} + \frac{(\gamma-1)A_w^2}{2}c^{-2} < 0$, hence $C_l^0 > C_h^0$ since $c_l < c_h$.

Since C_t^l is increasing in $\bar{L}$ which is increasing in p_t, the second part of the proof follows naturally. □

D Discussion of the model

Here we replace the assumption that only men enjoy the return from $\bar{L}$ by both men and women benefit from collaboration $\bar{L}$. However, it is not so natural to assume that women and men will benefit from $\bar{L}$ to the same degree. There are two aspects. At first, men may contribute more than women for collaboration work in rice production. For example, building irrigation facility was usually regarded as men's work in patrilineal and agricultural society. Second, if we assume that there is a gap between women and men's agricultural productivity, measured by γ, it should be kept so the improvement of technology by cooperation adds to labor productivity, rather than assume that it is independent of labor. This is not an issue in our baseline model, since $\gamma A_r + \bar{L}$ can be rewritten as $\gamma(A_r + \frac{\bar{L}}{\gamma})$, so it is equivalent to have a smaller benefit from collaboration multiplied by γ. For the full model, now the utility of type i household is:

$$u_i = A_w \gamma L^w_{mi} + A_w L^w_f + (A_r\gamma + \bar{L}\gamma)L^r_{mi} + (A_r + \bar{L})L^r_{fi} - c_i(\frac{1}{2} - L^w_{fi} - L^r_{fi})^2 - \tau 1\{L^r_{mi} + L^r_{fi} > 0\}. \tag{D.1}$$

Firstly, it is straightforward to see that a household will not have mix production. since both genders will work on wheat as long as $A_r + \bar{L} < A_w$. If $A_r + \bar{L} > A_w$, one may still prefer wheat due to the cost τ in rice production. However, whenever a household allocate one member on producing rice, the others must be allocated to produce rice as well when $A_r + \bar{L} > A_w$, since τ has already been paid.

Secondly, it is trivial to show that the proof of lemma 2 is unaffected with this setup, because now the externality in rice production works in the way that putting male on rice gets $\gamma\bar{L}$ (rather than $\bar{L}$), and putting female on rice gets $\bar{L}$ (rather than 0). In other words, our assumption guarantees that if low type households put all their labour into rice production, we always have $A_r + \bar{L} > A_w$, and it is optimal for the (low type) households to pay τ and produce rice. Hence, we still have that if the low type household gets to specialize in production, it always specialize in rice rather than wheat (one can still do all the math and show that $u^r_l > u^w_l$). So we only have two cases to be discussed:

- Case 1. Both type of households specialize in rice.
- Case 2. High type households specialize in producing wheat, while low type households produce rice only.

With case 1, we have equilibrium sex ratio equals to $SR = \frac{(\gamma-1)A_r+(\gamma-1)\bar{L}}{2c_l} + \frac{(\gamma-1)A_r+(\gamma-1)\bar{L}}{2c_h}$. We see that SR is increasing in A_r, and our assumption guarantees that when A_r increases $A_r - A_W$ always increases, making SR increases in $A_r - A_w$.

Case 2 happens when we have $A_r + \bar{L} > A_w$ but the high type still finds it not optimal to pay τ. In that case, $SR = \frac{(\gamma-1)A_r+(\gamma-1)\bar{L}}{2c_l} + \frac{(\gamma-1)A_w}{2c_h}$. We still have $SR'_{A_r} - SR'_{A_w} = \frac{\gamma-1}{2c_l} - \frac{\gamma-1}{2c_h} > 0$, so SR increases when A_r increases but A_w decreases. If A_r and A_w both increase, our assumption guarantees that $A_r - A_w$ always increases since A_r will change by a larger degree than A_w, which makes SR increase. Finally, when $A_r - A_w$ increases there could be a situation that we jump from case 2 to case 1, in this case it is easy to see that SR still increases. Overall, Proposition 2 still holds even if we allow that both male and female benefit from collaboration in rice production.

