The London School of Economics and Political Science

The Persistence of Institutions in Regions: Shocks, Public Policies, and Inequalities

Han Wang

A thesis submitted to the Department of Geography and Environment of the London School of Economics for the degree of Doctor of Philosophy

London, September 2022

Declaration

I certify that the Thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others.

I declare that this thesis consists of approximately 56,400 words.

Statement of conjoint work

I confirm that paper 1 was jointly co-authored with Professor Andrés Rodríguez-Pose and Professor Neil Lee and I contribute 70% of this work; paper 2 was jointly co-authored with Professor Andrés Rodríguez-Pose and I contribute 70% of this work; paper 3 was joint co-authored with Professor Neil Lee and I contribute 70% of this work.

Abstract

This thesis presents an empirical analysis of the influence of regional institutions on the social and economic inequalities through crises and policies changes. It is organised into four independent chapters.

Chapter 1 investigates how an indigenous government institution affect contemporary economic development. Government institutions have long been known as critical drivers of growth. However, it is still unknown whether the institution is resilient after experiencing large political shocks. Using the textual data from government records, we examine how differences in local government quality in Imperial times determined current differences in economic development.

Chapter 2 focuses on government institutions' role in explaining performance differences in pandemic prevention. Current research so far has primarily overlooked linking institutions to social outcomes. How institutions react to public health crises has attracted minimal attention. We collected and combined the Black Death's mortality rate and duration data with the location of autonomous cities to study whether a greater degree of autonomy could have allowed cities to react effectively after the pandemics.

Chapter 3 moves beyond public-order institutions to study the persistent economic effect of the private-order institutions. We used historical administrative data, the Statistics of Agriculture and Commerce, to digitalise the conditions of the chamber system of China in the 1910s. To measure contemporary business performance, we rely on the Firm Registration Database, which systemically documents start-ups' emergence since 1975. Using a difference-in-differences strategy, we identify cities heavily permeated with the historical chamber system as enjoying more entrepreneurs and flourishing with innovative performances when the market economy reopens. Our instrumental variable strategy has also verified these results.

Chapter 4 investigates the influence of culture. Within the background of son preference, we discuss how the family planning policy distorts a couple's decision-making about childbirth. Under natural conditions, parents satisfy potential son preferences by giving birth until a son is born. However, with restrictions on family size, the cultural preference for sons can no longer be smoothed out by adding additional children. Parents may practise prenatal and postnatal discrimination against daughters, lowering their probability of survival relative to sons. Drawing on the natural experiment of the Two-Child Policy, we examine whether inconsiderate policies could further trigger the dark side of a backward culture.

Acknowledgement

I am indebted to Andrés Rodríguez-Pose and Neil Lee for their supervision, support, and continued guidance.

I thank my colleagues and peers at the LSE and beyond for their helpful comments and fruitful discussions.

And finally, my gratitude goes to my wife, Iris, for the love, power, and encouragement during this process.

Table of Contents

ABSTRACT **2**

ACKNOWLEDGEMENT **3**

LIST OF TABLES **7**

LIST OF FIGURES **8**

SYNOPSIS **9**

1 Introduction 9
2 Themes of Research 12
2.1 Definition of institutions *13*
2.2 Formal institutions *13*
2.3 Informal institutions *15*
3 Summary of Chapters 15
4 Conclusion 17
References 19

CHAPTER 1: THE LONG SHADOW OF HISTORY IN CHINA: REGIONAL GOVERNANCE AND CHINESE TERRITORIAL INEQUALITY **22**

1 Introduction 22
2 Institutional Background 24
2.1 An alternative approach to measuring government quality *24*
2.2 Historical differences in local government quality in China *25*
2.3 The collapse of the Qing dynasty *27*
3 Data and descriptive analysis 28
3.1 Data on historical government quality *28*
3.2 Night-time satellite light density *29*
3.3 Control variables *30*
4 Empirical analyses 32
4.1 Empirical setups *32*
4.2 Baseline results *33*
4.3 Causal identification *37*
4.4 Instrumented results *40*
4.5 Robustness checks *41*
5. Understanding the mechanism 44
5.1 Decentralisation *44*
5.2 Good governance *46*
6. The effect of Western shocks 48
7 Conclusion 52
References 53
Appendix A 57
Appendix B 62

CHAPTER 2: LOCAL INSTITUTIONS AND PANDEMICS: CITY AUTONOMY AND THE BLACK DEATH **63**

1 Introduction 63
2 Historical context 65
2.1 The Black Death 65
2.2 Urban autonomy 67
3 The model and data 69
3.1 The research design and model 69
3.2 Data 71
4 Empirical analysis 76
4.1 Ordinary Least Square (OLS) estimates 76
4.2 Two-Stage least squares (2SLS) estimates 79
4.3 Placebo tests and robustness checks 81
5 Conclusion 89
References 91
Appendix A 97
Appendix B 103

CHAPTER 3: INCLUSIVE INSTITUTIONS, ENTREPRENEURSHIP AND INNOVATION IN CHINA 104

1 Introduction 104
2 Conceptual framework 108
3 Empirical background and data sources 110
3.1 Private-order institutions and chamber system in China 110
3.2 The fall and rise of market economy in China 113
3.3 Data sources 114
4 Chamber system and entrepreneurship 117
4.1 Baseline results 117
4.2 Results from instrumental variables 123
5 Innovative Effects of Chamber System 125
6 Conclusion 128
References 130
Appendix A 135

CHAPTER 4: BIRTH QUOTA AND GENDER BIAS: EVIDENCE FROM THE TWO-CHILD POLICY 149

1 Introduction 149
2 Institutional background 153
2.1 Fertility policies for ethnic minorities 153
2.2 Culture of son preference 157
3 Data sources and hypothesis testing 158
3.1 Micro-level population censuses 159
3.2 Rollout of fertility policies for ethnic minority groups 159
3.3 Introduction of ultrasound scans 160
3.4 Hypothesis testing 161
4 Empirical results: effects on the sex ratio 164
4.1 Baseline results 164
4.2 Different phases of policy implementation 167

4.3 Heterogeneous effect *168*
4.4 Birth spacing *170*
4.5 Ultrasound mechanism *171*
5 Empirical results: effects of artificial twinning 173
5.1 Baseline results *173*
5.2 Mechanism *175*
6 Conclusion 177
References 179
Appendix A 183

List of Tables

CHAPTER 1: THE LONG SHADOW OF HISTORY IN CHINA: REGIONAL GOVERNANCE AND CHINESE TERRITORIAL INEQUALITY

Table 1: Summary statistics of historical government quality, overall 31
Table 2: Impact of historical government quality on county night light luminosity in 2010: OLS estimate, overall 34
Table 3: Historical government quality on county night light luminosity in 2010: OLS estimate; variable of interest as ordinal variable 36
Table 4: The first stage result 39
Table 5: Historical governance quality on county night light luminosity in 2010: 2SLS estimate, overall 40
Table 6: OLS and 2SLS analysis, controlling more historical variables. 42
Table 7: OLS and 2SLS analysis excluding minority regions. 43
Table 8: Decentralization or centralization 45
Table 9: The effect of good governance 47
Table 10: The effect of Western shocks on contemporary economic development 50

CHAPTER 2: LOCAL INSTITUTIONS AND PANDEMICS: CITY AUTONOMY AND THE BLACK DEATH

Table 1: Descriptive Statistics 75
Table 2: City autonomy and mortality rate of Black Death 77
Table 3: City autonomy and Black Death duration 78
Table 4: City autonomy and mortality rate of the Black Death, 2SLS results 80
Table 5: Robustness check of the measure of city autonomy 86
Table 6: Robustness check of ruling out small cities 87
Table 7: Robustness check by using the boundary in the 14th century 88

CHAPTER 3: INCLUSIVE INSTITUTIONS, ENTREPRENEURSHIP AND INNOVATION IN CHINA

Table 1: Baseline results: the effect of the chamber system on entrepreneurship 120
Table 2: Baseline results: heterogeneous effects of chamber system on entrepreneurship 124
Table 3: The impact of chamber system on entrepreneurship: instrumental variable results 127

CHAPTER 4: BIRTH QUOTA AND GENDER BIAS: EVIDENCE FROM THE TWO-CHILD POLICY

Table 1: Summary statistics 160
Table 2: Validity of TCP on family size, household-level analysis 162
Table 3: Sex ratio at the second birth 166
Table 4: Examining sample selection: son-biased fertility stopping 167
Table 5: Sex ratio at the second birth: phases of policy implementation 168
Table 6: Sex ratio at the second birth: heterogeneous effect 169
Table 7: Birth spacing at the second birth 171
Table 8: Mechanism of ultrasound on sex ratio at the second birth 172
Table 9: Mechanism of ultrasound on birth spacing at the second birth 173
Table 10: Second birth are twins 175
Table 11: Birth spacing of twins at the second birth 177

List of Figures

SYNOPSIS

Figure 1: The link between logarithm GDP in 2020 and logarithm GDP in 1991 10
Figure 2: The link between logarithm population in 2020 and logarithm population in 1776 11
Figure 3: Themes of Research 13

CHAPTER 1: THE LONG SHADOW OF HISTORY IN CHINA: REGIONAL GOVERNANCE AND CHINESE TERRITORIAL INEQUALITY

Figure 1: Historical government quality between 1731 to 1911 29
Figure 2: Night light luminosity in 2010 30
Figure 3: The luminosity of prioritised counties and normal counties within the same prefecture 32
Figure 4:The location of garrisons in the Ming dynasty 38
Figure 5: Protestant missionary presence 49
Figure 6: Treaty ports 49

CHAPTER 2: LOCAL INSTITUTIONS AND PANDEMICS: CITY AUTONOMY AND THE BLACK DEATH

Figure 1: Mortality rate of the Black Death 71
Figure 2: Duration of the Black Death 72
Figure 3: The location of autonomous and nonautonomous cities 73
Figure 4: The distribution of t-statistics from randomly assigned mortality placebo tests 82
Figure 5: The distribution of t-statistics from randomly assigned autonomous city placebo tests 82
Figure 6: Distribution of t-statistics by randomly assigned mortality by certain percentage-points in one third of cities 83

CHAPTER 3: INCLUSIVE INSTITUTIONS, ENTREPRENEURSHIP AND INNOVATION IN CHINA

Figure 1: The rollout of chambers of commerce in the early 20th century China 108
Figure 2: The distribution of the chamber membership density in 1915 114
Figure 3: Average entrepreneurship high vs. low chamber density over time 117
Figure 4: The effect of private institutions on entrepreneurship over time 121
Figure 5: The effect of private institutions on newly registered SOEs over times 122
Figure 6: The heterogenous effect of chamber system on entrepreneurship over time 124
Figure 7: The chamber system and innovative performance across sectors over times 128
Figure 8: The merchant guild and innovative performance across sectors over times 129

CHAPTER 4: BIRTH QUOTA AND GENDER BIAS: EVIDENCE FROM THE TWO-CHILD POLICY

Figure 1: Geographic distribution of announcement of fertility policies for minority groups 156
Figure 2: Geographic distribution of legalisation of fertility policies for minority groups 157
Figure 3: Introduction of ultrasound and abortion rate 161
Figure 4: Sex ratio trends among TCP groups 163

Synopsis

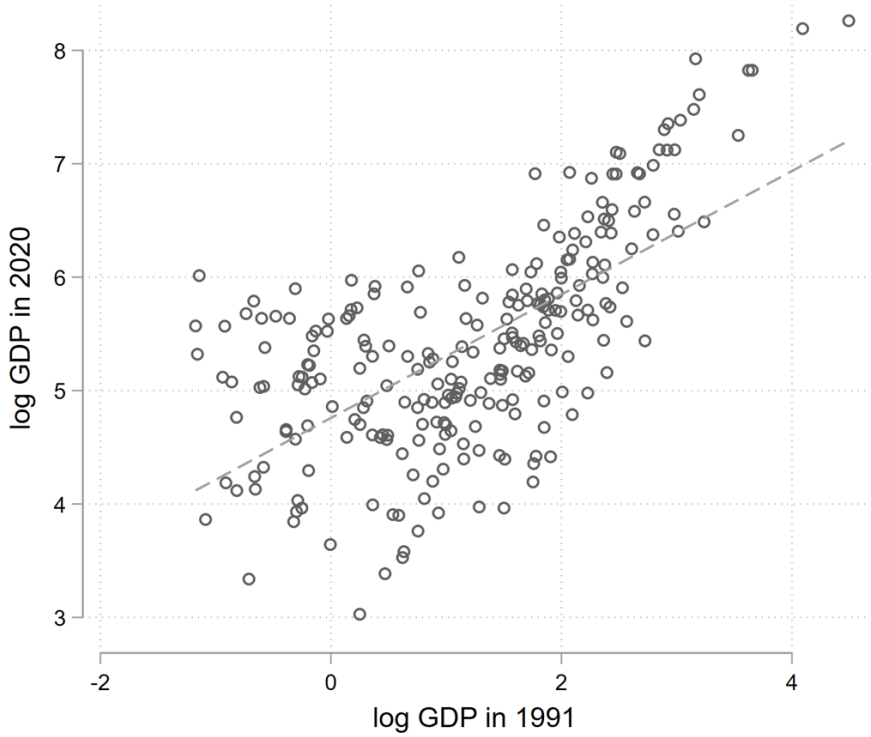
1 Introduction

The role of institutions in explaining determinants of development has been rediscovered in recent years. Theories have long predicted that countries with better institutions, such as secure property rights, lower levels of corruption, and fewer distortionary policies, tend to invest both physical and human capital more efficiently to achieve a greater level of income (Gerif, 1993; North, 1990; Rothstein, 2011). With the availability of state-of-the-art econometric techniques, a growing body of empirical research shows that institutions have significant impacts on contemporary economic outcomes (Acemoglu et al., 2001; Glaeser & Shleifer, 2002; Nunn, 2008). A consensus has gradually been reached that only by removing institutional factors from the residual terms can we gain a more comprehensive understanding of the rationale of development and underdevelopment.

This evidence is not restricted to the national level. Studies have uncovered that institutions are even more indispensable within countries at the local and regional levels. The most general approach to measuring organisations is the Regional Quality of Government Index. For example, by adopting this index within European regions, it has been found that contemporary local government quality heavily affects labour productivity, innovative performance and current regional diversification (Cortinovis et al., 2017; Rodríguez-Pose & Di Cataldo, 2015; Rodríguez-Pose & Ganau, 2022). Similar findings have been proved in other developing scenarios, such as the African and Chinese regions (Iddawela et al., 2021; Rodríguez-Pose & Zhang, 2019). Therefore, when we take the snippet of the modern era, the positive interaction between local institutions and development has always been harmonised. However, does the statement still hold for the long-term perspective? As we stretch out the timeline, further complications may arise.

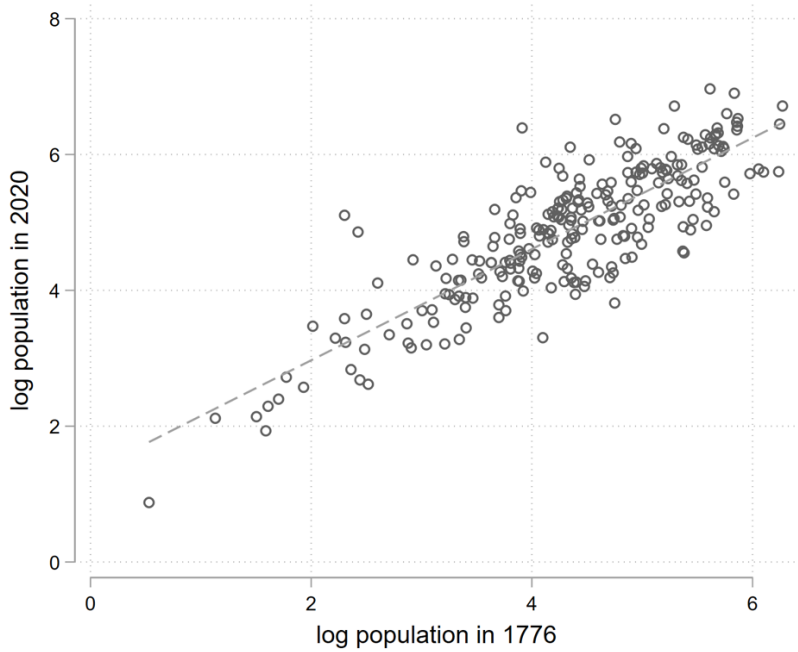
The relative development level among regions is quite persistent in the long term. Figures 1 and 2 have adopted the case of Chinese cities to illustrate this. Figure 1 provides a visualisation of the correlation of the logarithm of the GDP in 2019 and the respective statistics for 1991. It demonstrates that the in-sample fit is considerable even when adopting GDP level of Chinese cities. There have been success stories of rapid structural transformation and industrialisation during the past three decades, notably in cities such as Shenzhen and Suzhou. There are also episodes of decline, such as in some northeast cities. However, as Figure 1 reveals, such episodes are the exception rather than the rule; the relative GDP level of cities is quite persistent.

Figure 1: The link between logarithm GDP in 2020 and logarithm GDP in 1991



Even if one examines sustaining development from the more classic perspective of the regional population, the correlation remains very high. By merging historical Chinese city population data in 1776 with contemporary populations in 2020, the regression yields a close-to-unity elasticity, implying that cities that were 10% better than other places in 1776 are (on average), today, 8.18% more developed. The in-sample fit is considerable as well, with an R-squared over 0.71. In sum, while there are some success and failure cases, most Chinese cities are close to the regression line.

Figure 2: The link between logarithm population in 2020 and logarithm population in 1776



It is clear that using modern institutions to explain such persistent development would introduce the dilemma of endogeneity. Besides, for contemporary institutions *per se*, key pillars such as the rule of law, government effectiveness, or control of corruption take years, decades, or even centuries to construct. Once constructed, these organisational arrangements depreciate slowly. This decline makes it difficult to trace the factors contributing to city dynamics using contemporary data. While contemporary data can shed light on processes that determine the rise and fall of cities in modern times, much more can be learned through historical data and events that explain path dependencies. Therefore, exploring more persistent institutional explanations for regional development is imminent in current economic geography and development economics.

This PhD thesis, intending to trace the rationale of governments and corresponding development, draws on the lens of history and spotlights how institutional origins affect socio-economic inequalities at a regional level. Historical context provides several advantages to the analysis. First, historical contexts offer insights into distinct trajectories and institutions' processes of evolution that cannot be obtained from current data alone (Greif, 1998). History offers a range of natural experiments to help us identify critical parts of the development process. Some necessary institutional changes occur only infrequently, and, to learn their effect, one can only reach deep into the past. Second, studying institutional economics from a historical perspective also offers features that can help us better identify causal economic connections. Within a historical setting, there

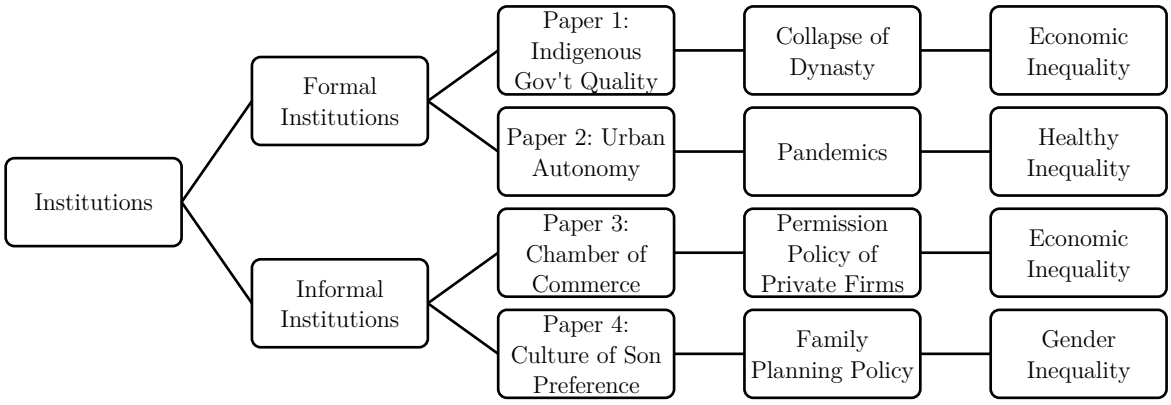
are fewer interdependencies between separate economic units, making the mechanisms of institutions more easily observable.

In this thesis, I aim to contribute to the literature by exploring underestimated historical contexts. In the first and third chapters, to excavate the effects of formal and informal historical institutions on economic inequality, a set of historical achieves and records were consulted to measure historical institutions. This is the first time both public and private-order institutional arrangements in developing regions were digitalised and examined empirically, to the best of my knowledge. The other two chapters add a new important perspective to the existing literature by researching social inequalities. Temporal and spatial differences in outcomes on public health and gender inequalities were detected and then linked with the institutional background of the times.

2 Themes of Research

The research topic of this thesis can be approached from various angles. First, the thesis focuses not only on a single type of institution. The first two chapters examine government institutions (also known as public-order institutions), a prototype of modern formal institutional arrangement. The last two chapters highlight the possible effects of informal institutions from the perspective of business networks and social customs. Second, to test the institution's persistence, the ideal approach is to examine the consequences of radical change through different mutations. Chapters One and Two explore the outcomes of shock events such as the dynasty's collapse and public health crises. Such political and health shocks are generally external and cannot be predicted by institutions per se. In addition to drawing on shock events, possible effects of public policies are explored in Chapters Three and Four. Finally, as possible outcomes of institutions, this thesis also explores different styles of inequalities. The institution's most well-known and perceptible impact is on the economic level. Chapters One and Three cover this issue and discuss spatial economic disparities at the county and city level by adopting novel data such as light-time luminosities and registration of universal enterprises. However, little is known about the role of institutions on the societal level: Chapters Two and Four endeavour to fill in this research blank. Figure 3 illustrates the research themes in this thesis.

Figure 3: Themes of Research



2.1 Definition of institutions

To examine the effects of institutions empirically, one first needs to understand what is meant by institutions. Defining institutions is an arduous task, and current literature on the topic has not agreed on a standard definition. This thesis adopts the relatively minimalist definition of institutions as a starting point to deal with the challenge. The most cited definition of institutions is given by North (1990), who defines institutions as 'the humanly devised constraints that structure human interactions. They comprise formal constraints (rules, laws, constitution), informal constraints (norms of behaviour, convention, and self-imposed codes of conduct), and their enforcement characteristics'.

Following this broad definition, formal institutions can be regarded as universal and transferable rules and generally include constitutions, laws, charters, bylaws, regulations, and elements such as the rule of law and property rights (Rodríguez-Pose, 2013). In contrast, informal institutions include a series of features of group life, such as culture, norms, traditions, social conventions, interpersonal contacts, relationships, and informal networks (Rodríguez-Pose & Storper, 2006). This dichotomous definition of institutions helps us avoid possible ambiguities and overlaps. Besides, it also explains that a comprehensive understanding can only be achieved after exploring both formal and informal institutions. Thus, this thesis's dichotomous definition of institutions has been accepted and analysed for formal and informal institutions.

2.2 Formal institutions

From the perspective of formal institutions, there has been ample research on the institutional origins of development. However, the two strands have followed somewhat different paths. Heavily affected by the study of Acemoglu et al. (2001, 2002) and La Porta

et al. (1997, 1998), empirical research has mainly focused on modern Western institutions' historical legacy (Acemoglu and Johnson, 2005). However, the Western world is not the only source of government institutions and anecdotal evidence and case studies have invariably stressed the role of deeply rooted indigenous institutions (Herbst, 2000). In this thesis, indigenous and universal institutional arrangements shaped in the pre-modern era are systematically explored.

In the first chapter, a Chinese indigenous institution, the *Chongfanpinan* system, is studied. This was a local governance institution set up by the ruler of the Qing Dynasty in 1731 and lasted until its downfall in 1912. The overarching aim of this institutional arrangement was to solve challenges in local governance by reducing complexity and allowing the imperial bureaucracy to better deal with the diverse conditions of large territories (Scott, 1998). Counties were divided into four classes, and well-trained civil servants were allocated to counties in the higher hierarchy (Liu, 2012). Since efficient officials and bureaucrats represent human capital and facilitate the efficient delivery of public goods and services, counties in higher tiers enjoyed better government quality (Xu, 2018). This study investigates whether pre-modern indigenous government institutions continue to affect economic performance in these regions.

The second chapter turns attention from the Eastern to the Western World. It investigates a universal institution, urban autonomy, which first occurred in Spain and Italy in the 11th century and gradually spread across Europe during the next few centuries (Van Zanden et al., 2012). Within autonomous cities, personal freedom was longer threatened by distant kings and was protected by law. Citizens who had been excluded from the urban government began taking part in the affairs of urban governance. City autonomy also brought about greater scrutiny of government officials' actions by local citizens and new forms of regulations. As a key political institution, urban autonomy effectively limited the monarchy's power, secured property rights, and stimulated economic development (Frank, 1978; North and Thomas, 1973). Thus, researchers have always regarded urban autonomy as a prototype of modern Western institutions and a key driver of continued economic development (Weber, 1956). However, we know next to nothing about the effects of urban autonomy in dealing with a social crisis. To fill this gap, this chapter examines how urban autonomy dealt with the Black Death across cities in Western Europe between 1347 and 1352.

2.3 Informal institutions

The category of informal institutions is more complex and are unlikely to cover all types in one thesis. Thus, I choose the two most representative issues, business networks and culture, to examine in the second part of the thesis.

The Chamber of Commerce system of the early 20th century of China is introduced in the third chapter to research business networks. Anecdotes and case studies have illustrated that business organisations and associations can take the place of a defective government in enabling the proper functioning of the market (Greif, 2006; North & Thomas, 1972). Throughout history, societies have lacked public authorities able and willing to enforce institutional rules governing economic activities. Substitute institutions like the Chamber of Commerce have emerged to replace the corresponding vacuums. There are heated debates on whether private-order institutions promote or inhibit overall economic output (Ogilvie, 2014; Prak & Van Zanden, 2013). The persistent impact of the chamber system is analysed by combining the historical record with the current firms' registration data to resolve the dispute.

The fourth chapter addresses the culture of son preference as another example of informal institutions. Like many other Asian countries, son preference is deeply rooted in China (Das Gupta et al., 2003). The sex-biased preference culture is nearly indelible in Chinese society, attributed to the patriarchal and old-age support systems. In the worst situation, the imprint of backward culture may change from implicit to explicit under the promotion of policies. The final chapter tests how the Two-Child policy affected the sex ratio at birth in the cultural background of son preference.

3 Summary of Chapters

Chapter 1: The long shadow of history in China: Regional governance reform and Chinese territorial inequality

Do external shocks affect local government quality and, consequently, long-term economic development? The collapse in 1911 of the Qing Dynasty in China was one of the greatest institutional shocks in world history, marking the end of more than 2,000 years of imperial rule. We exploit this shock to examine the impact of changes in historic local government quality on economic development today. By measuring variations in governance quality across 1,664 Chinese counties and examining their impact on long-term economic development, we show that historical differences in local governance quality are strong predictors of current geographical differences in economic development. This positive relationship is robust to a rich gamut of controls and checks. To further address

causality issues, we instrument historical government quality with the location of military towns in the preceding Ming dynasty. The analysis shows that history has left a deep legacy on governance differences across China that determine, to a considerable extent, current Chinese regional inequalities.

Chapter 2: Local institutions and pandemics: City autonomy and the Black Death

Local institutions have long been regarded as key drivers of economic development. However, little is known about the role of institutions in preparing places to cope with public health crises and pandemics. This chapter sheds light on how the nature of a local institution, city autonomy, influenced variations in the incidence of the Black Death — possibly the worst pandemic ever recorded— across cities in Western Europe between 1347 and 1352. We examine urban autonomy not only because it represented a major political shift in medieval times, but because, more importantly, it also represents a key prototype of modern political institutions. By exploiting data on the spatial variation of Black Death's mortality rates and duration using OLS and 2SLS methods, we uncover that city autonomy reduced mortality rates by, on average, almost 10 percent. Autonomous cities were in a better position to adopt swift and efficient measures against the pandemic than those governed by remote kings and emperors. This relationship has been confirmed by a series of placebo tests and robustness checks. In contrast, there is no evidence to suggest that city autonomy was a factor in reducing the duration of the pandemic in European cities.

Chapter 3: Inclusive institutions, entrepreneurship and innovation in China

What role do inclusive institutions play in the geography of entrepreneurship and innovation? In the third paper we test how local institutions in China shape the geography of new firm creation and innovation, using the period of the Mao era – when private ownership was effectively banned – as a structural break. We distinguish between Chambers of Commerce, inclusive institutions designed for open networking and opportunity creation, and merchant guilds, which were dominated by local elites and instead protected established firms. We use a difference-in-differences (DID) strategy and show that cities with inclusive business networks before Mao, proxied through Chamber of Commerce memberships, created more entrepreneurs and innovated more even after the market economy restarted in the 1980s. In contrast, the contribution of merchant guilds is negligible, a finding we argue is because chambers were more inclusive politically and to

non-elites. To further address causality concerns, we use the route of telegraph before the rollout of the chambers as a source of exogenous variation.

Chapter 4: Birth Quota and Gender Bias: Evidence from the Two-Child Policy

The fourth chapter examines whether birth quotas distort families' responses to childbirth. Beginning in the 1980s, 60 million people of ethnic minority in China were subject to the country's two-child policy (TCP). Drawing on variations in TCP implementation, we estimate that the second children born in families subject to the policy with a firstborn girl were 5.25% more likely to be boys. Further, the effect surged to 8.29% after the policy legalisation phase. The prolonged birth spacing additionally confirms the finding that the TCP encouraged sex discrimination at birth. We also show that the increased availability of ultrasound facilitated sex-selective abortions. Our study further indicates that Han–minority mixed families tended to avoid the TCP by officially faking twins during the second birth when enforcement was strengthened.

4 Conclusion

The above four chapters construct this PhD thesis and examine the multidimensional outcomes of local institutions. Drawing on the external shocks, the first two papers explore whether formal institutions could withstand crises and persistently place positive effects on regional development. Each article tests a different dimension, namely the economic prosperity proxied by the nightlight luminosity and social benefits measured by the mortality and duration of the pandemics. The third and fourth papers examine the influence of development policies on informal institutions and endeavoured to understand how the informal institutions could upheaval the spatial inequalities. Two national policies have been tested, the relaunch of private ownership and the Two-Child Policy. Facilitated with the rich time series data in the modern era, a complete impact process is provided in papers 3 and 4.

Across the first two research, I find consistent evidence to question the importance of inclusive formal institutions on the socio-economic inequalities for both the global north and south. Specifically, the result in the first chapter suggests that although the long-ruling dynasty demised, institutional inertia and path dependence have been well-preserved over hundreds of years. Central rulers can extensively manipulate the historical local government quality with huge investment of political resources. Counties assigned to higher administrative hierarchies had been favoured by decentralisation and good

governance and thus persistently perform better economically in the long run. This indigenous institution has long been effective for regional development, despite the impact of Western shocks.

These findings have substantial implications for the future of developing countries like China. They first suggest that, even in developing countries, good quality of government is the critical pillar for economic development. In contrast, weak institutions act as predators, which prevents any sort of potential and talent from emerging and prospering. Secondly, it implies that regional prosperity can be improved through strategy efforts. Nowadays, place-based programs, which aim to foster economic growth in a specific area with a jurisdiction, have grown popular and have been pursued by many governments worldwide over the past several decades. Although the policy has been implemented broadly, research shows its effects are mixed (Neumark & Simpson, 2015). The case of the Qing dynasty insights that place-based policies can forge a better institutional environment when two prerequisites are satisfied: 1) the implementation of the policy is long term and sustainable; 2) the content of the policy needs to focus on the provision of political resources. However, both elements are largely scarce in current regional policies.

The second paper shows that the benefits of good government quality are not only limited in the domain of economics but also in the social dimension. In the context of medieval Europe, we proxy the government quality with urban autonomy. A greater degree of autonomy could have allowed cities to react rapidly and swiftly introduce quarantines and lockdowns shortly after the arrival of the plague. Non-autonomous cities, by contrast, did not have the luxury of adopting their independent measures. With more discretion, autonomous cities can perform more flexibly when confronted with the Black Death. Our analysis finds that city autonomy is strongly correlated to lower Black Death mortality, even when other factors affecting variations in mortality during mediaeval times are considered. This paper enlightens policymakers in the current COVID-19 pandemic: improving the quality of local institutions was and can still be an efficient mechanism to fight pandemics. When greater local autonomy is matched with competent and resourceful local institutions that implement adequate policies, the chances of fighting current and future pandemics more efficiently are likely to increase.

The remaining two chapters cast attention on the impact of informal institutions. In the third chapter, drawing on the economic reform since 1980, I endeavour to seek the roots of regional entrepreneurship by exploring the persistence of an informal institution, the Chamber of Commerce in China. The empirical findings confirm the assumption of inclusiveness that when an increasing share of individuals are involved in the operation of non-exclusive institutions, the positive effects on entrepreneurship are more likely to

persist and spill-over when the external environment allows (Acemoglu et al., 2015). In contrast, the merchant guild, which limits benefits within the same group based on identity agreement, does not contribute in the long run. By examining the internal organisational structure of chambers, we further find that chambers with a higher proportion of non-elites could accelerate the process of entrepreneurship emergence, highlighting the importance of democratic participation and discussions within business organisations.

However, not all policy shocks and reforms can take advantage of institutions and drive positive results. Chapter 4 shows that inconsiderate policies may revive the dark side of informal institutions. Like many other developing countries, son preference was traditionally rooted in China's agricultural society (Das Gupta et al., 2003). This chapter finds that, after the family planning policy imposed a two-child restriction on minority families, the cultural preference for sons can no longer be smoothed out by additional fertility. Instead, parents applied prenatal and postnatal approaches such as abortion and infanticide against daughters' birth. The empirical findings show that, ten years after the policy was implemented, the second kids' male-to-female ratio reached 1.7.

Regarding policy choices and implementation, both the third and fourth papers highlighted the interactions between strategies and institutions. Both national and regional policies can awaken the influence of informal institutions from implicit to explicit. More importantly, China's reforms over the past four decades have proved that both good and evil can be triggered if the institutional context has not been considered in the policy design process.

References

Acemoglu, D., & Johnson, S. (2005). *Unbundling institutions. Journal of political Economy, 113*(5), 949-995.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review, 91*(5), 1369-1401.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly journal of economics, 117*(4), 1231-1294.

Alesina, A., & Giuliano, P. (2015). Culture and institutions. *Journal of Economic Literature, 53*(4), 898-944.

Chang, H. J. (2011). Institutions and economic development: theory, policy and history. *Journal of Institutional Economics, 7*(4), 473-498.

Cortinovis, N., Xiao, J., Boschma, R., & van Oort, F. G. (2017). Quality of government and social capital as drivers of regional diversification in Europe. *Journal of Economic Geography, 17*(6), 1179-1208.

Das Gupta, M., Zhenghua, J., Bohua, L., Zhenming, X., Chung, W., & Hwa-Ok, B. (2003). Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. *The Journal of Development Studies, 40*(2), 153-187.

Frank, A. G. (1979). *Dependent Accumulation* (Vol. 492). NYU Press

Glaeser, E. L., & Shleifer, A. (2002). *Legal origins. Quarterly Journal of Economics, 117*(4), 1193-1229.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition. *American Economic Review*, 525-548.

Greif, A. (1998). Historical and comparative institutional analysis. *American Economic Review, 88*(2), 80-84.

Greif, A. (2006). *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge University Press.

Herbst, J. (2014). *States and power in Africa*. Princeton University Press.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Legal determinants of external finance. *Journal of Finance, 52*(3), 1131-1150.

Liu, Z. (2012). *Archival documents form the ming-qing era*. Humanities Center of Chengchi University

Ogilvie, S. (2014). The economics of guilds. *Journal of Economic Perspectives, 28*(4), 169-92.

Prak, M., & Van Zanden, J. L. (2013). *Technology, Skills and the Pre-modern Economy in the East and the West*. Brill.

Rodríguez-Pose, A. (2013). Do institutions matter for regional development?. *Regional studies, 47*(7), 1034-1047.

Rodríguez-Pose, A., & Di Cataldo, M. (2015). Quality of government and innovative performance in the regions of Europe. *Journal of Economic Geography, 15*(4), 673-706.

Rodríguez-Pose, A., & Ganau, R. (2022). Institutions and the productivity challenge for European regions. *Journal of Economic Geography, 22*(1), 1-25.

Rodríguez-Pose, A., & Storper, M. (2006). Better rules or stronger communities? On the social foundations of institutional change and its economic effects. *Economic geography, 82*(1), 1-25.

Rodríguez-Pose, A., & Zhang, M. (2019). Government institutions and the dynamics of urban growth in China. *Journal of Regional Science, 59*(4), 633-668.

Rodríguez-Pose, A., & Burlina, C. (2021). Institutions and the uneven geography of the first wave of the COVID-19 pandemic. *Journal of Regional Science, 61*(4), 728-752.

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.

Spolaore, E., & Wacziarg, R. (2013). How deep are the roots of economic development?. *Journal of Economic Literature, 51*(2), 325-69.

Iddawela, Y., Lee, N., & Rodríguez-Pose, A. (2021). Quality of sub-national government and regional development in Africa. *The Journal of Development Studies, 57*(8), 1282-1302.

Neumark, D., & Simpson, H. (2015). Place-based policies. In *Handbook of regional and urban economics* (Vol. 5, pp. 1197-1287). Elsevier.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge university press.

North, D. C., & Thomas, R. P. (1973). *The rise of the western world: A new economic history*. Cambridge University Press.

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

Rothstein, B. O., & Teorell, J. A. (2008). What is quality of government? A theory of impartial government institutions. *Governance, 21*(2), 165-190.

Van Zanden, J. L., Buringh, E., & Bosker, M. (2012). The rise and decline of European parliaments, 1188–1789 1. *The Economic History Review, 65*(3), 835-861.

Weber, Max. (1958). *The City*, New York

Xu, G. (2018). The costs of patronage: Evidence from the British empire. *American Economic Review, 108*(11), 3170-98.

Chapter 1: The Long Shadow of History in China: Regional Governance and Chinese Territorial Inequality

1 Introduction

Institutional quality has long been considered one of the main pillars of economic development (Acemoglu et al., 2001; La Porta et al., 1998; Rodríguez-Pose, 2013). Good government institutions, such as the effective rule of law, low levels of corruption, and efficient protection of property rights, are at the root of differences in development (Acemoglu & Johnson, 2005; Fukuyama, 2006; Rothstein, 2011). Weak government institutions are also increasingly considered important across developing countries (e.g., Acemoglu et al., 2015; Iddawela et al., 2021; Jia, 2014; Michalopoulos & Papaioannou, 2014). Government quality matters at both the national and sub-national levels. However, how much social and political stability is needed for government quality to have an impact on economic performance? Can deeply rooted historical differences in government quality persist after experiencing large institutional shocks?

China provides an ideal setting to analyse the interaction between formal institutions and shocking events, not just because of its economic importance and recent dynamism, but also because of its long and rich history as a relatively well-organised and structured empire (Koss, 2017; Kuhn, 2002; Spence, 1990). The rise and fall of the empire undoubtedly had a huge impact on the entire country. The demise of China's last imperial dynasty happened during the 1911 Revolution (also known as the Xinhai Revolution), when China's last emperor, Pu Yi, announced his abdication on 12 February 1912. This marked the end of over 2,000 years of imperial rule and the beginning of China's republican era. A political system that had lasted more than two millennia suffered a sudden and abrupt collapse.

The fall of the Qing Dynasty unleashed events that have had a fundamental impact on the country and the lives of ordinary Chinese over the last century. The Empire's downfall brought about momentous political transformations, radical changes to traditional Chinese culture and values based on Confucianism, and, more recently, a complete overhaul of the country's economy. China was undoubtedly thrust into a completely different era after 1912. However, do these deep-seated changes mean that the influence of institutions brewed over two millennia of imperial rule no longer count for anything? Or do they still shape differences in development across the country?

This paper aims to address this issue by checking whether historical institutions continue to affect the country's long-term economic performance after the collapse of the last

imperial era. To assess historical government quality, we have digitalised one of the most influential institutions in the Qing dynasty: the *Chongfanpinan* system. This political institution, implemented in 1731, led to a durable civil service system, which was not abolished until the collapse of the Qing dynasty in 1911. By adopting this system, the central government actively contributed to determining the quality of local governments in China by allocating bureaucratic resources differentially. In other words, more and better human resources were awarded to selected counties to improve governance. *Chongfanpinan* contributed to an important divergence in local government quality, as the chosen counties were allocated the best administrators and attracted more resources.

The analysis examines the extent to which current differences in economic development are determined by differences in local government quality during the late imperial times. We have digitised the *Chongfanpinan* government quality data by hand using the historical records called 'Veritable Records of the Qing Emperors (*Qingshilu*)'. These records were later combined with indicators of current economic development at county level. The current development indicator is elaborated using satellite images obtained in 2010 of night-time light density. The results of the analysis show that the shadow of history in China is still very long. In spite of the substantial changes brought about by modernisation in the last century, differences in government quality in imperial times still determine variations in economic development this century. Chinese counties with better historical government quality as a result of the long-term implementation of the *Chongfanpinan* system are today more developed than those with lower-quality government a century ago. This connection between historical governance quality and current levels of economic development is robust to the inclusion of a rich gamut of covariates, including county-level geographic, historical socio-economic, and contemporary socio-economic controls, as well as prefecture-level fixed effects.

To further address causality concerns, we have instrumented differences in imperial-time government quality with the location of military towns during the Ming dynasty (1368–1644), which preceded the Qing dynasty). More specifically, we have constructed a measure of vulnerability to military conflicts, based on the location of garrisons in the Ming dynasty, to isolate exogenous variation in the decision to select counties. The imperial government of the time prioritised counties that were close to borders or in areas prone to military conflict.

This research contributes to the political economy literature on how political factors affect economic geography. In contrast to existing studies (Ades & Glaeser 1995; Campante & Do, 2014; Davis & Henderson, 2003), our focus has been on the effect of subnational historical conditions on contemporary economic outcomes. We revealed some new factors underlying the link between deep-rooted institutional structures and

economic development. We also contribute to a growing literature examining how historical institutions affect the process of development (Chen et al., 2017; Dell et al., 2018; Duranton et al., 2009; Michalopoulos & Papaioannou, 2013). More precisely, our study uses a historical indicator of governance quality that is, to the best of our knowledge, unique. As such, it relies on criteria first developed by bureaucrats and civil servants to construct the indicator of institutional quality.

The remainder of this paper proceeds as follows. Section 2 describes the literature and historical context of the imperial governance system and the shock of the collapse of the Qing dynasty. Section 3 introduces the data used in the analysis and presents the descriptive statistics. Section 4 reports the estimation results. Section 5 examines the two possible channels through which historical governance conditions continue to influence economic development patterns at the county level in China, while Section 6 further tests the effect of Western shocks on the imperial governance system. Section 7 offers the main conclusions.

2 Institutional Background

As background to our analysis, we first discuss the definition and perspective on the government quality we adopt, before describing our measurement of historical local government quality and the collapse of the Qing dynasty, the last imperial dynasty. This discussion aims to link historical institutional quality with the external shocks in the Chinese context.

2.1 An alternative approach to measuring government quality

Since at least the 1990s, institutions have generally been treated as key factors in shaping economic development (La Porta et al., 1998; North, 1990). Government quality, in particular, has attracted considerable attention. This is a multifaceted concept, encompassing all elements conditioning the government capacity to provide public goods and services. One of the most commonly used definitions of government quality comes from Rothstein and Teorell (2008: 165), who consider government quality as 'the impartiality of institutions that exercise government authority'. This concept links with that of good governance, viewed as 'the traditions and institutions by which authority in a country is exercised' (Kraay et al., 2010:5). Based on these and related definitions, the World Bank compiled the widely used Worldwide Governance Indicators, a set of measures of quality of government for 136 countries. This indicator inspired similar measurements at a regional level for the European Union.

Most government quality measurements rely on subjective indicators. Calculating government quality based on surveys and/or interviews is certainly a valid approach but also has some limitations. First, in certain political regimes it may not be feasible to obtain data on government quality simply through interviews, as those surveyed may not be able to express themselves freely. In some cases, academic institutions or international organisations may not be permitted by the local government to conduct interviews on government administrative capabilities. Moreover, historical records of government quality based on surveys are inexistent; therefore, relying solely on interview data is frequently only valid to examine either recent or current government quality.

To fill this gap, we follow Fukuyama's (2013) broader definition of government quality. He defines government quality as 'a government's ability to make and enforce rules, and to deliver services, regardless of whether that government is democratic or not'. Following this perspective, we consider that an efficient bureaucracy —according to a rationalistic Weberian approach— will produce better services than one that is highly discretionary or riven by nepotistic and clientelistic ties. Well trained and efficient officials and bureaucrats represent the human capital of a government and facilitate the efficient delivery of public goods and services. Several empirical studies have uncovered the connection between well-trained bureaucrats and efficient governments (Balan et al., 2020; Fukuyama, 2013; Xu, 2018). In this research, we follow this approach and measure government efficiency through the level of training and professionalisation of the civil service.

2.2 Historical differences in local government quality in China

It was not until a century into the Qing dynasty that a comprehensive reform aimed at discriminating among counties and improving the quality of local government was implemented. In 1731, Emperor Yong Zheng, the fourth emperor of the Qing dynasty, set up a local governance system called *Chongfanpinan*. The overall aim of the *Chongfanpinan* system was to improve the use of the limited human and financial resources by allocating more resources and, especially, better trained civil servants to those counties considered to be of strategic importance for the Empire (Liu, 2012). The objective of the reform was to solve challenges in local governance by reducing complexity and allowing the imperial bureaucracy to better deal with the diverse conditions of its territories (Scott, 1998).

The establishment of the *Chongfanpinan* system represented a fundamental bureaucratic reform of governance in China. The Chinese state selected locations to place more and better human and economic resources based on a series of suitability criteria. These criteria included: (i) the importance of transport, (ii) the complexity of government

affairs, (iii) the difficulty of tax collection, and (iv) the regional security situation (Liu, 1993). Different counties in China were classified according to these criteria. The more they met the conditions, the greater the likelihood of receiving more resources through the *Chongfanpinan* system. Meanwhile, the central government was also concerned about military factors. Maintaining social stability and territorial unity was of paramount importance to uphold peace and prosperity in what fundamentally was a centralised empire (Spence, 1990). Putting all these criteria together, a four-county tier system was established. Counties with a score of four were ranked highest; those with a score of one were ranked lowest.

The allocation of a county to a specific tier had consequences. First, the highest-ranked counties were deemed of strategic importance for the Empire and rewarded with additional human resources. The highest qualified and, in theory, more efficient bureaucrats were posted to top-tier counties —those awarded the greatest importance in the Chinese territorial hierarchy. To ensure that the Empire remained strong and prosperous, top-tier counties were be governed by bureaucrats with a wealth of experience in governance and the best educational background possible (Wang, 2007). Matching competent officials with top-tier places was believed to maximise the efficiency of governance.

Secondly, the allocation of well-prepared and high-ranking officials meant a de facto higher degree of decentralisation in top-tier counties. In counties that, according to the *Qingshilu*, were ranked higher than tier 2, provincial governors could participate in the process of appointing and dismissing local officials. In contrast, authority over civil servants in the remaining low-tier counties was kept under the strict control of the Ministry of Personnel in the central government. The rationale for this *de facto* political decentralisation in top-tier counties was that the provincial governors were more familiar with local conditions and understood the human capital needs of specific counties in their provinces. This practice limited the power of the distant and often aloof Ministry of Personnel in Beijing and promoted greater autonomy in the places deemed to be governed better (Hucker, 1985).

To sum up, the *Chongfanpinan* system represents the first known and widespread record associated with potential differences in the quality of local governance across a large number of sub-national administrations in a historical polity. The classification of different counties into different tiers had far-reaching consequences in terms of governance changes, as the top-ranked counties were allocated better resources for the handling of administrative affairs. This would have contributed to a greater spatial inequality in local governance quality in China, which grew over more than two centuries until the demise of the *Chongfanpinan* system with the downfall of the Qing dynasty in 1912.

2.3 The collapse of the Qing dynasty

The collapse of Imperial China at the beginning of the 20th century represented a complete break from the past. Imperial China, under the Qing dynasty, had grown increasingly inefficient. It had failed to modernise, leading to a weak state unable to confront the influence of Western powers. Discontent in elite groups, often established overseas, and the population in general, became more and more evident since the 1890s. The Revive China Society (*Xingzhonghui*) —founded by Sun Yat-sen in Honolulu in 1894[1] — was most prominent among these revolutionary groups. The early uprisings were quickly suppressed by the government. However, the insurrections kept on growing and, after the success of the Wuchang Uprising in Hubei province on 10 October 1911, the position of the Qing dynasty became untenable. Just two months after the Wuchang Uprising, fifteen provinces, including Hunan and Guangdong, had declared their independence. The revolution ended with the abdication of Emperor Pu Yi, marking the beginning of China's republican era.

The collapse of the Qing dynasty represents a fundamental turning point in Chinese history and one that has had a profound impact on the development of modern China, as it unleashed a series of transformations that turned Chinese traditional society and polity upside down. In politics, the demise of the Qing dynasty initiated China's political modernisation. Politicians tried to learn from the American political system and set up a presidential republic. Although the republican system was not implemented in the early years of the Republic of China, the values of democracy and republic were widely disseminated. From a cultural perspective, the authority of traditional Chinese culture and values based on Confucianism was severely shaken —a transition that was rapidly accelerated with the advance of communism. In economic terms, China has since embarked on several industrialisation and economic transformation waves, which are transforming China from the economic laggard that it had become at the end of the Qing dynasty into one of the main locomotives of the world economy. The demise of the Qing dynasty ushered a completely different era for China. However, did these changes do away with the institutional traditions established over centuries in Imperial China? In this paper, we aim to answer this question.

[1] The Revive China Society is the predecessor organisation of the Kuomintang (KMT).

3 Data and descriptive analysis

3.1 Data on historical government quality

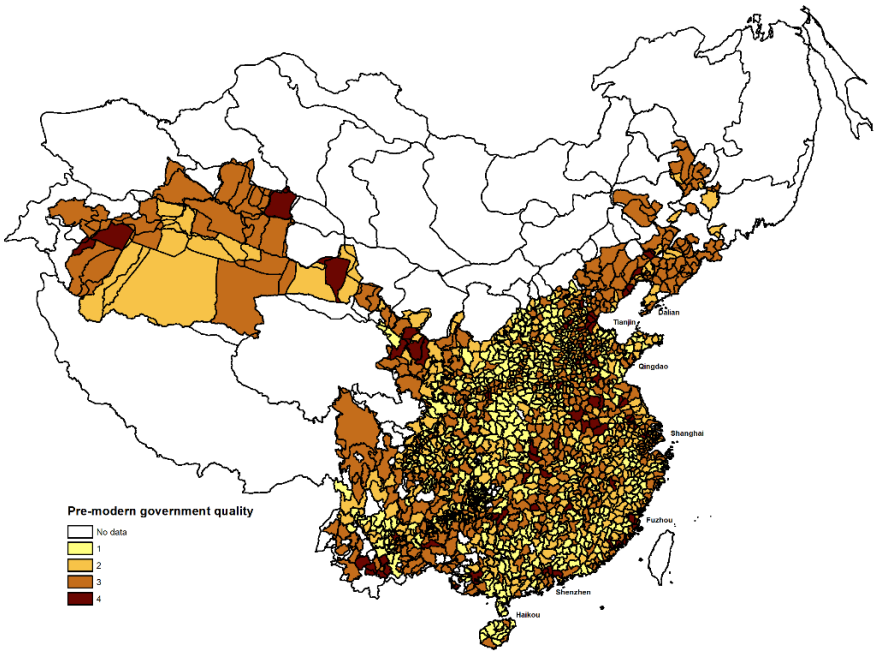
To explore the relationship between historical government quality and current territorial inequalities, we first map historical county differences in governance quality, according to the *Chongfanpinan* reform. This involved digitalising the textual data recorded in the Veritable Records of the Qing Emperors (*Qingshilu*).[2] The *Qingshilu* contains the official compilation of historical records of the Qing dynasty, including a large number of political, economic, cultural, and geographical aspects. The *Chongfanpinan* system is one of the many records in the *Qingshilu*. To better represent the data contained in the *Qingshilu*, we use a map containing the county borders of the Qing dynasty as they were in 1820.[3] The data of 1,664 counties in total were digitalised (Figure 1), containing the overwhelming majority of the population of modern China. The counties not included in the analysis (mostly in territories in present-day Tibet, Qinghai, and Inner and Outer Mongolia) were border areas where the Qing government did not implement the *Chongfanpinan* reform.[4] Interestingly, as Figure 1 shows, counties with higher state capacity were not, as might be expected, overwhelmingly concentrated in the economically developed coastal regions.

[2] The detailed approach to digitalising the historical data is illustrated in Appendix A.

[3] The border file was provided by CHGIS (2007).

[4] Moreover, some border regions, although formally part of the Chinese Empire, were not under the de facto jurisdiction of the Qing dynasty.

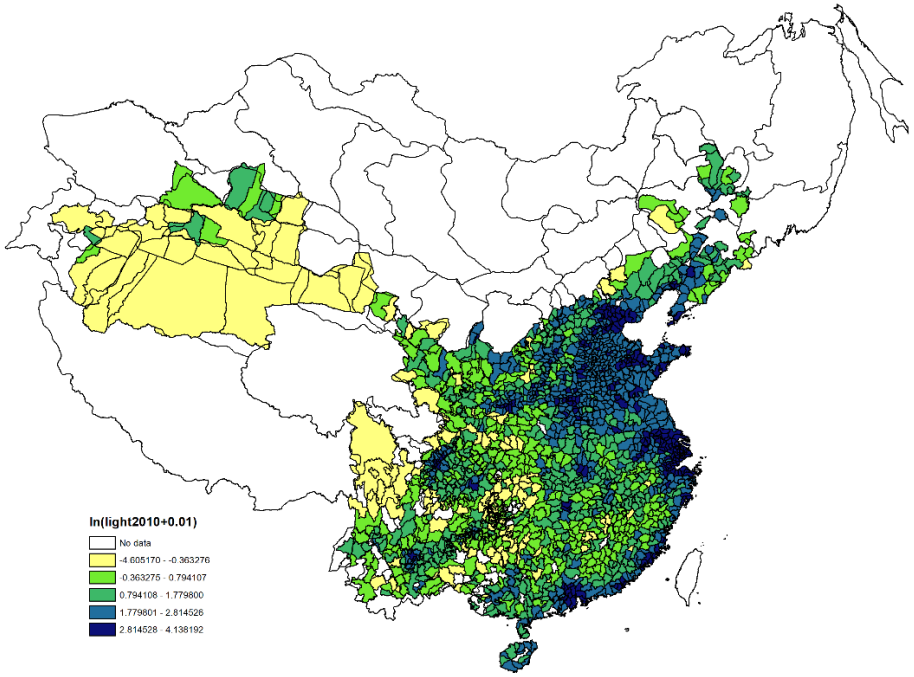
Figure 1: Historical government quality between 1731 to 1911



3.2 Night-time satellite light density

Given the lack of an exact correspondence between historic and present-day county boundaries, we follow Henderson et al. (2012) and use satellite data of night-time light luminosity in 2010 to measure current levels of economic development. Research using this method has proven that night light data is a reliable indicator for measuring local economic development (Michalopoulos & Papaioannou, 2013; Storeygard, 2016; Frick et al., 2019). In our case, night-time satellite light density has the advantage that a) we can combine historical boundaries to current levels of development in a more accurate way than by using official GDP data, and that b) it allows us to overcome the unreliability of some Chinese county-level GDP data (e.g., Chen et al., 2019; Michalski & Stoltz, 2013). Figure 2 provides an overview of the 2010 level of development in historical counties of China, using night light luminosity as a proxy.

Figure 2: Night light luminosity in 2010



3.3 Control variables

Current levels of development in China are not simply the result of historical differences in the quality of local governance. Indeed, a large number of other factors play a role in deciding why certain Chinese counties are more developed than others. To take these factors into account, we use four categories of control variables, which are likely to influence current local levels of prosperity in China. These categories include geographic factors, historical and contemporary socio-economic factors, and prefecture-level fixed effects.

Geographic factors. We consider several key geographical features: longitude, latitude, the size of land area, average slope, and elevation. Pomeranz (2009) argues that, in the case of China, natural conditions played a fundamental role in determining China's transition from the pre-modern to the modern era, leading to the great divergence in wealth between China and Europe. Longitude and latitude have also featured prominently in accounts of differences in economic development by several social scientists (Diamond, 1998; Sachs, 2001), as has also been the case of ruggedness (Nunn & Puga, 2012).

Historical socio-economic factors. We also control for a series of historical socio-economic factors, which might have influenced historical differences in development in China. We consider whether a county was a local capital during the Qing dynasty, its distance to the coast or the Grand Canal, its distance to the national capital (Beijing), and local agricultural yields. Being a local capital during the Qing era would have attracted resources and development. Similarly, distance to the national capital and the Grand Canal may have contributed to trade and prosperity (Bai

& James, 2015). The coast was the gateway to outside technology, knowledge and trade (Chen et al., 2017), giving coastal counties —and those nearby— a considerable economic advantage.

Contemporary socio-economic factors. To tease out the contemporary socio-economic effects, we use the population census data at county level in 2010.[5] We take into consideration the average years of schooling of the adult population, manufacturing and urbanisation ratios, and population density. As there are boundary changes between counties between the Qing dynasty and modern China, we employ GIS techniques to map contemporary data to the county boundary in the Qing dynasty. Appendix B explains the process of data construction in detail.

Prefecture fixed effects. The prefecture represents the division above the county level in the Chinese administrative structure, which encompasses between five and ten counties. Since China is a big and diverse country, it is difficult to compare counties across directly due to cultural and geographical differences. Therefore, introducing prefecture fixed effects, captures other unobserved effects. The sources and descriptive statistics of all the variables are summarised in Appendix Table A1.

Table 1: Summary statistics of historical government quality, overall

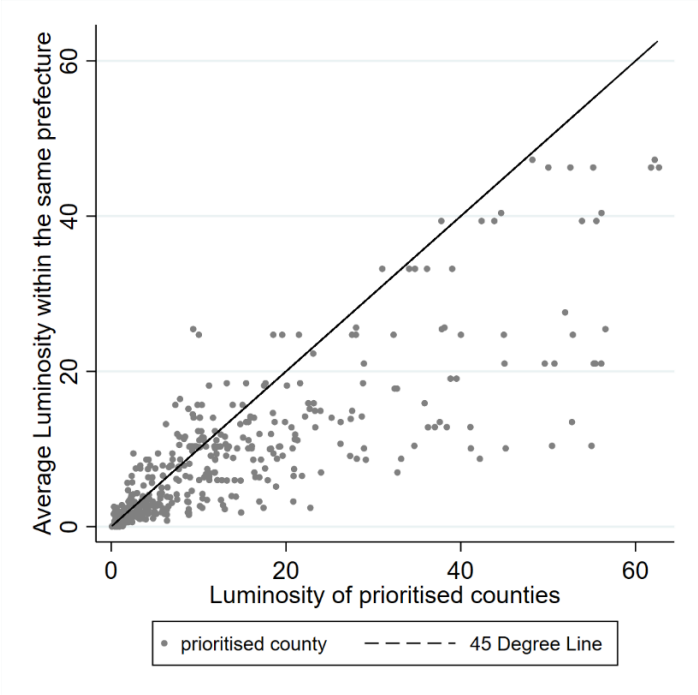
Variable	**obs.**	**mean**	**std**	**min**	**max**
		Panel A: All Observations			
nightlight	1,664	8.622	10.444	0.019	62.679
ln(0.01+nightlight)	1,664	1.502	1.252	-3.532	4.138
		Panel C: hierarchy=1			
nightlight	652	6.596	7.058	0.062	50.4
ln(0.01+nightlight)	652	1.364	1.1	-2.62	3.92
		Panel D: hierarchy=2			
nightlight	445	8.582	10.458	0.074	60.651
ln(0.01+nightlight)	445	1.519	1.212	-2.482	4.105
		Panel E: hierarchy=3			
nightlight	493	11.679	11.939	0.019	62.679
ln(0.01+nightlight)	493	1.569	1.385	-3.532	4.138
		Panel F: hierarchy=4			
nightlight	74	16.935	17.131	0.26	56.078
ln(0.01+nightlight)	74	2.173	1.331	-1.311	4.027

The descriptive statistics for the dependent variable (current levels of development) at different ranks of the independent variable of interest (Qing dynasty county governance tier) are presented in Table 1. The mean value of luminosity at the county level in 2010 is 8.622, while the median is significantly lower at 4.931. There are, however, large differences in luminosity across counties,

[5] The reliable county level data could only be achieved through population census. Generally, China conducts a population census every 10 years. The latest census data is in 2010.

which can be related to historical differences in local governance. The mean luminosity in the counties in tier 1, according to the *Chongfanpinan* system of measurement, is 6.596. For counties in tier 2 it is 8.582, and it increases successively, reaching 11.679 and 16.935 for tiers 3 and 4, respectively. This suggests that counties with better historical governance quality are economically far better off today. To ensure that the reported differences are not driven by a few outliers, we plot the mean luminosity of counties in the top historical *Chongfanpinan* ranks (those in tier 3 and tier 4) vis-à-vis counties with a lower rank within the same prefecture and find the economic development much better (to the left of the 45-degree line) in the great majority of cases (70%) (Figure 3).

Figure 3: The luminosity of prioritised counties and normal counties within the same prefecture



4 Empirical analyses

4.1 Empirical setups

In order to examine whether a higher score in the *Chongfanpinan*—indicating better local governance between one and three centuries ago—has influenced current territorial disparities in China, we use the following specification:

$$y_{i,p} = \beta Historical\ government\ quality_{i,p} + \gamma_1 X^g_{i,p} + \gamma_2 X^h_{i,p} + \gamma_3 X^c_{i,p} + \alpha_p + \varepsilon_i \qquad (1)$$

where the dependent variable $y_{i,p}$, reflects the level of economic development of county i in prefecture p in 2010, proxied by night light density. *Historical government quality*$_{i,p}$ is the explanatory variable of interest. It denotes local governance quality in county i in prefecture p under the *Chongfanpinan* reform during the later period of the Qing dynasty between 1731 and 1911. The equation includes several controls, which can affect differences in present-day levels of development across China. $X^g_{i,p}$represents a series of geographic controls, including land area, longitude, latitude, slope, and the ruggedness of the county topography i in prefecture p. $X^h_{i,p}$ controls for historical socio-economic factors, such as whether the county was the seat of a prefectural capital in historical times, its distance to the national capital (Beijing), distance to the coast, distance to the Grand Canal, and agricultural yields of county i in prefecture p. $X^c_{i,p}$ considers several contemporary socio-economic factors, including population density, years of schooling of the adult population, and the urbanisation and manufacturing ratios of county i in prefecture p. All regressions are conducted with prefecture-level fixed effects.

4.2 Baseline results

Table 2 reports the baseline results conducted using OLS. These show that there is a positive and highly significant association between historical governance quality and present-day development levels in China. In Column (1), including only prefecture-level fixed effects, the coefficient on the historical government quality is positive and highly significant (0.210). The coefficient also remains significant when geographic controls are included in column (2). Physical geography factors, such as latitude and longitude as well as terrain ruggedness, do not determine current differences in economic prosperity, which is slightly inconsistent with previous research (Diamond, 1998; Sachs, 2001), but aligned with the literature that has highlighted the primacy of institutions over geographical factors (Ketterer and Rodríguez-Pose, 2018; Rodrik et al., 2004). Column (3) shows that governance quality in the late Qing period is robust after controlling for historical socio-economic factors. Places located close to Beijing or the coast are wealthier than more distant ones. Chen et al (2017) also found that being close to the coast matters for economic prosperity in China. The results are also robust to the introduction of contemporary socio-economic factors (Column (4)). Greater shares of urbanisation and manufacturing and a more educated population are all connected with higher levels of GDP per head today. The introduction of additional controls —and fundamentally and as expected, of contemporary socio-economic controls— lowers the coefficient of historical governance quality. However, it remains positive and significant at a 99% confidence level throughout.

Table 2: Impact of historical government quality on county night light luminosity in 2010: OLS estimate, overall

	(1)	(2)	(3)	(4)
VARIABLES	Night light luminosity in 2010			
Historical government quality	0.210***	0.235***	0.218***	0.110***
	(0.0256)	(0.0223)	(0.0221)	(0.0233)
Land area (logged)		-0.329***	-0.313***	-0.224***
		(0.0516)	(0.0496)	(0.0441)
Longitude		0.0796	0.0868	0.00785
		(0.0606)	(0.0551)	(0.0398)
Latitude		0.0935	0.0760	0.0739
		(0.0665)	(0.0596)	(0.0543)
Slope		-2.162***	-2.406***	-1.762***
		(0.327)	(0.322)	(0.289)
Elevation (logged)		-0.0903	-0.0248	-0.0761
		(0.0554)	(0.0497)	(0.0481)
Prefectural capital			0.101	0.0314
			(0.0758)	(0.0627)
Distance to the coast			-0.199***	-0.0939**
			(0.0548)	(0.0415)
Distance to the Grand Canal			-0.0490	0.00663
			(0.0457)	(0.0391)
Distance to Beijing			-0.161***	0.0223
			(0.0449)	(0.0454)
Agricultural yields (logged)			0.0806	0.185
			(0.181)	(0.118)
Human capital (logged)				2.061***
				(0.306)
Urbanization ratio				1.207***
				(0.235)
Manufacturing ratio				0.0155***
				(0.00300)
Population density				149.5
				(112.3)
Constant	0.433***	-9.778*	-5.928	-6.933*
	(0.0767)	(5.568)	(5.156)	(3.640)
Geographic controls	N	Y	Y	Y
Historical socio-economic controls	N	N	Y	Y
Current socio-economic controls	N	N	N	Y
Prefecture Fixed Effects	Y	Y	Y	Y
Observations	1,664	1,664	1,638	1,456
R-squared	0.737	0.811	0.811	0.858

Robust standard errors adjusted for clustering at the prefecture level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In Table 3, we repeat the exercise by treating our variable of interest, historical government quality, as an ordinal variable. Treating a variable of interest as an ordinal variable could facilitate the interpretation of results, particularly if the hierarchical levels have different impacts. The reference group is the counties with the lowest historical government quality (those historically awarded a score of one). The results show that county night light luminosity in 2010 increases as historical government quality levels improves. For example, in column (1) —including only prefecture-level fixed effects—, the average night light luminosity of counties that were historically ranked in the top government quality tier is US$1,086.44 higher than that of counties at the bottom of the government quality scale.[6] Including all the control variables narrows the gap between the two groups to US$661.60 (column (4) of Table 3), but the gap remains significant and robust.

Overall, the results point to a strong legacy of past local governance in China on present-day economic outcomes. Counties that were deemed more strategically important and/or better governed and that, as a consequence, were allocated better trained and more able bureaucrats during the Qing dynasty are today far more developed than those that had weaker governance and were not prioritised in terms of resources. This result is robust to the introduction of many controls that could have helped make the county better off in both historical and present times.

[6] This figure is based on a back-of-envelope calculation: 4,550.45*e^0.280 /8.622. USD 4,550.45 was China's per capita GDP in 2010; 8.659 was the average luminosity in that year.

Table 3: Historical government quality on county night light luminosity in 2010: OLS estimate; variable of interest as ordinal variable

	(1)	(2)	(3)	(4)
VARIABLES	Night light luminosity in 2010			
Hierarchy 2	0.240***	0.239***	0.232***	0.167***
	(0.0547)	(0.0477)	(0.0477)	(0.0466)
Hierarchy 3	0.391***	0.455***	0.420***	0.236***
	(0.0606)	(0.0500)	(0.0500)	(0.0470)
Hierarchy 4	0.722***	0.750***	0.711***	0.226**
	(0.114)	(0.103)	(0.104)	(0.105)
Land area (logged)		-0.328***	-0.313***	-0.222***
		(0.0515)	(0.0496)	(0.0439)
Longitude		0.0793	0.0860	0.00324
		(0.0607)	(0.0553)	(0.0393)
Latitude		0.0923	0.0758	0.0732
		(0.0666)	(0.0595)	(0.0538)
Slope		-2.155***	-2.396***	-1.751***
		(0.325)	(0.321)	(0.289)
Elevation (logged)		-0.0922*	-0.0268	-0.0731
		(0.0556)	(0.0499)	(0.0491)
Prefectural capital			0.105	0.0249
			(0.0758)	(0.0623)
Distance to the coast			-0.200***	-0.0963**
			(0.0551)	(0.0410)
Distance to the Grand Canal			-0.0473	-0.00400
			(0.0471)	(0.0351)
Distance to Beijing			-0.158***	0.0181
			(0.0459)	(0.0457)
Agricultural yields (logged)			0.0813	0.190*
			(0.178)	(0.115)
Human capital (logged)				2.097***
				(0.310)
Urbanization ratio				1.198***
				(0.235)
Manufacturing ratio				0.0154***
				(0.00300)
Population density				165.7
				(116.7)
Constant	0.673***	-9.461*	-5.666	-6.214*
	(0.0606)	(5.562)	(5.213)	(3.602)
Geographic controls	N	Y	Y	Y
Historical socio-economic controls	N	N	Y	Y
Current socio-economic controls	N	N	N	Y
Prefecture Fixed Effects	Y	Y	Y	Y
Observations	1,664	1,664	1,638	1,456
R-squared	0.737	0.811	0.811	0.859

Robust standard errors adjusted for clustering at the prefecture level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.3 Causal identification

To what extent does the fact that better-governed counties during the Qing dynasty could have been those that were already better off affect our results? The *Chongfanpinan* reform may have fundamentally favoured rich counties and, thus, contributed to accentuating already existing historical territorial disparities in China. Moreover, there is also the risk that, despite the large number of controls included in the analysis, omitted variables may bias the results. To deal with these concerns, we employ an instrumental variable (IV) approach.

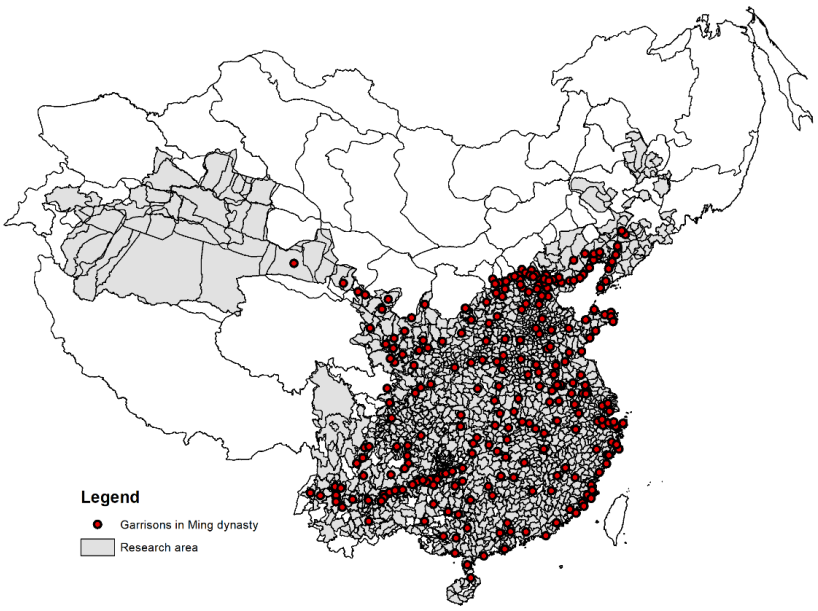
4.3.1 The location of garrisons as the instrument variable

To isolate exogenous variation in the decision to select counties, we construct measures of vulnerability to military conflicts based on the location of garrisons during the Ming dynasty (1368–1644), the dynasty that preceded the Qing dynasty.[7] During the Ming period, the Chinese government granted more autonomous powers to counties that were close to the military front lines. This was considered a measure that would allow them to respond quickly to military threats. In border areas provincial governors could freely choose local officials who were suitably acquainted with local characteristics. The arrival of the Qing dynasty led to the demise of the Ming-era garrison system in 1711 (Mao, 2018). It was replaced by a new military order, known as the Eight Banners stationary military system (Ding, 2003). Thus, the distance to the garrison in the Ming regime was completely exogenous from decisions to classify counties at particular levels of government quality in the *Chongfanpinan* system.

Figure 4 reports the location of military garrisons during the Ming dynasty. The density of garrisons is far higher along the Sino-nomadic frontier (Bai & Kung, 2011). The coast as well as strategic trade routes towards Central Asia and India was also heavily fortified (Figure 4). At the time, the Chinese government was more likely to award more autonomous resources and better officials to counties close to military sites. Using the location of garrisons during the Ming dynasty as our IV strategy, they can, therefore, be considered as orthogonal to proximity to counties, conditional to fixed effects and controls.

[7] The source of location of military garrisons during the Ming dynasty is obtained from http://worldmap.harvard.edu/data/geonode:Ming_Garrisons

Figure 4:The location of garrisons in the Ming dynasty



The relevance of our chosen instrument is confirmed by the first stage analysis results (Table 4). We begin by regressing in column (1) the military presence during the Ming dynasty —proxied by the number of garrisons in a county— on pre-modern government quality, including only prefecture-level fixed effects as controls. The coefficient is positive (0.349) and highly significant. The inclusion of geographic and historical and contemporary socio-economic controls in columns (2), (3), and (4), respectively, lowers somewhat the dimension of the coefficient, but it remains strongly positive and significant. Hence, the chosen instrument represents a good predictor of local governance quality during the Qing dynasty.

Table 4: The first stage result

	(1)	(2)	(3)	(4)
VARIABLES	Historical governance quality			
Garrison in Ming dynasty	0.349***	0.318***	0.281***	0.220***
	(0.0525)	(0.0473)	(0.0416)	(0.0394)
Land area (logged)		0.470***	0.521***	0.562***
		(0.0658)	(0.0525)	(0.0489)
Longitude		-0.0192	-0.0374	-0.0414
		(0.0338)	(0.0409)	(0.0429)
Latitude		-0.152***	-0.236***	-0.255***
		(0.0494)	(0.0566)	(0.0553)
Slope		-1.010***	-1.032***	-0.494
		(0.367)	(0.343)	(0.348)
Elevation (logged)		-0.127*	-0.120**	-0.116*
		(0.0700)	(0.0602)	(0.0637)
Prefectural capital			0.538***	0.465***
			(0.0889)	(0.0916)
Distance to the coast			-0.0162	0.0891
			(0.0715)	(0.0821)
Distance to the Grand Canal			-0.333***	-0.325***
			(0.0489)	(0.0708)
Distance to Beijing			-0.228**	-0.124*
			(0.103)	(0.0750)
Agricultural yields (logged)			0.119	0.137
			(0.102)	(0.103)
Human capital (logged)				0.383
				(0.408)
Urbanization ratio				0.796**
				(0.331)
Manufacturing ratio				-0.00294
				(0.00377)
Population density				418.8***
				(119.0)
Constant	3***	10.20***	20.83***	17.75***
	(0)	(3.304)	(4.519)	(4.759)
Geographic controls	N	Y	Y	Y
Historical socio-economic controls	N	N	Y	Y
Current socio-economic controls	N	N	N	Y
Prefecture Fixed Effects	Y	Y	Y	Y
Observations	1,665	1,664	1,638	1,456
R-squared	0.374	0.442	0.480	0.520

Robust standard errors adjusted for clustering at the prefecture level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.4 Instrumented results

We now report the 2SLS estimates in Table 5. The F-statistics suggest that our instrument is strong. The F-statistic remains over 30, even after including a rich gamut of covariates (Column (4)). More importantly, the IV estimates are also highly significant. Specifically, column (1) shows that historical governance quality itself has a significant impact on local economic outcomes. Adding the geographic controls in column (2) further increases the magnitude of the coefficient. In columns (3) and (4), the causal relationship between historical governance quality and levels of development in 2010 remains, even after introducing both historical and current socio-economic controls. Interestingly, the dimension of the instrumented coefficient (0.404, column (4) of Table 5) is much larger than the equivalent OLS estimate (0.110, column (4) of Table 2), which suggests that the OLS analysis may understate the influence of historical governance quality on present-day development. Overall, the 2SLS results demonstrate that differences in historical governance quality across counties in China influence, to a considerable extent, current differences in wealth.

Table 5: Historical governance quality on county night light luminosity in 2010: 2SLS estimate, overall

	(1)	(2)	(3)	(4)
VARIABLES	Night light luminosity in 2010			
Historical government quality	0.644***	0.627***	0.725***	0.404***
	(0.0661)	(0.0852)	(0.0921)	(0.103)
Land area (logged)		-0.525***	-0.596***	-0.398***
		(0.0758)	(0.0685)	(0.0713)
Longitude		0.0858*	0.102**	0.0203
		(0.0519)	(0.0486)	(0.0377)
Latitude		0.149***	0.199***	0.149***
		(0.0576)	(0.0633)	(0.0560)
Slope		-1.665***	-1.772***	-1.597***
		(0.335)	(0.342)	(0.281)
Elevation (logged)		-0.0453	0.0343	-0.0423
		(0.0463)	(0.0515)	(0.0445)
Prefectural capital			-0.195*	-0.112
			(0.0997)	(0.0796)
Distance to the coast			-0.202***	-0.127***
			(0.0660)	(0.0449)
Distance to the Grand Canal			0.130**	0.106*

			(0.0613)	(0.0581)
Distance to Beijing			-0.0303	0.0497
			(0.0468)	(0.0452)
Agricultural yields (logged)			0.0362	0.149
			(0.160)	(0.103)
Human capital (logged)				1.880***
				(0.279)
Urbanization ratio				0.942***
				(0.230)
Manufacturing ratio				0.0164***
				(0.00290)
Population density				13.92
				(89.66)
Constant	-0.868***	-13.56***	-16.52***	-11.92***
	(0.198)	(4.897)	(4.972)	(3.940)
Geographic controls	N	Y	Y	Y
Historical socio-economic controls	N	N	Y	Y
Current socio-economic controls	N	N	N	Y
Prefecture Fixed Effects	Y	Y	Y	Y
Observations	1,664	1,664	1,638	1,456
R-squared	0.665	0.759	0.728	0.831
First stage F	44.363	45.095	45.435	31.117

Robust standard errors adjusted for clustering at the prefecture level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.5 Robustness checks

To test the validity and strength of the above results we conduct several robustness checks. First, to further reduce the concern that our results may be biased by the political, economic, and social conditions, we collect additional information regarding these key criteria for location selection and additional variables. These include population data for 1600 (the only population data available at the county level before 1730), data regarding the number of natural disasters, the security situation, and local taxation before the government of the *Chongfanpinan* system. Table 6 shows that the results fundamentally remain unchanged even after the introduction of additional controls.

We then assess whether the inclusion of some minority regions in the analysis affects the coefficients. In fact, most of China's ethnic minorities lived near military fortresses. Based on our above-mentioned distribution principle of government quality, ethnic minority areas are more likely to have been allocated more and better government resources. Therefore, there might be a concern that our results will be affected by the ethnic minority areas. However, excluding ethnic minority areas, for most of Han China, the influence of the political system on the economy may be ineffective. Moreover, because of

their different culture and norms, especially in historical times, the inclusion of minority regions may affect both the measurement of historical institutional quality under the *Chongfanpinan* reform and its link to current levels of GDP per capita (Bai & Kung, 2011). Hence, in Table 7, we re-run the analysis including only those counties in Han China and without considering minority regions, such as Xinjiang and Gansu. The results indicate that the connection between governance quality during the Qing dynasty and night light luminosity at the county level in 2010 is robust to the exclusion of minority regions.

Table 6: OLS and 2SLS analysis, controlling more historical variables.

	(1)	(2)	(3)	(4)
VARIABLES	Night light luminosity in 2010			
	OLS	OLS	OLS	2SLS
Historical government quality	0.194***	0.191***	0.0932***	0.264***
	(0.0237)	(0.0248)	(0.0260)	(0.0965)
Crime (logged)	0.0812	0.0568	-0.0565	-0.0424
	(0.0972)	(0.0993)	(0.0950)	(0.0886)
Disaster (logged)	-0.00533	0.0259	-0.0276	-0.0424
	(0.0637)	(0.0621)	(0.0517)	(0.0480)
Tax (logged)	0.0655	0.0731	0.0535	0.0434
	(0.0544)	(0.0555)	(0.0501)	(0.0436)
Historical population (logged)	0.145***	0.136***	0.0739*	0.0399
	(0.0484)	(0.0477)	(0.0379)	(0.0379)
Land area (logged)	-0.396***	-0.399***	-0.254***	-0.341***
	(0.0507)	(0.0523)	(0.0513)	(0.0672)
Longitude	0.0517	0.0182	-0.0266	-0.0124
	(0.0486)	(0.0479)	(0.0436)	(0.0392)
Latitude	0.0728	0.106	0.0909	0.128**
	(0.0715)	(0.0746)	(0.0680)	(0.0615)
Slope	-1.780***	-1.824***	-1.372***	-1.313***
	(0.353)	(0.357)	(0.316)	(0.292)
Elevation (logged)	-0.0531	-0.0109	-0.0717	-0.0555
	(0.0526)	(0.0522)	(0.0515)	(0.0445)
Prefectural capital		-0.0150	-0.0115	-0.117
		(0.0761)	(0.0706)	(0.0902)
Distance to the coast		-0.219***	-0.0984**	-0.106**
		(0.0638)	(0.0464)	(0.0449)
Distance to the Grand Canal		-0.00781	0.0147	0.0633
		(0.0429)	(0.0403)	(0.0519)
Distance to Beijing		-0.122**	0.0530	0.0741
		(0.0538)	(0.0528)	(0.0490)
Agricultural yields (logged)		-0.0329	0.0654	0.0712
		(0.139)	(0.142)	(0.133)
Human capital (logged)			1.828***	1.705***
			(0.317)	(0.277)
Urbanization ratio			1.282***	1.134***
			(0.231)	(0.205)
Manufacturing ratio			0.0153***	0.0160***

			(0.00287)	(0.00267)
Population density			137.5	73.74
			(113.5)	(99.11)
Constant	-7.919*	-0.677	-3.496	-6.867
	(4.265)	(4.415)	(4.050)	(4.222)
Additional controls	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y
Historical socio-economic controls	N	Y	Y	Y
Current socio-economic controls	N	N	Y	Y
Prefecture Fixed Effects	Y	Y	Y	Y
Observations	1,396	1,395	1,234	1,234
R-squared	0.811	0.815	0.862	0.853
First stage F				34.39

Robust standard errors adjusted for clustering at the prefecture level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: OLS and 2SLS analysis excluding minority regions.

	(1)	(2)	(3)	(4)
VARIABLES	Night light luminosity in 2010			
	OLS	OLS	OLS	2SLS
Historical government quality	0.234***	0.222***	0.127***	0.378***
	(0.0217)	(0.0224)	(0.0228)	(0.0907)
Land area (logged)	-0.304***	-0.303***	-0.288***	-0.428***
	(0.0487)	(0.0496)	(0.0442)	(0.0639)
Longitude	0.140***	0.114**	-0.00712	0.0189
	(0.0504)	(0.0549)	(0.0379)	(0.0360)
Latitude	0.0432	0.0526	0.106*	0.164***
	(0.0591)	(0.0617)	(0.0568)	(0.0540)
Slope	-2.371***	-2.412***	-1.515***	-1.418***
	(0.317)	(0.317)	(0.250)	(0.244)
Elevation (logged)	-0.0525	-0.0184	-0.0828*	-0.0297
	(0.0546)	(0.0494)	(0.0438)	(0.0427)
Prefectural capital		0.0632	-0.0345	-0.183**
		(0.0811)	(0.0681)	(0.0863)
Distance to the coast		-0.189***	-0.111**	-0.125**
		(0.0608)	(0.0475)	(0.0520)
Distance to the Grand Canal		-0.0476	0.00664	0.0953
		(0.0458)	(0.0506)	(0.0624)
Distance to Beijing		-0.169***	-0.00397	0.0563
		(0.0437)	(0.0387)	(0.0460)
Agricultural yields (logged)			0.121*	0.0838
			(0.0632)	(0.0549)
Human capital (logged)			1.923***	1.653***
			(0.433)	(0.409)
Urbanization ratio			-0.00637	0.0145
			(0.132)	(0.113)
Manufacturing ratio			0.302**	0.235**

			(0.129)	(0.110)
Population density			0.0137***	0.0132***
			(0.00281)	(0.00256)
Constant	-15.33***	-7.682	-7.639**	-13.56***
	(4.616)	(5.185)	(3.851)	(4.019)
Geographic controls	Y	Y	Y	Y
Historical socio-economic controls	N	Y	Y	Y
Current socio-economic controls	N	N	Y	Y
Prefecture Fixed Effects	Y	Y	Y	Y
Observations	1,572	1,572	1,393	1,393
R-squared	0.810	0.814	0.863	0.842
First stage F				42.449

Robust standard errors adjusted for clustering at the prefecture level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Furthermore, to show that our results are not affected by economic outcomes in a single year, in the Appendix Figure A3 and Figure A4, we repeat the analysis by replacing the luminosity in the other years from 1992 to 2012. Both OLS and 2SLS results are not affected by individual years. We also conduct the regression using province fixed effects, rather than prefectural-level ones. The results remain similar.

5. Understanding the mechanism

Having established a causal relationship between historical local governance institutions and contemporary local development, we now endeavour to dismantle the mechanism. We lay out two hypotheses. The first concerns '*the role of political decentralisation*': current variations in development across China may be driven by the long-term decentralisation triggered by historical institution. The second hypothesis is '*good governance led by qualified officials*': the *Chongfanpinan* institution positively affected economic outcomes at county level via the higher performance of the officials allocated to counties with a higher score. In the following sections, we present evidence for or against each hypothesis.