One concern would be that what will happen if $\bar{L}$ is not multiplied by γ, but additive to productivity. i.e. women on rice gives $A_r + \bar{L}$ while men have $\gamma A_r + \bar{L}$. This will not affect the proof of lemma 1 as well. However, there will be an interval of $A_r - A_w$ in which female will work on rice while men work on wheat. To see this, notice that women will prefer wheat if $A_w > A_r + \bar{L} \Rightarrow A_w - A_r > \bar{L}$. Men will prefer wheat if $\gamma A_w > \gamma A_r + \bar{L} \Rightarrow A_w - A_r > \frac{\bar{L}}{\gamma}$. If the former inequality holds, so does the latter since $\gamma > 1$. So if $\frac{\bar{L}}{\gamma} < A_w - A_r < \bar{L}$, women will prefer rice while men still prefer wheat. In other words, we do not adopt this assumption of labor productivity because it generates a situation where females have a comparative advantage to men on rice production as well as cooperation, which is quite far from the reality in agricultural societies. Also, the existence of the situation where women work on rice while men work on wheat lacks support of real world evidence. Last but not least, on the theory side, the spirit of our model is that sex ratio is increasing in the gap between rice and wheat productivity, given the condition that men have comparative advantage over women on rice (and women on wheat). Any case that violates the condition is not the interest of this paper and should be excluded.

Bibliography

Agassiz, AR (1894), "Our commercial relations with chinese manchuria." *The Geographical Journal*, 4, 534–556.

Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi (2020), "Just a few seeds more: value of network information for diffusion." *Available at SSRN 3062830*.

Alatas, Vivi, Abhijit Banerjee, Arun G Chandrasekhar, Rema Hanna, and Benjamin A Olken (2016), "Network structure and the aggregation of information: Theory and evidence from indonesia." *American Economic Review*, 106, 1663–1704.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013), "On the origins of gender roles: Women and the plough." *The Quarterly Journal of Economics*, 128, 469–530.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2018), "Traditional agricultural practices and the sex ratio today." *PloS one*, 13, e0190510.

Alsan, Marcella, Marianne Wanamaker, and Rachel R Hardeman (2020), "The tuskegee study of untreated syphilis: a case study in peripheral trauma with implications for health professionals." *Journal of general internal medicine*, 35, 322–325.

Alvard, Michael S (2003), "Kinship, lineage, and an evolutionary perspective on cooperative hunting groups in indonesia." *Human Nature*, 14, 129–163.

Ambrus, Attila, Markus Mobius, and Adam Szeidl (2014), "Consumption risk-sharing in social networks." *American Economic Review*, 104, 149–182.

Arnold, Fred and Liu Zhaoxiang (1992), "Sex preference, fertility, and family planning in china." In *The Population of Modern China*, 491–523, Springer.

Atwell, Paul and Noah L Nathan (2022), "Channels for influence or maps of behavior? a field experiment on social networks and cooperation." *American Journal of Political Science*, 66, 696–713.

Baber, Ray Erwin (1934), "Marriage in ancient china." *The Journal of Educational Sociology*, 8, 131–140.

Bandiera, Oriana and Imran Rasul (2006), "Social networks and technology adoption in northern mozambique." *The Economic Journal*, 116, 869–902.

Banerjee, Abhijit, Emily Breza, Arun G Chandrasekhar, Esther Duflo, Matthew O Jackson, and Cynthia Kinnan (2021), "Changes in social network structure in response to exposure to formal credit markets." Technical report, National Bureau of Economic Research.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson (2019), "Using gossips to spread information: Theory and evidence from two randomized controlled trials." *The Review of Economic Studies*, 86, 2453–2490.

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan (2015), "The miracle of microfinance? evidence from a randomized evaluation." *American economic journal: Applied economics*, 7, 22–53.

Banerjee, Abhijit V (1992), "A simple model of herd behavior." *The Quarterly Journal of Economics*, 107, 797–817.