5.1 Decentralisation

The effect of decentralisation has long been controversial. Although increasing evidence shows that decentralisation could boost economic prosperity, both at subnational and national level, several works suggest, on the contrary, that it may also undermine overall economic performance (Bo, 2020; Dell, 2010; Rodríguez-Pose & Ezcurra, 2011). Studying historical institutions provides us with an ideal setting to further test the effects of decentralisation.

Based on the principle of the local governance institution in the Qing dynasty, counties that were above tier 2 had the autonomy to appoint local officials. The appointment of local officials in these counties was partially a prerogative of provincial governors.[8] Conversely, local officials in counties below tier 2 in the political hierarchy were fully chosen by the Ministry of personnel of central administration. To examine the first hypothesis, the key independent variable, decentralisation, is a dummy equal to 1, where provincial governors have the right to appoint and remove officials in the local, and 0 otherwise. The other key independent variable, centralisation, is a dummy equal to 1, where local officials are appointed by the central government directly, and 0 otherwise. As before, our dependent variable is luminosity in 2010.

Table 8: Decentralization or centralization

	(1)	(2)	(3)	(4)
VARIABLES	Night light luminosity in 2010			
	OLS	2SLS	OLS	2SLS
Decentralization	0.163***	0.9529***		
	(0.0590)	(0.2051)		
Centralization			0.173	-1.1642***
			(0.113)	(0.3173)
Land area (logged)	-0.212***	-0.3121***	-0.176**	-0.2997***
	(0.0717)	(0.0807)	(0.0702)	(0.1092)
Longitude	-0.0497***	-0.0450***	-0.0510***	-0.0492***
	(0.00942)	(0.0103)	(0.00939)	(0.0101)
Latitude	0.0395***	0.0396***	0.0405***	0.0327***
	(0.00920)	(0.0105)	(0.00881)	(0.0119)
Slope	-0.909***	-0.8416***	-0.954***	-0.7124*
	(0.308)	(0.3094)	(0.292)	(0.4050)
Elevation (logged)	-0.0709*	-0.0412	-0.0807**	-0.0524
	(0.0383)	(0.0382)	(0.0390)	(0.0468)
Prefectural capital	0.0493	-0.1728*	0.129	-0.1304
	(0.0713)	(0.0962)	(0.0802)	(0.1068)
Distance to the coast	-0.228***	-0.2284***	-0.230***	-0.2158***
	(0.0396)	(0.0444)	(0.0370)	(0.0550)
Distance to the Grand Canal	-0.151***	-0.1383***	-0.148***	-0.1908***
	(0.0316)	(0.0352)	(0.0283)	(0.0460)
Distance to Beijing	-0.0672	-0.0547	-0.0648	-0.1036
	(0.0490)	(0.0526)	(0.0453)	(0.0686)
Agricultural yields (logged)	0.348***	0.3328***	0.341***	0.4168***
	(0.0837)	(0.0809)	(0.0818)	(0.1091)
Human capital (logged)	1.609***	1.3304***	1.679***	1.5860***
	(0.321)	(0.3504)	(0.324)	(0.3877)
Urbanization ratio	1.490***	1.4919***	1.552***	1.0714**

[8] Provincial governors still needed to have a discussion with central administration.

	(0.400)	(0.3897)	(0.408)	(0.5422)
Manufacturing ratio	0.00920**	0.0079*	0.00903**	0.0124***
	(0.00377)	(0.0041)	(0.00365)	(0.0047)
Population density	578.1***	410.3851***	641.5***	421.0634***
	(140.8)	(130.6963)	(151.4)	(139.9461)
Constant	4.425***	3.8472**	4.379***	5.6608***
	(1.569)	(1.6351)	(1.581)	(1.8198)
Geographic controls	Y	Y	Y	Y
Historical socio-economic controls	Y	Y	Y	Y
Current socio-economic controls	Y	Y	Y	Y
Prefecture Fixed Effects	Y	Y	Y	Y
Observations	1,580	1,580	1,580	1,580
R-squared	0.640	0.5873	0.641	0.4405
First stage F		55.1		40.4

Robust standard errors adjusted for clustering at the prefecture level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results are reported in Table 8. Column (1) includes the OLS results with all sets of controls, which shows that decentralisation is both positively and significantly associated with economic development today. The OLS result is further confirmed by the 2SLS estimate shown in column (2). In fact, the magnitude of the coefficient corrected by the instrument is over five times larger than the original estimate. Column (3) and (4) illustrate the effect of centralisation on current development. Interestingly, the instrumented result shows that centralisation does significantly lead to lower levels of development. In sum, we find that, compared with centralised counties, the autonomous counties in historical times currently enjoy higher levels of development.

5.2 Good governance

Good governance is regarded as one of the key indicators of institutions (Besley & Persson, 2011; Rodríguez-Pose & Storper, 2006). However, although research has found that good governance is a fundamental driver of development in developed countries, there is less evidence confirming this is also the case for authoritarian regimes in historical times. Here we measure good governance through the qualifications of local officials. The hypothesis is that counties continuously governed by more highly qualified officials create a good governance ecosystem. Consequently, these counties with better historical governance would have ended up more developed today.

To quantify the official's qualifications and the extent of good governance, we draw on the China Government Employee Database (*Jinshenlu*) (CGED), provided by Campbell et al. (2019). CGED introduces the comprehensive information about 638,153 local

officials in the Qing dynasty.[9] We use the local official's scores in the civil service examination (*Keju*) as the key proxy for their qualifications. The result of the civil service examination is a good indicator because the main purpose and content of this examination were to train and select suitable imperial officials (Bai & Jia, 2016). According to this line of reasoning, we can assume that more highly qualified officials would be those with higher scores. We follow the classification of the civil service examination: (1) officials with the lowest scores are classified as 1; (2) those that pass the licensing exam at the prefectural level (*Yuankao*) as 2; (3) those that pass the qualifying exam at the provincial level (*Xiangshi*) as 3; and (4) those that pass the academy exam at the national level (*Huikao*) as 4. The specific structure of the civil service exam is shown in Appendix A. To match other variables, we then aggregate the results of the civil service examination for the officials working in any given county at county level.

We first report the interaction between local governance and the qualification of local officials in Column (1) and (2) of Table 9. Unsurprisingly, our result confirms that more qualified officials were highly likely to be appointed to counties deemed to be of greater strategic importance. This result is highly statistically significant. Columns (3) and (4) report the effect of good governance. The interaction of the scores of local civil servants with historical government quality is positive and significant (column (3)), suggesting that in places with high government quality, good governance is at the root of a higher level of GDP per capita. The 2SLS estimates of column (4) also support this view. Overall, the results confirm the intuition that historically persistent good governance in historical times is an important driver of current differences in development across China.

Table 9: The effect of good governance

	(1)	(2)	(3)	(4)
			Night light luminosity	
VARIABLES	Official		in 2010	
	OLS	2SLS	OLS	2SLS
Historical government quality	0.0924***	0.3820***	-0.0694	-0.0683
	(0.0210)	(0.1219)	(0.0829)	(0.2757)
Good governance			-0.156**	-0.3311**
			(0.0736)	(0.1409)
Historical government quality*good governance			0.0638**	0.1373**
			(0.0310)	(0.0687)
Land area (logged)	-0.0439	-0.1984**	-0.314***	-0.4330***
	(0.0319)	(0.0776)	(0.0476)	(0.0808)
Longitude	0.00418	0.0024	-0.0507***	-0.0539***

[9] The current time span of CGED is from 1900 to 1912.

	(0.00694)	(0.0074)	(0.0110)	(0.0106)
Latitude	-0.0136**	-0.0155**	0.0384***	0.0378***
	(0.00659)	(0.0074)	(0.0102)	(0.0109)
Slope	0.0341	0.1528	-1.552***	-1.3461***
	(0.175)	(0.2114)	(0.253)	(0.2532)
Elevation (logged)	-0.0215	-0.0110	0.0150	0.0252
	(0.0235)	(0.0285)	(0.0291)	(0.0284)
Prefectural capital	-	-	-	-
Distance to the coast	0.0251	0.0225	-0.287***	-0.3052***
	(0.0243)	(0.0266)	(0.0355)	(0.0417)
Distance to the Grand Canal	-0.0200	-0.0148	-0.148***	-0.1361***
	(0.0239)	(0.0264)	(0.0276)	(0.0296)
Distance to Beijing	0.0203	0.0109	-0.0631	-0.0735
	(0.0322)	(0.0412)	(0.0428)	(0.0479)
Agricultural yields (logged)	-0.0231	-0.0078	0.184***	0.2229***
	(0.0596)	(0.0620)	(0.0699)	(0.0841)
Human capital (logged)	0.414	0.3420	1.444***	1.4069***
	(0.269)	(0.2974)	(0.321)	(0.3226)
Urbanization ratio	0.0242	-0.1546	1.741***	1.4308***
	(0.214)	(0.2570)	(0.267)	(0.2894)
Manufacturing ratio	0.00183	0.0021	0.00574	0.0056
	(0.00246)	(0.0025)	(0.00380)	(0.0039)
Population density	-51.58	-202.6159**	412.6***	270.5656**
	(44.69)	(84.4557)	(129.3)	(115.2557)
Constant	1.392	0.9892	6.771***	7.1966***
	(1.262)	(1.4649)	(1.673)	(1.7511)
Geographic controls	Y	Y	Y	Y
Historical socio-economic controls	Y	Y	Y	Y
Current socio-economic controls	Y	Y	Y	Y
Prefecture Fixed Effects	Y	Y	Y	Y
Observations	1,154	1,154	1,154	1,152
R-squared	0.045	-0.1003	0.722	0.6864
First stage F		44.9		9.05

Robust standard errors adjusted for clustering at the prefecture level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6. The effect of Western shocks

How does the influence of local governance quality during the Qing era on present-day territorial disparities in China compare with that of the institutional changes introduced during the 19th century as a consequence of the expansion of the influence of Western powers in the country? The waxing Western influence after the First Opium Wars in the 1840s has traditionally been considered by historians as a fundamental factor behind the pervasiveness of regional inequality in China (Cameron, 1933; Jia, 2014; Long et al., 2019). However, the evidence about the extent to which the shocks linked to the

intervention by Western powers affected local economic growth in China remains scarce. To examine whether the shocks of Western intervention had a similar or greater effect than that of pre-existing Qing era differences in local governance quality on the distribution of wealth in China, we select two representative elements of Western intervention in China: protestant missionary activities and treaty ports. This type of interventions grew during the 19th century and are considered to have deeply shaped modern China (Jia, 2014; Long et al., 2019). We use protestant missionary activities as a proxy for Western cultural intervention in Chinese society, while the establishment of treaty ports represents the more political and economic dimension of the intervention.

Figure 5: Protestant missionary presence

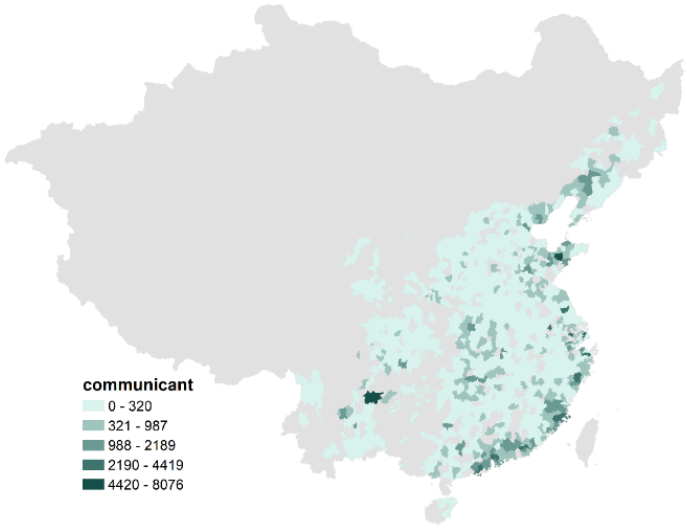
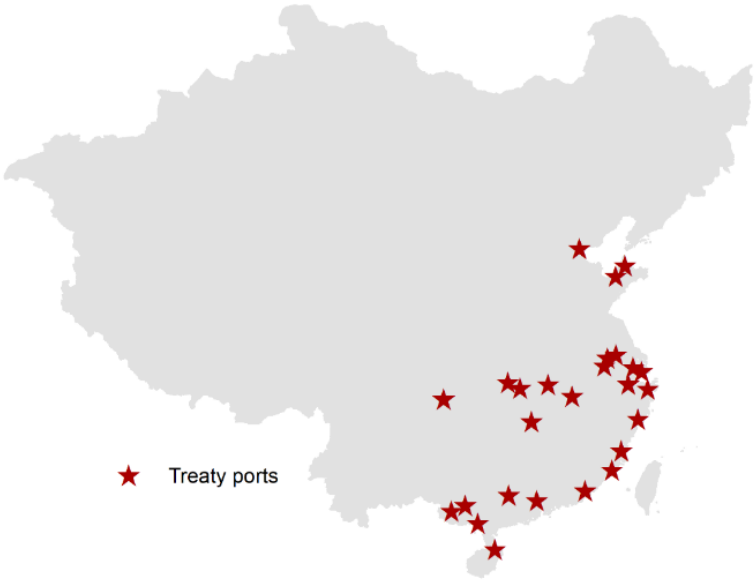


Figure 6: Treaty ports



Protestant missions in China started in the early 19th century, but their presence grew exponentially towards the end of the century. According to Stauffer (1922), by the beginning of the 20th century protestant missionaries were present in over 94% of counties in China proper and Manchuria (Figure 5). Research has shown that, in China, Protestantism had a significant causal effect on economic prosperity as measured by urbanization in 1920 and industrialisation between 1841 and 1915 (Bai & Kung, 2015). Treaty ports were port cities established after the First Opium War and the ensuing Treaty of Nanking of 1842 granting Britain, first, and —following a number of other 'unequal' treaties— other Western powers, free access to trade with China through these cities. By the end of the 19th century close to 80 free ports had been set up in China, most on the Eastern seaboard and along the Yang-Tze and Pearl rivers (Figure 6). The Western countries that managed the treaty ports established municipal authorities, police, and separate judicial systems.

To what extent has the sprawling Western influence over China in the 19th century left a legacy on economic development differences when compared to the earlier local differences in government quality? Did the shocks associated with Western intervention erase the influence of pre-established institutions? We test whether that is the case by regressing the effect of these shocks alongside differences in Qing era local governance quality following the same logic as in model (1). We conduct both OLS and 2SLS regressions.

The results of Table 10 clearly show that differences in historical local governance quality in China survived the disruption of the 19th and early 20th century Western shocks. Differences in governance quality associated with the *Chongfanpinan* reform have left more of a trace on current levels of development in China than the cultural and institutional influence of Protestant missionaries. The coefficient for historical governance quality is significant in columns (1) and (2), while that denoting the presence of Protestant missionaries in a county is insignificant. The coefficient of the interaction term between both variables is also insignificant. This implies that the activities of protestant missionaries did not weaken the long-term impact of deeply seated governance differences across China.

Table 10: The effect of Western shocks on contemporary economic development

	(1)	(2)	(3)	(4)
VARIABLES	Night light luminosity in 2010			
	OLS	2SLS	OLS	2SLS
Historical gov't quality	0.121***	0.3236***	0.123***	0.4689***
	(0.0275)	(0.0999)	(0.0244)	(0.1180)

Missionary presence	9.43e-05	0.0001		
	(7.88e-05)	(0.0001)		
Historical gov't quality*Missionary presence	-4.73e-06	-0.0000		
	(2.79e-05)	(0.0001)		
Treaty ports			-0.311	0.1939
			(0.300)	(0.3448)
Historical gov't quality* Treaty ports			-0.0949*	-0.3095***
			(0.0557)	(0.1043)
Land area (logged)	-0.247***	-0.3826***	-0.224***	-0.4134***
	(0.0523)	(0.0822)	(0.0438)	(0.0748)
Longitude	0.0550	0.0449	0.00531	0.0130
	(0.0389)	(0.0375)	(0.0396)	(0.0372)
Latitude	-0.00382	0.0628	0.0720	0.1483***
	(0.0489)	(0.0527)	(0.0538)	(0.0550)
Slope	-1.456***	-1.2193***	-1.747***	-1.5361***
	(0.361)	(0.3382)	(0.286)	(0.2830)
Elevation (logged)	-0.0872	-0.0734	-0.0790*	-0.0491
	(0.0533)	(0.0458)	(0.0475)	(0.0444)
Prefectural capital	0.0909	-0.0178	0.0193	-0.1623*
	(0.101)	(0.1077)	(0.0637)	(0.0886)
Distance to the coast	-0.0733**	-0.0975**	-0.0985**	-0.1440***
	(0.0359)	(0.0385)	(0.0413)	(0.0462)
Distance to the Grand Canal	0.0125	0.0754	0.00930	0.1225**
	(0.0399)	(0.0525)	(0.0390)	(0.0613)
Distance to Beijing	-0.00932	0.0181	0.0249	0.0600
	(0.0420)	(0.0419)	(0.0460)	(0.0462)
Agricultural yields (logged)	0.257	0.3620*	0.187	0.1513
	(0.244)	(0.1910)	(0.119)	(0.1062)
Human capital (logged)	2.038***	1.8113***	2.054***	1.8458***
	(0.306)	(0.2735)	(0.305)	(0.2836)
Urbanization ratio	0.927***	0.7719***	1.228***	0.9897***
	(0.239)	(0.2145)	(0.239)	(0.2352)
Manufacturing ratio	0.0155***	0.0158***	0.0156***	0.0166***
	(0.00339)	(0.0031)	(0.00298)	(0.0029)
Population density	98.30	37.3901	159.9	37.5564
	(92.09)	(75.7029)	(114.7)	(92.6930)
Constant	-9.344***	-12.4332***	-6.661*	-11.4157***
	(3.270)	(3.6019)	(3.627)	(3.8430)
Geographic controls	Y	Y	Y	Y
Historical socio-economic controls	Y	Y	Y	Y
Current socio-economic controls	Y	Y	Y	Y
Prefecture Fixed Effects	Y	Y	Y	Y
Observations	986	986	1,456	1,456
R-squared	0.884	0.8704	0.858	0.8245
First stage F		8.17		13.5

Robust standard errors adjusted for clustering at the prefecture level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Treaty ports left more of a trace. As seen in columns (3) and (4) of Table 10, the coefficients for both historical governance quality and the establishment of a treaty port are both positive and significant. Better governed counties in historical times and

counties that contained a treaty port are —once other factors are also taken into consideration— significantly richer today than areas of the country with weaker governance records and that did not contain a port. The negative and significant interaction term between both variables signals that the presence of a treaty port meaningfully weakened the influence of historical Chinese governance institutions over subsequent economic development.

7 Conclusion

In this paper, we have analysed the extent to which historical differences in government quality developed close to four centuries ago have survived the political, economic, and cultural shocks that followed the demise of two millennia of imperial China and still affect differences in county development across China today. Few studies to date have examined the inertia of imperial institutions in a country that, by all accounts, has had what can be considered a turbulent last century.

Our results show that variations in historical local government quality developed during the Qing dynasty still determine, to a considerable extent, differences in economic performance today. They demonstrate that the shadow of history in China is very long and that not even the traumatic events linked to the collapse of Imperial China and the wars and changes in political and economic regimes that ensued have managed to do away with institutions that were developed and became deeply rooted in China over centuries. Our results emphasise the consequential impact of governance reforms of local institutions on economic development, even centuries after the reforms took place and the practices behind them are abandoned. While this point has been overlooked in much of recent development economics, it shows the remaining influence of historical institutional reforms in current levels of development (Duranton et al., 2009). Differences in the quality of local government created by historical reforms still matter today and continue to play a significant, if uneven, role in China's development process (Ang, 2016).

References

Acemoglu, D., Johnson, S., & Robinson, J. (2005). The rise of Europe: Atlantic trade, institutional change, and economic growth. *American Economic Review, 95*(3), 546-579.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review, 91*(5), 1369-1401.

Acemoglu, D., García-Jimeno, C., & Robinson, J. A. (2015). State capacity and economic development: A network approach. *American Economic Review, 105*(8), 2364-2409.

Ades, A. F., & Glaeser, E. L. (1995). *Trade and circuses: explaining urban giants. The Quarterly Journal of Economics, 110*(1), 195-227.

Ang, Y. Y. (2016). *How China escaped the poverty trap*. Cornell University Press.

Bai, Y., & Jia, R. (2016). Elite recruitment and political stability: the impact of the abolition of China's civil service exam. *Econometrica, 84*(2), 677-733.

Bai, Y., & Kung, J. K. S. (2011). Climate shocks and Sino-nomadic conflict. *Review of Economics and Statistics, 93*(3), 970-981.

Bai, Y., & Kung, J. K. S. (2015). Diffusing knowledge while spreading God's message: Protestantism and economic prosperity in China, 1840–1920. *Journal of the European Economic Association, 13*(4), 669-698.

Balan, P., Bergeron, A., Tourek, G., & Weigel, J. (2020). Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the DR Congo. CEPR Discussion Papers 15138, C.E.P.R. Discussion Papers.

Besley, T., & Persson, T. (2011). *Pillars of prosperity: The political economics of development clusters*. Princeton University Press.

Bo, S. (2020). Centralization and regional development: Evidence from a political hierarchy reform to create cities in China. *Journal of Urban Economics, 115*, 103182.

Cameron, M.E. (1931). *The reform movement in China, 1898-1912*. Stanford University Press.

Campante, F. R., & Do, Q. A. (2014). Isolated capital cities, accountability, and corruption: Evidence from US states. *American Economic Review, 104*(8), 2456-81.

Campbell, C., Chen, B., Ren, Y. & Lee, J. (2019). China Government Employee Database-Qing (CGED-Q) Jinshenlu 1900-1912 Public Release, https://doi.org/10.14711/dataset/E9GKRS. DataSpace@HKUST, V1

Chen, T., Kung, J. K. S., & Ma, C. (2017). Long live Keju! The persistent effects of China's imperial examination system. *The Persistent Effects of China's Imperial Examination System (June 26, 2017).*

Chen, W., Chen, X., Hsieh, C. T., & Song, Z. (2019). A Forensic Examination of China's National Accounts (No. w25754). National Bureau of Economic Research.

Davis, J. C., & Henderson, J. V. (2003). Evidence on the political economy of the urbanization process. *Journal of urban economics, 53*(1), 98-125.

Dell, M. (2010). The persistent effects of Peru's mining mita. *Econometrica, 78*(6), 1863-1903.

Dell, M., Lane, N., & Querubin, P. (2018). The historical state, local collective action, and economic development in Vietnam. *Econometrica, 86*(6), 2083-2121.

Diamond, J. M. (1998). *Guns, germs and steel: a short history of everybody for the last 13,000 years*. Random House.

Duranton, G., Rodríguez-Pose, A., & Sandall, R. (2009). Family types and the persistence of regional disparities in Europe. *Economic Geography, 85*(1), 23-47.

Frick, S.A., Rodríguez-Pose, A., & Wong, M.D. (2019). Toward economically dynamic special economic zones in emerging countries. *Economic Geography, 95*(1), 30-64.

Fukuyama, F. (2006). *The end of history and the last man*. Simon and Schuster.

Fukuyama, F. (2013). What is governance?. *Governance, 26*(3), 347-368.

Guy, R. K. (2017). *Qing governors and their provinces: the evolution of territorial administration in China, 1644-1796*. University of Washington Press.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American economic review*, 102(2), 994-1028.

Hucker, C. O. (1985). *A Dictionary of Official Titles in Imperial China*. Stanford University Press.

Iddawela, Y., Rodríguez-Pose, A., & Lee, N. (2021). Quality of sub-national government and regional development in Africa. *Journal of Development Studies*, https://doi.org/10.1080/00220388.2021.1873286

Jia, R. (2014). The legacies of forced freedom: China's treaty ports. *Review of Economics and Statistics, 96*(4), 596-608.

Jones, S. and Tarp, F. (2016). Does foreign aid harm political institutions?. *Journal of Development Economics, 118*, 266-281.

Ketterer, T. D., & Rodríguez-Pose, A. (2018). Institutions vs. 'first-nature' geography: What drives economic growth in Europe's regions?. *Papers in Regional Science, 97*, S25-S62.

Koss, D. (2017). Political Geography of Empire: Chinese Varieties of Local Government. *The Journal of Asian Studies, 76*(1), 159-184.

Kraay, A., Kaufmann, D., & Mastruzzi, M. (2010). *The worldwide governance indicators: methodology and analytical issues*. The World Bank.

Kuhn, P. A. (2002). *Origins of the modern Chinese state*. Stanford University Press.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (1999). The quality of government. *The Journal of Law, Economics, and Organization, 15*(1), 222-279.

Liu, Z. (1993). Chong, Fan, Pi and Nan: An Exploration of the Ranking of Qing Administrative Units. Bulletin of Institute of History and Philology, *Academia Sinica, 64*(1), 175-204.

Liu, Z. (2012). *Archival documents form the ming-qing era*. Humanities Center of Chengchi University

Long, C., Murrell, P., & Yang, L. (2019). Memories of colonial law: The inheritance of human capital and the location of joint ventures in early-reform China. *China Economic Review, 58*, 101284.

Michalopoulos, S., & Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary African development. *Econometrica, 81*(1), 113-152.

Michalopoulos, S., & Papaioannou, E. (2014). National institutions and subnational development in Africa. *The Quarterly journal of economics, 129*(1), 151-213.

Michalski, T., & Stoltz, G. (2013). Do countries falsify economic data strategically? Some evidence that they might. *Review of Economics and Statistics, 95*(2), 591-616.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge university press.

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

Pomeranz, K. (2009). *The great divergence: China, Europe, and the making of the modern world economy (Vol. 28).* Princeton University Press.

Rodríguez-Pose, A. (2013). Do institutions matter for regional development? *Regional studies, 47*(7), 1034-1047.

Rodríguez-Pose, A., & Ezcurra, R. (2011). Is fiscal decentralization harmful for economic growth? Evidence from the OECD countries. *Journal of Economic Geography, 11*(4), 619-643.

Rodríguez-Pose, A., & Storper, M. (2006). Better rules or stronger communities? On the social foundations of institutional change and its economic effects. *Economic Geography, 82*(1), 1-25.

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of Economic Growth, 9*(2), 131-165.

Rothstein, B. O., & Teorell, J. A. (2008). What is quality of government? A theory of impartial government institutions. *Governance, 21*(2), 165-190.

Rothstein, B. (2011). *The quality of government: Corruption, social trust, and inequality in international perspective.* Chicago: University of Chicago Press.

Xu, G. (2018). The costs of patronage: Evidence from the British empire. *American Economic Review, 108*(11), 3170-98.

Sachs, J. D. (2001). *Tropical underdevelopment* (No. w8119). National Bureau of Economic Research.

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed.* Yale University Press.

Spence, J. D. (1990). *The search for modern China.* WW Norton & Company.

Storeygard, A. (2016). Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. *The Review of economic studies, 83*(3), 1263-1295.

Thornton, P. M. (2007). *Disciplining the state: Virtue, violence and state-making in modern China.* Cambridge. Harvard University Press.

Zhou, Z. (2005). *The History of China's Local Administrative System.* Shanghai People's Press

Appendix A

Figure A1: The political hierarchy of the Qing dynasty

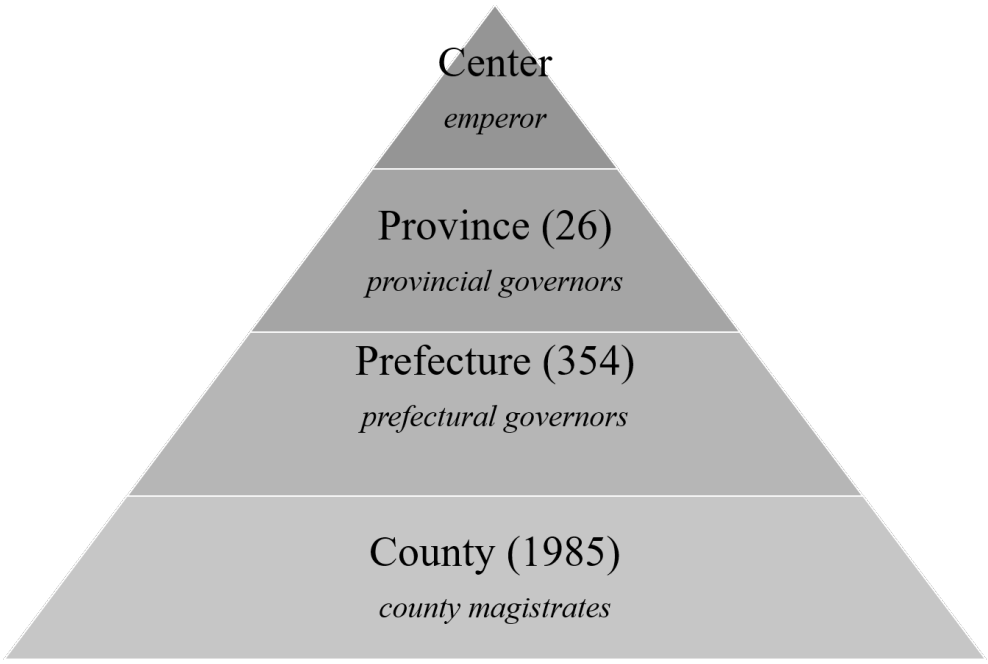


Figure A2: The example of digitalising the local governance institution

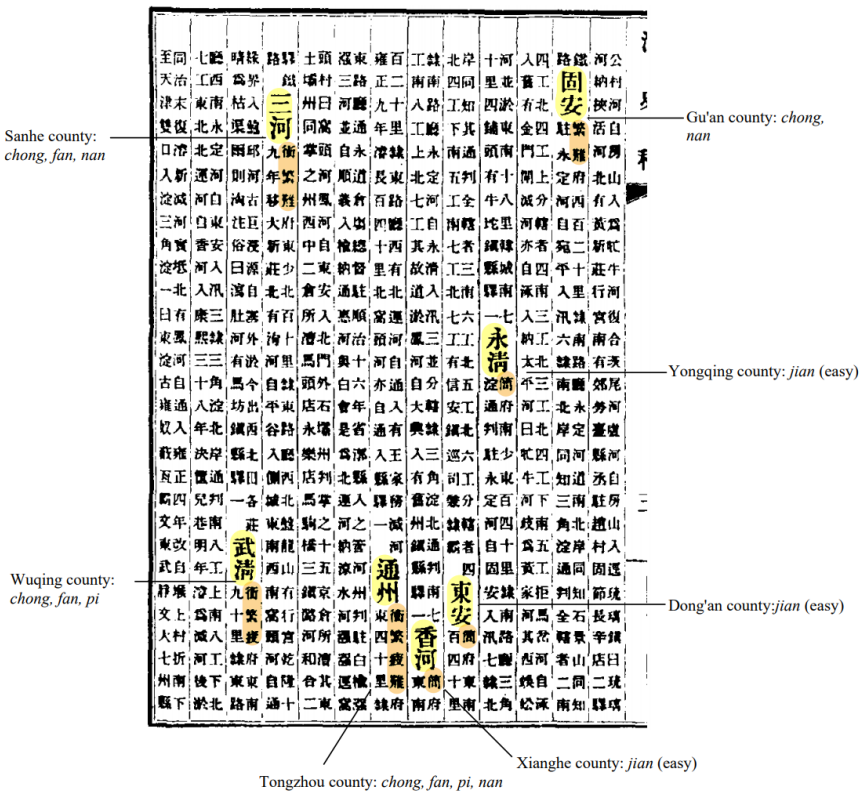


Figure A3: The effect of the per-modern local governance institution over time (OLS results)

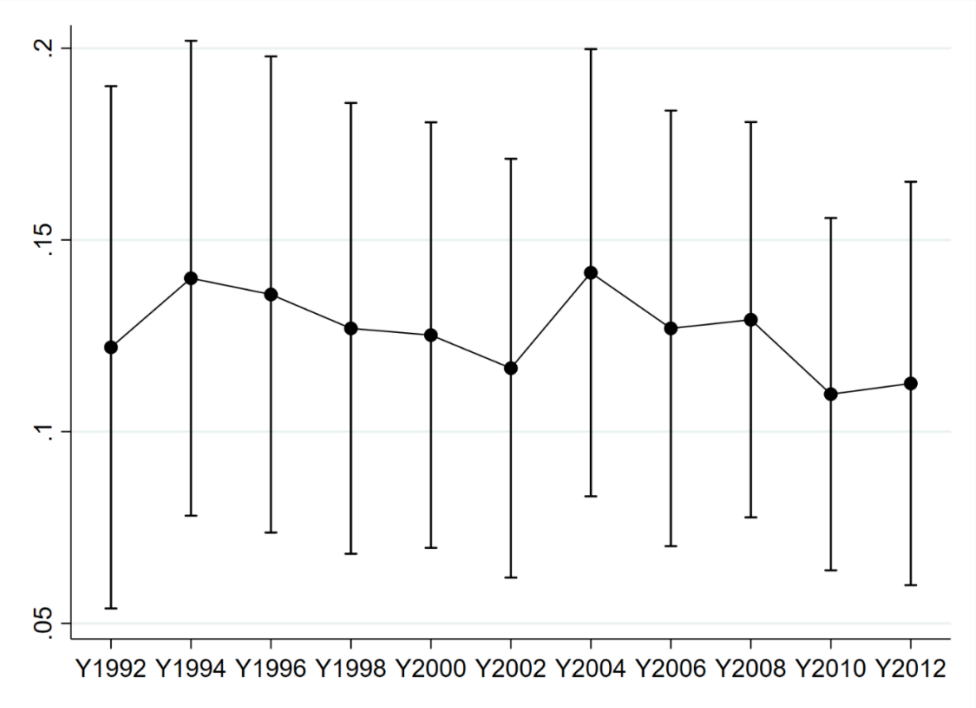


Figure A4: The effect of the per-modern local governance institution over time (2SLS results)

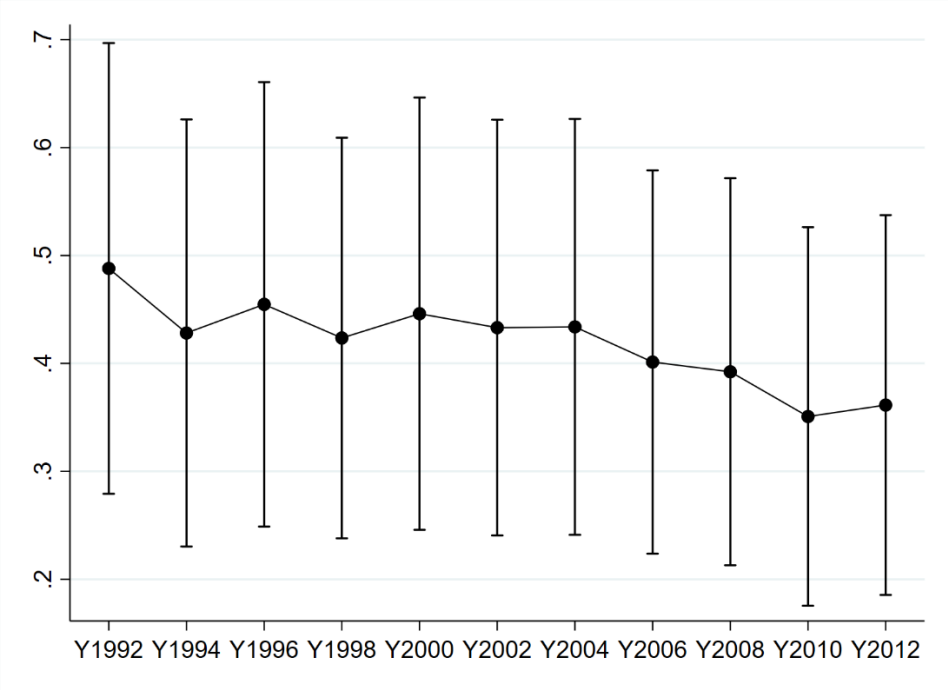


Table A1: Sources and definition

Variable	Definition	Sources	Obs	Mean	S.D
	(a) economic development				
nightlight	nightlight	1	1,969	7.463	10.087
lnnightlight	log nightlight	1	1,909	1.079	1.814
lnnightlight001	log (nightlight+0.01)	1	1,969	0.947	1.930
	(b) Chongfanpinan institution				
local government quality	Government quality	2	1,665	1.993	0.931
	(c) geographic control				
longitude	longitude	3	1,985	110.935	8.163
latitude	latitude	3	1,985	33.034	6.981
land area	log of area	3	1,985	-1.547	1.177
slope	slope	3	1,969	0.212	0.187
elevation	log of elevation	3	1,969	5.749	1.670
	(d) historical socio-economic control				
prefecture capital	prefectural capital	3	1,985	0.057	0.232
distance to coast	log of distance to coast	3	1,985	12.777	1.305
distance to canal	log of distance to canal	3	1,985	13.164	1.225
distance to capital	log of distance to national capital	3	1,985	13.712	0.945
agricultural yields	log of agricultural suitability	4	1,859	8.013	0.845
	(e) current socio-economic control				
human capital	log of schooling year	5	1,679	2.112	0.148
manufacturing ratio	employment ratio in manufacturing	5	1,840	17.772	12.627
urbanization ratio	urbanization ratio	5	1,840	0.378	0.143
population density	population density	5	1,840	0.000	0.001
	(f) instrument				
garrison	# of garrisons in Ming dynasty	6	1,985	0.182	0.629
	(g) mechanism				
official	local official's education	7	1,323	2.621	0.614
	(h) western shock				
missionary	# of communicants	8	1,137	266.601	576.897
treaty port	treaty port dummy	9	1,977	0.110	0.313

Data sources:

1: Defense meteorological satellite program's operational linescan system

2: Veritable Records of the Qing Emperors (Qingshilu)

3: Harvard Yenching Institution (2007), CHGIS, Version 4

4: Galor & Ozak (2016), The caloric suitability indices

5: Population census by county in 2010

6: Szonyi & Michael, Ming Garrisons (1363-1644)

7: Campbell et al (2019), China Government Employee Database in Qing dynasty (Jinshenlu)

8: Stauffer, Milton T. (1922). The Christian Occupation of China: A General Survey of the Numerical Strength and Geographical Distribution of the Christian Forces in China. China Continuation Committee, Shanghai

9: Jia, R. (2014). The legacies of forced freedom: China's treaty ports. Review of Economics and Statistics, 96(4), 596-608.

Table A2: OLS and 2SLS analysis with province fixed effect

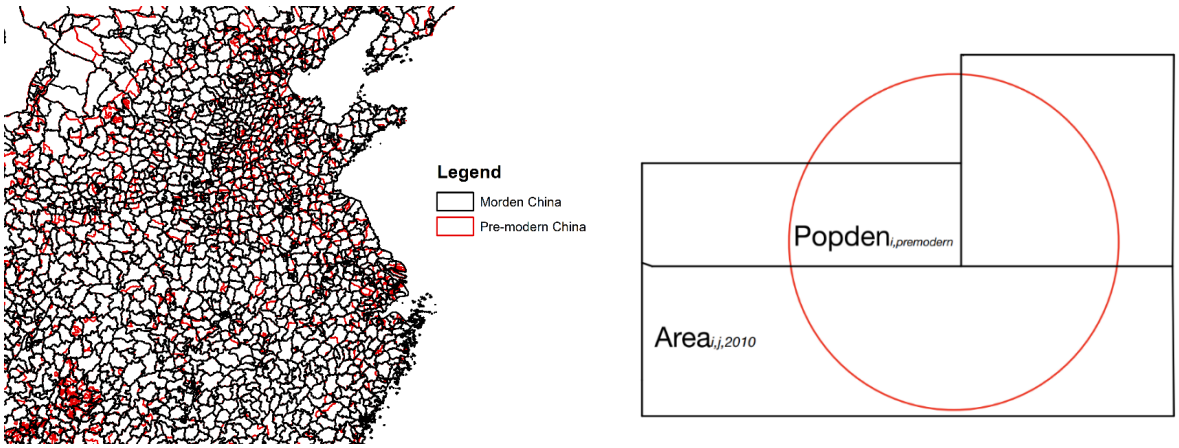
VARIABLES	(1)	(2)	(3)	(4)
	Night light luminosity in 2010			
Historical government quality	0.247***	0.227***	0.0840***	0.511***
	(0.0282)	(0.0351)	(0.0248)	(0.145)
Land area (logged)	-0.484***	-0.455***	-0.236***	-0.447***
	(0.0660)	(0.0679)	(0.0417)	(0.0752)
Longitude	0.0409	-0.00534	-0.0465	-0.0348
	(0.0299)	(0.0335)	(0.0276)	(0.0285)
Latitude	0.0167	0.0150	0.0138	0.0353
	(0.0360)	(0.0415)	(0.0272)	(0.0323)
Slope	-1.961***	-2.272***	-1.617***	-1.505***
	(0.335)	(0.309)	(0.334)	(0.289)
Elevation (logged)	-0.0920	-0.000519	-0.0433	-0.00230
	(0.0623)	(0.0631)	(0.0478)	(0.0438)
Prefectural capital		-0.0466	-0.0258	-0.206**
		(0.0632)	(0.0558)	(0.0850)
Distance to the coast		-0.249***	-0.169***	-0.170***
		(0.0411)	(0.0353)	(0.0326)
Distance to the Grand Canal		-0.0734*	-0.0417	0.0296
		(0.0387)	(0.0326)	(0.0466)
Distance to Beijing		-0.153***	-0.00413	0.00367
		(0.0514)	(0.0369)	(0.0451)
Agricultural yields (logged)		0.0373	0.127*	0.157*
		(0.0923)	(0.0621)	(0.0829)
Human capital (logged)			2.129***	2.025***
			(0.329)	(0.314)
Urbanization ratio			1.135**	0.820**
			(0.421)	(0.400)
Manufacturing ratio			0.0160***	0.0171***
			(0.00372)	(0.00365)
Population density			372.2***	149.9
			(122.6)	(96.99)
Constant	-3.251	7.368**	2.864	-1.443
	(2.562)	(3.001)	(2.004)	(2.358)
Geographic controls	Y	Y	Y	Y
Historical socio-economic controls	N	Y	Y	Y
Current socio-economic controls	N	N	Y	Y
Province Fixed Effects	Y	Y	Y	Y
Observations	1,664	1,638	1,456	1,456
R-squared	0.665	0.675	0.765	0.683
First stage F				42.719

Robust standard errors adjusted for clustering at the province level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix B

We collect information on socio-economic conditions at county level. The information includes urbanization rate, manufacturing ratio, population density, and average schooling year. The information is extracted from the population census in 2010. As county boundaries in China have changed over last 100 years, we adapt the census data to the historical county boundaries. This allows us to match recent data with historical information from China. The process of data adaptation is illustrated by Figure B1.

Figure B1. The difference of boundary between pre-modern and modern China



Here, the red circle represents county i in the Qing dynasty (pre-modern China). County i overlaps with county j in 2010 (depicted by the rectangle). County i is divided into j polygons by the boundaries of county j. The area of each polygon is depicted as $Area_{i,j,2010}$. In order to transform the data, we first calculate the 2010 indicators — population density, urbanization ratio, manufacturing ratio, and human capital (j=1,2,…)— for each rectangle of county j, using the 2010 boundaries. We assume that each of these indicators are evenly distributed in each rectangle. We then calculate the population in each small polygon as $Area_{i,j,2010} \times popden_{j,2010}$. This calculation facilitates estimating the population density of each county i in historical times, using the following formula:

$$popden_{i,pre-modern\ China} = \frac{\sum_{j=1}^{j} Area_{i,j,2010} \times popden_{j,2010}}{Area_i}$$

Chapter 2: Local institutions and pandemics: City autonomy and the Black Death

1 Introduction

While the role of cities in promoting economic development has been widely acknowledged, its negative externalities, such as air pollution, crime and pandemics, have only been taken seriously when they endanger citizens' health and lives (Duranton et al, 2015). The recent outbreak of Covid-19 has cast the role of pandemics for cities in another light. It has become evident that, as a consequence of the agglomeration of individuals and geographic proximity in cities (often combined with a lack of preventative measures), urban dwellers are a highly vulnerable group to the spread of disease. The high urban mortality and infection rates of SARS, Ebola, and COVID-19 have reminded everyone how vulnerable cities are to pandemics, even in modern times (Atkeson, 2020; Bowles et al., 2016; Desjardins et al., 2020; Qiu et al., 2018). This observation has been confirmed through reviewing the impacts of historical pandemics, including Justinian's plague, the Black Death, and the 1918 influenza pandemic, on human lives.

The sheer density of cities is considered by many a factor that limits the efficiency of any measure taken against pandemics (Almond, 2006; Jedwab et al., 2019). Often, cities are at the mercy of disease and their variations in performance during pandemics is often considered to be dependent on luck. This has led some researchers on pandemics to treat them as exogenous shocks (Weil, 2010; Young, 2005). In the field of applied economics, it is becoming increasingly popular to treat the outbreak of epidemics as a natural experiment to examine the impact of external shocks on the local economy (Almond, 2006; Dittmar & Meisenzahl, 2020; Donaldson & Keniston, 2016; Jedwab et al., 2019). In this respect, there is little differentiation between anti-pandemic efforts made by cities and those by their inhabitants. However, are the short-term consequences of pandemics truly random? Are institutional factors really unimportant to the control of pandemics? More importantly, are changes to city governance while fighting epidemics futile?

In this paper, we intend to answer these questions by systematically reviewing how differences in urban institutions affected the spread of the Black Death during its initial outbreak between 1347 and 1352. It should be noted that, although the biggest outbreak of the plague took place during this period, there were recurrent outbursts later in the century and in the following centuries. Numerous European cities, in particular,

witnessed repeated episodes of the Black Death, as was the case of London between the 14th and 17th centuries (Earn et al., 2020). Nevertheless, the biggest shock and incidence took place during the initial outbreak. Taking this into account, as well as a lack of complete pan-European data on the incidence of subsequent outbreaks, we focus our attention on the initial mid-14th century outbreak.

The Black Death has been chosen for three reasons. First, it is possibly the largest demographic shock in the history of mankind. It is estimated that one- to two-thirds of Europe's population died from the plague between 1347 and 1352 (Ziegler, 2013). Entire villages disappeared, while many cities went into permanent decline. Although the Black Death has attracted attention from many disciplines, research on the political rationale behind the large variation in mortality rates across Western European cities remains limited. Most political research on the Black Death emphasizes the interaction between wars and the plague in Eurasia (Caferro, 2018; Jackson, 2018). Only a few pieces of research (Ormrod, 1996; Palmer, 2000) have specifically investigated into the plague from the perspective of rule of law or national government. However, while all of these works examine how political influence evolved after the outbreak of an epidemic, the political system's response to the Black Death has largely been ignored. Some cities were almost completely devastated, while others were mostly spared by the plague. Differences in local institutions may have played a role in these variations.

Second, the Black Death is the first pandemic in history that was recorded in detail, with information on mortality rates, durations of specific waves, and the impact on cities' populations (Alfani & Murphy, 2017; Benedictow, 2004). The data available make it possible to comprehensively analyse whether the different plague prevention measures adopted by cities made a difference for the impact of the pandemic.

Third, the Black Death is often identified as a crucial turning point in history, which may still have a profound impact on modern development (Herlihy, 1997). There is a popular view that the Black Death is the key element in Europe's divergence from the rest of the world (Diamond, 2013; Epstein, 2000; North & Thomas, 1973). Some even declare that the plague set Western Europe on a path to faster economic development by contributing to the onset of a high-income Malthusian equilibrium (Clark, 2008). However, without a deep understanding of what caused significant variations in the Black Death's impact from city to city, it is difficult to assess the validity of these arguments.

We use new city level data stemming from Christakos et al. (2005) to uncover what has led to the considerable variations in the Black Death's impact across Western European cities between 1347 and 1352. Based on our knowledge, although explorations of the

Black Death differ, few papers directly focus on the role of institutions in the Black Death, especially from the perspective of the local government. In fact, we know next to nothing about the effects of local institutions in dealing with the spread of the pandemic and its impact. However, in this period, the autonomy and governance of European cities varied significantly from one city to another. Did cities with greater autonomy, which may have enabled them to react faster and differently against the spread of the plague, perform better than cities that were under the tight control of remote kings and emperors?

To answer these questions, we will look at the link between city autonomy in Europe during the Middle Ages and Black Death mortality rates during the initial outbreak of the plague. Our results show that greater autonomy has led to a considerable reduction in pandemic death rates. This finding is robust to the inclusion of a gamut of geographic, social, and political covariates. The impact of city autonomy was therefore significant, but not seconded by the political factors considered in this analysis. Urban autonomy facilitated a more nimble and effective adoption of measures against the diffusion of the disease within the city. Nevertheless, the duration of the Black Death pandemic in different cities was relatively random. Based on our results, none of the variables considered have any impact on reducing the epidemic duration.

Our search contributes to literature on the geography of plague by bringing into light the local institutional dimension. Most of the perspectives driven by politics focus on the direct impact of the pandemic on local politics and political economy. In contrast, our research question explores the political rationale behind variations in the impact of the plague. In addition, by considering possible errors in historical local data, our paper explores whether this is the case by means of different experiments and trials to ensure that the estimates are reliable.

The remainder of this paper proceeds as follows: section 2 describes the historical context of the Black Death and the diversity of autonomous cities at that time. Section 3 introduces the model and data used in this study. Section 4 presents the estimated results, and section 5 offers a conclusion.

2 Historical context

2.1 The Black Death

The Black Death was one of the most devastating pandemics recorded in human history. The plague reached Europe in October 1347, when 12 ships from the Black Sea were moored in the Sicilian port of Messina (Hajar, 2012). The people on the dock found that most of the sailors on the ship were dead and those still alive were dying, covered in

blood and pus. The Sicilian government decided to expel these ships immediately, but it was too late. Based on estimations, over the following five years, the Black Death killed over 20 million people in Medieval Europe, which is around one to two-thirds of the population at that time (Benedictow, 2004).

Scientists have understood the Black Death from an epidemiological perspective. It was spread by a bacillus called *Yersinia pestis* (named after Alexandre Yersin, who discovered the germ at the end of the 19th century). Under normal conditions, Yersinia pestis transmits poorly from person to person, but it can have a high transmission rate under exceptional circumstances, as has happened in the past in Manchuria (Nishiura 2006), Madagascar (Randremanana et al., 2019), and arguably also in Europe during the Black Death (Dols, 1979). A series of recent paleo-genomic studies have explored the paths of transmission and found that during the Black Death, both natural rodent-based foci within Europe and repeated introductions from Central/Eastern Asia contributed to its spread (Demeure et al., 2019).

The heterogeneous effects of the play remain, however, still inconclusive. Recent bioarcheological research has shown that the impacts of the Black Death could have been influenced by people's age and pre-existing health status (DeWitte & Wood, 2008). Godde et al. (2020) support this, suggesting that frail individuals were more vulnerable to the illness. However, they find that the risk of death in the elderly did not increase at all. Gender could have played a greater role in differences in incidence. Using data from Hainaut, Curtis and Roosen (2017) found the Black Death was sex-selective, killing all women than men in its initial stages. However, a synthetic study by Godde and Hens (2021) has come to the conclusion that in London the plague killed both females and males indiscriminately. Overall, variations in the transmission patterns of the plague may have been shaped by specific conditions in different cities and countries.

Medieval cities were notorious for high levels of ill-health and bad living conditions, compared to rural and nomadic lifestyles. Pest infestation was common. Water-supply infrastructure built to ensure public access to clean water only existed in a few cities, such as medieval London (Salzman, 2017). In most cities, water was transported from outside and then shared between many people for different uses. Disease vectors, such as fleas and bedbugs, easily carried bacteria between humans. Roundworms and whipworms were not restricted to the poor, who might not have clean water or clean clothes in their house. Even Richard III, the king of England, was infected by roundworms. Although privies have been built in Medieval Europe for a long time, they were not in sufficient supply, meaning that faecal contamination of food was rife (Mitchell, 2015). Furthermore, sewage ran in the gutter and people threw their faeces and urine directly out of the window (Gottfried, 2010). These bad living environments in the cities directly

affected people's health both before and during the Black Death era. Bioarcheological evidence has shown that health conditions in London generally decreased in the 13th century, which possibly contributed to the high mortality rates in the city during the Black Death (DeWitte, 2015). By measuring heath conditions considering adult stature, DeWitte and Hughes-Morey (2012) found that poor health —proxied by short stature— was associated with higher mortality rate during the Black Death.[10] Insalubrious living conditions caused cities in Medieval Europe facilitated the spread of the Black Death.

The incidence of the Black Death across cities in Europe in its initial stages were, however, far from homogenous. From 1347 to 1352, mortality rates varied greatly from city to city. In some cities, such as Leuven, Arras, and Douai, reported mortality rates were lower than 10 percent. In contrast, mortality rates in places including Yarmouth, Scicli in Sicily, and Ciudad Real exceeded 70 percent of the city population. What factors explain such large variations in mortality? Some scholars argue that ports and centres of trade routes were hit the hardest by the Black Death (Yue et al., 2017). However, counterexamples abound and make it impossible to reach a definite conclusion. For example, the Italian port of Genoa reportedly had a Black and Death mortality of about 35 percent. This was lower than that of inland Verona (45 percent), but also lower than in the port of Venice (60 percent) (Christakos et al., 2005: 163). Although there are some rough descriptive statistics and case studies, historical accounts have not rationalized the patterns in the data (Cohn & Alfani, 2007; Theilmann & Cate, 2007). More importantly, when the Black Death appeared, European societies were unprepared to face the threat. However, as it became clear that the plague was there to stay, a process of institutional adaptation occurred. This institutional adaptation became an example of how humans react to a change in their environment.

2.2 Urban autonomy

One factor that has generally been overlooked is the role of local institutions in confronting the plague. Urban governments in Western Europe had been diverging for more than 200 years before the Black Death hit.[11] Following the collapse of the Carolingian Empire, some cities started to develop forms of local participative government and demand representation in national policymaking. The first occurrences of autonomous cities were in Spain and Italy in the 11th century and quickly spread across Europe in the following

[10] has been argued that stature is positively related to the health condition of adults (DeWitte & Hughes-Morey, 2012).

[11] In this study, we have followed the definition and data provided by Bairoch et al. (1988), Christakos et al. (2005) and Stasavage (2014). The Western European countries considered include modern-day Austria, Belgium, Denmark, France, Germany, Italy, Netherlands, Portugal, Spain, Switzerland, and the UK.

centuries (Van Zanden et al., 2012). Urban elites such as lawyers, merchants, and entrepreneurs became increasingly influential in these cities (Belloc et al., 2016). Members of the elite soon began to form associations and make agreements to cooperate on issues of common concern (Guiso et al., 2016). As a result, autonomous cities started to slowly evolve (Jones, 2003). Citizens who had previously been excluded from urban governance began to partake in the sessions of the city government. Within the processes, citizens gradually learned how to regulate economic and social issues and resolve their disputes with a decentralized approach. Citizen participation improved across all of Europe and, in some cases, the need for central authority and authoritarian leaders became redundant (Belloc et al., 2016). Within autonomous cities, personal freedom was no longer threatened by distant kings and was protected by local laws. City autonomy also brought about a greater scrutiny of government officials' actions by local citizens and new forms of regulation. It is worth noting that not all the people living in autonomous cities had the qualification of citizenship and were allowed to participate in the political process. Citizenship in Italy only awarded to men who owned a house after coming of age. Women, servants and minorities (mostly Jews and Muslims at the time) were excluded (Belloc et al., 2016). In some places citizenship was also linked to either patronage (by religious authorities or nobles) or membership of associations (mainly guilds) (Alsayyad & Roy, 2016).

The increase in the number of politically autonomous cities in the late Middle Ages has been considered a factor in the economic and political rise of Europe at the time (Weber, 1956). Research by Acemoglu et al. (2005), De Long and Shleifer (1993), and Jacob (2010) also points to the positive effect that autonomous cities had on urban development. Acemoglu et al. (2005) argued that rapid economic development mainly occurred in places with non-absolutist initial institutions. The greater the level of local authority and the fewer the constraints on economic activity imposed by the state, the higher the incentives and opportunities for economic and urban expansion. As a key political institution, urban autonomy could effectively limit the power of the monarchy, secure property rights, and stimulate economic development (Frank, 1978; North, 1978; North & Thomas, 1973).

However, research so far has been limited to linking urban autonomy to economic outcomes. How city autonomy affected reactions to a severe public health crises, such as the Black Death, has attracted limited attention. Our knowledge of the topic is, so far, restricted to a number of anecdotes that point to a more rapid and active reaction of self-governing cities relative to other cities (Geltner, 2020). Autonomous cities seemed to have led the introduction of more effective health measures and institutions. At a local level, within an infected community, human contact was limited by quarantines

and other temporary restrictions on freedom of movement (Alfani & Murphy, 2017). Venice, considered a classic example of a well-organized autonomous city, took a series of measures that illustrated they understood that the Black Death was not just bad air. Specifically, officials in the city inspected wine, fresh meat, and the water; ships were boarded and searched; new burial regulations were put in place, meaning that corpses were thrown into barges; measures were also introduced to restrict population clustering. Houses affected by the plague were painted with vinegar and fumigated with sulphur, while religious processions were banned unless a license was granted. The city also established a command-and-control centre where all decisions concerning the pandemic were centralised and communicated to the population (CitiesX, 2018). Non-autonomous cities, by contrast, did not have the luxury of adopting their own measures and had to wait for decisions to be made by distant kings in, say, Paris, London, or Toledo, or by the Pope in Rome (Hohenberg & Lees, 1995). It is not always the case that autonomous cities made rapid and/or suitable decisions against a pandemic everyone knew little about, but they at least had more initiative relative to cities with no or a far lower degree of autonomy. In view of this we can formulate two basic hypotheses:

H1: Within the European cities, autonomous cities fared better than non-autonomous cities in terms of reported mortality rates during the initial outbreak of the Black Death.

H2: Within the European cities, autonomous cities adopted measures that shortened the duration of the pandemic than those adopted by non-autonomous cities.

3 The model and data

3.1 The research design and model

As discussed, the primary objective of this paper is to investigate whether and to what extent the urban autonomy could affect the severity of the Black Death. Considering that our data structure is cross-sectional, to test the two hypotheses above, omitted variable bias and reverse causality could be the most important problems in the empirical stage of our research. To relieve these concerns, our empirical strategy takes both Ordinary Least Squares (OLS) analysis and instrument variable strategy. Basic OLS regression helps us observe changes in the coefficient of urban autonomy when different covariates are gradually added to the model. Thus, Ordinary Least Squares (OLS), with a set of controlled variables and spatial fixed effects, are introduced to explore the interactions between the Black Death and autonomous cities. After the OLS analysis, we will use a suitable instrument to deal with the endogeneity more thoroughly. The model adopts the following form:

$$Mortalityrate/Duration_{i,1347-1352} = \beta_1 Autonomous\ city_i + \beta_2 X_i + \beta_3 Z_i + \alpha_c + \varepsilon_i \quad (1)$$

where i represents a city, and the outcome variables are, $Mortalityrate_{i,1347-1352}$, which stands for the Black Death-related mortality rates in percentages (the proportion of deaths in the overall population) in the years between 1347 and 1352; $Duration_{i,1347-1352}$, which depicts how long the Black Death lasted, measured in months. The independent variable of interest, $Autonomous\ city_i$, is a dummy variable that measures whether a city had a high degree of self-governance during the years of the plague.

X_i represents a vector of economic, political, and socio-demographic characteristics of the city (including whether the city was a capital city, or if it had a bishop, archbishop, or a university). Specifically, urban population is chosen because we consider that the epidemic could spread faster in populous cities; capital cities usually control abundant political and economic resources and may be able to cope well when facing the plague; a parliament, a bishop and an archbishop were all important political and religious institutions in medieval Europe. By considering these factors, we compare the impacts of the plague with the impacts of urban autonomy. Universities, as the proxy variable of human capital at the regional level, are likely to provide more feasible strategies when dealing with diseases.

Z_i corresponds to a vector that includes the city's geographic factors, such as whether the city was on a river, the sea, along a Roman road, as well as indications of the soil quality and elevation above sea level. Transportation methods and routes such as rivers, seas, hubs and roman roads are included in the analysis because they tend to accelerate the spread of pandemics, which lead to increases in mortality and in their duration. Soil quality and elevation can help us, to some extent, analyse the impact of natural resources on the spread of the Black Death.

To control regionally correlated unobserved variables, we include country dummies to add the spatial fixed effects α_c into the regressions. Europe being a vast and diverse continent makes it difficult to compare cities across countries directly due to weather, cultural, and geographical differences. Introducing country fixed effects may help us capture other unobserved effects. Furthermore, to deal with the potential within-country correlation of the error term, we cluster the standard error at the country level for all specifications.

3.2 Data

In this section we outline the sources and attributes of the data. In our research, one of the challenges is to quantify reasonably both explanatory and outcome variables — namely the severity of the incidence of the Black Death and urban autonomy— in medieval times. Both mortality and duration data of the Black Death is provided by Christakos et al. (2005), who collected, processed, and verified the Black Death data for a large number of European cities systematically. Regarding our explanatory variable, urban autonomy, we follow Bosker et al. (2013) and Stasavage (2014), and use a dummy variable to distinguish whether a city is autonomous or not. In the following section, we will introduce the data sources of the dependent, independent, and control variables in more detail.

The Black Death. Data on the Black Death has been taken from Christakos et al. (2005). These authors compiled data on reported mortality rates based on information from a wide array of historical sources. They carefully examined each data point and judged between conflicting estimates based on the best available information. They also collected information on the duration of the Black Death in certain cities. According to their dataset, the average mortality rate in European cities was approximately 42.57 percent, with the average duration of a wave of the plague lasting 6.3 months. The distributions of mortality rates and duration within Western European cities are shown in Figures 1 and 2. Although no geographical pattern emerges in terms of mortality, Figure 1 identifies a number of hotspots where the wreck associated with the Black Death was higher than elsewhere. The highest incidence was found in Sicily, central Italy, southern France, Catalonia, and southern England.[12] The incidence was lower in southern Germany, Belgium, and most of central and northern France. In terms of the duration of the plague, the Black Death lasted longer in Belgium, northern France, central Germany, Andalusia, and in isolated cities such as Toledo, Barcelona, and Naples (Figure 2). There seems to be no correlation between reported mortality rates in cities and the duration of the plague there (Figure A2 in appendix). Naples, for example, suffered the plague for a long time, but its mortality rate was low. In the Sicilian cities of Scicli and Syracuse, the pandemic was shorter, but its impact was far more devastating.

Figure 1: Mortality rate of the Black Death

[12] To have a more intuitive understanding of the countries and boundaries we are discussing, we use the boundaries of modern countries.

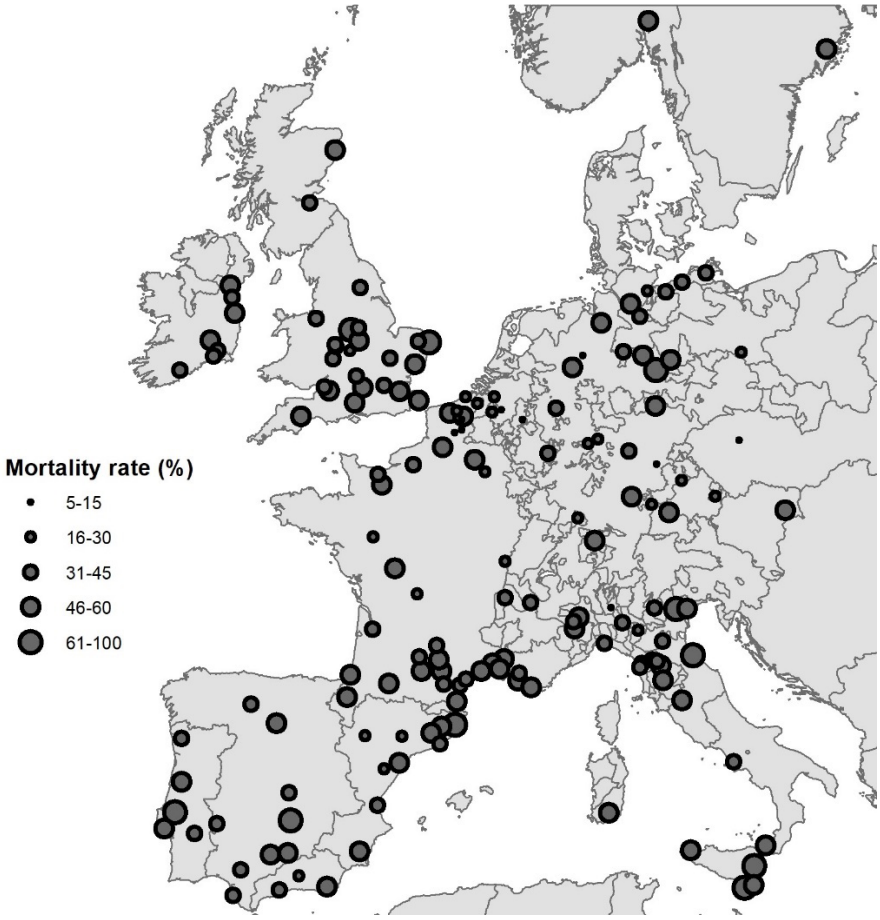
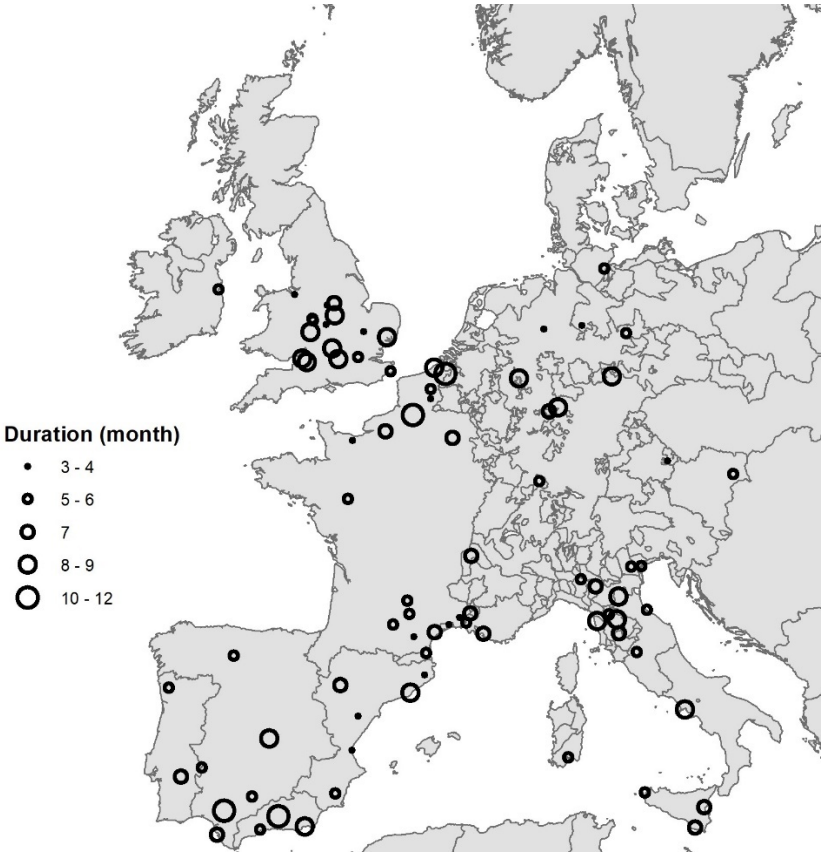
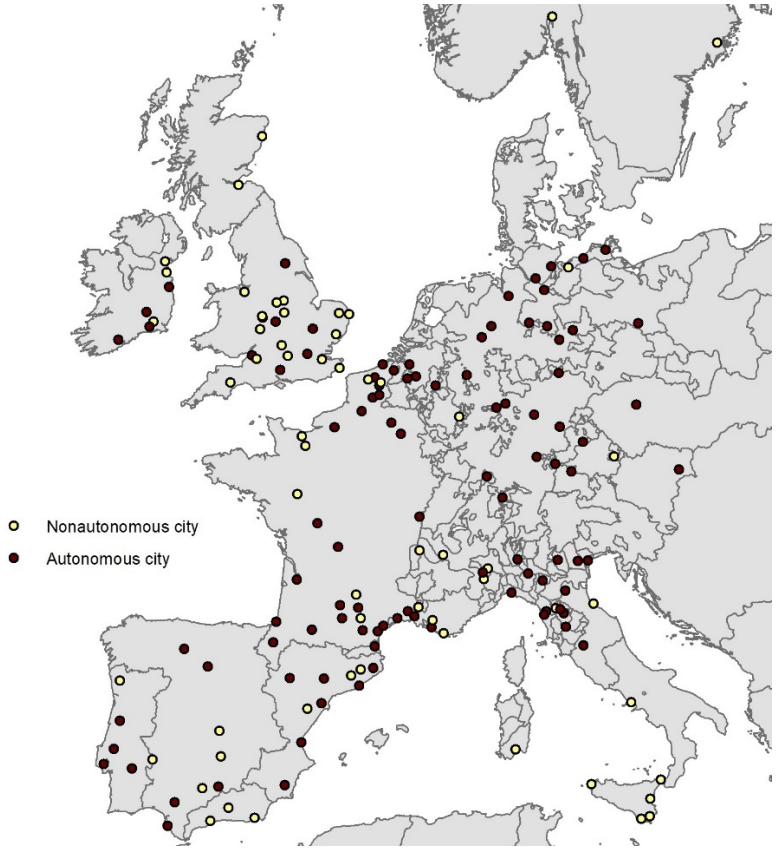


Figure 2: Duration of the Black Death



Autonomous cities. To measure the autonomy of cities, we use data provided by Bosker et al. (2013) (Figure 3). These authors classified cities according to their degree of local participative government. Cities were classed as autonomous if they had institutions for self-governance and the members of these institutions were chosen by the inhabitants of the city itself, not by outside rulers, i.e., if there was a local urban participative organization that made decisions about local urban affairs. Evidence of the presence of consuls, official documents like notarial acts signed by the representatives of cities, and the presence of imperial charters to grant self-governance to cities, were the criteria used to indicate whether a city was autonomous. To further verify the accuracy of our data and as a means to ensure robustness, we also adopt a stricter definition of urban autonomy provided by Stasavage (2014).

Figure 3: The location of autonomous and nonautonomous cities



City population. The main source of city population data is Bairoch (1988). This source reported population estimates for cities in 1300. The criterion for inclusion in the dataset is a city population greater than 1,000 inhabitants at the beginning of the 14th century. To expand the sample size, we also add the population of 14 cities that were mentioned by Christakos et al. (2005) (Figure A1 in appendix).

Social and political controls. We consider whether a city was a capital city. We also register whether it was a bishopric or an archbishopric, respectively, based on the information provided by McEvedy and Jones (1977). Capitals and cities high up in the ecclesiastical hierarchy were important seats of power and, therefore, more likely to attract people and economic activity based on the presence of a sovereign or bishop. We use information provided by Bosker et al. (2013) to establish whether a city had a university or not. In Europe, universities started to appear from the twelfth century onwards across the continent. As a breeding ground for knowledge and development, the effect of universities on urban development can be considered essential (Cantoni & Yuchtman, 2010). Furthermore, we also document whether a city belonged to a political entity where it could participate in the political process by having representatives in an active parliament (Van Zanden et al., 2012).

Geographic controls. The geographical characteristics included in the analysis concern a city's opportunities for long-distance trade and its degree of agricultural development. Using data from Nussli (2011), we capture a city's potential for water-based trade in medieval times by documenting whether it was located within ten kilometres of the sea or a river. In the case of land-based transportation, we resort to Bosker et al.'s (2013) work to document whether a city was on a former Roman road or a hub of a Roman road. An advantage of using Roman roads was their uniformity across Europe. To capture the agricultural potential of a city, data from Ramankutty et al. (2002) is utilised. These authors combine information on soil quality into one index, indicating the probability of viable agriculture in a given place in medieval times. Finally, we also collect each city's elevation from Jarvis et al. (2008). The elevation is reported in metres. It should be noted that these covariates were taken from data for the year 1300 because it is closest to the start of the pandemic.

Our primary sample consists of a total of 162 cities that had more than 1000 inhabitants in 1300 and for which records of Black Death mortality rates are available. These cities made up close to 60 percent of the urban population of Western Europe in this period. Out of the 162 cities, we have plague duration information for 83. The descriptive statistics of variables are summarized in Table 1.

Table 1: Descriptive Statistics

Variable	**Obs.**	**Mean**	**STD**	**Min**	**Max**
Mortality rate	162	42.56864	14.96201	5	100
Duration	83	6.349398	1.889856	3	12
Autonomous city	162	0.62963	0.484401	0	1
River	162	0.759259	0.428859	0	1
Sea	162	0.234568	0.425042	0	1
Hub roman road	162	0.351852	0.479029	0	1
Roman road	162	0.234568	0.425042	0	1
Soil quality	162	0.724261	0.227047	0.117	0.999
Elevation	162	123.8148	168.0637	0	934
Population (logged)	162	2.164298	1.32459	-0.69315	5.010635
Capital	162	0.12963	0.336937	0	1
Parliament	162	0.345679	0.477064	0	1
Bishop	162	0.462963	0.500173	0	1
Archbishop	162	0.135803	0.343641	0	1
University	162	0.092593	0.29076	0	1

4 Empirical analysis

4.1 Ordinary Least Square (OLS) estimates

Table 2 reports the baseline results of the mortality rates of the Black Death. We begin by considering the autonomy of a city as our variable of interest. To provide a benchmark, at first, the only control introduced is country fixed effect in medieval times, which eliminates the spatial correlations in column (1), before controlling for other covariates. The coefficient for the autonomy of a city is negative and significant (Table 2, column (1)). City autonomy is strongly related to lower Black Death mortality. Geographic, social, and political controls were gradually added in columns (2) to (7). The result of adding these controls does not undermine the significance of the coefficient for city autonomy. Once factors such as the size of the city, its accessibility, the agricultural potential of surrounding areas, and its position in the political and/or ecclesiastical hierarchy are controlled for, city autonomy remains a strong factor connected to lower mortality rates during the Black Death. As column (7) suggests, being an autonomous city could have reduced the mortality associated with the plague by 5.22 percent on average, compared to non-autonomous cities.

When it comes to geographic factors, elevation is significantly linked to lower Black Death mortality rates (column 2). Cities located at high altitudes are frequently less accessible and have less convenient transportation than cities located along trade roads or at crossroads. This is crucial for preventing the spread of a pandemic, as the cost of exchanging goods and trade is higher and this limits the influx of travellers who may bring the disease in their carts. A lower accessibility to other cities may have become an advantage during the Black Death and elevation may have provided natural protection against the spread of the disease. Soil quality is positively related to mortality in some of the regressions (Columns (5) - (7)), meaning that areas with fertile land and markets for agricultural produce were more exposed to the plague (Heinonen-Tanski & van Wijk-Sijbesma, 2005). Other geographic factors, such as being a port or trade centre, do not appear connected with Black Death mortality during the first outbreak. This is consistent with Christakos et al.'s (2005). However, it is worth noting that, according to some authors (e.g., Yue et al., 2017), ports and trade centres remained more vulnerable to the plague.

Columns (3) - (7) of Table 2 consider social and political factors. Column (3) shows that the population size is both significantly and negatively related to mortality rates. This implies that larger cities were, in contrast to expectations, less affected by high reported mortality rates statistically. This is in line with Roosen (2017: 45-47), who using the Hainaut mortmain accounts, rejected the previous widespread belief of an urban/rural divide in the incidence rates of the plague. Our results support Roosen's (2018) view

that rural areas did not fare better during the Black Death. Carozzi et al. (2020) also report similar results during the COVID-19 pandemic, when checking the interactions between urban density and the severity of COVID-19. Larger cities were often the seats of power and may have had better resources to combat the consequences of the plague. Nevertheless, we find no association whatsoever between a raft of political and religious factors and the incidences of the Black Death (Table 2, columns (4) to (7)). While some institutionalists have argued that an active parliament is another key element behind Europe's rise (De Long & Shleifer, 1993; North & Weingast, 1989), we find no evidence that hosting a parliament provided any protection against the spread of the Black Death, once city autonomy has controlled for. Similar results emerge in the case of universities (Table 2, column (7)). The fact that universities at the time were mostly concerned with teaching theology and philosophy, rather than science and medicine, may help to explain this result.

Table 2: City autonomy and mortality rate of Black Death

	Black death mortality rate (percent, 1347-1352)						
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Autonomous city	-6.187**	-6.278***	-4.598***	-4.526***	-4.855***	-5.200***	-5.222***
	(2.334)	(1.254)	(0.799)	(0.732)	(0.832)	(0.662)	(0.652)
River		-2.318	-1.289	-1.292	-1.178	-1.252	-1.277
		(3.236)	(3.391)	(3.409)	(3.181)	(3.043)	(3.088)
Sea		-0.600	0.844	1.019	1.357	1.378	1.380
		(3.048)	(2.647)	(2.907)	(2.592)	(2.574)	(2.603)
Hub roman road		-3.512	-2.001	-2.035	-1.794	-2.561	-2.563
		(2.019)	(2.078)	(2.030)	(2.061)	(2.929)	(2.943)
Roman road		-0.414	-0.282	-0.289	0.0860	0.0122	0.0563
		(2.137)	(2.155)	(2.139)	(1.920)	(2.205)	(2.147)
Soil quality		8.312	8.952	8.924	10.25**	10.92*	10.95*
		(6.079)	(5.795)	(5.765)	(4.419)	(5.456)	(5.448)
Elevation		-2.157**	-2.456***	-2.405**	-2.293**	-2.257**	-2.254**
		(0.751)	(0.786)	(0.944)	(0.777)	(0.868)	(0.874)
Population			-2.463***	-2.401***	-2.292***	-2.416***	-2.427***
			(0.394)	(0.503)	(0.441)	(0.362)	(0.393)
Capital				-0.763	-0.613	-0.928	-0.897
				(3.393)	(3.689)	(4.124)	(4.176)
Parliament					3.308	3.191	3.222
					(4.850)	(4.336)	(4.191)
Bishop						1.609	1.586
						(2.015)	(1.968)
Archbishop						2.959	2.982
						(4.865)	(4.921)
University							0.386

							(2.688)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	162	162	162	162	162	162	162
R-squared	0.276	0.336	0.366	0.367	0.371	0.374	0.374

Robust standard errors adjusted for clustering at the country level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In Table 3, we estimate the factors that may be connected to the duration of the Black Death pandemic in different cities. In this case the information available is limited to 83 out of the 162 cities included in the original analysis. Here, virtually none of our variables are linked to how long the city was affected by the Black Death, implying that the duration of the pandemic in each city was relatively random. With the exception of the presence of an archbishopric, none of the variables are significant. The negative sign for city autonomy points to a potential reduction in the duration of the Black Death, but the coefficients of the OLS results are not significant at any level. Interestingly, in Christakos et al.'s (2005) research, there is a positive correlation between city size and the duration of an epidemic, which contradicts our findings. In Table 3A (in appendix), we find that a positive relationship between urban population size and the duration only exists when the regression is estimated without country FE or other controls (shown in column (1) of Table A1). Thus, the positive relationship between population and duration is likely to be biased and influenced by other characteristics of cities.

Table 3: City autonomy and Black Death duration

	Duration of Black Death (month, 1347-1352)						
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Autonomous city	-0.0509	0.00730	-0.257	-0.314	-0.343	-0.384	-0.392
	(0.427)	(0.635)	(0.690)	(0.676)	(0.681)	(0.660)	(0.669)
River		0.104	0.0221	-0.0309	0.00423	-0.0151	-0.0307
		(0.836)	(0.803)	(0.771)	(0.765)	(0.770)	(0.757)
Sea		0.136	-0.0513	-0.208	-0.160	-0.112	-0.114
		(0.827)	(0.807)	(0.769)	(0.775)	(0.692)	(0.690)
Hub roman road		0.813	0.615	0.647	0.644	0.387	0.391
		(0.768)	(0.751)	(0.756)	(0.773)	(0.918)	(0.925)
Roman road		-0.000975	0.0141	0.0104	0.0341	0.146	0.174
		(0.362)	(0.356)	(0.358)	(0.361)	(0.180)	(0.147)
Soil quality		0.677	-0.128	0.00167	0.259	0.502	0.559
		(1.049)	(1.356)	(1.257)	(1.148)	(1.442)	(1.414)
Elevation		0.0746	0.0962	0.0415	0.0562	0.0912	0.0906
		(0.112)	(0.0929)	(0.101)	(0.0836)	(0.0615)	(0.0633)
Population			0.433*	0.374	0.395	0.313	0.297

			(0.199)	(0.210)	(0.225)	(0.172)	(0.177)
Capital				0.559	0.539	0.303	0.333
				(0.339)	(0.325)	(0.371)	(0.399)
Parliament					0.388	0.331	0.353
					(0.423)	(0.293)	(0.301)
Bishop						0.324	0.306
						(0.596)	(0.606)
Archbishop						1.439*	1.443*
						(0.747)	(0.764)
University							0.168
							(0.354)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	83	83	83	83	83	83	83
R-squared	0.107	0.152	0.193	0.199	0.204	0.249	0.250

Robust standard errors adjusted for clustering at the country level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.2 Two-Stage least squares (2SLS) estimates

Our variable of interest, whether a city was autonomous or not, is likely to be associated with a myriad of factors. Although we have already controlled for many possible confounding ones, the problem of omitted variables may remain. To tackle this challenge, we adopt a strategy introduced by Persson and Tabellini (2009) and Acemoglu et al. (2014) in their analysis of democratic transitions of countries. We instrument city autonomy with the proportion of other cities with an urban autonomy status in the same region, leaving out the observation of the city concerned. The regional polygon data represents a second-level administrative division of Europe in the 14th century, which is the most precise regional level data available[13]. Transitions into or out of autonomous cities occur in regional waves that reflect learning and spillover effects in neighbouring localities and create a power vacuum at the regional level of government. This variable meets the exclusion restriction that is conditional for other covariates and country fixed effects, as we assume that regional waves of the autonomous transitioning of political institutions could have influenced the impact of the Black Death. As a result, the potential influence of omitted variable bias on the regression coefficients is largely minimized.

Specifically, we posit that a city's autonomous status, c, was influenced by the degree of autonomy of neighbouring cities. Therefore, we consider R_c as the geographic region where a city is located. The regional influence that city c experienced to become autonomous is defined by:

[13] Both the regional definition and data are provided by Nussli (2011) (http://www.euratlas.com/about.html).

$$I_c = \{c' : c' \neq c, R_{c'} = R_c\} \tag{2}$$

$$Z_c = \frac{1}{|I_c|} \sum_{c'} \in I_c \tag{3}$$

The 2SLS estimates are shown in Table 4. The instrument, city autonomy in the region, is always significant, both in the first stage and in the reduced form regression (F-statistics for the excluded instrument range from 13.6–22.4). Specifically, column (1) in Table 4 shows that city autonomy itself has a significant impact on Black Death mortality rates when country fixed effects are introduced. Adding additional controls slightly decreases the magnitude of the coefficient, but the causal relationship between the autonomy of a city and lower Black Death mortality rates remains strong and highly statistically significant (columns (2) to (7)). The instrumented coefficient for city autonomy is larger than the OLS estimate, indicating that the Black Death mortality rate was 9.29 percent lower in cities with a substantial degree of self-rule. This suggests that the OLS result underestimates the effect of the city autonomy on the incidence of the pandemic. Overall, our 2SLS results are consistent with the OLS results and demonstrate that autonomous cities performed much better than their more centrally governed counterparts in their tackling of the pandemic.

Table 4: City autonomy and mortality rate of the Black Death, 2SLS results

	Black death mortality rate (percent, 1347-1352)						
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	-	-	-	-	-	-	-
Autonomous city	11.6243***	8.7177***	8.7944***	8.5551***	9.1506***	9.2885***	9.2885***
	(3.9847)	(2.5383)	(2.9166)	(3.0114)	(2.8389)	(3.0135)	(3.0135)
River		-0.6611	-0.6510	-0.6106	-0.6906	-0.7322	-0.7322
		(3.0064)	(2.8405)	(2.8393)	(2.6348)	(2.6109)	(2.6109)
Sea		-1.6484	-1.6362	-1.4121	-2.3933	-2.3943	-2.3943
		(1.9530)	(1.8157)	(1.7387)	(2.7338)	(2.7331)	(2.7331)
Hub roman road		0.6220	0.5831	1.0443	1.0072	1.0036	1.0036
		(2.3518)	(2.6566)	(2.3103)	(2.2972)	(2.3191)	(2.3191)
Roman road		0.8159	0.8337	1.1092	0.9173	1.0331	1.0331
		(2.2049)	(1.8990)	(2.1177)	(2.4616)	(2.4622)	(2.4622)
Soil quality		10.5854*	10.6156*	11.9200**	12.7423**	12.8609**	12.8609**
		(5.9784)	(5.7034)	(5.0321)	(6.0030)	(6.0658)	(6.0658)
		-	-	-	-	-	-
Elevation		2.4543***	2.4647***	2.3248***	2.3101***	2.3040***	2.3040***
		(0.6937)	(0.8373)	(0.6818)	(0.7375)	(0.7438)	(0.7438)
		-	-	-	-	-	-
Population		2.0836***	2.0906***	2.0090***	2.1514***	2.1693***	2.1693***

		(0.6879)	(0.6859)	(0.5735)	(0.4250)	(0.4112)	(0.4112)
Capital			0.1556	0.1975	-0.0848	0.0019	0.0019
			(3.3322)	(3.2087)	(3.4935)	(3.4933)	(3.4933)
Parliament				3.8850	3.6355	3.7141	3.7141
				(3.5677)	(3.0940)	(2.9035)	(2.9035)
Bishop					2.5127	2.4841	2.4841
					(2.0286)	(1.9889)	(1.9889)
Archbishop					3.4961	3.5582	3.5582
					(4.6012)	(4.6684)	(4.6684)
University						0.8283	0.8283
						(1.8221)	(1.8221)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	162	162	162	162	162	162	162
R-squared	0.2519	0.3549	0.3545	0.3622	0.3644	0.3639	0.3639
First stage F	22.4	15.5	13.5	13	19	21.2	21.2

Robust standard errors adjusted for clustering at the country level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

4.3 Placebo tests and robustness checks

4.3.1 Placebo tests

So far, we have established that city autonomy in medieval times was an institution that helped protect cities from the devastation of the Black Death. However, it is worth noting that dealing with any type of data referring to medieval times involves a certain degree of uncertainty. Although the data we use are the best available from a comparative perspective, the mortality rate and duration of the Black Death may have been underestimated.[14] For example, Roosen (2017) considers that the data by Christakos et al. (2005) for the case of Belgium may have incurred in some errors during the process of extrapolation. This may challenge Christakos et al.'s (2005) argument that cities in Belgium had the lowest mortality rates in the early stages of the Black Death. As both of our mortality and duration data for the Black Death is extracted from Christakos et al (2005), our estimates may also be affected by potential inaccuracies in the data. To address concerns that our baseline results are not related to potential estimation errors of historical data, we resort to Monte Carlo simulations to conduct a series of placebo tests.