Baron, Reuben M and David A Kenny (1986), "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." *Journal of personality and social psychology*, 51, 1173.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak (2021), "Can network theory-based targeting increase technology adoption?" *American Economic Review*, 111, 1918–43.

Bengtsson, Tommy, Cameron Campbell, James Z Lee, et al. (2004), *Life under pressure: Mortality and living standards in Europe and Asia, 1700-1900*. Mit Press.

BenYishay, Ariel and A Mushfiq Mobarak (2019), "Social learning and incentives for experimentation and communication." *The Review of Economic Studies*, 86, 976–1009.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan (2015), "Gender identity and relative income within households." *The Quarterly Journal of Economics*, 130, 571–614.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992), "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of Political Economy*, 100, 992–1026.

Boserup, Ester (2007), *Woman's role in economic development*. Earthscan.

Bourlès, Renaud, Yann Bramoullé, and Eduardo Perez-Richet (2021), "Altruism and risk sharing in networks." *Journal of the European Economic Association*, 19, 1488–1521.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet (2015), "Social networks and the decision to insure." *American Economic Journal: Applied Economics*, 7, 81–108.

Callaway, Brantly and Pedro HC Sant'Anna (2021), "Difference-in-differences with multiple time periods." *Journal of econometrics*, 225, 200–230.

Cameron, Lisa, Xin Meng, and Dandan Zhang (2019), "China's sex ratio and crime: Behavioural change or financial necessity?" *The Economic Journal*, 129, 790–820.

Campbell, Cameron and James Lee (2008), "Kin networks, marriage, and social mobility in late imperial china." *Social Science History*, 32, 175–214.

Campbell, Cameron and James Z Lee (2011), "Kinship and the long-term persistence of inequality in liaoning, china, 1749-2005." *Chinese Sociological Review*, 44, 71–103.

Carlsson, Hans and Eric Van Damme (1993), "Global games and equilibrium selection." *Econometrica: Journal of the Econometric Society*, 989–1018.

Carranza, Eliana (2014), "Soil endowments, female labor force participation, and the demographic deficit of women in india." *American Economic Journal: Applied Economics*, 6, 197–225.

Çelen, Boğaçhan and Shachar Kariv (2004), "Distinguishing informational cascades from herd behavior in the laboratory." *American Economic Review*, 94, 484–498.

Chandrasekhar, Arun G, Cynthia Kinnan, and Horacio Larreguy (2018), "Social networks as contract enforcement: Evidence from a lab experiment in the field." *American Economic Journal: Applied Economics*, 10, 43–78.

Chen, Shuang, Cameron Dougall Campbell, and James Lee (2014), "Categorical inequality and gender difference: Marriage and remarriage in northeast china, 1749-1913." In *Similarity in Difference: Marriage in Europe and Asia, 1700-1900*.

Chen, Ting, James Kai-sing Kung, and Chicheng Ma (2020), "Long live keju! the persistent effects of china's civil examination system." *The economic journal*, 130, 2030–2064.

Chenoweth, Erica and Zoe Marks (2022), "Revenge of the patriarchs: Why autocrats fear women." *Foreign Aff.*, 101, 103.

Cohen, Myron L (1990), "Lineage organization in north china." *The Journal of Asian Studies*, 49, 509–534.

Conley, Timothy G and Christopher R Udry (2010), "Learning about a new technology: Pineapple in ghana." *American Economic Review*, 100, 35–69.

Coontz, Stephanie (2006), *Marriage, a history: How love conquered marriage*. Penguin.

Cooper, Russell, Douglas V DeJong, Robert Forsythe, and Thomas W Ross (1992), "Communication in coordination games." *The Quarterly Journal of Economics*, 107, 739–771.

Croll, Elisabeth (1985), "Introduction: Fertility norms and family size in china." In *China's One-Child Family Policy*, 1–36, Springer.