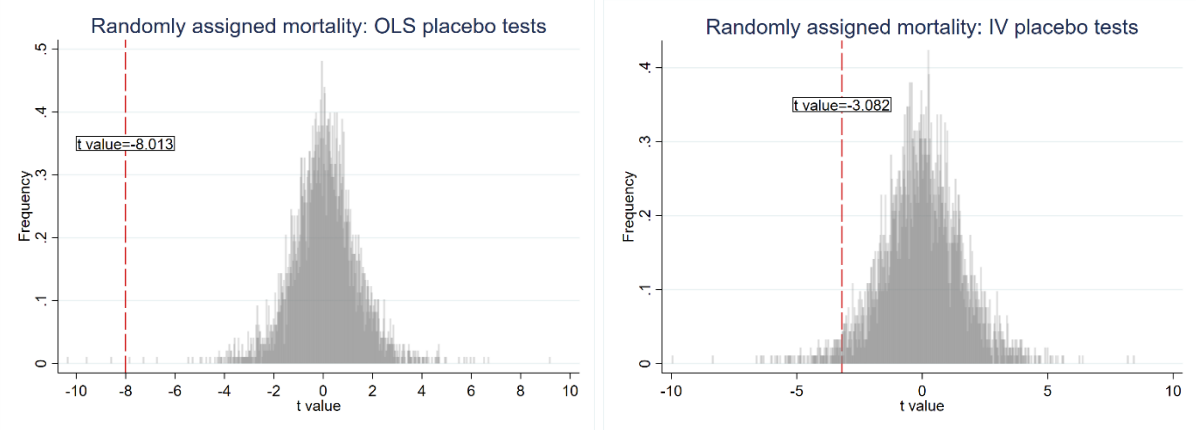
In the first set of Monte Carlo simulation trials, we compare the estimated effects with the distribution of placebo treatment effects when the mortality rates of cities were randomly assigned. Specifically, we randomly scramble the mortality rates of the 162 cities included in the sample. We then estimate placebo treatment effects according to both OLS and 2SLS models in the baseline analysis. Figures 4(a) and 4(b) plot the

[14] We are grateful for reviewers' kind reminders on the possible flaws in the data quality of the Black Death.

distribution of t-statistics from the placebo treatment effects, after running the regressions 5,000 times. The vertical lines mark the location of the t-statistics of the actual treatment effect. Among the 5,000 trials, we find that cases in which the corresponding t value exceeds the baseline OLS and 2SLS results are relatively few. The share of the placebo t-statistics that is larger than the actual statistic (P (t<=T)) can be interpreted as analogous to a p-value. It represents the probability that a randomly assigned mortality rate will present an effect at the same or higher level of significance than the actual mortality rate. Our results show that both P(t<=T) for OLS and 2SLS are smaller than 0.001. These results are also supported by Figures 4(a) and 4(b). Consequently, we can safely reject the null that our result is indifferent to the placebo treatment effects at all significance levels.

Figure 4: The distribution of t-statistics from randomly assigned mortality placebo tests

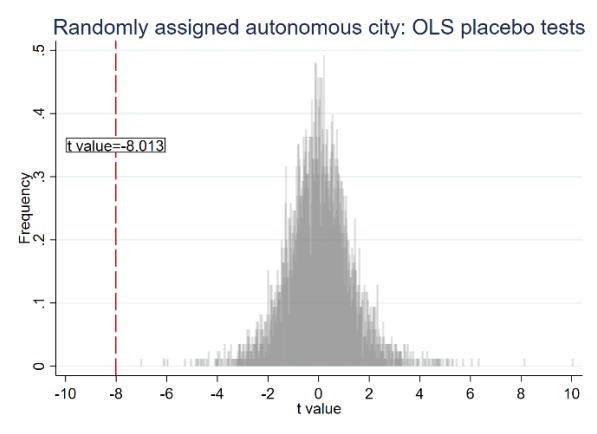
(a) (b)

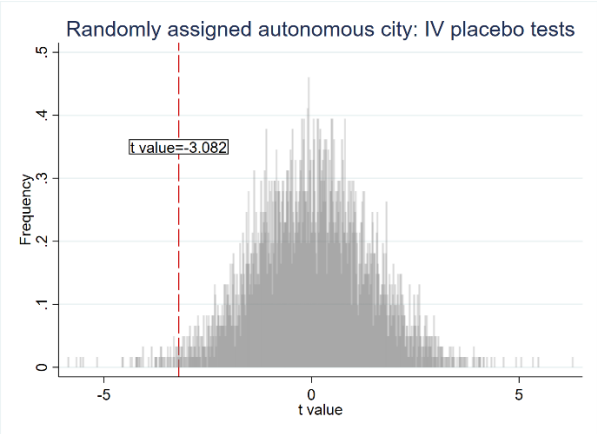


Another potential concern is that the same systematic data bias may also exist in our treatment variable: city autonomy. To address this concern, we verify our baseline results by randomly allocating the autonomous cities following the same approach. The simulations show that our baseline results are stable as well (Figure 5) and that P(t<=T) for both OLS and 2SLS are smaller than 0.001. This implies that our causal results are extremely unlikely to be caused by systematic data errors.

Figure 5: The distribution of t-statistics from randomly assigned autonomous city placebo tests

(a) (b)

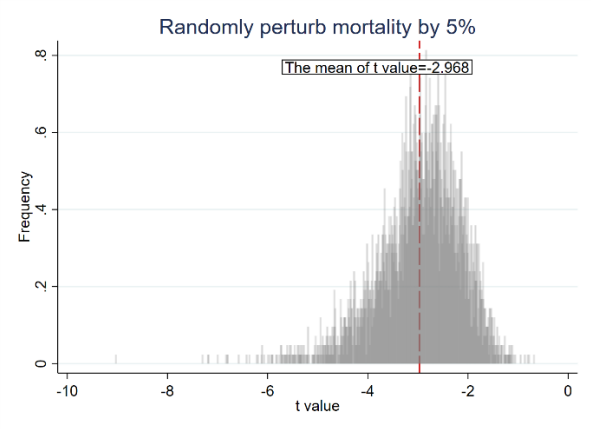


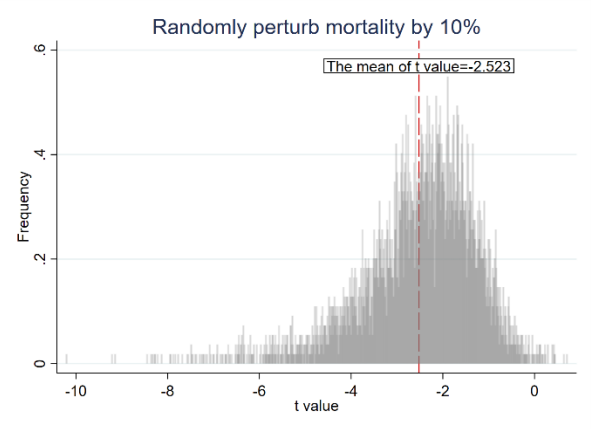


However, it should be noted that even if no systematic errors in the data are detected, this does not imply that the results of the baseline estimations are completely robust. It is common in historical data that some records of observations may deviate from reality. To address this concern, we conduct a second set of placebo tests. The logic behind this set of tests is to randomly select a proportion of observations (i.e., 20 or 30 percent) in the sample and alter the mortality data by a certain percentage (i.e., 1 - 20 percent). In practice, we first randomly select one-third of observations and modify their value by a certain percentage. For instance, if we hypothesize that one-third of the observations have a 5 percent measurement error, the alteration (or perturbation) level is set at 5 percent. In this case, if a region's reported mortality rate is 10 percent, once being selected, its mortality rate would increase to 15 percent. Second and to get the distribution of the random events, we perform Monte Carlo simulations, running the regressions 5,000 times. Executing this trial helps us examine the distribution of t-statistics by randomly altering the mortality rate by different percentages then. We can test how robust our results are in the case where there is a certain degree of errors in the mortality data. If manipulating the mortality rate by a small range (i.e., 1-3 percent) makes our results insignificant, it means that the estimations are fragile and possibly affected by inaccuracies in the mortality data. In Figures 6(a)-(d) we report the results of randomly altering one third of reported mortality rates by 5 - 20 percent, respectively.

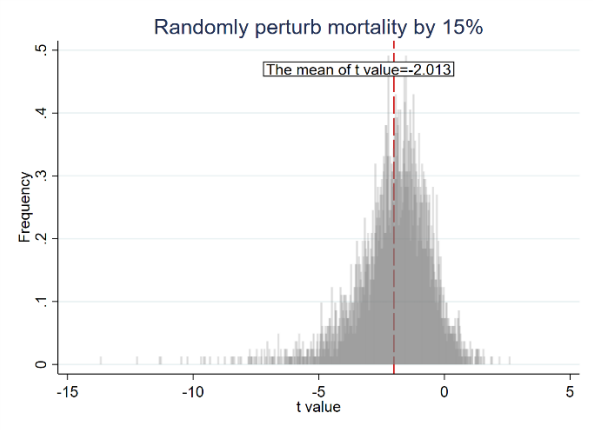
Figure 6: Distribution of t-statistics by randomly assigned mortality by certain percentage-points in one third of cities

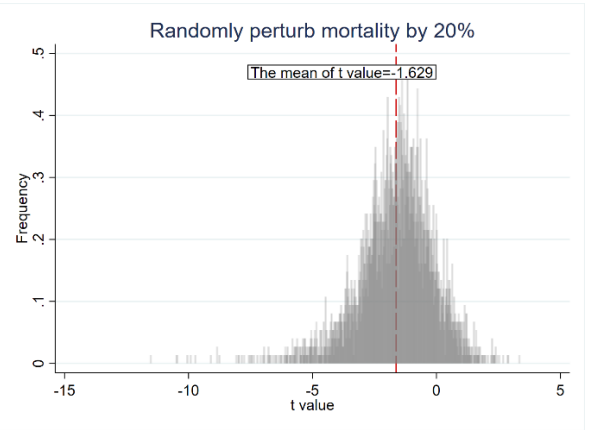
(a) (b)





(c) (d)





Figures 6(a) - (c) confirm that our baseline estimates are firmly robust when altering one third of the mortality data randomly by between 5 and 15 percent. The mean results of a random perturbation on morality by 5 and 10 percent are both greater than 2.5 (Figures 6(a) - (b)). As Figure 6(c) shows, the mean of the distribution of t-value is greater than 2 even after changing the mortality rate by 15 percent. According to the simulations, our results will only become invalid when the mortality data is manipulated by 20 percent. Figure 6(d) shows, the average of the distribution of the t-value is 1.63 when altering city mortality by 20 percent. However, it is unlikely that the mortality data for one-third of the cities included in the sample have errors greater than 20 percent, given the mean and standard deviation of mortality in the Black Death. Therefore, this second set of placebo tests further confirms that our results are reliable.

Similar experiments on the relationship between local institutions and the duration of the Black Death are conducted in Figure A3 (in appendix), and we have also found that our baseline results are credible.

Figures 6(a) - (c) confirm that our baseline estimates are firmly robust when altering one third of the mortality data randomly by between 5 and 15 percent. The mean results of a random perturbation on morality by 5 and 10 percent are both greater than 2.5 (Figures 6(a) - (b)). As Figure 6(c) shows, the mean of the distribution of t-value is greater than 2 even after changing the mortality rate by 15 percent. According to the simulations, our results will only become invalid when the mortality data is manipulated by 20 percent. Figure 6(d) shows, the average of the distribution of the t-value is 1.63 when altering city mortality by 20 percent. However, it is unlikely that the mortality data for one-third of the cities included in the sample have errors greater than 20 percent, given the mean and standard deviation of mortality in the Black Death. Therefore, this second set of placebo tests further confirms that our results are reliable.

Similar experiments on the relationship between local institutions and the duration of the Black Death are conducted in Figure A3 (in appendix), and we have also found that our baseline results are credible.

4.3.2 Robustness checks

Concerns that our baseline results could be driven by bias and errors may not be fully solved by the above placebo tests. To further verify our results, we implement several robustness checks to resolve the concerns from the perspective of data measurement, sample selection, and model design.

Firstly, there are two slightly different criteria being used currently to measure autonomous cities. Our current measurement of city autonomy so far employed was collected by Bosker et al. (2013). In this study, cities were considered as autonomous if there was an indication of the presence of a local urban participative organization that made decisions about local urban affairs. Stasavage (2014) provides an alternative indicator, using a more restrictive definition of autonomy than Bosker et al. (2013). Therefore, the concern is that our significant results could be false when we use a more stringent definition of urban autonomy.

To relieve this concern, Table 5 replicates both OLS and 2SLS estimates of Tables 3 and 4 by replacing the variable of interest, autonomous city, with Stasavage's (2014) indicator. The estimated coefficient of city autonomy remains negative and significant across all different OLS and 2SLS specifications. The magnitude of the 2SLS coefficient of the autonomous city is larger when using Stasavage's (2014) measure. The coefficients for the control variables, by and large, do not change.

Table 5: Robustness check of the measure of city autonomy

	Black death mortality rate (percent, 1347-1352)					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	OLS	OLS	2SLS	2SLS	2SLS
Autonomous city	-5.0368**	-4.4568***	-4.0081***	-13.8824**	-14.9301***	-10.1545***
	(1.8674)	(1.1838)	(0.5651)	(6.5676)	(4.6051)	(2.2042)
River		-2.7820	-1.3861		-0.8676	-0.4797
		(3.5536)	(3.3497)		(2.4577)	(2.7031)
Sea		-0.7974	1.4926		-1.2613	0.9239
		(3.0169)	(2.4686)		(2.9523)	(2.3525)
Hub roman road		-3.8534*	-2.5212		-2.4176	-2.1250
		(1.9946)	(2.9784)		(2.3392)	(3.0228)
Roman road		-0.8405	-0.1232		2.4302**	1.5250
		(2.2279)	(2.0609)		(1.2269)	(1.9406)
Soil quality		6.6294	9.7097		9.2953	11.5606*
		(6.6541)	(5.8718)		(6.2422)	(6.0379)
Elevation		-2.0814**	-2.1896**		-2.0711***	-2.1908***
		(0.7457)	(0.8573)		(0.6795)	(0.7413)
Population			-2.6179***			-2.4033***
			(0.4469)			(0.3509)
Capital			-1.4888			-0.6256
			(4.0361)			(3.7141)
Parliament			3.2076			4.1530
			(4.2674)			(3.0223)
Bishop			1.2892			2.6042
			(2.0680)			(1.9680)
Archbishop			2.6840			3.3629
			(5.1796)			(4.8453)
University			-0.2098			-0.2512
			(2.9098)			(1.4790)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	162	162	162	162	162	162
R-squared	0.2666	0.3228	0.3687	0.1991	0.2377	0.3420
First stage F				49.8	21.2	12.8

Robust standard errors adjusted for clustering at the country level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The second issue is that the population dataset from Bairoch (1988) only includes cities with more than 1,000 inhabitants. For 14 cities, which were smaller than 1,000 inhabitants, we relied on the population information provided by Christakos et al. (2005). There might be a concern that our results are driven by the inclusion of these small cities. By

ruling out cities that were not included in Bairoch's (1988) urban population records, Table 6 shows that excluding these cities which were less than 1000 people indeed reduce the magnitude of our coefficients. The coefficient of 2SLS estimation decrease from 9.29 to 6.03. However, the negative relationship between the city autonomy and Black Death mortality rates still holds and the results are not biased towards small cities.

Table 6: Robustness check of ruling out small cities

	Black death mortality rate (percent, 1347-1352)					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	OLS	OLS	2SLS	2SLS	2SLS
Autonomous city	-6.2666**	-6.6982***	-6.3029***	-16.6175***	-8.5776***	-6.0257**
	(2.3523)	(1.1259)	(0.9851)	(4.4059)	(2.8146)	(2.5797)
River		-1.9550	-1.6520		-1.6823	-1.6831
		(3.2170)	(2.7570)		(2.7615)	(2.4933)
Sea		0.2185	1.7892		0.2212	1.8130
		(3.3485)	(2.8923)		(3.0228)	(2.5112)
Hub roman road		-2.0202	-2.1487		-1.7699	-2.1601
		(2.2691)	(3.2965)		(1.9251)	(2.7654)
Roman road		1.8357	2.0992		2.3957*	2.0340
		(1.6431)	(1.8190)		(1.3589)	(1.7586)
Soil quality		9.7289*	14.0955**		10.6314**	13.9735***
		(4.7881)	(4.7473)		(5.1689)	(5.0634)
Elevation		-2.3483**	-2.3186**		-2.3446***	-2.3157***
		(0.8621)	(1.0307)		(0.7624)	(0.8871)
Population			-2.7916***			-2.8095***
			(0.4833)			(0.3871)
Capital			-0.7178			-0.7800
			(4.4259)			(3.6542)
Parliament			2.4999			2.4787
			(3.1898)			(2.7262)
Bishop			2.2978			2.2346
			(2.2088)			(2.1071)
Archbishop			4.2004			4.1609
			(4.8487)			(4.6225)
University			1.1840			1.1618
			(2.7566)			(2.5823)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	148	148	148	148	148	148
R-squared	0.2661	0.3359	0.3663	0.0304	0.3335	0.3663
First stage F				76.3	55.6	38.5

Robust standard errors adjusted for clustering at the country level are given in parentheses

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

Furthermore, it is also possible that our findings are affected by a lack of weather data from the medieval era. For example, by drawing on the case of Azerbaijan, Morris et al. (2013) find the climate and plagues are likely to be historically related. To reduce this concern, we replace the contemporary country fixed effect with the historical sovereign state fixed effect. Compared to modern European countries, there are more sovereign states in the Middle Ages. Thus, replacing the current country fixed effect with the historical spatial fixed effects could help us better control for weather-related effects . Table 7 shows that, although the magnitude of our variable of interest has decreased to some extent, the results are still valid when using historical boundaries as the fixed effect.

In addition to the impact of weather, the specific season of the initial plague outbreak may also affect our estimates. We have further collected the data related to the starting dates of plague released by Christakos et al. (2005). In Table A1, we find that the reported mortality rates in cities where the plague hit dropped significantly in the summer, when compared with rates in the spring. More importantly, adding the season dummies confirms our 2SLS results and further enlarges the magnitude of the coefficient of autonomous cites by more than 7 percent compared to 2SLS results in the baseline.

Table 7: Robustness check by using the boundary in the 14th century

	Black death mortality rate (percent, 1347-1352)					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	OLS	OLS	22SLS	2SLS	2SLS
Autonomous city	-4.4949**	-5.2393**	-3.1416*	-4.7561***	-5.8014***	-3.0388*
	(1.5309)	(1.7070)	(1.6739)	(1.5933)	(1.7132)	(1.6875)
River		0.4937	1.2626		0.5810	1.2535
		(2.7741)	(3.0736)		(2.2722)	(2.5345)
Sea		1.8597	2.3713		1.9928	2.3545
		(2.6710)	(2.6173)		(2.0731)	(1.8751)
Hub roman road		-1.1269	0.1372		-1.0629	0.1375
		(3.9400)	(4.8908)		(3.1202)	(3.7703)
Roman road		-0.3340	0.3096		-0.2460	0.2988
		(3.0486)	(2.9872)		(2.2730)	(2.2049)
Soil quality		13.7428	13.1637		13.8268**	13.1451*
		(7.7321)	(9.1442)		(6.3222)	(7.2614)
Elevation		-1.0075	-1.5040		-1.0184	-1.5022
		(1.2182)	(1.3416)		(0.9940)	(1.0673)
Population			-3.8821***			-3.8880***
			(1.0604)			(0.7493)
Capital			-1.0864			-1.1052
			(4.3550)			(3.4661)
Parliament			2.1280			2.1410
			(3.1664)			(2.7543)
Bishop			1.4125			1.3892

			(2.2477)			(1.7791)
Archbishop			5.1373			5.1222
			(5.9267)			(4.5941)
University			-3.9977			-4.0077
			(4.8697)			(3.6105)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	162	162	162	162	162	162
R-squared	0.4233	0.4510	0.5162	0.4233	0.4508	0.5162
First stage F				16	16.3	15.6

Robust standard errors adjusted for clustering at the country level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5 Conclusion

By proposing that both cities and their citizens are vulnerable to the short-term consequences of pandemics, a number of economists have considered the variations in mortality rates from infectious diseases to be largely random (Almond, 2006; Donaldson & Keniston, 2016). Hence, it has often been the case that reported differences in the Black Death incidence have been attributed mainly to chance and not to institutional or any other factors that may have influenced the preparation and reaction of cities in medieval Europe, when confronted with a plague of this dimension. However, were geographic and institutional factors really irrelevant in determining the differential incidences of the Black Death across cities in Europe?

In this paper we have provided empirical evidence of the links between local institutions at the time of the Black Death and the variation of the incidences of the pandemic across Western Europe. We have shown that city autonomy was at the centre of differences in pandemic-related mortality across cities. Comparing to their counterparts, autonomous cities performed better in the face of one of the most devastating pandemics the Western world has ever faced. Specifically, cities with a high degree of autonomy reduced — depending on estimations— Black Death mortality rates by between 5.22 and 9.29 percent. A large set of placebo tests and robustness checks confirms the robustness of this result.

City autonomy in medieval times was, moreover, a far stronger tool in mitigating the effects of the plague than nearly all other geographical and political characteristics of cities at the time considered. Being a capital city, hosting a parliament, or having a bishop or an archbishop at the beginning of the 14th century did not reduce the incidence of the plague. In other words, among various political factors at the time, only city autonomy and the powers it granted local citizens have provided some protection against the deadliest of pandemics. In our analysis, we also find that larger cities performed better when confronting the pandemics in the case of Black Death. Finally, the elevation

of the city was the only physical geographic factor linked to a lower Black Death mortality rate.

While city autonomy influenced reported mortality rates, it did not affect the lengths of the pandemic in different cities. None of the other controls were connected to a reduction in the duration of the Black Death wave in a given city. Hence, it seems that the factors behind the duration of the pandemic in different cities are relatively random in contrast to the mortality rate results.

There are, of course, several limitations to the research presented here. Firstly, in this paper, we refer to the autonomous city as if autonomy is a binary indicator. However, in practice, it is important to realize that the situations were far more complex. As Stasavage (2014) argues, autonomy was certainly a question of degree. Cities like Venice had complete autonomy over almost all government affairs. In contrast, cities like Ghent enjoyed a substantial degree of autonomy over certain types of affairs, but were still subject to princely intervention. However, due to the limited number of historical records, the full extent of the influence of a city's autonomy on the pandemic cannot yet be analysed. The second problem is understanding the mechanisms through which an autonomous institution helped limit the incidence rates of the Black Death. Further research is needed in this respect. Thirdly, due to data limitations, we only focus our research on the interaction between local institutions and the initial outbreak of the plague. It should be noted that the plague did not end in 1352 and that numerous European cities had recurrent outbreaks in the following decades and centuries. Studying the influence of city autonomy across recurrent outbreaks and the overall impact of the Black Death over time would be very interesting, but requires a substantial improvement in data availability.

This paper suggests, in line with Rodríguez-Pose and Burlina (2021), that good local government institutions play an essential role in a pandemic, even in medieval times when the objective conditions to fight disease were far weaker than today. When confronted with the challenge of what is possibly the worst pandemic ever recorded, autonomous city governments were more effective in adopting better prevention measures. Thus, improving the quality of local institutions can be a good way of fighting pandemics. It helped self-governing cities in medieval Europe and it could also deliver better results today.

References

Acemoglu, D., Johnson, S., & Robinson, J. (2005). The rise of Europe: Atlantic trade, institutional change, and economic growth. *American economic review*, *95*(3), 546-579.

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, *127*(1), 47-100.

Alfani, G., & Murphy, T. E. (2017). Plague and lethal epidemics in the pre-industrial world. *the Journal of Economic History*, *77*(1), 314-343.

Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population. *Journal of political Economy*, *114*(4), 672-712.

Alsayyad, N., & Roy, A. (2006). Medieval modernity: On citizenship and urbanism in a global era. *Space & Polity*, *10*(1), 1-20.

Atkeson, A. (2020). *What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios* (No. w26867). National Bureau of Economic Research.

Bairoch, P. (1988). *Cities and economic development: from the dawn of history to the present*. University of Chicago Press.

Benedictow, O. J. (2004). *The Black Death, 1346-1353: the complete history*. Boydell & Brewer.

Belloc, M., Drago, F., & Galbiati, R. (2016). Earthquakes, religion, and transition to self-government in Italian cities. *The Quarterly Journal of Economics*, *131*(4), 1875-1926.

Bosker, M., Buringh, E., & Van Zanden, J. L. (2013). From Baghdad to London: Unraveling urban development in Europe, the Middle East, and north Africa, 800–1800. *Review of Economics and Statistics*, *95*(4), 1418-1437.

Bowles, J., Hjort, J., Melvin, T., & Werker, E. (2016). Ebola, jobs and economic activity in Liberia. *Journal of Epidemiology & Community Health 70*(3), 271-277.

Caferro, W. (2018). *Petrarch's War: Florence and the Black Death in Context*. Cambridge University Press.

Cantoni, D., & Yuchtman, N. (2014). Medieval universities, legal institutions, and the commercial revolution. *The Quarterly Journal of Economics*, *129*(2), 823-887.

Carozzi, F., Provenzano, S., & Roth, S. (2020). Urban Density and COVID-19. *IZA Discussion Paper,* 13440.

Christakos, G., Olea, R. A., Serre, M. L., Wang, L. L., & Yu, H. L. (2005). *Interdisciplinary public health reasoning and epidemic modelling: the case of black death* (pp. 110-14). New York: Springer.

CitiesX. (2018, Jan 26). *Venetian Responses to Plague*. [Video]. YouTube. https://www.youtube.com/watch?v=TTM-WfSBNkQ&t=2s

Clark, G. (2008). *A farewell to alms: a brief economic history of the world* (Vol. 25). Princeton University Press.

Cohn, Jr, S. K., & Alfani, G. (2007). Households and plague in early modern Italy. *Journal of Interdisciplinary History*, *38*(2), 177-205.

Cox, G. W. (2017). Political institutions, economic liberty, and the great divergence. *The Journal of Economic History*, *77*(3), 724-755.

Curtis, D. R., & Roosen, J. (2017). The sex-selective impact of the Black Death and recurring plagues in the Southern Netherlands, 1349–1450. *American Journal of Physical Anthropology*, 164(2), 246-259.

De Long, J. B., & Shleifer, A. (1993). Princes and merchants: European city growth before the industrial revolution. *The Journal of Law and Economics*, *36*(2), 671-702.

Demeure, C. E., Dussurget, O., Fiol, G. M., Le Guern, A. S., Savin, C., & Pizarro-Cerdá, J. (2019). Yersinia pestis and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. *Genes & Immunity*, 20(5), 357-370.

DeWitte, S. N., & Wood, J. W. (2008). Selectivity of Black Death mortality with respect to preexisting health. *Proceedings of the National Academy of Sciences*, 105(5), 1436-1441.

DeWitte, S. N., & Hughes-Morey, G. (2012). Stature and frailty during the Black Death: the effect of stature on risks of epidemic mortality in London, AD 1348–1350. *Journal of Archaeological Science*, 39(5), 1412-1419.

DeWitte, S. N. (2015). Setting the stage for medieval plague: Pre-black death trends in survival and mortality. American Journal of Physical Anthropology, 158(3), 441-451.

Desjardins, M. R., Hohl, A., & Delmelle, E. M. (2020). Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography*, 102202.

Diamond, J. (2013). *Guns, germs and steel: a short history of everybody for the last 13,000 years*. Random House.

Dittmar, J. E., & Meisenzahl, R. R. (2020). Public goods institutions, human capital, and growth: Evidence from German history. *The Review of Economic Studies*, *87*(2), 959-996.

Dols, M. W. (1979). The second plague pandemic and its recurrences in the Middle East: 1347-1894. *Journal of the Economic and Social History of the Orient/Journal de l'Histoire Économique et Sociale de l'Orient*, 162-189.

Donaldson, D., & Keniston, D. (2016). Dynamics of a Malthusian Economy: India in the Aftermath of the 1918 Influenza. *Unpublished Manuscript*.

Earn, D. J., Ma, J., Poinar, H., Dushoff, J., & Bolker, B. M. (2020). Acceleration of plague outbreaks in the second pandemic. *Proceedings of the National Academy of Sciences*, *117*(44), 27703-27711.

Epstein, S. R. (2000). *Freedom and growth: the rise of states and markets in Europe, 1300-1750* (Vol. 17). Routledge.

Frank, A. G. (1979). *Dependent accumulation* (Vol. 492). NYU Press.

Hajar, R. (2012). The air of history (part II) medicine in the Middle Ages. *Heart views: the official journal of the Gulf Heart Association*, 13(4), 158.

Geltner, G. (2020). The Path to Pistoia: Urban Hygiene Before the Black Death. *Past & Present*, *246*(1), 3-33.

Glaeser, E. (2011). *Triumph of the City*. Pan.

Godde, K., Pasillas, V., & Sanchez, A. (2020). Survival analysis of the Black Death: Social inequality of women and the perils of life and death in Medieval London. *American Journal of Physical Anthropology*, 173(1), 168-178.

Godde, K., & Hens, S. M. (2021). An epidemiological approach to the analysis of cribra orbitalia as an indicator of health status and mortality in medieval and post-medieval London under a model of parasitic infection. *American journal of physical anthropology*, 174(4), 631-645.

Gottfried, R. S. (2010). *Black death*. Simon and Schuster.

Guiso, L., Sapienza, P., & Zingales, L. (2016). Long-term persistence. *Journal of the European Economic Association*, *14*(6), 1401-1436.

Heinonen-Tanski, H., & van Wijk-Sijbesma, C. (2005). Human excreta for plant production. *Bioresource technology*, *96*(4), 403-411.

Herlihy, D. (1997). *The Black Death and the transformation of the West*. Harvard University Press.

Hohenberg, P. M., & Lees, L. H. (1995). The Making of Urban Europe, 1000–1994: *With a New Preface and a New Chapter*. Harvard University Press.

Jackson, P. (2018). *The Mongols and the Islamic World*. Yale University Press.

Jacob, M. (2010). Long-term persistence: The free and imperial city experience in Germany. *Available at SSRN 1616973*.

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4. *available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org)*, *15*, 25-54.

Jedwab, R., Johnson, N. D., & Koyama, M. (2019). Negative shocks and mass persecutions: evidence from the Black Death. *Journal of Economic Growth*, *24*(4), 345-395.

Jedwab, R., Johnson, N. D., & Koyama, M. (2019). Pandemics, places, and populations: evidence from the Black Death.

Jones, E. (2003). *The European miracle: environments, economies and geopolitics in the history of Europe and Asia*. Cambridge University Press.

McEvedy, C., & Jones, R. (1978). *Atlas of world population history*. Penguin Books Ltd, Harmondsworth, Middlesex, England.

Mitchell, P. D. (2015). Human parasites in medieval Europe: lifestyle, sanitation and medical treatment. *Advances in parasitology*, *90*, 389-420.

Morris, L. R., Blackburn, J. K., Talibzade, A., Kracalik, I., Ismaylova, R., & Abdullahyev, R. (2013). Informing surveillance for the lowland plague focus in Azerbaijan using a historic dataset. *Applied Geography*, *45*, 269-279.

Nishiura, H. (2006). Epidemiology of a primary pneumonic plague in Kantoshu, Manchuria, from 1910 to 1911: statistical analysis of individual records collected by the Japanese Empire. *International journal of epidemiology*, 35(4), 1059-1065.

North, D. C. (1981). *Structure and change in economic history*. Norton.

North, D. C., & Thomas, R. P. (1973). *The rise of the western world: A new economic history*. Cambridge University Press.

North, D. C., & Weingast, B. R. (1989). Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England. *The journal of economic history*, *49*(4), 803-832.

Nussli, Christos, (2011). URL: http://www.euratlas.com/about.html.

North, D. C. (1978). Structure and performance: The task of economic history. *Journal of Economic literature*, *16*(3), 963-978.

Ormrod, M. (1996). The politics of pestilence. Government in England after the Black Death. *PAUL WATKINS MEDIEVAL STUDIES*, 147-181.

Palmer, R. C. (2000). *English Law in the Age of the Black Death, 1348-1381: A Transformation of Governance and Law*. Univ of North Carolina Press.

Persson, T., & Tabellini, G. (2009). Democratic capital: The nexus of political and economic change. *American Economic Journal: Macroeconomics*, *1*(2), 88-126.

Qiu, Wuqi, Cordia Chu, Ayan Mao, and Jing Wu. "The impacts on health, society, and economy of SARS and H7N9 outbreaks in China: a case comparison study." *Journal of environmental and public health* 2018 (2018).

Ramankutty, N., Foley, J. A., Norman, J., & McSweeney, K. (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Global Ecology and biogeography*, *11*(5), 377-392.

Randremanana, R., Andrianaivoarimanana, V., Nikolay, B., Ramasindrazana, B., Paireau, J., Ten Bosch, Q. A., ... & Rajerison, M. (2019). Epidemiological characteristics of an urban plague epidemic in Madagascar, August–November, 2017: an outbreak report. *The Lancet Infectious Diseases*, 19(5), 537-545.

Rodríguez-Pose, A., & Burlina, C. (2021). Institutions and the uneven geography of the first wave of the Covid-19 pandemic. *Journal of Regional Science*, *61*, 728-752.

Roosen, J. (2017). Severity and Selectivity of the Black Death and Recurring Plague in the Southern Netherlands (1349-1450). *TSEG-The Low Countries Journal of Social and Economic History*, *14*(4), 25-55.

Salzman, J. (2017). *Drinking Water: A History (Revised Edition)*. Abrams.

Stasavage, D. (2014). Was Weber right? The role of urban autonomy in Europe's rise. *American Political Science Review*, *108*(2), 337-354.

Theilmann, J., & Cate, F. (2007). A plague of plagues: The problem of plague diagnosis in medieval England. *Journal of Interdisciplinary History*, *37*(3), 371-393.

Van Zanden, J. L., Buringh, E., & Bosker, M. (2012). The rise and decline of European parliaments, 1188–1789 1. *The Economic History Review*, *65*(3), 835-861.

Weber, Max. (1958). *The City*, New York

Weil, D. N. (2010). Endemic diseases and African economic growth: Challenges and policy responses. *Journal of African Economies*, *19*(suppl_3), iii81-iii109.

Young, A. (2005). The gift of the dying: The tragedy of AIDS and the welfare of future African generations. *The Quarterly Journal of Economics*, *120*(2), 423-466.

Yue, R. P., Lee, H. F., & Wu, C. Y. (2017). Trade routes and plague transmission in pre-industrial Europe. *Scientific reports*, *7*(1), 1-10.

Ziegler, P. (2013). *The black death*. Faber & Faber.

Appendix A

Figure A1: The urban population in 1300

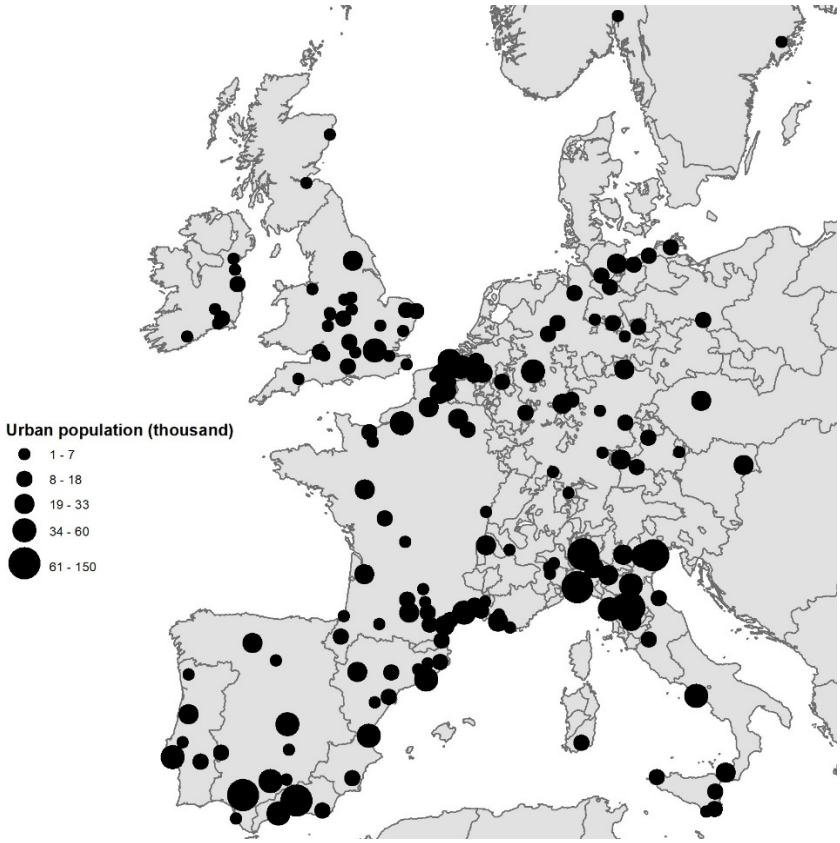


Figure A2: Mortality rate and duration of the Black Death

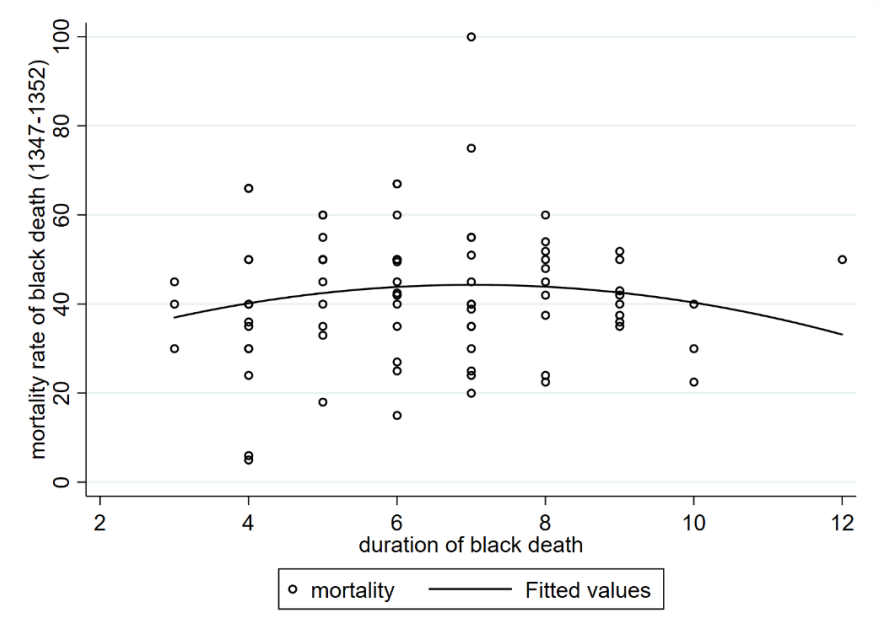


Figure A3 Distribution of t-statistics by randomly assigned duration by certain months in one third of cities

(a) (b)

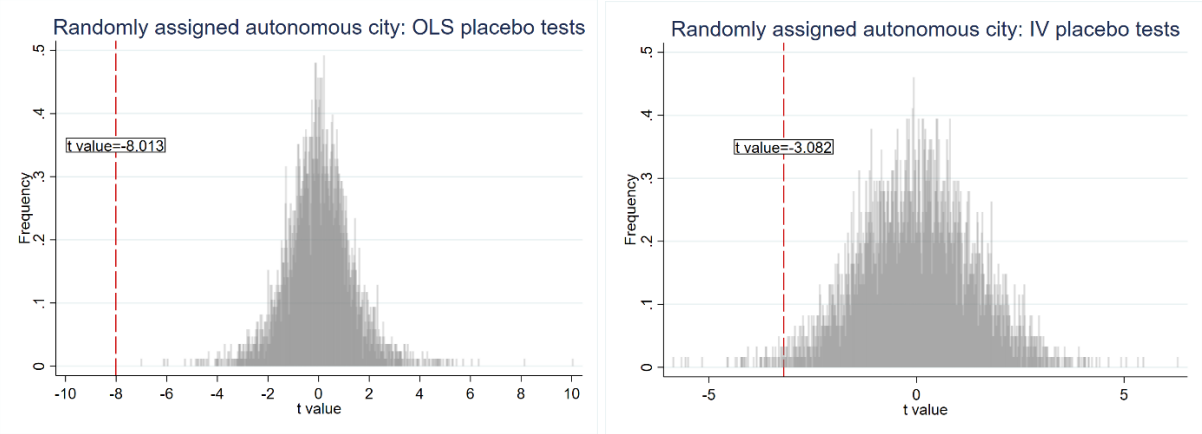
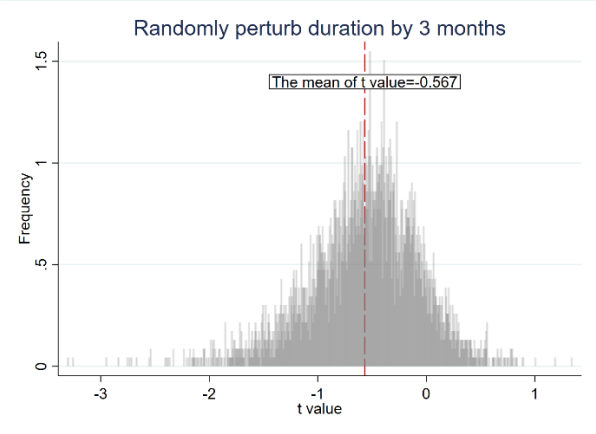
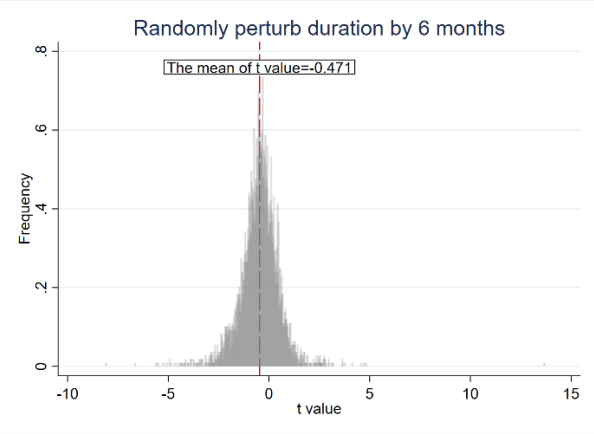


Figure A4 Distribution of t-statistics by randomly assigned duration by certain months in one third of cities

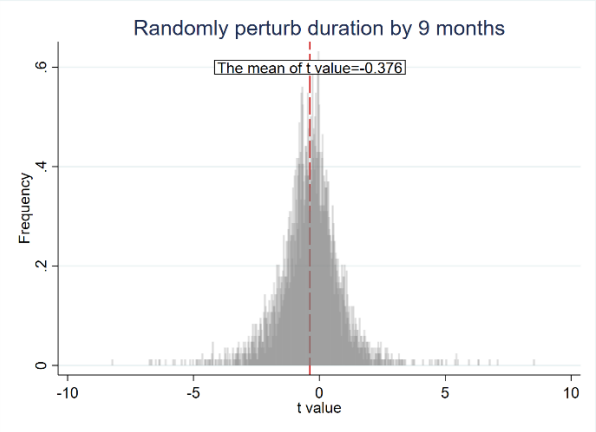
(a)



(b)



(c)



(d)

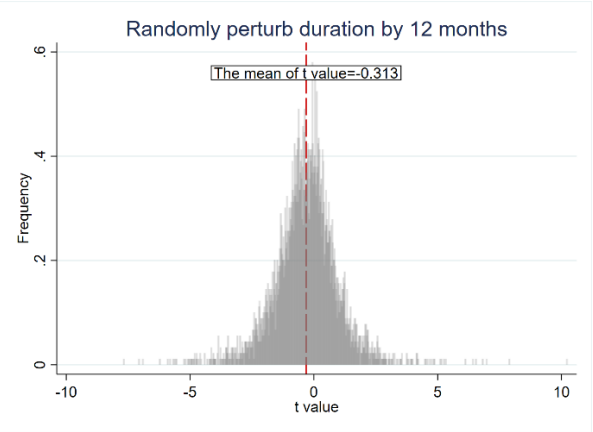


Table A1: Robustness check by adding season of the initial outbreak

Black death mortality rate (percent, 1347-1352)						
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	OLS	OLS	2SLS	2SLS	2SLS
Autonomous city	-5.3303	-5.0968*	-4.5161	-12.2863***	-13.5922*	-16.7747**
	(3.3068)	(2.5335)	(2.8691)	(4.1114)	(7.1185)	(7.8730)
Summer	-7.5837**	-6.1900**	-6.2956**	-8.0935***	-6.8651***	-7.1305***
	(3.1355)	(2.3664)	(2.5870)	(3.0189)	(2.2549)	(2.2924)
Autumn	-1.0766	-0.2566	-1.7139	-2.3161	-1.4018	-2.6956
	(3.9265)	(2.8164)	(2.5837)	(3.9219)	(3.1447)	(1.9727)
Winter	-3.3282	-4.1989	-4.6306	-3.7396	-4.7596	-5.7260
	(4.5067)	(4.6288)	(5.2878)	(3.9486)	(3.9220)	(4.2611)
River		-1.1406	-0.6715		1.1184	1.6921
		(3.7896)	(3.8836)		(4.5414)	(3.5225)
Sea		-1.9536	-1.5753		-1.4385	-2.9276
		(4.3063)	(4.6870)		(4.2226)	(3.5993)
Hub roman road		-2.3389	-1.8430		-1.5634	-1.7804
		(2.8702)	(3.3297)		(2.7349)	(3.4216)
Roman road		0.4484	1.1231		2.7157	3.0008
		(3.2099)	(3.0513)		(3.4343)	(3.1639)
Soil quality		7.6065	9.9916		10.8757	13.7428
		(8.9627)	(9.8860)		(9.9293)	(9.6359)
Elevation		-2.5756**	-2.9488**		-2.3935**	-2.8457***
		(0.9485)	(1.1018)		(0.9697)	(1.0488)
Population			-1.9504			-0.6566
			(1.3929)			(1.4666)
Capital			1.8877			5.3458
			(6.6008)			(3.9841)
Parliament			1.2865			1.9927
			(4.2621)			(2.2631)
Bishop			2.0795			3.1252
			(1.6145)			(2.3200)
Archbishop			3.9595			3.8957
			(4.8428)			(3.7960)
University			-2.3045			-2.8454**
			(3.7351)			(1.2865)
country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	122	122	122	122	122	122
R-squared	0.2743	0.3444	0.3684	0.2268	0.2820	0.2552
First stage F				24.5	38.8	14.7

Robust standard errors adjusted for clustering at the country level are given in parentheses

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

Table A2: Robustness check by adding season of the initial outbreak

Duration of Black Death (month, 1347-1352)						
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	OLS	OLS	2SLS	2SLS	2SLS
Autonomous city	-0.1688	-0.1567	-0.5133	-0.5855	-0.8549	-4.9226
	(0.4140)	(0.6582)	(0.7395)	(0.8614)	(1.2105)	(7.0475)
Summer	0.1242	-0.0233	-0.2837	0.0991	-0.0759	-0.5450
	(0.5776)	(0.6414)	(0.5842)	(0.4428)	(0.4951)	(0.5671)
Autumn	-0.2144	-0.1140	0.0119	-0.2497	-0.1555	-0.0703
	(0.2193)	(0.1784)	(0.3280)	(0.2455)	(0.1651)	(0.7642)
Winter	0.6553*	0.7356*	0.5649	0.6672**	0.7253**	0.4458
	(0.3202)	(0.3397)	(0.7959)	(0.3202)	(0.3664)	(0.9086)
River		0.1598	0.1420		0.2112	0.2570
		(0.8594)	(0.8114)		(0.7026)	(0.5425)
Sea		0.2301	-0.1022		0.1380	-1.5944
		(0.7077)	(0.7532)		(0.5319)	(1.4641)
Hub roman road		0.9577	0.6561		0.9971	0.3855
		(0.8854)	(1.0943)		(0.7269)	(1.0730)
Roman road		0.1851	0.2597		0.3867	0.9216
		(0.3302)	(0.2301)		(0.5199)	(0.9557)
Soil quality		0.3021	0.3647		0.6848	2.2471
		(1.1392)	(0.9776)		(1.3887)	(3.0007)
Elevation		0.1539	0.1583		0.1508*	0.0197
		(0.0970)	(0.1778)		(0.0820)	(0.3748)
Population			0.2594			0.8014
			(0.2584)			(1.1283)
Capital			0.4017			2.0896
			(0.5926)			(2.8899)
Parliament			0.3806			0.2912
			(0.2655)			(0.5096)
Bishop			0.0841			0.3866
			(0.5320)			(0.9626)
Archbishop			1.1912			1.0163***
			(0.8892)			(0.3541)
University			-0.1716			-0.4659
			(0.2567)			(1.3095)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76	76	76	76	76	76
R-squared	0.1262	0.1802	0.2592	0.1167	0.1566	-0.5150
First stage F				2.85	2.94	0.34

Robust standard errors adjusted for clustering at the country level are given in parentheses

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

Table A3: The relationship between city size and duration

Duration of Black Death (month, 1347-1352)		
	(1)	(2)
VARIABLES	duration	duration
Population	0.340**	0.297
	(0.170)	(0.177)
Autonomous city		-0.392
		(0.669)
River		-0.0307
		(0.757)
Sea		-0.114
		(0.690)
Hub roman road		0.391
		(0.925)
Roman road		0.174
		(0.147)
Soil quality		0.559
		(1.414)
Elevation		0.0906
		(0.0633)
Capital		0.333
		(0.399)
Parliament		0.353
		(0.301)
Bishop		0.306
		(0.606)
Archbishop		1.443*
		(0.764)
University		0.168
		(0.354)
Country FE	No	Yes
Observations	83	83
R-squared	0.047	0.250

Robust standard errors adjusted for clustering at the country level are given in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix B

Table B1: Variable descriptions and sources

Variable	Description	Source
Outcome variables		
Mortality rate	mortality rate of the Black Death from 1347 to 1352	Christakos et al. (2005)
Duration	duration of the Black Death from 1347 to 1352	Christakos et al. (2005)
Explanatory variables		
Autonomous city	whether city had institution for self-governance	Bosker et al. (2013); Stasavage (2014)
River	whether city is within 10km from a river	Nussli (2011)
Sea	whether city is within 10km from a coast	Nussli (2011)
Hub roman road	whether city locates on the hub of Roman road	Bosker et al. (2013)
Roman road	whether city locates on a former Roman road	Bosker et al. (2013)
Soil quality	The probability that certain location could be cultivated	Ramankutty et al. (2002)
Elevation	elevation	Jarvis et al (2008)
Population (logged)	urban population in 1300	Bairoch (1988); Christakos et al. (2005)
Capital	whether city was a capital in 1300	McEvedy & Jones (1978).
Parliament	whether city had representatives in an active parliament in 1300	(Van Zanden et al., 2012)
Bishop	whether city had the presence of the bishop in 1300	Bosker et al. (2013)
Archbishop	whether city had the presence of the archbishop in 1300	Bosker et al. (2013)
University	whether city had the presence of the university in 1300	Bosker et al. (2013)

Chapter 3: Inclusive institutions, entrepreneurship and innovation in China

1 Introduction

Why do some regions develop while others lag behind? One of the dominant explanations, dating back North (1990), has been that institutions - the 'rules of the game' in a society or, more formally, the humanly devised constraints that shape human interactions – play a crucial role. Such rule making is normally seen as happening through government action, and current research has quantified institutions through a general government quality index (Rodríguez-Pose & Di Cataldo, 2015; Wang et al., 2021) or specific political arrangements such as decentralisation (Dell et al., 2018), parliament (Acemoglu & Robinson, 2012), colonialism (Baruah et al., 2021), etc. However, societies have often lacked public authorities able to enforce institutional rules governing economic activities. Instead, private order institutions – those formed through voluntary collective action by private agents – stepped in and took the place of defective governments in enabling the proper functioning of the market (Greif, 2006; North & Thomas, 1972).

This paper considers how private order institutions influence the geography of entrepreneurship and innovation in China. We categorise chambers of commerce, guilds, private judges and serfdom as private-order institutions that play the dominant role when the public sector does not. We build on the insight, developed by Acemoglu and Robinson (2012), amongst others, that institutions can be either inclusive or exclusive. The former is likely to focus on networking and opportunity creation; the latter, on rent extraction and protection of existing firms. Conceptually, the inclusiveness of the institution promotes the economic participation of a large segment of the population, encourages individuals to make the best use of their skills and choose their own jobs, maintains impartial public contracting authorities, and enables the entry of new businesses (Acemoglu & Robinson, 2012). According to the argument, inclusive private-order institutions, whose rules apply uniformly to all people regardless of gender, religion, race or social class, should persistently create a friendly entrepreneurial environment. In contrast, exclusive private-order institutions, whose application is differentiated for different subsets of economic agents, tend to be negative for development (Hillmann, 2013; Ogilvie & Carus, 2014).

China is a particularly interesting case for this study: the fall and rise of the market economy in the second half of the 20th century provide us with a unique empirical setting to examine the persistence of this economic institution. After several rounds of

collectivisation movements such as the Joint public-private management and the Cultural Revolution, private enterprise was severely suppressed in China from the 1950s to the 1970s (Teiwes, 1987).[15] It was not until 1980 that the identity of entrepreneurs was formally regained and private property rights were recognised,[16] allowing us to precisely examine the regional resilience of local institutions (Martin & Sunley, 2015). This structural break provides a discontinuity which allows us to separate out the effect of entrepreneurial persistence from year-to-year, with the effect of inclusive networks which is more likely to persist over time.

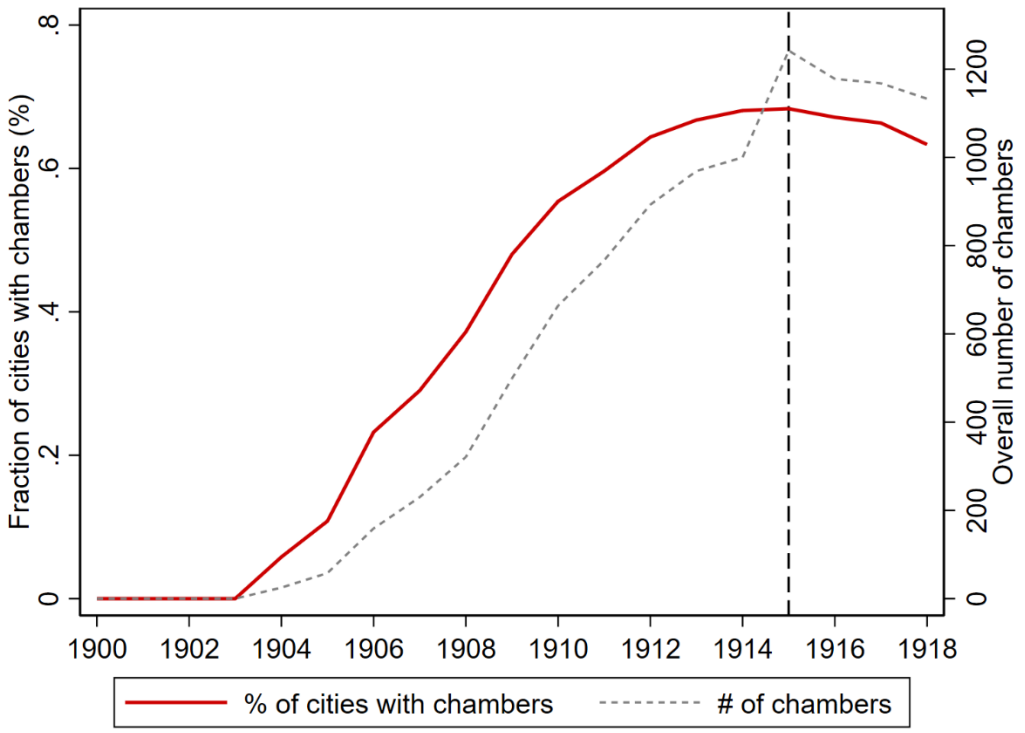
The inclusive institution we consider is the Chinese Chamber of Commerce. Unlike traditional sectoral or regional organisations, chambers accepted applications from all social classes and identity groups (Bennett, 1996). They rapidly evolved into community-oriented organisations beyond the framework of merchant guilds since their commencement in 1904 (Chen, 2011). Chamber membership was entirely voluntary and so participation rates varied considerably across cities (Maennig et al., 2015). After a decade of development, there were over a thousand chambers across China in the 1910s, and their total number had been stabilised nationwide in the 1910s (Figure 1).[17] This steady historical cross-city variation allows us to identify how institutions affect entrepreneurship and innovation at the city level.

[15] Figure A1 depicts the extent of constrain on the private businesses from 1950s to 1970s.

[16] In the period between Mao's death and the formal launch of market economy (1976-1979), subject to the great pressure from employment, private business was acquiesced by both central and regional government. This precondition allows us to examine pre-parallel trend assumption before the legalisation of private ownership in 1980.

[17] Figure A2 shows that chamber membership is also stable at the provincial level from 1912 and 1920.

Figure 1: The rollout of chambers of commerce in early 20th century China



To measure the chamber system, we draw on historical administrative records dating back to 1915 and manually collected information on the location of chambers, number of members, types of chambers and memberships, frequency of meetings, number of disputes resolved, and micro-level income and expenditure. To measure contemporary entrepreneurship outcomes, we use the Firm Registration Database, which tracks the footprint of all existing private business and then aggregates to our city-year-level panel from 1975 to 1990. We also collect patent data from the China National Intellectual Property Administration (CNIPA) to measure local innovation capacity in the 1980s. As far as we know, this is the first use of this data in empirical work. We then use a Difference-in-Difference estimation strategy to identify the impact of chamber system before and after the market reform.

We find firstly that inclusive institutions matter for entrepreneurship in the long run. When the market resumed, the chamber system guaranteed an open network of entrepreneurs, settling disputes within merchant groups and imparting knowledge of doing business fostered the creation of private business. Our results show that one standard deviation increase in logged chamber density translates into a 5.35 percentage point increase in the level of entrepreneurship between 1975 and 1990 at the city level. This does not extend to exclusive institutions and we find that the merchant guilds, whose

membership was limited to specific groups, have no impact. We further test the credibility of our baseline results by the placebo test replacing the outcome variable as the number of newly registered State-Owned Enterprises (SOE) birth. The invalidity of SOE creation further emphasises that the influence of the chamber system on entrepreneurship is exclusive.

Being an open-source network, the elite group was also allowed to join in the chamber system. Theories predict that institutions with a higher non-elite ratio subsequently underwent major institutional improvements. In contrast, the inclusiveness of institutions is likely to be eroded if organisations are dominated by elites (Acemoglu et al., 2011; Blum, 1978). We examine it by exploring the internal organisational structure of chambers. We find that chambers with a higher proportion of non-elites could dissolve the power of local elites and accelerate the process of entrepreneurship emergence, highlighting the importance of democratic participation and discussions (Ogilvie, 2021).

To further address endogeneity concerns, we use an instrumental variable based on the shortest distance from the city to the nearest telegraph station in 1903, one year before the expansion of chambers. Qualitative evidence shows that Chinese merchants in the early 20th century used the telegraph to transmit information (Baark, 1997). Telegraph construction facilitates communication between trading groups and promotes the expansion of local chambers. However, the location of the telegraph depended entirely on military needs rather than other socio-economic factors (Gao & Lei, 2021). This crucial fact helps us meet the exclusion restriction of our IV. Our two-stage least squares (2SLS) regressions produce estimates that are consistent with ordinary least squares (OLS) results.

The inclusive institution also established an innovation-friendly environment and increased the incentives to innovate. By further introducing the promotion of China's patent law in 1985, we find evidence that the chamber system significantly improves the overall innovation capacity of regions. In addition to our main results, we also analyse whether our findings differ across industries. Our estimates suggest that inclusive institutions are particularly pronounced for high-tech innovations, which were those associated with sectors such as chemistry and engineering (Donges et al., 2022). The development in these high-tech areas has provided the continued impetus for economic prosperity in the long run.

The remainder of this paper proceeds as follows. The next section presents a conceptual framework. Section 3 discusses the empirical settings and data sources. Section 4 presents the empirical results regarding the impact of the chamber system on entrepreneurship. Section 5 examines the corresponding innovative effects. Section 6 concludes the study.

2 Conceptual framework

The literature on institutions has tended to focus on formal, governmental institutions such as legal or government structure. However, voluntary private sector organisations and associations can replace low-capacity governments in enabling markets to function (Greif, 2006; North & Thomas, 1973). Throughout most of history, societies have lacked public authorities able and willing to enforce institutional rules governing economic activities. Substitute institutions, such as coalitions, networks, chambers, guilds and communities, have emerged to fill the gap. Norms and informal rules of behaviour are not limited to government institutions and can also develop through private sector organisations (Dixit, 2009; Helpman, 2009). In this paper, we follow Katz (1996, 2000) and define private institutions as those formed through the voluntary collective actions of private agents without the engagement of public authorities. With this conceptual clarification, we place chambers of commerce, guilds and private judges in the category of private institutions. Public institutions, in contrast, are those associated with the formal public authorities of society, such as local governments, bureaucracies, courts and parliaments. Our focus in this paper is on the characteristics and impact of private institutions.

The impact of private institutions on economic development is ambiguous. Some researchers have seen private institutions as agents of progress and technological innovation (Epstein & Prak, 2008; Lucassen et al., 2008). Yet private institutions are diverse, and they can have very different effects – to see this, consider the different effects of guilds or serfdom, both of which are private institutions of different forms. Instead, we use Acemoglu and Robinson's (2012) framework and distinguish between inclusive and extractive systems. Inclusive systems promote the economic participation of a large proportion of the population, motivate individuals to make the best use of their skills and choose their own jobs, allow people to make free choices, ensure secure private property and allow new businesses to enter. Extractive institutions, by contrast, are defined as those that are not inclusive. A small group of individuals exploits the rest of the population (Acemoglu & Robinson, 2012). Inclusive institutions enable more people to participate in political, economic, and societal discussions. They create good economic incentives for innovation and good political incentives for the rule of law, which in turn promote economic growth. Extractive institutions allow privileged groups to expropriate resources from the rest of society, undermining property rights and preventing innovation (Ogilvie & Carus, 2014). Based on this, we would expect inclusive private order institutions to be growth enhancing, and to aid entrepreneurship and innovation; exclusive ones will have a detrimental impact.

The current literature tends to portray the inclusiveness of institutions as a binary, with institutions categorised simply as either inclusive or exclusive (Acemoglu & Robinson, 2012; Donges et al., 2022; North et al., 2009). However, this ignores a richer history. In mediaeval Europe, for example, merchant guilds created greater inclusiveness than feudal regimes and craft guilds. However, urban guild regimes still excluded most of society. Merchant guilds used their involvement in politics to exclude craftsmen, and craft guilds used it to exclude groups below them (Ogilvie & Carus, 2014). In other words, there are no perfectly inclusive institutions. Even the historical states that came closest to the rule of law often allowed themselves obvious weaknesses and inconsistencies. Therefore, the inclusiveness of private-order institutions should not be treated as a dichotomous variable. It is best to think of inclusiveness as a continuum along which historical institutions are distributed. We argue that inclusiveness exists across the lens of the spectrum and is a matter of degree. If more people participate in these institutions, the greater their positive (or negative) effects will be.

Our focus in this study is the Chamber of Commerce. The first chambers originated in the 1760s in Jersey, New York and North Staffordshire and then spread throughout the world (Bennett, 2011). The chamber system evolved to address constraints in the socio-economic environment (Maennig et al., 2015). As an intermediary organisation between government and business, they have expanded along with the market economy (Brockmann & Lacho, 2015). However, compared to their public counterparts, chambers have unique advantages in that they are democratically elected by the business community, have transparent governance that offers legitimacy, and, importantly, are locally rooted (Bennett, 2011).

Chambers aim to establish favourable conditions for businesses. These include influencing economic and social actors to create and seize new business opportunities, working with organisations that set rules for international trade and foreign investment, avoiding potential conflicts with interest groups and minimising political risks, and actively engaging the media and other sectors to protect the image and reputation of businesses (Saner et al., 2000; Noel & Luckett, 2014). Furthermore, by building and maintaining social relationships and networking, local chambers create relational and social capital and may help innovation (Maennig & Ölschläger, 2011). In short, chambers provide platforms for entrepreneurs to exploit synergy and cooperation opportunities. In line with these findings, we put forward the first hypothesis that we want to test empirically:

H1: As an inclusive private institution, the Chamber of Commerce is positively related to the long-term growth of entrepreneurship.

The chamber system is also distinct from other conventional industrial or clan-based organisations. They accept applications from a wider range of groups and offer more diverse services to their members (Bennett, 1996). In other words, instead of gathering people with the same identity, chambers connect merchants operating in different fields of business activities in a wider scope. People of different birthplaces, gender, kinships or social class can join chambers without any restrictions or prohibitions. Moreover, joining the chambers is completely voluntary in most countries. Because of this, the participation rate can vary greatly from place to place. In some cities, merchants might be more enthusiastic about joining the local chamber. In other cities, it could be lower. Cross-city variations in chamber membership density allow us to examine how differences in institutions affect entrepreneurial performance at the city level. In conjunction with our assumption 2, accordingly, we construct the second hypothesis:

H2: Higher participation in the chamber system is positively related to the persistent entry of entrepreneurs into the market.

If both hypotheses are confirmed, we can expect that, as an inclusive private-order institution, the Chamber of Commerce, when it includes more people in its organisation, is more likely to shape a favourable business environment and generate durable benefits. To empirically examine our hypotheses, we apply our empirical setting of the chamber system in China in the following sections.

3 Empirical background and data sources

3.1 Private-order institutions and chamber system in China

The Chamber of Commerce system stands in contrast to the less-inclusive guilds. Chinese guilds had existed since the Song dynasty and gradually spread (Skinner & Baker, 1977). They were formed in several stages: individual merchants mobilised their colleagues based on common birthplaces, raised funds for a meeting place, and sought opportunities to conspire with local governments. The group then negotiates and formalises procedures for the protection and use of their common property. Unlike in Europe, membership in Chinese guilds was not linked to citizenship. They did, however, impose strict requirements on members' birthplace and clan affiliation (Moll-Murata, 2008). For example, a Cantonese grain guild could only accept applicants who were in the grain business and born in Canton.

Guilds usually served to protect themselves as merchants and mutual aid by prohibiting competition, preserving monopolies, and providing social and cultural services. They achieved these goals mainly by controlling prices and wages or regulating the quality of

products and services in the region. Although guilds promoted the development of their members, the focus on the biologically based clan to pool resources led local elites to neglect the development of impersonal institutions for modernisation (Chen et al., 2021). With the dominance of exclusive institutions in the society, extensive economic support cannot be supplied for entrepreneurs when the need arose because of industrialisation efforts in the late 19th century.

Political uncertainty and failed interventions by western powers during the late Qing triggered a reshuffle of institutions. Reformists first introduced and promoted the Chamber of Commerce in China in 1904 with the aim of reducing the economic dominance of foreign firms. As the first state-legitimised association in China's history, chambers developed into community-based organisations that went beyond the scope of merchant guilds and officially controlled institutions. It is easy to see that the Chamber of Commerce is not a reproduction of a merchants' guild. Their structure and function, especially their social and political status, were totally different. While guilds only accepted members based on the specific uniformity, chambers deliberately recruited both elite and non-elite merchants, thus developing a city-wide merchant organisation (Chen, 2001). Chamber networks enable merchants to expand social relations beyond the limits and their kinship first time in Chinese history.[18] Joining a private institution no longer depended on an individual's social, political, or economic identity. Any merchant could join the local chambers if they could pay the membership fee. Chambers, therefore, promoted the network revolution through both relational diversification and expansion.

Unlike the guilds, the success of the chambers was not based on monopoly rights or granting licences. Within an individual chamber, merchants in the same market towns and cities were able to freely connect and form networks. The merchant group in China was previously divided by social status. Gentry merchants and guild leaders are generally defined as the elite group controlling varied material and social resources. Non-elite merchants, such as small and medium-sized factory owners and peddlers, often lacked the means to communicate with the elite (Esherick & Rankin, 1990; Rankin, 1986). Local merchant networks were therefore consistently fragmented. The local chamber system established interrelations through joint meetings, interlocking leaderships and other forms of contact. Such network development initiated by the local chamber brought elite and non-elite merchants into an institutionalised hierarchy. Across chambers, they strengthened merchant solidarity by developing similar structures and expanding their institutional contacts and personnel (Tang, 2017). Moreover, the chamber system also unified merchants and governors and provided a link with the government. The

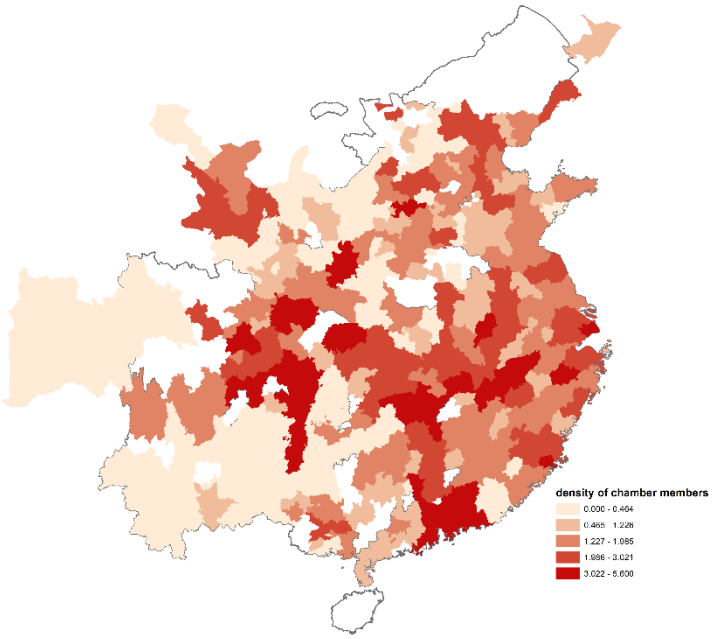
[18] Taking two local chambers - Shanghai and Shengze chambers - for example, Table A1 shows that non-natives usually accounted for the majority within the organisations even in the early stage of establishment.

chambers operated as an intermediary institution between the government and merchants, exercising their authority both as official deputies and as leaders of the merchants in promoting businesses and handling commercial affairs (Chen, 2011).

More importantly, there were no commercial laws and no corresponding judicial system in historical China. Generally, commercial disputes and litigation were handled as civil cases in a local administrative office. Due to the state policy of stressing agriculture and restraining commerce, officials tend to snub business cases. Since commercial disputes could not be resolved accurately and precisely, it was difficult to protect the legitimate rights and interests of entrepreneurs. Therefore, merchants wanted a specific commercial agency to handle commercial dispute cases. The Chamber of Commerce took responsibility for the mediation of commercial disputes. Specifically, the Chamber of Commerce established institutions such as appraisal and case resolution offices to deal with commercial disputes (Ma & Fu, 2009). To ensure impartiality, the personnel were democratically elected by the board of directors. The mediation agencies also formulate specific rules to regulate the case handling procedure in accordance with the provisions of the Articles of Association, thus standardising the process of dispute resolution by the chambers (Wang, 2009). In Suzhou, 70% of commercial disputes were resolved by the Chamber of Commerce, with the rest resolved by the local administrative office. In short, in the absence of the rule of law and government efficiency, the Chamber of Commerce provides qualified judicial resources for commercial development by institutionalising issues of commercial disputes.

With the advantages mentioned above, the chamber system experienced rapid development in both the core and peripheral areas of China. By 1915, more than a thousand local chambers were operating in China Proper. The variations between cities in historical chamber members allow us to examine how the variation in institutions affected entrepreneurial performance at the city level (Figure 2).

Figure 2: The distribution of the chamber membership density in 1915



3.2 The fall and rise of market economy in China

In 1954, the Chinese government launched the Socialist Reform of Capitalist Enterprises, which aimed to expropriate entrepreneurs and transfer ownership from private to public. This movement profoundly changed the corporate landscape in China. Under pressure, numerous entrepreneurs ceased operations and closed businesses. Others opted for a public-private partnership, and their enterprises were transformed into collective ownership. By 1956, the transformation process was essentially complete. Private enterprises were restructured into joint public-private management entities (Teiwes, 1987). After the enterprises were disenfranchised, entrepreneurs and their family members were further labelled and politically and socially suppressed in subsequent political movements such as the Great Leap Forward and the Cultural Revolution (Chen et al., 2016).

It was only with economic reform in the late 1970s that entrepreneurs regained their legitimacy. After the Cultural Revolution in 1976, the private economy was implicitly accepted by governments. One year later, organisations of local chambers were resumed and allowed to carry out activities and communications among members. In 1980, applications for business licences were formally opened to individuals, meaning that the political and economic status of entrepreneurs was recognised for the first time in modern China (MacFarquhar & Fairbank, 1991).[19] The abolition and reestablishment of the

[19] The process of private economy policy has been noted in the official website of Chinese central government, as well: http://www.gov.cn/xinwen/2018-04/20/content_5284602.htm

market economy provided a unique opportunity to study how the chamber system as an institution could persist despite an unprecedentedly long economic disruption.

3.3 Data sources

To analyse the long-term impact of the chamber system, we construct a city-year panel dataset combining the historical record of the chamber system in 1915 and entrepreneurship from 1975 to 1990. Firm Registration Data provides the information on the annual birth of the private business, which helps us measure city entrepreneurship dynamically. We discuss the construction of our treatment variables, outcome variables and control variables accordingly.

To capture the effect of the chamber system, we require information on the chamber development at the city level. We use the recently released Statistics of Agriculture and Commerce, compiled by the Ministry of Agriculture and Commerce of the Republic of China in the 1910s. We collected information including chamber name, type (e.g., whether it is a general or branch chamber), establishment year, the number of elite and non-elite members, the frequency of meetings, the number of disputes settled, the funding conditions, etc. by hand. Although the Chamber of Commerce updated this information annually, the data from 1915 is the most comprehensive as it covered most market towns and cities. Figure 1 and Figure A1 show that the number of chambers and their memberships stabilised during 1910s, which means that using data from 1915 presents a good picture. Our primary measure of chamber size is the number of chamber members. Knowing that the total number of members in the city largely depends on the local population, we construct a measure of chamber density per capita using historical population data from 1910, the year closest to our chamber data. Since the most granular population data is available at the city level, we aggregate our data and analysis at the city level.[20] In total, we obtained 254 cities (Figure 2).

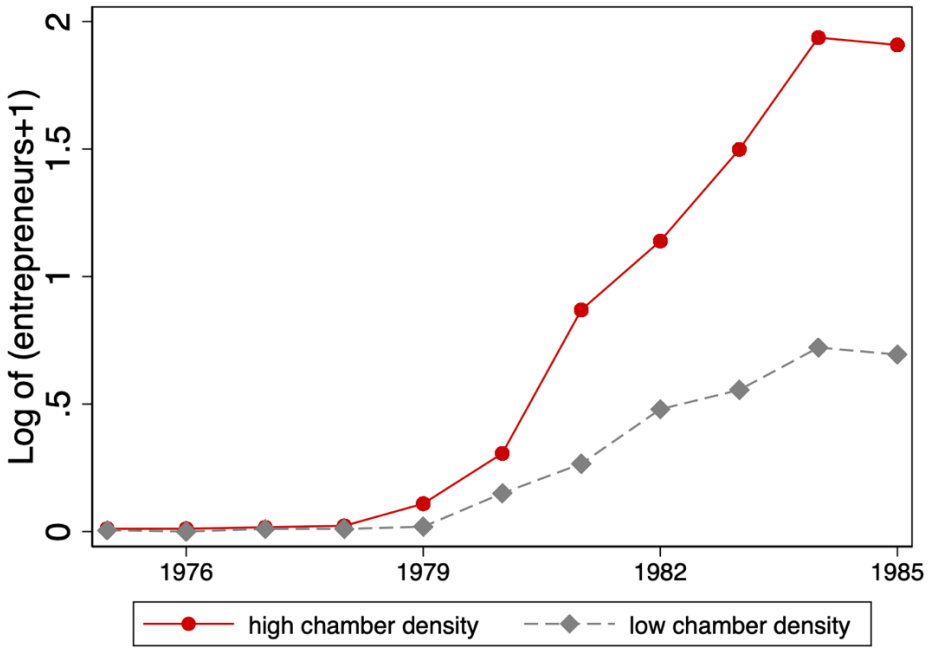
To measure entrepreneurial activity, we follow the approach of the Entrepreneurship Database developed by the World Bank (2010) and use the number of newly registered businesses as the proxy of entrepreneurship. With the advantage of data availability, adopting the annual business registries could facilitate comparing our results with other developing countries (Klapper & Love, 2011). Besides, considering China as a transition economy in the 1980s, the number of SOEs is not inappreciable within the overall

[20] Cities refer to prefectures in China's context. It is the second hierarchy of regional government. China's regional administration has been stable as the form of three level hierarchy 'province-city/prefecture-county' for a long time. Our population data shows that the average city population was 1.57 million in 1910.

number of registered businesses. Since the generation of SOEs is irrelevant and even incompatible with the spirit of entrepreneurship, we exclude all the newly registered SOEs and only consider new private business registration as entrepreneurship in our analysis. The State Administration for Market Regulation systemically documents the Firm Registration Data. All Chinese firms, whether small or big, are legally obligated to register with this public authorise (Bai et al., 2020). Businesses that haven't enlisted are considered illegal and immediately banned from operating in the market. Therefore, our data to measure entrepreneurship is a population of all registered private firms in China. They are also unique in that they contain information about the founders, such as location, type of business and industry. We geocoded the address information and then aggregated the number of private firms at the city level.

It should be noted that although private business licenses were only introduced in 1980, individuals were able to start private sector businesses with government approval in the late 1970s. Figure A3, for example, shows that the date of opening of the private economy was earlier than the date when licencing started. Our Firm Registration Data also recorded the opening date of each entrepreneur, which helps us conduct a preliminary trend test for our DID estimation in the following sections. Connecting with the raw data of historical chambers and contemporary entrepreneurs, Figure 3 divides the density of local chambers into high and low by comparing the mean and plots the number of entrepreneurs over the years. It shows that cities with a high density of chambers created a much higher number of entrepreneurs after the economic reform in 1980. This trend increased dramatically in the first five years of the opening of the market economy.

Figure 3: Average entrepreneurship high vs. low chamber density over time



Clearly, the observed gap in Figure 3 could be biased by other factors. Therefore, we consider a number of covariates in our analysis. First, we also consider merchant guilds as another common business organisation. Since merchant guilds were formed through members' birthplace, we identify them based on their geographic locations.[21] We then include city characteristics such as longitude and latitude, the nature of the terrain, distance from the coast and agricultural suitability. Geographical information comes from the historical Chinese project GIS, and agricultural quality is determined using the Caloric Suitability Index (Galor & Özak, 2016). China was an agriculture-dominated economy in the 1970s, so the weather has influenced people's occupational choices. Severe weather and natural disasters may force farmers to enter the private economy. To address this, we control the impact of floods and droughts on entrepreneurship. The historical annual local weather data are based on records from 267 weather stations. We assign the records to the nearest cities and define two separate binary indicators for exceptional floods and droughts. The summary statistics for all variables can be found in Table A2.

[21] Where there are merchant guilds, there may also have chambers. The existence of the two private institutions does not conflict geograhically. See more detailed information about the introduction of China's Merchant guilds in Wikipedia: https://en.wikipedia.org/wiki/Ten_Great_Merchant_Guilds

4 Chamber system and entrepreneurship

4.1 Baseline results

Our baseline specification uses a difference-in-differences strategy to examine the link between the chamber system and new firm registrations before and after the market economy reform:

$$Y_{i,t} = \beta_1 \ln\left(\frac{Chamber\ members}{City\ Population}\right)_i \times Post_t + \beta_2 \boldsymbol{X}_i \times Post_t + \beta_3 \delta_p \times Time_t + \beta_4 \boldsymbol{Z}_{i,t} + \gamma_i + \vartheta_t + \epsilon_{it} \quad (2)$$

Where $Y_{i,t}$ denotes the number of newly registered private businesses in city i in year t. $\ln\left(\frac{Chamber\ members}{City\ Popoulation}\right)_i$ is the logged chamber member density (members as a share of population) in the city i. The dummy variable $Post_t$ is 0 for the year before 1980 and 1 for 1980 to 1990. City dummies γ_i and year dummies ϑ_t capture differences over years and across cities. We allow the time-invariant city features to vary over time by interacting them with $\boldsymbol{X}_i \times Post_t$. We include a provincial dummy interacting with the time trend, $\delta_p \times Time_t$, to allow all province-specific characteristics to vary over time. Last, we control for time-varying city features, $\boldsymbol{Z}_{i,t}$, including extreme weather events. Standard errors are clustered at the city level.

The results are presented in Table 1 Columns (1) - (3) show the results after accounting for city and year fixed effects. Column (1) shows that cities with higher historical chamber density have higher levels of entrepreneurship after the reopening of the market economy. Columns (2) and (3), by contrast, show that the merchant guild, known as an extractive private order institution, does not contribute to the long-term growth of business human capital. Column (4) reports the specification by adding the province-time trend. The result indicates that the institution of the Chamber of Commerce is positive and significant. On average, a one standard deviation increase in logged chamber density implies a nearly 6% higher level of entrepreneurship. However, this estimate could pick up the effects of other potential covariates of chamber density. Therefore, in columns (5) and (6), we allow the time-invariant city characteristics to vary over time and add time-varying weather controls. After adding all these controls, the magnitude of the estimate remains 0.362. Our preferred specification in column (6) shows that a one-unit increase in chamber density is associated with a 5.35 percentage point increase in entrepreneurship. In short, Table 1 indicates that the penetration level of the historical chamber system is positively correlated with the entry of new firms in contemporary China, reflecting our hypotheses. It is also consistent with the narratives of anecdotes. Interviews and memoirs described that these local chambers resumed most of their original functions

after the allowance of private ownership in the 1980s (Zhang & Cao, 2016). Historically developed chambers carried out their inclusive business traditions. They provided places and networking opportunities for citizens, organised experienced entrepreneurs to teach business skills and also negotiated policy flexibility with the local governments. As a result, local entrepreneurship has been reawakened where chambers of commerce once flourished.