Dawkins, Richard (2016), *The selfish gene*. Oxford university press.

De Brauw, Alan, Jikun Huang, Linxiu Zhang, and Scott Rozelle (2013), "The feminisation of agriculture with chinese characteristics." *The Journal of Development Studies*, 49, 689–704.

De Brauw, Alan, Qiang Li, Chengfang Liu, Scott Rozelle, and Linxiu Zhang (2008), "Feminization of agriculture in china? myths surrounding women's participation in farming." *The China Quarterly*, 194, 327–348.

de Martí, Joan and Pau Milán (2019), "Regime change in large information networks." *Games and Economic Behavior*, 113, 262–284.

DeGroot, Morris H (1974), "Reaching a consensus." *Journal of the American Statistical Association*, 69, 118–121.

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel (2003), "Persuasion bias, social influence, and unidimensional opinions." *The Quarterly Journal of Economics*, 118, 909–968.

Desmet, Klaus and Stephen L Parente (2014), "Resistance to technology adoption: The rise and decline of guilds." *Review of economic dynamics*, 17, 437–458.

Ding, Yizhuang, Songyi Guo, James Z Lee, and Cameron Campbell (2004), *Immigration and Eight Banner Society in Liaodong (Liaodong yimin zhong de* *qiren shehui)*. Shanghai Academy of Social Sciences Press (Shanghai Kexueyuan Chubanshe).

Dong, Hao (2018), "Extended family and reproductive success: Comparative evidence from east asian household registration data, 1678–1945." In *PAA 2018 Annual Meeting*, PAA.

Drăgulescu, Adrian and Victor M Yakovenko (2001), "Exponential and power-law probability distributions of wealth and income in the united kingdom and the united states." *Physica A: Statistical Mechanics and its Applications*, 299, 213–221.

Driessen, Theo SH (2013), *Cooperative games, solutions and applications*, volume 3. Springer Science & Business Media.

Eastin, Joshua and Aseem Prakash (2013), "Economic development and gender equality: Is there a gender kuznets curve?" *World Politics*, 65, 156–186.

Eastman, Lloyd E et al. (1989), "Family, fields, and ancestors: constancy and change in china's social and economic history, 1550-1949." *OUP Catalogue*.

Fan, Joseph PH, Qiankun Gu, and Xin Yu (2021), "Collectivist cultures and the emergence of family firms." *Forthcoming in Journal of Law and Economics*.

Farrell, Joseph (1987), "Cheap talk, coordination, and entry." *The RAND Journal of Economics*, 34–39.

Farrell, Joseph and Garth Saloner (1988), "Coordination through committees and markets." *The RAND Journal of Economics*, 235–252.

Fei, Hsiao-Tung (1946), "Peasantry and gentry: an interpretation of chinese social structure and its changes." *American Journal of Sociology*, 52, 1–17.

Forsythe, Nancy, Roberto Patricio Korzeniewicz, and Valerie Durrant (2000), "Gender inequalities and economic growth: A longitudinal evaluation." *Economic development and cultural change*, 48, 573–617.

Freedman, Maurice (2021), *Lineage organisation in South-eastern China*. Routledge.

Freedman, Ronald, Lolagene C Coombs, Ming-Cheng Chang, and Te-Hsiung Sun (1974), "Trends in fertility, family size preferences, and practice of family planning: Taiwan, 1965-1973." *Studies in family planning*, 5, 270–288.

Galor, Oded and Ömer Özak (2016a), "The agricultural origins of time preference." *American Economic Review*, 106, 3064–3103.

Galor, Oded and Ömer Özak (2016b), "The agricultural origins of time preference." *American Economic Review*, 106, 3064–3103.

Genius, Margarita, Phoebe Koundouri, Céline Nauges, and Vangelis Tzouvelekas (2014), "Information transmission in irrigation technology adoption and diffusion: Social learning, extension services, and spatial effects." *American Journal of Agricultural Economics*, 96, 328–344.