Table 1: Baseline results: the effect of the chamber system on entrepreneurship

	(1)	(2)	(3)	(4)	(5)	(6)
	Log of (entrepreneurs+1)					
post_x_chamber density	0.484***		0.481***	0.362***	0.328***	0.502***
	(0.0806)		(0.0806)	(0.0651)	(0.0636)	(0.144)
post_x_guild		0.199	0.153	-0.169	-0.216	-0.204
		(0.178)	(0.165)	(0.138)	(0.146)	(0.145)
City FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Province × time	N	N	N	Y	Y	Y
Time-invariant controls × post	N	N	N	N	Y	Y
Time-varying controls	N	N	N	N	N	Y
Observations	4,064	4,064	4,064	4,064	4,064	4,064
R-squared	0.721	0.695	0.722	0.793	0.798	0.799

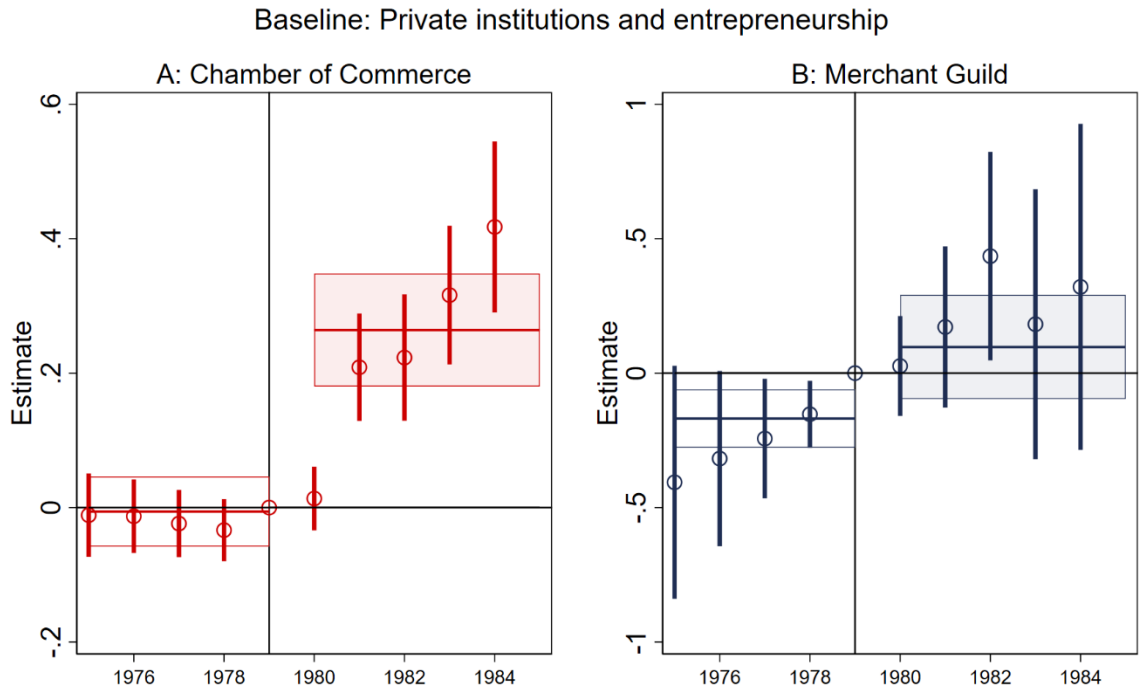
Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1975 to 1990. Entrepreneurship, as outcome variable, is proxied by the newly registered private business in each year. Post is denoted as 0 for the year before the reform in 1980 and 1 from 1980 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

To scrutinise our estimations, we also examine the link between the historic chamber system and contemporary entrepreneurship year by year. In doing so, we directly test whether there are already different trends for cities with different chamber densities even before the 1980 reform. More importantly, it allows us to show how quickly the effect of the chamber system took hold and whether it persists. We estimate regressions:

$$Y_{i,t} = \sum_{s=1975}^{1990} \beta_5 \ln\left(\frac{Chamber\ members}{City\ Population}\right)_i \times Post_t + \sum_{s=1975}^{1990} \beta_6 \boldsymbol{X}_i \times Post_t + \beta_7 \delta_p \times Time_t + \beta_8 \boldsymbol{Z}_{i,t} + \gamma_i + \vartheta_t + \epsilon_{it} \quad (2)$$

The parameters of interest are the β_5, which capture the year-specific entrepreneurial advantage of cities with higher chamber density. Figure 4A presents our estimate. It shows that there is no significant difference in the pre-trends for high- or low-density cities, while the positive impact of chamber density occurred since 1980, when economic reform was launched. This confirms that the revival of entrepreneurship only starts when national policies and rules allow it. Since 1981, the magnitude of the impact increased to over 0.2 percentage points, showing a steady upward trend. The coefficient rose to over 0.4 percentage points. By contrast, when replacing the chamber with the merchant guild, the buoyancy of the private institution quickly dissipated. Figure 4B shows that, although the guild appeared to be a stimulant in the early stages of reform, its year-to-year effect was trivial and insignificant. This stark contrast shows again that inclusiveness is matters if institutions are to have a positive effect.

Figure 4: The effect of private institutions on entrepreneurship over time

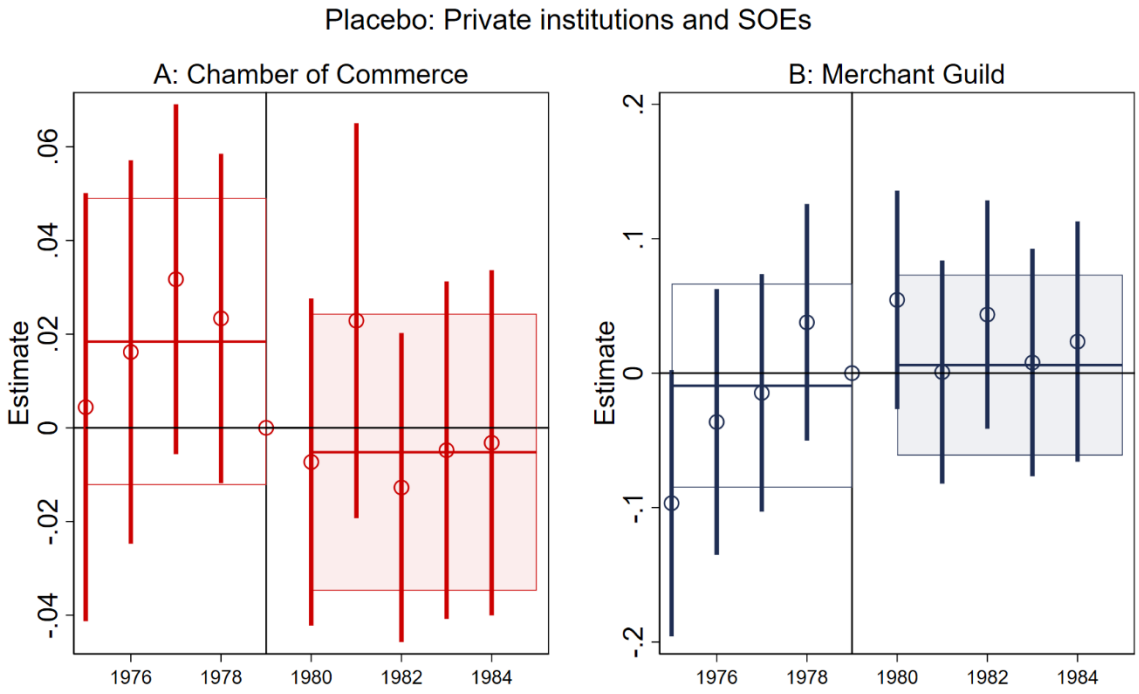


Note: Coefficient estimates on the year interactions are plotted as dots with their 90% confidence intervals indicated with vertical lines. Coefficient estimates on the aggregate interactions are shown with horizontal lines, and their 90% confidence intervals are indicated as boxes. Standard errors are clustered at the city level.

Could our estimation coincide with the economic outcomes led by other reforms taking place simultaneously? It is known that China was still largely dominated by the state-owned economy in the 1980s. Thus, to verify the exclusive effect of private institutions

on entrepreneurship, we conduct a placebo outcome test by adopting the newly registered SOEs as outcome variables (Eggers et al., 2021). Private institutions should not directly impact the development of the state-owned economy. If it turns out that the chamber of commerce has a positive effect on the engagement with SOEs, its impact on entrepreneurship could be questioned as a coincidence of timing. Estimations presented in Figure 5 dispel those concerns. Both coefficients of chamber and guild demonstrate that private institutions are irrelevant to state-led development, whether inclusive or not. These results confirm the peculiar effect of inclusive private institutions on entrepreneurship.

Figure 5: The effect of private institutions on newly registered SOEs over times



Note: Coefficient estimates on the year interactions are plotted as dots with their 90% confidence intervals indicated with vertical lines. Coefficient estimates on the aggregate interactions are shown with horizontal lines, and their 90% confidence intervals are indicated as boxes. Standard errors are clustered at the city level.

Furthermore, worrying that positive effects come from the way of measuring the chamber system, we also test using other characteristics of the chamber system: frequency of meetings, number of disputes resolved, revenue and expenditure as proxies for the historical chamber system. We also normalise the local population in 1910. Table A3 and

Figure A5 show that the significant and positive effects of chambers are consistent.[22] To further confirm the credibility of our baseline results, we also run a set of robustness checks such as excluding provincial capital cities, excluding treaty ports, excluding cities without chambers and standardising chamber density by city land size from Table A4 to A7, all of the estimates support our baseline findings.

In addition to the main effects of the chamber system, we examine whether there may be additional effects due to differences in political power and the internal structure of chambers of commerce. First, within chambers, private order institutions theoretically increase the degree of inclusiveness by diluting the dominance of elite groups (Ogilvie, 2021). Higher participation of non-elites could lead to more inclusive democratic discussions and decisions. So it is the share of non-elites that really has a positive impact on new entry in the long run. Second, as theories by political scientists and economists show, chambers with higher political status should have stronger managerial effects compared to chambers with lower status (Acemoglu et al., 2005). Given the superior political status and power of general chambers, their persistent influence over entrepreneurship may be even greater than that of the ordinary branch chamber.

By distinguishing and resolving the participation rate of elite and non-elite, we can construct non-elite chamber density and elite chamber density. They are measured through $\ln\left(\frac{non-elite\ chamber\ members}{city\ population}\right)_i$ and $\ln\left(\frac{elite\ chamber\ members}{city\ population}\right)_i$, respectively.[23] In columns (1) and (2) of Table 2, we evaluate the effects of non-elite and elites chamber densities separately and find both are statistically positive. To find out which is more effective, we run a horse race and find that chambers dominated by elite merchants become invalid. The outcome for non-elite chamber density, by contrast, is maintained and becomes larger, implying that our results are driven by inclusive chambers with a greater share of non-elite members. Our year-by-year estimates (Figure 6A) also show that chambers, which is characterised by a high proportion of non-elites, has gained significant economic results at the beginning of reform. In contrast, the effect of local elites has always been muted in the process of reform (Figure 6B).

We further test whether general chambers have a greater impact compared to their industry counterparts by dichotomising the chamber type and testing for political

[22] Figure A6 conducts the placebo tests between the chamber density and SOEs by replacing chamber density with the frequency of meetings, the number of disputes resolved, revenue and expenditure. The exclusive effect of the chamber on entrepreneurship had not fluctuated when we select the different proxies for the local chamber.

[23] We map the distributions of non-elite chamber density and elite chamber density in Figure A7 and A8.

status.[24] By comparing the magnitudes in column (4) to (6) in Table 2 and annual estimates in Figure 6C and 6D, we find that while the impact of both general chambers and branch chambers is favourable, the impact of general chambers is much larger than that of branch chambers. This points to the importance of the political standing of an institution, even for an inclusive organisation in the long run. In short, our results in Table 2 underline the importance of political status and de-elitism rate for an inclusive private-order institution.

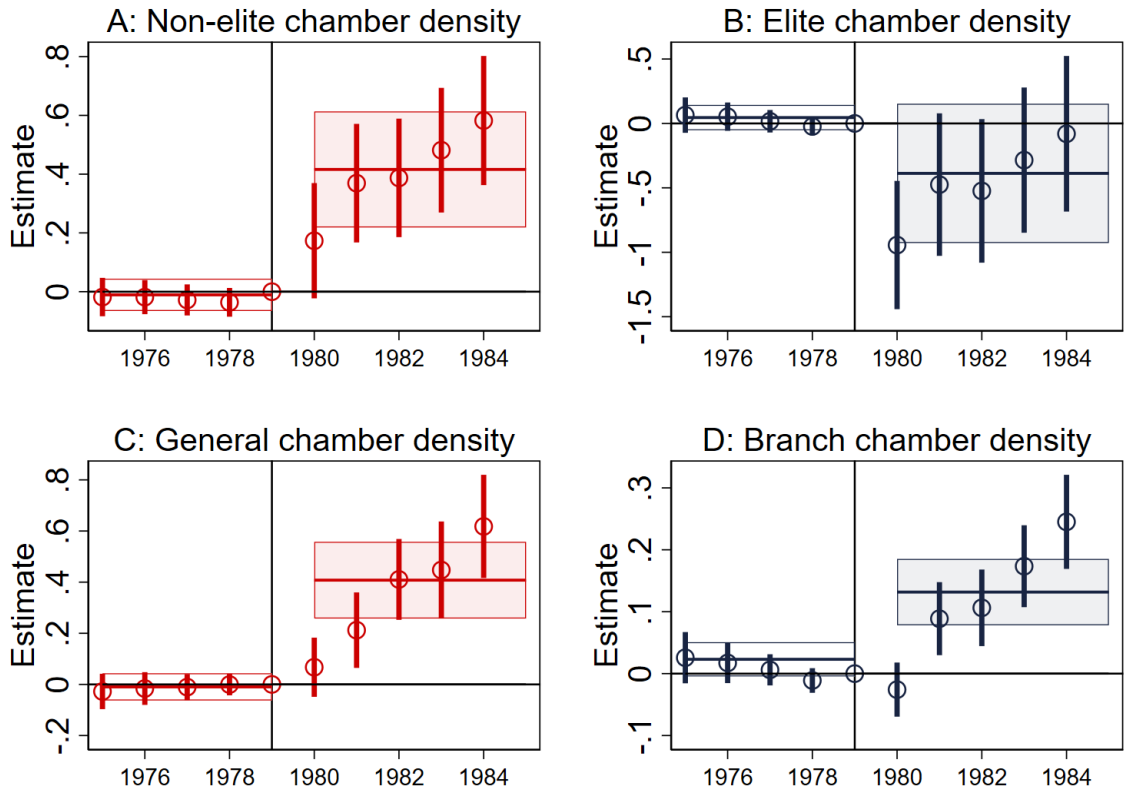
Table 2: Baseline results: heterogeneous effects of chamber system on entrepreneurship

	(1)	(2)	(3)	(4)	(5)	(6)
	Log of (entrepreneurs+1)					
post_x_non-elite chamber density	0.334***		0.465***			
	(0.0658)		(0.136)			
post_x_elite chamber density		0.669***	-0.370			
		(0.169)	(0.331)			
post_x_general chamber density				0.554***		0.532***
				(0.0919)		(0.0953)
post_x_branch chamber density					0.162***	0.142***
					(0.0432)	(0.0370)
post_x_guild	-0.217	-0.241	-0.207	-0.257*	-0.201	-0.226
	(0.145)	(0.149)	(0.145)	(0.143)	(0.156)	(0.147)
City FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Province × time	Y	Y	Y	Y	Y	Y
Time-invariant controls × time	Y	Y	Y	Y	Y	Y
Time-varying controls	Y	Y	Y	Y	Y	Y
Observations	4,064	4,064	4,064	4,064	4,064	4,064
R-squared	0.798	0.795	0.799	0.800	0.792	0.803

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1975 to 1990. Entrepreneurship, as outcome variable, is proxied by the newly registered private business in each year. Post is denoted as 0 for the year before the reform in 1980 and 1 from 1980 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Figure 6: The heterogenous effect of chamber system on entrepreneurship over time

[24] We also map the distributions of general chamber density and branch chamber density in Figure A9 and A10.



Note: Coefficient estimates on the year interactions are plotted as dots with their 90% confidence intervals indicated with vertical lines. Coefficient estimates on the aggregate interactions are shown with horizontal lines, and their 90% confidence intervals are indicated as boxes. Standard errors are clustered at the city level.

4.2 Results from instrumental variables

The DID strategy used above allows us to exclude the effects of omitted variables when their effects do not change before and after the relaunch of the private economy. However, the initial treatment, a higher density of chambers, may be endogenous to some other omitted factor which subsequently increases entrepreneurship. For example, the approval of the establishment of chambers from the Qing dynasty onwards allowed cities with different underlying characteristics to choose different penetration levels of chambers of commerce. In this case, the differences in contemporary entrepreneurship could be partly due to these underlying characteristics. To address this potential issue, we use the telegraph route before the chamber's implementation as an instrumental variable to isolate exogenous variation in adoption.

Before the telegraph, communication between merchants relied on slow trade routes along waterways. The telegraph network, which marked a significant reduction in communication time over long distances, became popular in China in the 1880s. Merchants quickly learned its value. Some studies show that sending telegrams became a daily task

for merchant groups after its implementation (Lin, 2016). Thus, when chamber systems were introduced in the early 20th century, cities that were closer to the site of telegraph stations were expected to have greater advantages in preparing, establishing and expanding chambers due to the catalytic effect of the telegraph. This opinion has been verified by the historical record (Baark, 1997).

A natural concern is that the placement of the telegraph might be associated with other economic factors. However, telegraph locations were determined by military, rather than commercial factors. The decision to lay new lines was often triggered by warfare and diplomatic conflicts because telegraph lines were only extended nationwide when the government reassessed the importance of the telegraph for national defence and state administration (Baark, 1997; Gao & Lei, 2021). For example, the first landline outside the treaty ports was built to connect Tianjin and Dagu Fort to exchange urgent military messages during the Second Opium War. Decisions on the placement of telegraph stations were completely dominated by government requirements (Wang, 1988).

The Qing dynasty post map provides us with the routes of the telegraph network in 1903, allowing us to measure the distance between the city and the telegraph offices built.[25] Using the telegraph map drawn before the rollout of the chamber system in 1904 helps us to further rule out endogenous issues. For our instrumental variable design, we thus estimate the following first-stage regression:

$$\ln\left(\frac{Chamber\ members}{City\ Population}\right)_i \times Post_t = \sigma_1 Telegraph_{i,1903} \times Post_t + \sigma_2 \boldsymbol{X}_i \times Post_t + \sigma_3 \delta_p \times Time_t + \sigma_4 \boldsymbol{Z}_{i,t} \gamma_i + \vartheta_t + \epsilon_{it} \quad (3)$$

Table 3 presents our IV results. Panel B shows that the first-stage F-statistics in the excluded IV are always above the conservative cut-off value of 10 in all estimated specifications. Columns (1) and (2) of Panel B show that $Telegraph_{i,1903}$ is a strong predictor of chamber density adoption and that an additional unit reduction in distance to the telegraph increases chamber density by 3.77 to 4.84 percentage points. In Panel A, IV estimates of column (1) and (2) waves from 0.817 to 0.996 depend on the control conditions, implying that cities with higher chamber density significantly improve the level of entrepreneurship. In columns (3) and (4), we also find strong positive effects of non-elite chamber density. In contrast, the effect of elite chambers turns significantly negative. This set of IV results further verifies the importance of non-elite chamber participation. Furthermore, our instrumented results in the last two columns reflect a similar picture

[25] We digitalise postal map attached in A11.

as our OLS results: chambers with higher bargaining power are more influential than their industry counterparts. All in all, our instrumented results in Table 3 confirm OLS results in Tables 1 and 2.

Table 3: The impact of chamber system on entrepreneurship: instrumental variable results

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A	Log of (entrepreneurs+1)					
post_x_chamber density	0.817**	0.996***				
	(0.400)	(0.340)				
post_x_non-elite chamber density			0.998***	1.963**		
			(0.341)	(0.928)		
post_x_elite chamber density				-3.670*		
				(2.106)		
post_x_general chamber density					0.932***	0.800***
					(0.308)	(0.299)
post_x_branch chamber density						0.159***
						(0.0443)
post_x_guild	-0.123	-0.152	-0.151	-0.0628	-0.223	-0.189
	(0.175)	(0.183)	(0.180)	(0.191)	(0.151)	(0.152)
Panel B	*First Stage*					
post_x_distance to telegraph	-0.0377**	-0.0484***	-0.0483***	-0.0201***	-0.0517***	-0.0481***
	(0.0147)	(0.0149)	(0.0148)	(0.00682)	(0.0136)	(0.0114)
City FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Province × time	N	Y	Y	Y	Y	Y
Time-invariant controls × post	N	Y	Y	Y	Y	Y
Time-varying controls	N	Y	Y	Y	Y	Y
Observations	4,064	4,064	4,064	4,064	4,064	4,064
R-squared	0.810	0.834	0.829	0.964	0.746	0.782
F-stat	107.998	167.832	173.575	141.08	509.08	483.845

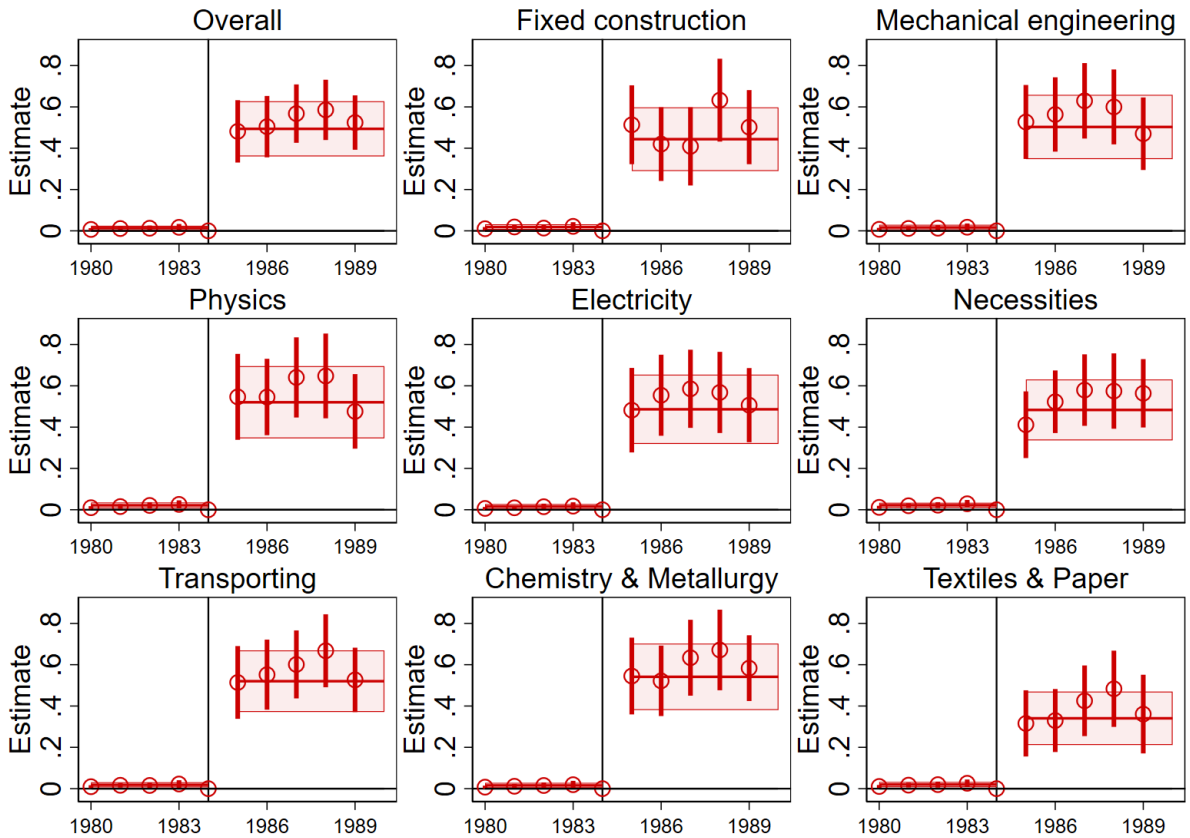
Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1975 to 1990. Entrepreneurship, as outcome variable, is proxied by the newly registered private business in each year. Post is denoted as 0 for the year before the reform in 1980 and 1 from 1980 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

5 Innovative Effects of Chamber System

The focus of this paper has, so far, been on entrepreneurship. But the mechanisms linking inclusive institutions to innovation may also apply to other measures of economic performance. Because of this, we turn to innovation – an outcome closely related to entrepreneurship which can also have a long-term impact on regional growth. An inclusive institution could foster growth-favouring incentives for innovative activities, as emphasised in endogenous growth models (Aghion & Howitt, 1990; Grossman & Helpman,

1991). The inclusive institutional rules that favour collective action could further promote technological innovations (Parente & Prescott, 1999). Institutions are seen as essential for innovation (Rodríguez-Pose & Di Cataldo, 2015; Donges et al., 2022), yet the literature has yet to test the influence of private sector institutions. To encourage invention and drive innovation, China introduced the first patent law in its history in 1985 and established the corresponding application system to authorise and protect intellectual property rights,[26] helping us to study the effect of chamber institutions on innovative performances in detail. We collect micro-level patent data from the China National Intellectual Property Administration (CNIPA) to measure creative performance between 1985 and 1990. To identify the impact over time, we again run a flexible model that allows us to examine the impact year by year.

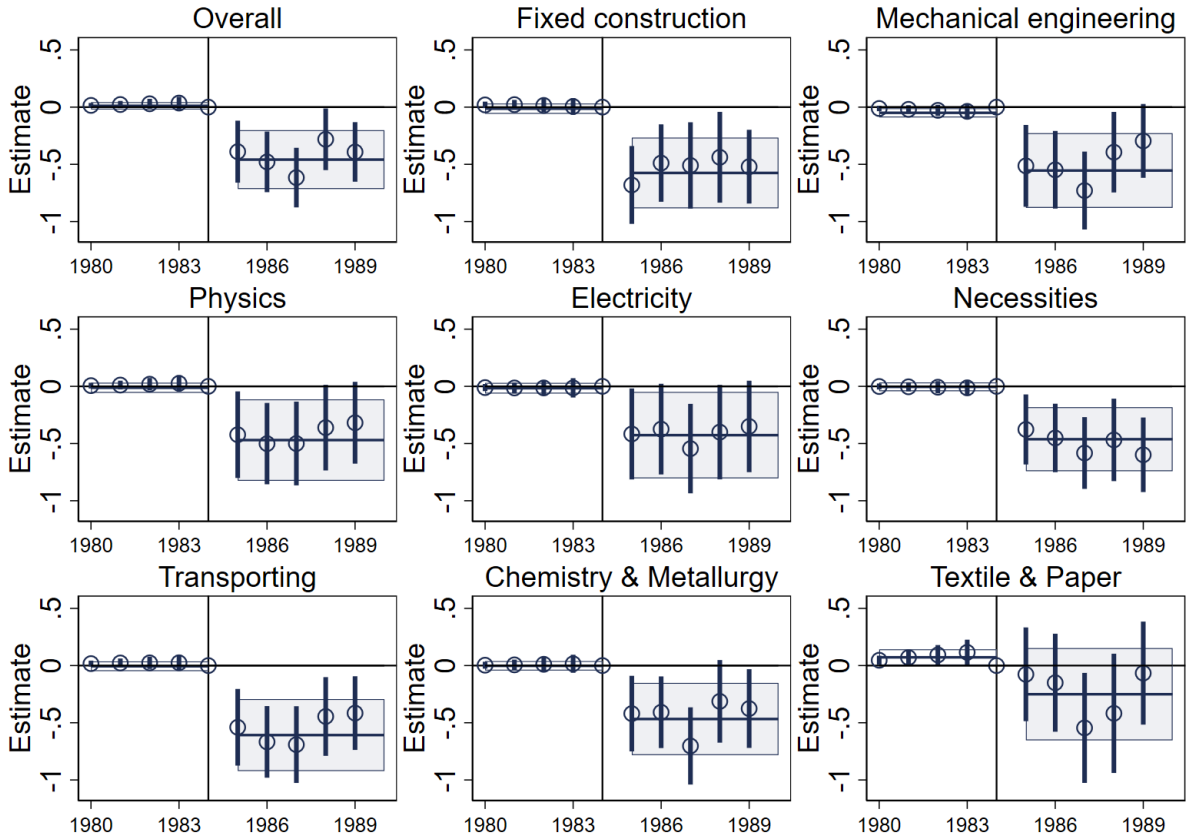
Figure 7: The chamber system and innovative performance across sectors over times



[26] On March 12th, 1984, the National People's Congress passed the patent law of China, which came into effect on April 1st, 1985. See more detailed information about China's patent law in https://en. wikipedia.org/wiki/Patent_law_of_China

Note: Coefficient estimates on the year interactions are plotted as dots with their 90% confidence intervals indicated with vertical lines. Coefficient estimates on the aggregate interactions are shown with horizontal lines, and their 90% confidence intervals are indicated as boxes. Standard errors are clustered at the city level.

Figure 8: The merchant guild and innovative performance across sectors over times



Note: Coefficient estimates on the year interactions are plotted as dots with their 90% confidence intervals indicated with vertical lines. Coefficient estimates on the aggregate interactions are shown with horizontal lines, and their 90% confidence intervals are indicated as boxes. Standard errors are clustered at the city level.

The general and sectoral impacts of the chamber system are shown in Figure 7. The outcome variable is the number of patents generated by private businesses at the city level. To make the innovation performance comparable across the industrial level, we conduct normalisation on each of them. The result shows that patenting increased in cities with higher chamber density immediately after the patent application was adopted. Furthermore, the effect has been relatively stable over the years. Our sectoral results below show that the level of innovation in the different sectors also increased after the adoption of the patent application system, and this response was very immediate and

statistically significant over time.27 Our results also confirm that inclusive institutions have a more pronounced impact on high-tech industries such as chemistry and engineering. To better compare the differences between inclusive and exclusive institutions on innovative performances, Figure 8 depicts the effect of the guild on innovation. It shows that exclusive institutions are ineffective for creating and disseminating new ideas in both high-tech and low-tech industries. The prevalence of negative coefficient across industries illustrates the hindrance of local guild atmosphere to innovation in the long run.

We further specify our findings on chamber system in the OLS and IV estimates in Table A11. Basically, the estimates show the same story, that cities with well-established chamber systems experienced a surge in innovative activity when rules and regulations allowed it. In short, our results are consistent with the claim that chambers are organisations that influence regional innovativeness (Fromhold-Eisebith, 2004) and show that innovative returns are higher in the better institutions (Rodriguez-Pose & Di Cataldo, 2015).

6 Conclusion

This paper explores the institutional roots of regional entrepreneurship and innovation. We hypothesise that open and inclusive private-order institutions may persist over long time periods and leads to long-term entrepreneurial growth even after structural breaks.

We analyse our hypotheses by adopting China's chamber system in 1915 and using a city year panel for the period 1975–1990. Our novel data and empirical setup allows us to go beyond that existing body of literature that focuses on the static long-term relationships between institutions and relevant outcomes. Our methodology allows us to track the dynamic effect over time.

We have three core results. First, our findings show that the overall effect of the chamber system, an inclusive regional private order institution, positively affected new firm creation even after the structural break of the Mao era. In contrast, the effect of less-inclusive merchant guilds faded away. Second, noting the complexity of institutions, we further explore heterogeneity across and within chambers. Our analysis suggests that the higher non-elite participation rate and political power of chambers are critical to the persistently better development of entrepreneurship. Finally, we show that these effects generalise to another key input into local economic development, innovation. In

[27] The sectoral classification follows the same line with the official definition given by CNIPA: https://www.cnipa.gov.cn/art/2020/12/8/art_2152_155480.html

summary, our findings confirm that the chamber system, as the inclusive private institution, can have a persistent effect on the geography of entrepreneurship and innovation, even despite considerable change in wider public institutions.

Our article contributes to several strands of literature. First, we provide the first rigorous and comprehensive empirical evidence on the long-term institutional influence on entrepreneurship and innovation. Although there is emerging literature on how local spirits affect the level of entrepreneurship, they have mainly focused on factors driven by geography, culture and historical level of self-employment (Fritsch & Wyrwich, 2017; Glaeser et al., 2015). Our article provides new insight into the rationale for regional entrepreneurship. Second, our article adds to the long-standing discussion on the persistence effect of institutions (Acemoglu & Robinson, 2008). By documenting the Chamber of Commerce in the framework of inclusiveness, we provide direct evidence of the importance of private-order institutions for long-term development in the developing country. Third, we also join the discussion of the rationale of China's economic miracle in the last four decades, provide an institutional perspective and show that the rationale of growth has already laid a foreshadowing in history (Ang, 2016; Lin et al., 2004).

Although the paper provides some novel insights, several aspects require further analysis, our analysis has not touched on the mechanisms of chamber system on entrepreneurship because of data limitations. More evidence is required to better understand how the institution affects merchants' specific activities, such as re-establishing merchant groups or even lobbying after the reform. Second, affected by the restriction of the Hukou system, the population was very unlikely to migrate across cities in China in the 1980s. Thus, the growth of entry of entrepreneurs should be interpreted as the effects of native people, and our context cannot explore the differential effects on native and migrant entrepreneurs. Finally, our work builds on the sectoral classification of firm creations and innovations and should be complemented by analyses looking at the relationship between institutions and specific skills and tasks as better data become available. We underline these as important and interesting directions for future research.

References

Acemoglu, D., Cantoni, D., Johnson, S., & Robinson, J. A. (2011). The consequences of radical reform: The French Revolution. *American Economic Review*, 101(7), 3286-3307.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. *Handbook of Economic Growth*, *1*, 385-472.

Acemoglu, D., & Robinson, J. A. (2008). Persistence of power, elites, and institutions. *American Economic Review*, *98*(1), 267-93.

Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail: The origins of power, prosperity, and poverty*. Currency.

Aghion, P., & Howitt, P. (1990). A model of growth through creative destruction.

Ang, Y. Y. (2016). *How China escaped the poverty trap*. Cornell University Press.

Baark, E. (1997). *Lightning Wires: the telegraph and China's technological modernisation, 1860-1890* (Vol. 1). Greenwood Publishing Group.

Bai, C. E., Hsieh, C. T., Song, Z. M., & Wang, X. (2020). *Special deals from special investors: The rise of state-connected private owners in China* (No. w28170). National Bureau of Economic Research.

Baruah, N. G., Henderson, J. V., & Peng, C. (2021). Colonial legacies: shaping African cities. *Journal of Economic Geography*, *21*(1), 29-65.

Bennett, R. J. (1996). Can transaction cost economics explain voluntary chambers of commerce?. *Journal of Institutional and Theoretical Economics*, 654-680.

Bennett, R. J. (2011). *Local business voice: the history of chambers of commerce in Britain, Ireland, and revolutionary America, 1760-2011*. Oxford University Press.

Blum, J. (1978). *The end of the old order in rural Europe*. Princeton University Press.

Brockmann, E. N., & Lacho, K. J. (2015). Benefits of chamber of commerce membership: Large vs. small population centers. *The Entrepreneurial Executive*, *20*, 31.

Chen, Y., Wang, H., & Yang, D. (2016). Salience of History and the Preference for Redistribution. *Available at SSRN 2717651*.

Chen, Z., Ma, C., & Sinclair, A. J. (2021). Banking on the Confucian Clan: Why China Developed Financial Markets So Late. *The Economic Journal*.

Chen, Z. (2001). The Origins of Chinese Chambers of Commerce in the Lower Yangzi Region. *Modern China*, *27*(2), 155-201.

Chen, Z. (2011). *Modern China's Network Revolution: Chambers of Commerce and Sociopolitical Change in the Early Twentieth Century*. Stanford University Press.

Dell, M., Lane, N., & Querubin, P. (2018). The historical state, local collective action, and economic development in Vietnam. *Econometrica*, *86*(6), 2083-2121.

Dixit, A. (2009). Governance institutions and economic activity. *American Economic Review*, *99*(1), 5-24.

Donges, A., Meier, J. M., & Silva, R. C. (2022). *The impact of institutions on innovation*. Management Science.

Eggers, A. C., Tuñón, G., & Dafoe, A. (2021). Placebo tests for causal inference. *Working Paper, University of Chicago.*

Esherick, J., & Rankin, M. B. (Eds.). (1990). *Chinese local elites and patterns of dominance* (Vol. 11). Univ of California Press.

Epstein, S. R., & Prak, M. (2008). Introduction: guilds, innovation and the European economy, 1400-1800. *Guilds, innovation and the European economy*, *14001800*.

Fritsch, M., & Wyrwich, M. (2017). The effect of entrepreneurship on economic development—an empirical analysis using regional entrepreneurship culture. *Journal of Economic Geography*, *17*(1), 157-189.

Fromhold-Eisebith, M. (2004). Innovative milieu and social capital—complementary or redundant concepts of collaboration-based regional development?. *European Planning Studies*, *12*(6), 747-765.

Galor, O., & Özak, Ö. (2016). The agricultural origins of time preference. *American Economic Review*, *106*(10), 3064-3103.

Gao, P., & Lei, Y. H. (2021). Communication infrastructure and stabilising food prices: evidence from the telegraph network in China. *American Economic Journal: Applied Economics*, *13*(3), 65-101.

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics*, *97*(2), 498-520.

Greif, A. (2006). *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge University Press.

Grossman, G. M., & Helpman, E. (1991). *Innovation and growth in the global economy*. MIT press.

Helpman, E. (2009). *The mystery of economic growth.* Harvard University Press.

Hillmann, H. (2013). Economic institutions and the state: insights from economic history. *Annual Review of Sociology*, *39*, 251-273.

Katz, A. (1996). Taking private ordering seriously. *University of Pennsylvania law review*, *144*(5), 1745-1763.

Katz, E. D. (2000). Private order and public institutions: comments on McMillan and Woodruff's Private order under dysfunctional public order. *Michigan law review*, 2481-2493.

Klapper, L., & Love, I. (2011). The impact of the financial crisis on new firm registration. *Economics Letters*, 113(1), 1-4.

Lin, Y. 2016. Communication and trade: The message transmission of jiao merchants in Taiwan and Ningbo at the end of the nineteenth century. *History Inquiry, (58), 157-193.*

Lin, J. Y., Cai, F., & Li, Z. (2004). *The China miracle: Development strategy and economic reform.* The Chinese University of Hong Kong Press.

Lucassen, J., De Moor, T., & Van Zanden, J. L. (2008). The return of the guilds: Towards a global history of the guilds in pre-industrial times. *International Review of Social History*, *53*(S16), 5-18.

Ma, Min., & Fu, H. (2010). Research on the History of the Chinese Chamber of Commerce in the Past 20 Years (1990-2009). *Modern History Research, 2.*

MacFarquhar, R., & Fairbank, J. K. (1991). Revolutions within the Chinese Revolution, 1966-1982.

Maennig, W., & Ölschläger, M. (2011). Innovative milieux and regional competitiveness: The role of associations and chambers of commerce and industry in Germany. *Regional Studies*, *45*(4), 441-452.

Maennig, W., Ölschläger, M., & Schmidt-Trenz, H. J. (2015). Organisations and regional innovative capability: the case of the chambers of commerce and industry in Germany. *Environment and Planning C: Government and Policy*, 33(4), 811-827.

Martin, R., & Sunley, P. (2015). On the notion of regional economic resilience: conceptualisation and explanation. *Journal of Economic Geography*, *15*(1), 1-42.

Moll-Murata, C. (2008). Chinese guilds from the seventeenth to the twentieth centuries: An overview. *International Review of Social History*, *53*(S16), 213-247.

Noel, N. M., & Luckett, M. (2014). The benefits, satisfaction, and perceived value of small business membership in a chamber of commerce. *International Journal of Nonprofit and Voluntary Sector Marketing*, *19*(1), 27-39.

North, D. C., & Thomas, R. P. (1973). *The rise of the western world: A new economic history*. Cambridge University Press.

North, D. C., Wallis, J. J., & Weingast, B. R. (2009). *Violence and social orders: A conceptual framework for interpreting recorded human history*. Cambridge University Press.

North, D.C. 1990. *Institutions, Institutional change, and Economic Performance*. Cambridge University Press.

Ogilvie, S. (2021). Thinking carefully about inclusiveness: evidence from European guilds. *Journal of Institutional Economics*, *17*(2), 185-200.

Ogilvie, S., & Carus, A. W. (2014). Institutions and economic growth in historical perspective. *Handbook of Economic Growth*, *2*, 403-513.

Parente, S. L., & Prescott, E. C. (1999). Monopoly rights: A barrier to riches. *American Economic Review*, *89*(5), 1216-1233.

Rodríguez-Pose, A., & Di Cataldo, M. (2015). Quality of government and innovative performance in the regions of Europe. *Journal of Economic Geography*, *15*(4), 673-706.

Rankin, M. B. (1986). *Elite activism and political transformation in China: Zhejiang Province, 1865-1911*. Stanford University Press.

Saner, R., Yiu, L., & Søndergaard, M. (2000). Business diplomacy management: A core competency for global companies. *Academy of Management Perspectives*, *14*(1), 80-92.

Skinner, G. W., & Baker, H. D. (1977). *The city in late imperial China*. Stanford University Press.

Tang, L. (2017). *Merchants and Society in Modern China: From Guild to Chamber of Commerce*. Routledge.

Teiwes, F. C. (1987). Establishment and consolidation of the new regime. *The Cambridge History of China*, *14*(Part 1).

Wang, E. (1988). *Sheng Xuanhuai and the Chinese Telegraph Business.* Academia

Sinica.

Wang, H. (2009). Standardisation of Mediating Disputes at the Chamber of Commerce in Late Qing Dynasty. *Journal of Yancheng Institute of Technology, 22(1), 40-44.*

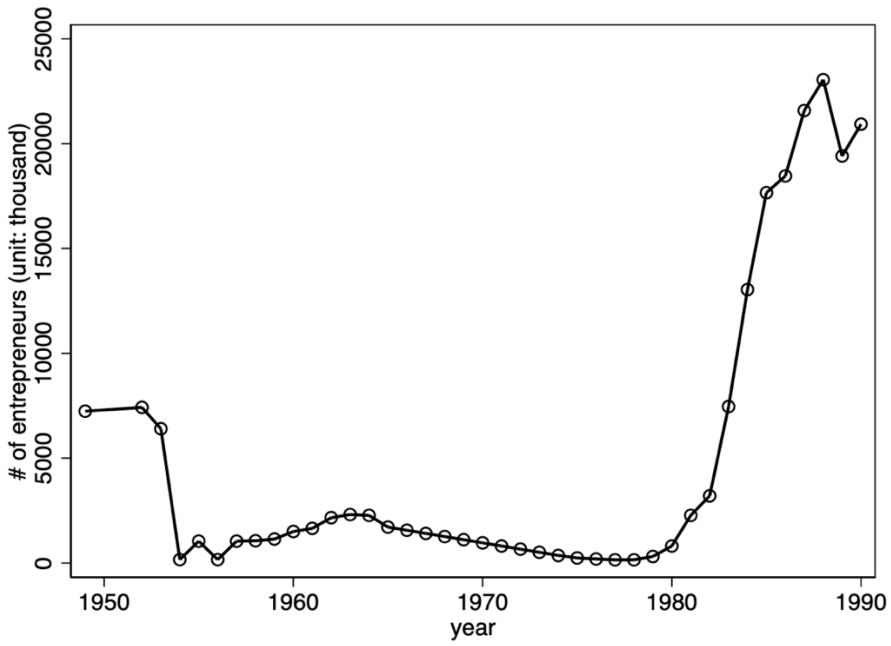
Wang, H., Rodríguez-Pose, A., & Lee, N. (2021). The long shadow of history in China: Regional governance reform and Chinese territorial inequality. *Applied Geography, 134*, 102525.

World Bank. (2010). *Entrepreneurship Snapshots 2010: Measuring the Impact of the Financial Crisis on New Business Registration.* The World Bank.

Zhang, Z., & Cao, Y. (2016). *Wenzhou Chamber of Commerce and Wenzhou Business.* Zhonghua Book Company.