Goldin, Claudia (1994), "The u-shaped female labor force function in economic development and economic history."

Goldin, Claudia (1995), "Career and family: College women look to the past." *NBER working paper*.

Golub, Benjamin and Matthew O Jackson (2010), "Naive learning in social networks and the wisdom of crowds." *American Economic Journal: Microeconomics*, 2, 112–149.

Golub, Benjamin and Evan Sadler (2017), "Learning in social networks." *Available at SSRN 2919146*.

Greif, Avner and Guido Tabellini (2010a), "Cultural and institutional bifurcation: China and europe compared." *American economic review*, 100, 135–140.

Greif, Avner and Guido Tabellini (2010b), "Cultural and institutional bifurcation: China and europe compared." *American economic review*, 100, 135–140.

Hamon, Raeann R and Bron B Ingoldsby (2003), *Mate selection across cultures*. Sage.

Harrell, Stevan and Thomas W Pullum (1995), "Marriage, mortality, and the developmental cycle in three xiaoshan lineages." *Chinese historical microdemography*, 141–62.

Hausfater, Glenn (1984), "Infanticide: comparative and evolutionary perspectives." *Current anthropology*, 25, 500–502.

He, Quqiong, Ying Pan, and Sudipta Sarangi (2018), "Lineage-based heterogeneity and cooperative behavior in rural china." *Journal of Comparative Economics*, 46, 248–269.

He, Yu (1989), "Two key gateways shaping the economy of northeast china: a discussion on the rise and fall of yingkou and dalian in modern times (qiandong dongbei jingji de liang da menhu - cong jindai yingkou yu dalian de shengshuai tan qi)." *The Qing History Journal (Qingshi Yanjiu)*, 4, 19–24.

Heß, Simon, Dany Jaimovich, and Matthias Schündeln (2021), "Development projects and economic networks: Lessons from rural gambia." *The Review of Economic Studies*, 88, 1347–1384.

Hopkins, Ed, V Bhaskar, et al. (2011), "Marriage as a rat race: Noisy pre-marital investments with assortative matching." Technical report, Edinburgh School of Economics, University of Edinburgh.

Hudson, Valerie M and Andrea Den Boer (2002), "A surplus of men, a deficit of peace: Security and sex ratios in asia's largest states." *International Security*, 26, 5–38.

Inglehart, Ronald and Paul R Abramson (1994), "Economic security and value change." *American political science review*, 88, 336–354.

Jackson, Matthew O and Evan C Storms (2017), "Behavioral communities and the atomic structure of networks." *arXiv preprint arXiv:1710.04656*.

Jayachandran, Seema (2015), "The roots of gender inequality in developing countries." *Annu. Rev. Econ*, 7, 63–88.

Jia, Ruixue (2014), "The legacies of forced freedom: China's treaty ports." *Review of Economics and Statistics*, 96, 596–608.

Jiang, Quanbao, Yanping Zhang, and Jesús J Sánchez-Barricarte (2015), "Marriage expenses in rural china." *China Review*, 15, 207–236.

Johnson, David (1977), "The last years of a great clan: the li family of chao chün in late t'ang and early sung." *Harvard Journal of Asiatic Studies*, 5–102.

Kübler, Dorothea and Georg Weizsäcker (2004), "Limited depth of reasoning and failure of cascade formation in the laboratory." *The Review of Economic Studies*, 71, 425–441.

Kung, James Kai-sing and Nan Li (2011), "Commercialization as exogenous shocks: The effect of the soybean trade and migration in manchurian villages, 1895–1934." *Explorations in Economic History*, 48, 568–589.

Langer, William L (1974), "Infanticide: A historical survey." *The Journal of Psychohistory*, 1, 353.

Lee, James and Cameron Dougall Campbell (2005), "Living standards in liaoning, 1749-1909: Evidence from demographic outcomes." In *Living standards in the past: New perspectives on well-being in Asia and Europe*.

Lee, James, Cameron Dougall Campbell, and Shuang Chen (2010), "China multigenerational panel dataset, liaoning (cmgpd-ln), 1749-1909." *Data Sharing for Demographic Research (DSDR).*

Lee, James and Wang Feng (1999), "Malthusian models and chinese realities: The chinese demographic system 1700–2000." *Population and Development Review*, 25, 33–65.

Lee, James, Wang Feng, and Cameron Campbell (1994), "Infant and child mortality among the qing nobility: Implications for two types of positive check." *Population Studies*, 48, 395–411.

Lee, James Z, Cameron D Campbell, et al. (1997), *Fate and fortune in rural China: social organization and population behavior in Liaoning 1774-1873*. Cambridge University Press.

Li, Shuzhuo (2007), "Imbalanced sex ratio at birth and comprehensive intervention in china."

Ligon, Ethan, Jonathan P Thomas, and Tim Worrall (2000), "Mutual insurance, individual savings, and limited commitment." *Review of Economic Dynamics*, 3, 216–246.

Malthus, Thomas Robert (1986), *The Works of Thomas Robert Malthus: An essay on the principle of population, 1826 (I)*, volume 2. Pickering.

Mammen, Kristin and Christina Paxson (2000), "Women's work and economic development." *Journal of economic perspectives*, 14, 141–164.

Mao, Liping (2007), *Research on Dowry in Qing Dynasty (Qingdai Jiazhuang Yanjiu)*. China Renmin University Press (Renmin Daxue Chubanshe).

Marx, Karl (1973), *Karl Marx on society and social change: With selections by Friedrich Engels*. University of Chicago Press.

Massey, Cade and George Wu (2005), "Detecting regime shifts: The causes of under-and overreaction." *Management Science*, 51, 932–947.

McCleary, Rachel M and Robert J Barro (2006), "Religion and economy." *Journal of Economic perspectives*, 20, 49–72.

Meng, Xiangdan (2014), *Feminization of agricultural production in rural China: A sociological analysis*. Ph.D. thesis, Wageningen University.

Morris, Stephen and Hyun Song Shin (2003), "Global games: Theory and applications." In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume 1*, 56–114, Cambridge University Press.

Munshi, Kaivan and Mark Rosenzweig (2016), "Networks and misallocation: Insurance, migration, and the rural-urban wage gap." *American Economic Review*, 106, 46–98.

Murendo, Conrad, Meike Wollni, Alan De Brauw, and Nicholas Mugabi (2018), "Social network effects on mobile money adoption in uganda." *The Journal of Development Studies*, 54, 327–342.

Nabli, Mustapha K, Jeffrey B Nugent, and Lamine Doghri (1989), "The size distribution and ownership type of firms in tunisian manufacturing." In *Contributions to Economic Analysis*, volume 183, 200–232, Elsevier.

Neelakantan, Urvi and Michele Tertilt (2008), "A note on marriage market clearing." *Economics Letters*, 101, 103–105.

Olden, Andreas and Jarle Møen (2022), "The triple difference estimator." *The Econometrics Journal*, 25, 531–553.

Ownby, David (2002), "Approximations of chinese bandits: perverse rebels, romantic heroes, or frustrated bachelors?" *Chinese femininities/Chinese masculinities: A reader*, 226–250.

Parravano, Melanie and Odile Poulsen (2015), "Stake size and the power of focal points in coordination games: Experimental evidence." *Games and Economic Behavior*, 94, 191–199.

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

Qian, Nancy (2008), "Missing women and the price of tea in china: The effect of sex-specific earnings on sex imbalance." *The Quarterly Journal of Economics*, 123, 1251–1285.

Redding, Gordon (2013), "The spirit of chinese capitalism." In *The Spirit of Chinese Capitalism*, de Gruyter.

Robinson, James A (2006), "Economic development and democracy." *Annu. Rev. Polit. Sci.*, 9, 503–527.

Rousseau, Jean-Jacques (1775), "Discourse on the origin and foundations of inequality among men in "the social contract" and "discourses," translated by gdh cole, rev." *JH Brumfitt, and JC Hall (London: Everyman's Library, 1973)*.

Sanders, Rebecca and Laura Dudley Jenkins (2022), "Control, alt, delete: Patriarchal populist attacks on international women's rights." *Global Constitutionalism*, 11, 401–429.

Schelling, TC (1960), "The strategy of conflict." Harvard University Press.

Sen, Amartya (1992), "Missing women." *BMJ: British Medical Journal*, 304, 587.

Shiue, Carol H (2017), "Human capital and fertility in chinese clans before modern growth." *Journal of Economic Growth*, 22, 351–396.

Shiue, Carol H and Wolfgang Keller (2022), "Marriage matching over five centuries in china." Technical report, National Bureau of Economic Research.

Shurtleff, William and Akiko Aoyagi (2022), *History of Soybeans and Soyfoods in China, in Chinese Cookbooks and Restaurants, and in Chinese Work with Soyfoods Outside China (Including Taiwan, Manchuria, Hong Kong & Tibet)(1949-2022): Extensively Annotated Bibliography and Sourcebook*. Soyinfo Center.

Smil, Vaclav (1993), *China's environmental crisis: an inquiry into the limits of national development*. ME Sharpe.

Smith, Lones and Peter Sørensen (2000), "Pathological outcomes of observational learning." *Econometrica*, 68, 371–398.

Stone, Linda and Diane E King (2018), *Kinship and gender: An introduction*. Routledge.

Szołtysek, Mikołaj, Sebastian Klüsener, Radosław Poniat, and Siegfried Gruber (2017), "The patriarchy index: A new measure of gender and generational inequalities in the past." *Cross-Cultural Research*, 51, 228–262.

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

Talhelm, Thomas, Xiao Zhang, Shige Oishi, Chen Shimin, Dechao Duan, Xiaoli Lan, and Shinobu Kitayama (2014), "Large-scale psychological differences within china explained by rice versus wheat agriculture." *Science*, 344, 603–608.

Tang, Can and Zhong Zhao (2023), "Informal institution meets child development: Clan culture and child labour in china." *Journal of Comparative Economics*, 51, 277–294.

Telford, Ted A (1995), "Fertility and population growth in the lineages of tongcheng county, 1520-1661." In *Chinese historical microdemography*, 48–93, University of California Press.

Thomas, Matthew Gwynfryn, Ting Ji, Jiajia Wu, QiaoQiao He, Yi Tao, and Ruth Mace (2018), "Kinship underlies costly cooperation in mosuo villages." *Royal Society Open Science*, 5, 171535.

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

Weber, Max (1951), *The Religion of China: Confucianism and Taoism.* ,Translated by Hans H. Gerth, London: Collier-Macmillan.

Williamson, Nancy E (1976), "Sex preferences, sex control, and the status of women." *Signs: Journal of Women in Culture and Society*, 1, 847–862.

Xu, Yiqing and Yang Yao (2015), "Informal institutions, collective action, and public investment in rural china." *American Political Science Review*, 109, 371–391.

Xue, Melanie Meng (2018), "High-value work and the rise of women: The cotton revolution and gender equality in china."

Yi, Zeng, Tu Ping, Gu Baochang, Xu Yi, Li Bohua, and Li Yongpiing (1993), "Causes and implications of the recent increase in the reported sex ratio at birth in china." *Population and development review*, 283–302.

Zhang, Shizun (2020), *Research on Liaohe Shipping and Northeast Economic Integration (Liaohe hangyunshi yu dongbei jingji yitihua yanjiu)*. Social Sciences Literature Press (Shehui Kexue Wenxian Chubanshe).

Zheng, Zhenman (2001), *Family lineage organization and social change in Ming and Qing Fujian*. University of Hawaii press.