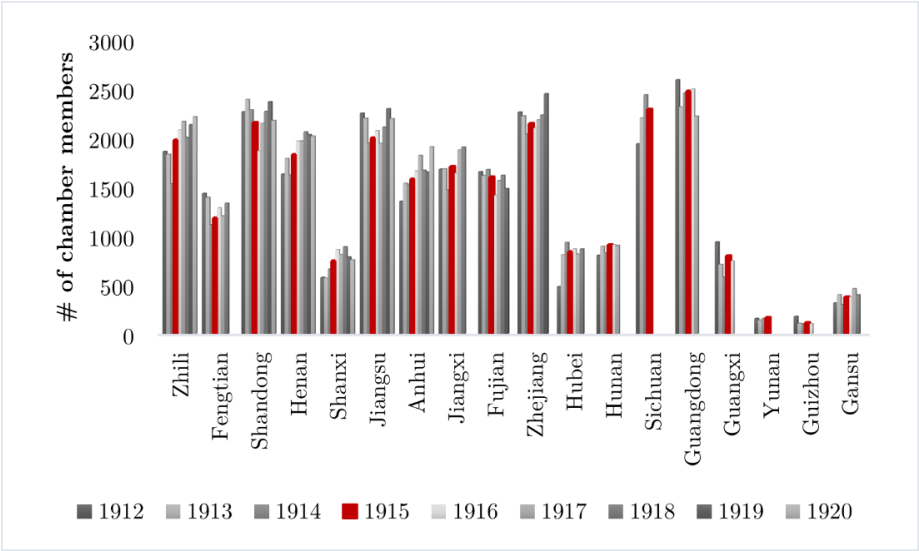
Appendix A

Figure A1. Number of private businesses in the national wide, 1949-1990



Note: The figure depicts the number of private businesses over years. The data is collected from National Bureau of Statistics and State Administration for Industry and Commerce. Information for 1950 and 1951 have not been recorded because of missing data. It echoed our main context that the private ownership was severely restrained from 1950s to 1970s.

Figure A2. Number of chamber memberships at the provincial level, 1912-1920



Note: The chart compares the number of memberships across provinces from 1912 to 1920. The data is collected from the Statistics of Agriculture and Commerce which compile by the Ministry of Agriculture and Commerce of the Republic of China. Unfortunately, chamber membership data is not available before 1912 and the data of chambers in certain provinces such as Sichuan and Yunnan are missing in the statistics in the later period of 1910s. However, it is clear that the situation of chamber tends to be stable at the regional level.

Figure A3: A sample of private business license in 1980s

Note: The figure provides a sample of private business license in 1980s. The license reflects information such as founder's name and address, opening date and registration date. Prior to 1980, tacit acceptance of the private ownership led to the gap between the opening and registration dates. Firm Registration Data provides the record of opening date of each business which help us to conduct the parallel trend test before the market economy reform in 1980.

Figure A4: Correlations between chamber member density and its other characteristics

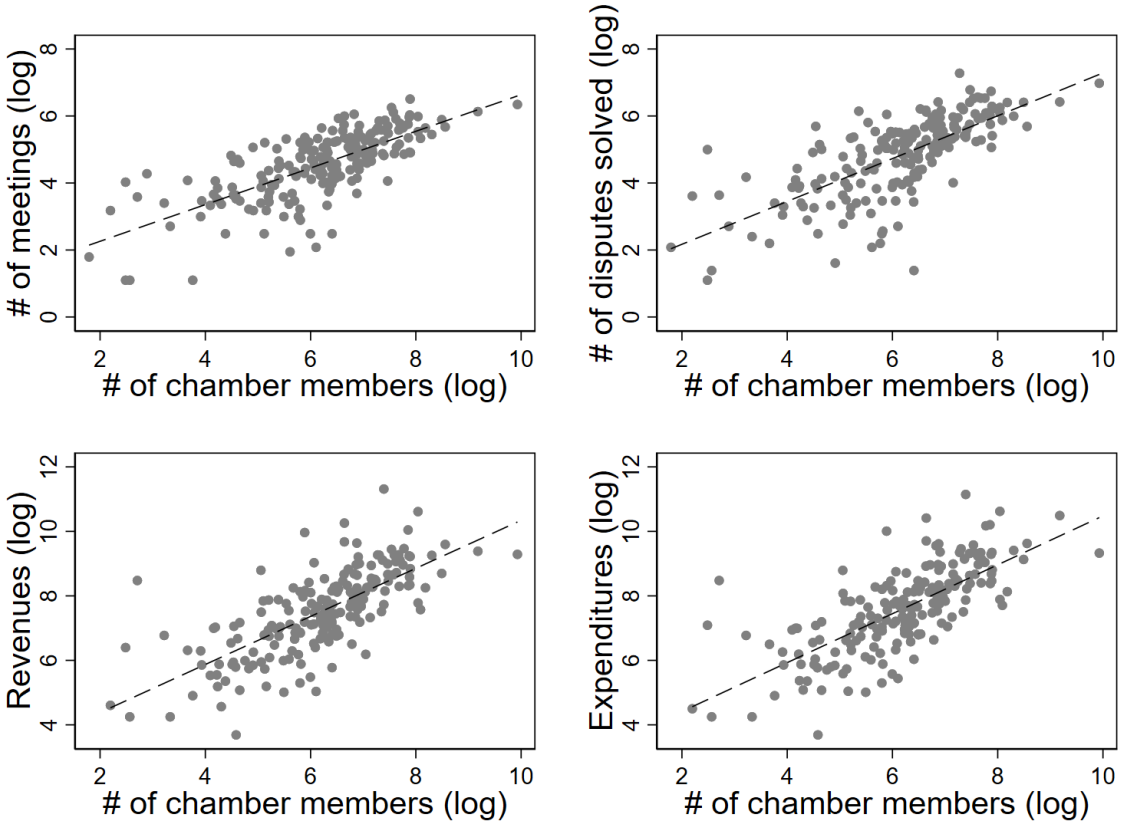


Figure A5: Robustness check: Chamber system and entrepreneurship over times

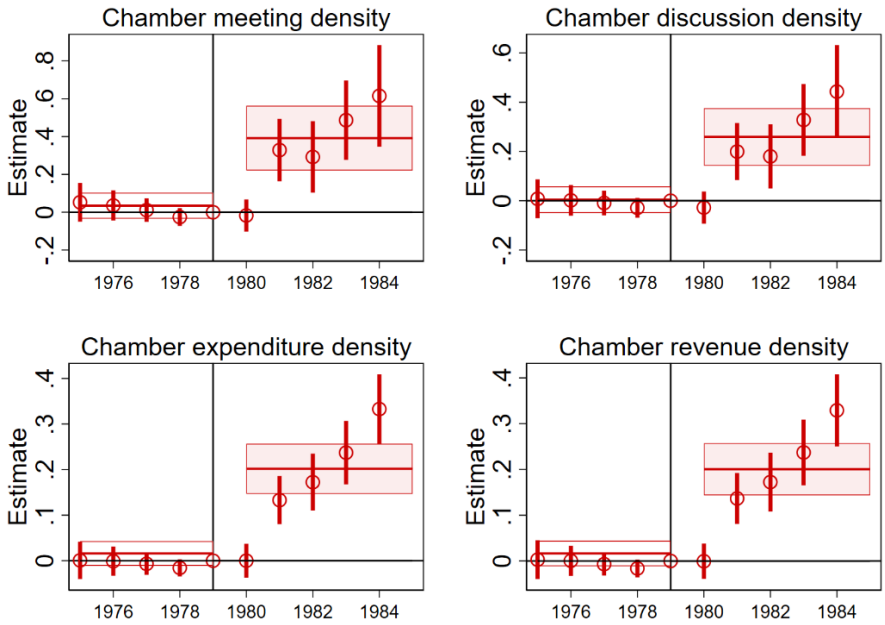


Figure A6: Placebo test: Chamber system and newly registered SOEs over times

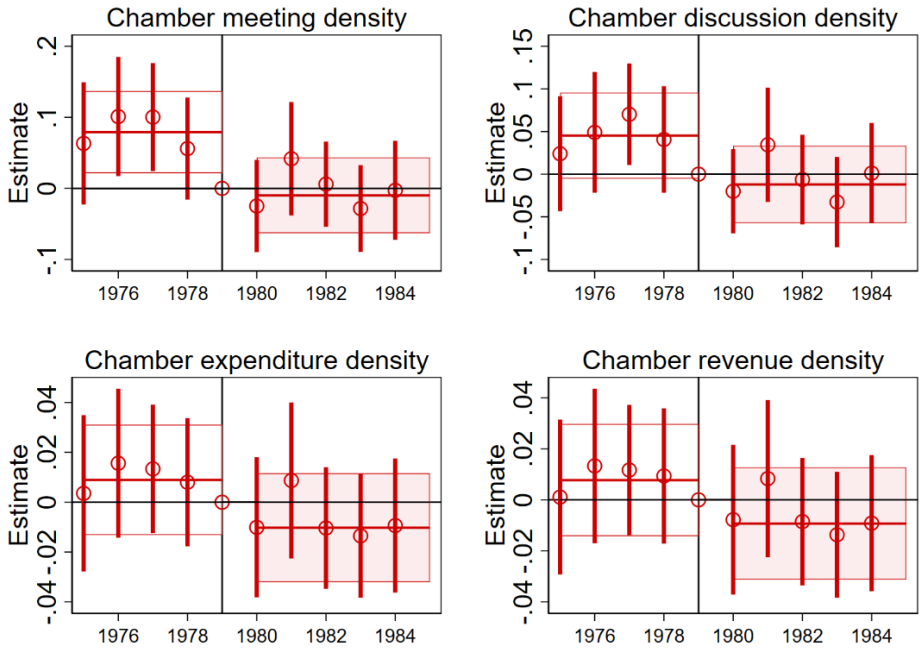


Figure A7: The distribution of non-elite chamber system in 1915

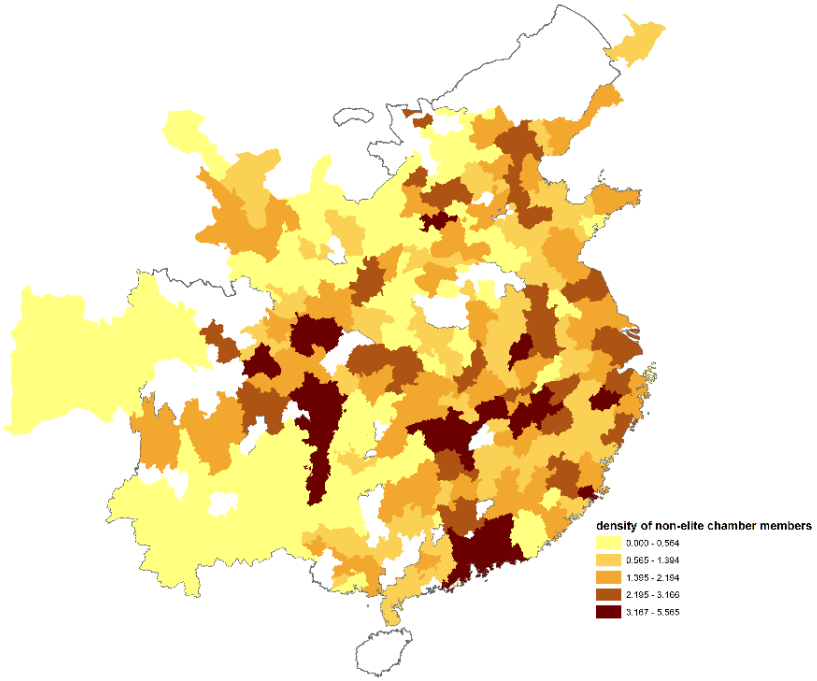


Figure A8: The distribution of non-elite chamber system in 1915

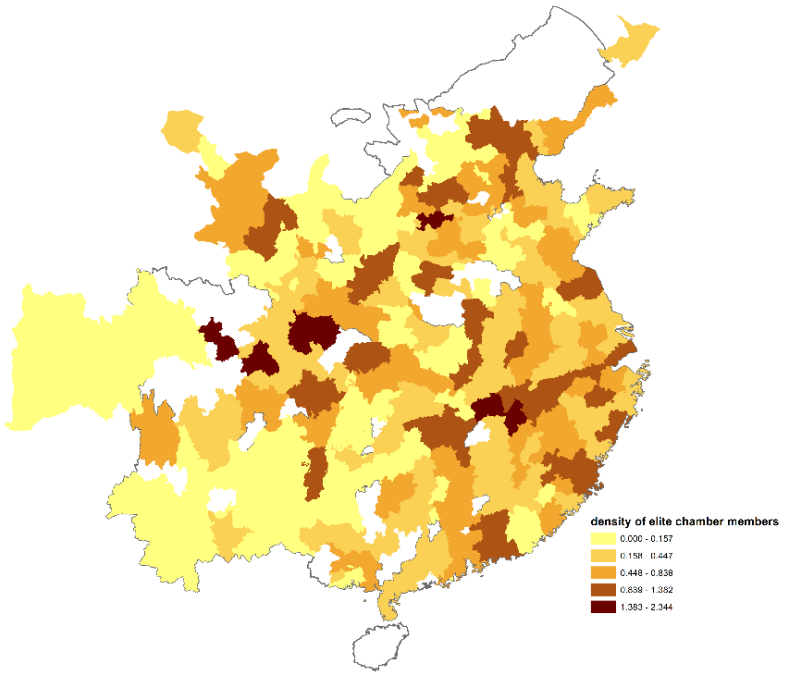


Figure A9: The distribution of general chamber system in 1915

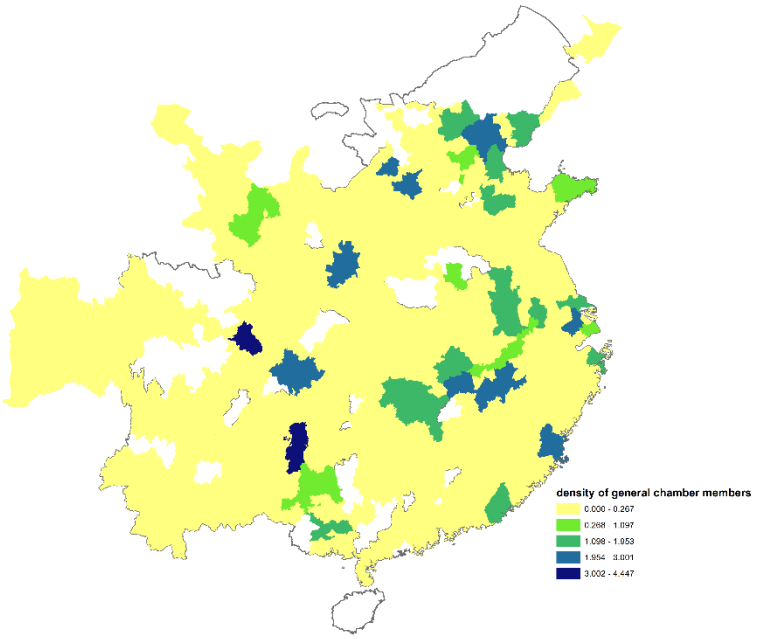


Figure A10: The distribution of branch chamber system in 1915

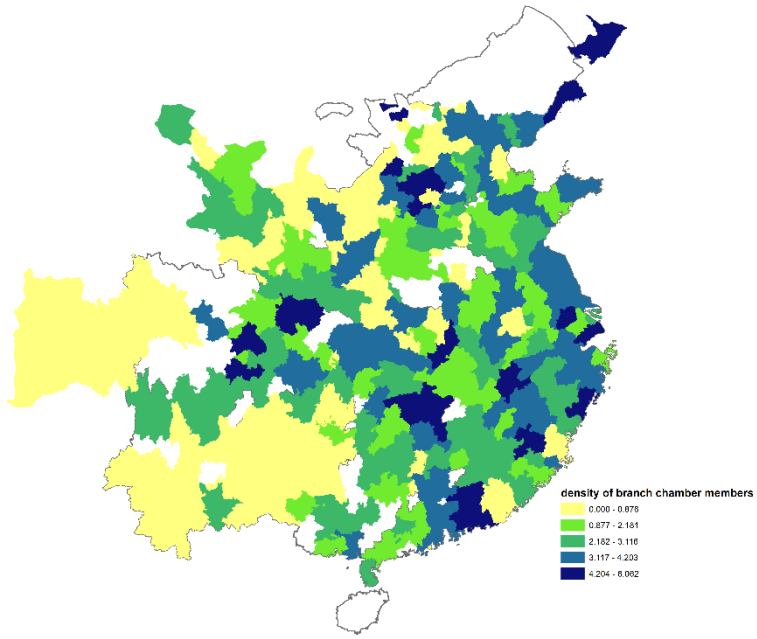


Figure A11: The Post Route Map of the Qing Dynasty in 1903

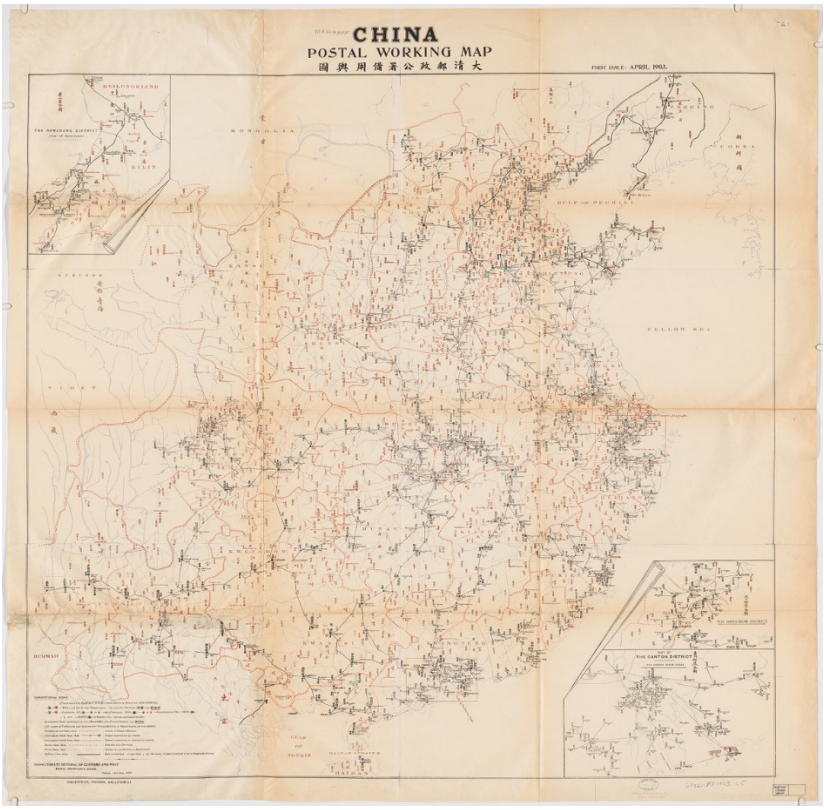


Figure A12: Minutes of meeting of Shanghai General Commercial Chamber

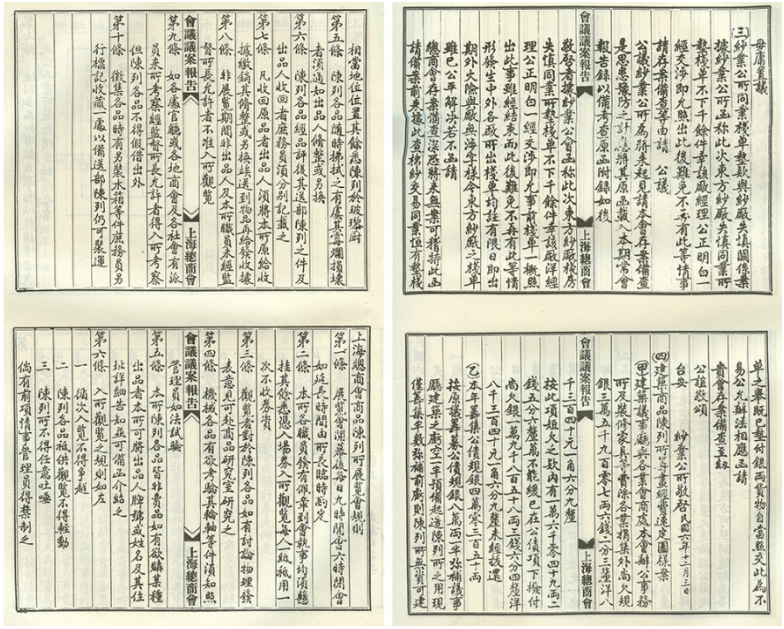


Table A1. Non-native rate of elites of non-elites in Shanghai and Shengze chambers

	Shanghai Chamber			Shengze Chamber		
Year	elites	non-elites	total	elites	non-elites	total
1906	76.19%	68.97%	70.89%	44.44%	40.00%	41.18%
1907				54.55%	51.61%	52.38%
1908	66.67%	67.14%	67.03%	54.55%	51.61%	52.38%
1909	57.14%	62.12%	60.92%	63.64%	58.06%	59.52%
1910	57.14%	68.12%	65.56%			
1911	57.14%	74.00%	69.01%			

Note: This table shows that non-native rates in Shanghai and Shengze chambers respectively. The data is collected from Chen (2011). It proves that, even in the early stage of establishment, chamber system was not restricted their memberships by the clans and birthplaces. Both elite and non-elite groups have a very high percentage of non-natives across years.

Table A2: Summary statistics for variables

Variable	Observations	Mean	SD
Panel A. Outcome variable			
Entrepreneurs	4,064	0.971	1.562
Panel B. Variables of interest			
post_x_chamber density	4,064	0.918	1.133
post_x_non-elite chamber density	4,064	0.860	1.097
post_x_elite chamber density	4,064	0.276	0.400
post_x_general chamber density	4,064	0.156	0.562
post_x_branch chamber density	4,064	1.428	1.597
Panel C. Control variables			
post_x_guild	4,064	0.238	0.426
post_x_agricultural suitability	4,064	5.577	3.810
post_x_coast	4,064	8.692	5.950
post_x_ruggedness	4,064	4.013	2.980
post_x_longitude	4,064	21.291	14.980
post_x_latitude	4,064	76.857	52.050
flood	4,064	0.265	0.441
drought	4,064	0.469	0.499

Table A3: Robustness checks: the effect of chamber system on entrepreneurship, OLS, replace chamber density with other chamber features

	(1)	(2)	(3)	(4)
	Log of (entrepreneurs+1)			
post_x_chamber meeting density	0.547***			
	(0.130)			
post_x_chamber discussion density		0.395***		
		(0.0926)		
post_x_chamber expenditure density			0.267***	
			(0.0417)	
post_x_chamber revenue density				0.263***
				(0.0429)
post_x_guild	-0.213	-0.221	-0.149	-0.158
	(0.156)	(0.157)	(0.155)	(0.155)
City FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Province × time	Y	Y	Y	Y
Time-invariant controls × post	Y	Y	Y	Y
Time-varying controls	Y	Y	Y	Y
Observations	4,064	4,064	4,064	4,064
R-squared	0.788	0.787	0.793	0.792

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1975 to 1990. Entrepreneurship, as outcome variable, is proxied by the newly registered private business in each year. Post is denoted as 0 for the year before the reform in 1980 and 1 from 1980 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Table A4: Robustness checks: the effect of chamber system on entrepreneurship, OLS, exclude capital cities

	(1)	(2)	(3)	(4)	(5)	(6)
	Log of (entrepreneurs+1)					
post_x_chamber density	0.419***		0.417***	0.262***	0.223***	0.223***
	(0.0848)		(0.0852)	(0.0623)	(0.0607)	(0.0608)
post_x_guild		0.177	0.150	-0.155	-0.203	-0.205
		(0.174)	(0.166)	(0.142)	(0.150)	(0.150)
City FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Province s× time	N	N	N	Y	Y	Y
Time-invariant controls × post	N	N	N	N	Y	Y
Time-varying controls	N	N	N	N	N	Y
Observations	3,808	3,808	3,808	3,808	3,808	3,808
R-squared	0.702	0.682	0.702	0.781	0.786	0.786

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1975 to 1990. Entrepreneurship, as outcome variable, is proxied by the newly registered private business in each year. Post is denoted as 0 for the year before the reform in 1980 and 1 from 1980 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Table A5: Robustness checks: the effect of chamber system on entrepreneurship, OLS, exclude treaty ports

	(1)	(2)	(3)	(4)	(5)	(6)
	Log of (entrepreneurs+1)					
post_x_chamber density	0.386***		0.385***	0.295***	0.266***	0.266***
	(0.0756)		(0.0762)	(0.0588)	(0.0591)	(0.0592)
post_x_guild		0.120	0.0860	-0.284**	-0.286**	-0.289**
		(0.166)	(0.160)	(0.112)	(0.128)	(0.128)
City FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Province × time	N	N	N	Y	Y	Y
Time-invariant controls × post	N	N	N	N	Y	Y
Time-varying controls	N	N	N	N	N	Y
Observations	3,728	3,728	3,728	3,728	3,728	3,728
R-squared	0.697	0.678	0.698	0.776	0.780	0.781

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1975 to 1990. Entrepreneurship, as outcome variable, is proxied by the newly registered private business in each year. Post is denoted as 0 for the year before the reform in 1980 and 1 from 1980 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Table A6: Robustness checks: the effect of chamber system on entrepreneurship, OLS, exclude cities without chambers

	(1)	(2)	(3)	(4)	(5)	(6)
	Log of (entrepreneurs+1)					
post_x_chamber density	0.420***		0.422***	0.368***	0.350***	0.349***
	(0.105)		(0.106)	(0.0797)	(0.0783)	(0.0784)
post_x_guild		0.0681	0.102	-0.159	-0.251	-0.254
		(0.200)	(0.190)	(0.158)	(0.162)	(0.162)
City FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Province × time	N	N	N	Y	Y	Y
Time-invariant controls × post	N	N	N	N	Y	Y
Time-varying controls	N	N	N	N	N	Y
Observations	3,168	3,168	3,168	3,168	3,168	3,168
R-squared	0.736	0.722	0.736	0.807	0.810	0.810

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1975 to 1990. Entrepreneurship, as outcome variable, is proxied by the newly registered private business in each year. Post is denoted as 0 for the year before the reform in 1980 and 1 from 1980 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Table A7: Robustness checks: the effect of chamber system on entrepreneurship, OLS, standardised chamber density by land size

	(1)	(2)	(3)	(4)	(5)	(6)
	Log of (entrepreneurs+1)					
post_x_chamber density	0.191***		0.190***	0.122***	0.112***	0.112***
	(0.0267)		(0.0269)	(0.0257)	(0.0279)	(0.0279)
post_x_guild		0.199	0.120	-0.177	-0.195	-0.197
		(0.178)	(0.165)	(0.140)	(0.150)	(0.150)
City FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Province × time	N	N	N	Y	Y	Y
Time-invariant controls × post	N	N	N	N	Y	Y
Time-varying controls	N	N	N	N	N	Y
Observations	4,064	4,064	4,064	4,064	4,064	4,064
R-squared	0.720	0.695	0.720	0.787	0.793	0.793

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1975 to 1990. Entrepreneurship, as outcome variable, is proxied by the newly registered private business in each year. Post is denoted as 0 for the year before the reform in 1980 and 1 from 1980 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Table A8: The effect of chamber system on innovative performance

	(1)	(2)
	DID	DID+IV
Outcome variables	**Variable of interest**: *post_x_chamber density*	
Overall patents	0.465***	2.078***
	(0.0800)	(0.563)
Fixed construction patents	0.121***	0.573***
	(0.0264)	(0.183)
Mechanical engineering patents	0.215***	1.021***
	(0.0419)	(0.298)
Physics patents	0.233***	0.977***
	(0.0492)	(0.293)
Electricity patents	0.200***	0.891***
	(0.0430)	(0.269)
Necessities patents	0.243***	1.135***
	(0.0467)	(0.327)
Transporting patents	0.271***	1.132***
	(0.0490)	(0.322)
Chemistry & Metallurgy patents	0.330***	1.510***
	(0.0620)	(0.413)
Textiles & Paper patents	0.0792***	0.441***
	(0.0185)	(0.153)
City FE	Y	Y
Year FE	Y	Y
Province × time	Y	Y
Time-invariant controls × post	Y	Y
Time-varying controls	Y	Y
Observations	2,794	2,794

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1980 to 1990. Innovation, as outcome variable, is proxied by the number of patents in each year. Post is denoted as 0 for the year before the reform in 1985 and 1 from 1985 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Table A9. The effect of chamber system on innovative performance: non-elite and elite chamber density, DID and DID + IV

	(1) DID	(2) DID+IV	(3) DID	(4) DID+IV
Variable of interest	*(non-elite chamber density)*		*(elite chamber density)*	
Outcome variables:				
Overall patents	0.237***	1.024***	-0.191	-8.788*
	(0.0422)	(0.267)	(0.201)	(5.132)
Fixed construction patents	0.0624***	0.269***	-0.0446	-2.312
	(0.0137)	(0.0847)	(0.0582)	(1.422)
Mechanical engineering patents	0.108***	0.484***	-0.0684	-4.203*
	(0.0215)	(0.135)	(0.105)	(2.532)
Physics patents	0.118***	0.462***	-0.106	-3.853*
	(0.0249)	(0.135)	(0.123)	(2.289)
Electricity patents	0.100***	0.425***	-0.0505	-3.633*
	(0.0220)	(0.124)	(0.115)	(2.127)
Necessities patents	0.129***	0.548***	-0.0818	-4.685*
	(0.0240)	(0.151)	(0.110)	(2.812)
Transporting patents	0.138***	0.559***	-0.0807	-4.701*
	(0.0257)	(0.153)	(0.122)	(2.808)
Chemistry & Metallurgy patents	0.169***	0.732***	-0.160	-6.288*
	(0.0320)	(0.191)	(0.158)	(3.644)
Textiles & Paper patents	0.0412***	0.207***	-0.0484	-1.850*
	(0.00957)	(0.0703)	(0.0628)	(1.108)
City FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Province × time	Y	Y	Y	Y
Time-invariant controls × post	Y	Y	Y	Y
Time-varying controls	Y	Y	Y	Y
Observations	4,064	4,064	4,064	4,064

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1980 to 1990. Innovation, as outcome variable, is proxied by the number of patents in each year. Post is denoted as 0 for the year before the reform in 1985 and 1 from 1985 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Table A10. The effect of chamber system on innovative performance: general and branch chamber density, DID and DID + IV

	(1) DID	(2) DID+IV	(3) DID	(4) DID+IV
Variable of interest	*(general chamber density)*		*(branch chamber density)*	
Outcome variables:				
Overall patents	0.493***	0.956***	0.0773***	0.708**
	(0.0581)	(0.194)	(0.0236)	(0.352)
Fixed construction patents	0.137***	0.252***	0.0172***	0.175*
	(0.0284)	(0.0640)	(0.00658)	(0.0979)
Mechanical engineering patents	0.231***	0.452***	0.0365***	0.338*
	(0.0392)	(0.0950)	(0.0114)	(0.174)
Physics patents	0.276***	0.432***	0.0312**	0.238
	(0.0454)	(0.0938)	(0.0123)	(0.146)
Electricity patents	0.238***	0.397***	0.0250**	0.244*
	(0.0424)	(0.0876)	(0.0117)	(0.137)
Necessities patents	0.284***	0.511***	0.0319**	0.348*
	(0.0405)	(0.106)	(0.0127)	(0.186)
Transporting patents	0.301***	0.522***	0.0404***	0.338*
	(0.0446)	(0.111)	(0.0133)	(0.184)
Chemistry & Metallurgy patents	0.368***	0.684***	0.0492***	0.482*
	(0.0462)	(0.134)	(0.0163)	(0.246)
Textiles & Paper patents	0.0972***	0.193***	0.00760	0.234**
	(0.0224)	(0.0535)	(0.00576)	(0.110)
City FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Province × time	Y	Y	Y	Y
Time-invariant controls × post	Y	Y	Y	Y
Time-varying controls	Y	Y	Y	Y
Observations	4,064	4,064	4,064	4,064

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1980 to 1990. Innovation, as outcome variable, is proxied by the number of patents in each year. Post is denoted as 0 for the year before the reform in 1985 and 1 from 1985 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Table A11: The effect of chamber system on innovative performance, normalised value

	(1)	(2)
	DID	DID+IV
Outcome variables	**Variable of interest**: *post_x_chamber density*	
Overall patents	0.489***	2.078***
	(0.0808)	(0.514)
Fixed construction patents	0.438***	1.983***
	(0.093)	(0.588)
Mechanical engineering patents	0.499***	2.239***
	(0.094)	(0.603)
Physics patents	0.517***	2.075***
	(0.106)	(0.580)
Electricity patents	0.486***	2.067***
	(0.102)	(0.580)
Necessities patents	0.479***	2.131***
	(0.089)	(0.564)
Transporting patents	0.514***	2.056***
	(0.090)	(0.539)
Chemistry & Metallurgy patents	0.534***	2.329***
	(0.098)	(0.587)
Textiles & Paper patents	0.336***	1.785***
	(0.077)	(0.574)
City FE	Y	Y
Year FE	Y	Y
Province × time	Y	Y
Time-invariant controls × post	Y	Y
Time-varying controls	Y	Y
Observations	2,794	2,794

Note: Robust standard errors clustered at the city level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of panel is from 1980 to 1990. Innovation, as outcome variable, is proxied by the number of patents in each year. To make the performance more comparable across industries, we also normalise the outcome variables. Post is denoted as 0 for the year before the reform in 1985 and 1 from 1985 and 1990. Controls include agricultural suitability, distance to coast, ruggedness, longitude, latitude, flood and drought.

Chapter 4: Birth Quota and Gender Bias: Evidence from the Two-Child Policy

1 Introduction

The 'missing women' phenomenon has persisted and intensified over the past 50 years, with female deficit rising from 61 million in 1970 to 142.6 million in 2019 (Chao et al., 2019). Current literature has explored various factors contributing to the issue, including sex-diagnostic technology (Chen et al., 2013), desired fertility (Jayachandran, 2017), land reform (Almond et al., 2019) and inheritance rights (Bhalotra et al., 2020). One frequently speculated culprit is family planning policy. Under natural conditions, parents satisfy potential son preferences by giving birth until a son is born. In such cases, the gender ratio should be balanced rather than skewed towards males (Xu & Pak, 2015). However, with restrictions on family size, the cultural preference for sons can no longer be smoothed out by additional fertility.[28] As a response, parents may practise prenatal and postnatal discrimination against daughters, lowering their probability of survival relative to that of sons. Thus, in theory, the female deficit may be increased by the application of family planning policies.

However, empirically, the relationship between fertility policies and sex selection is still an open question. The main challenge to answering it is finding a suitable exogenous variable with which to perform the analysis. The seminal works relied on the One-Child Policy (OCP) applied to Han-ethnic people in China. For example, based on the strictness of the OCP, Ebenstein (2010) attempts to prove that it led to a sex imbalance. By comparing the sex ratio of Han and ethnic minorities using a difference-in-differences (DID) framework, Li et al. (2010) demonstrate there was a positive relationship between the OCP and the sex ratio at birth. However, one critical point that has been misrepresented is that the OCP was not a gender-blind policy *per se* (Jayachandran, 2017). Especially when implementing OCP in practice, two-thirds of provinces allowed firstborn-girl families living in rural areas to have a second child (Scharping, 2013; Zhang, 2017).[29] In a more granular policy division, this revision was thus considered by some as the '1.5-child policy' (Hvistendahl, 2010; Loh and Remick, 2015). Based on the official estimation, nearly 52.9% of the overall population in China was exposed to the 1.5-child policy. Mixing consideration of the 1.5-child policy and the OCP could lead to severe

[28] One classic example is provided by Jayachandran (2017): 'Suppose a couple strongly wants to have at least one son. If they wish to have six children, there is only a 1% chance they will be without a son, but if they wish to have only two children, there is a 24% chance'.

[29] Around 75%–80% of people were living in the rural area in China at the time.

sample selection bias. From the second parity, the DID estimation would compare firstborn-girl Han families with both firstborn-girl and firstborn-boy ethnic minority families. Knowing that families with firstborn girls have a much higher tendency to have boys in second births, the estimation of the sex ratio would be overestimated.

To overcome this challenge, in this paper, we propose the consideration of an understudied family planning policy – the two-child policy (TCP) – to seek causal evidence linking birth quota and family reactions to childbirths. In its literal meaning, the TCP stipulates that households are allowed to have only two children. If couples do not obey the fertility rule and give birth to over two children, they will be penalised for the above-quota deliveries. Punishments usually take the form of cancellations of government subsidies, ineligibility for government jobs, or financial penalties, among others.

Globally, the application of the TCP is more conventional than that of the OCP, since the associated fertility quota is closer to the desired family size and thus leads to less resistance against governments. Up to this point, the TCP has been applied in various countries, such as Vietnam, Myanmar, Singapore, China, Iran and Egypt.[30] Well-known slogans, such as 'Two is Enough' and 'Stop at Two', have also been popularised over several rounds of fertility campaigns. Aside from the popularity of this policy, what is more important for our research is that it no longer treats families differently based on the first child's gender. Due to the policy's gender-blindness, research on the TCP can avoid the limitations of the OCP and directly compare gender-related outcomes.

This paper draws on fertility policies applied to ethnic minority groups in China. It is well known that China implemented a nationwide TCP from 2016 to 2020. However, little attention has been given to early TCP experiments among ethnic minority groups. Unlike the Han majority, 54 out of 55 ethnic minority groups were exempt from the OCP beginning in 1980 due to their relatively small share of the population along with political considerations (Scharping, 2013).[31] However, this did not mean that minority groups were exempt from family planning. Following government policies, minorities were divided into two categories: groups subject to the TCP and groups subject to the '2+ child policy'. For example, Yao, Miao and Dong people living in the province of Guangxi were allowed to have two children. In contrast, Daur, Evenki and Oroqen people

[30] See more detailed information in the brief introduction of application of the TCP worldwide given on Wikipedia: https://en.wikipedia.org/wiki/Two-child_policy

[31] The only exception is the Zhuang ethnic minority. The overall population of the Zhuang was over 10 million before the fertility policy implementation for minority groups. Thus, like the Han majority, the Zhuang minority was required to abide by the OCP. Since the Zhuang have followed OCP, we excluded them in our primary estimations. We discuss the categorisation of China's fertility policies after 1980 in detail in the Institutional Background section.

in the province of Inner Mongolia were allowed to have more than two children (Scharping, 2013). Based on our estimations, over 60 million minority people were subject to the TCP beginning in the 1980s. In comparison, 10 million minority people in China were allowed to have more than two children.

Knowing that differential application of the TCP across ethnic groups has been embodied in various regulations, we carefully exploit variations across provinces in terms of the timing by which the TCP was enforced beginning in 1980 to estimate the causal effect of the TCP on sex selection at birth (Banister, 2004; Park & Han, 1990; Peng, 1996; Scharping, 2013). We collect and verify novel data on the provincial implementation of the TCP at the ethnic-group level from multiple sources and link this information with the 1990 and 2000 micro-level population census data. Fertility policies for minority groups have always been issued simultaneously within provinces, enabling us to measure minority groups following the 2+ child policy to examine the effects of the TCP.

Our empirical strategy draws on the demographic law that the firstborn child's gender will be biologically normal, while any family interventions on their children's sex happen at higher birth orders (Anukriti et al., 2020; Bhalotra et al., 2020). This principle has been verified by the stylised facts in China (Almond et al., 2019; Zeng et al., 1993). In our paper, instead of taking this law for granted, we also verify the randomness of firstborn child gender according to empirical data. The fundamental assumption underlying our approach is that without the enforcement of the TCP, outcomes relevant to the second childbirth would be identical between firstborn-girl and firstborn-boy families.

Our analysis yielded the following salient findings. We first provide causal evidence on the relationship between the TCP and the sex ratio. We found that after the launch of the fertility policy, for ethnic minority families subject to the TCP, there was a significant increase of 5.2% in the probability of having a son if the firstborn child was a daughter. In contrast, there was no substantial increase in the sex ratio for families with a firstborn son. This male-biased sex ratio was persistent when we included additional household variables representing time-invariant factors and province-by-year of birth fixed effects to absorb time-varying province characteristics. We also performed a placebo test to demonstrate that the ratio of boys to girls at the second birth was unaltered for minority families with more than two children pre- and post-TCP. We then go on to specify the effects of the policy implementation at different stages. Our triple-differences estimation shows that during the legalisation phase, the magnitude of the coefficient was more than doubled when compared with the announcement phase, increasing from 3.2% to 8.3%.

Second, to confirm the sex selection of births, we studied the birth spacing between the first two births as an outcome. The direction and magnitude of the measured coefficient were the synthetic results of two forces. First, the TCP may have stimulated in families with firstborn girls a stronger desire to have a second child at a shorter birth interval. Second, if the second child in firstborn-girl families was a daughter, families may have resorted to gender intervention as a response, and the observed interval between births might tend to be longer. In this manner, a positive coefficient would indicate that gender selection activities were widespread given the implementation of the TCP. Our results support this assumption and reveal a significant increase in birth spacing in firstborn-girl families relative to firstborn-boy families post-TCP. This finding further strengthens our argument that, given the prevalent culture of son preference in Chinese society, policy constraints on family size are detrimental to the gender ratio at birth at higher birth orders.

Third, we further explored the possible mechanisms of gender selection. In China, the diffusion of ultrasound technology began in the 1970s, facilitating prenatal sex detection and sex-selective abortion. Following Chen et al. (2013) and Almond et al. (2019), we used the time at which the ultrasound machine was first introduced in provinces and found that the diffusion of the ultrasound machine significantly boosted the sex imbalance during the phase of TCP legalisation. However, this tool for prenatal sex detection did not promote female foeticide during the period of TCP announcement and legalisation. This observation was confirmed by examining the relationship between the TCP, ultrasound and birth spacing.

Finally, we investigated a potential increase in the number of artificial twins in second births after the TCP. To legally have more children under the TCP, one possible approach is to falsely claim second and third children as twins. We observed that during the legalisation phase, the probability of having twins during the second birth increased significantly among firstborn-girl families, indicating that having artificial twins became a popular strategy. This result was mainly associated with Han–minority mixed-marriage families, who are heavily influenced by Confucianism and patriarchy. Generally, we might consider a case in which an elder child is registered along with a younger one as a twin. In this case, the reported birthdate of the twins is likely to follow the birthday of the younger child since the parents must wait until the younger child has been born before registering the artificial twins together. As a result, the interval between the two observed births would be longer. This was also reflected in our estimations.

Our paper speaks to several strands of the literature. First and foremost, we provide rigorous and comprehensive empirical evidence on policy constraints regarding family size and parents' responses to these constraints. Although there have been several

empirical studies on how the OCP affects sex imbalance at birth, the results are limited by the major defect in the OCP design and are thus more likely to be biased (Banister, 2004; Das Gupta, 2005; Ebenstein, 2020; Li et al., 2011). Some works have discussed the TCP in different contexts; however, these studies have focused on positive association or qualitative case study analyses (Ngo, 2020; Wei & Zhang, 2014). A rigorous and systematic study discussing the TCP and sex imbalance remains a conspicuous gap in the literature. To fill this gap, we investigated one of the most gender-biased countries in the world and highlight the enormous costs of gender inequality. Second, our paper also adds to the discussion on the interaction between policies and culture. The emerging literature shows that culture may be responsive to procedures or policies able to correct unreasonable behaviour (Ashraf et al., 2020; Bau, 2021; Huang & Zhang, 2021). In line with this, our paper indicates that ill-considered policies may stimulate the practices of pre-modern cultures within modern society.

The rest of the paper is structured as follows. Section 2 describes the institutional background for the study. Section 3 introduces the data source and hypothesis testing. Section 4 presents the findings on the relation between the TCP and sex selection. Section 5 investigates the relationship between the TCP and artificial twins. Finally, Section 6 concludes the paper.

2 Institutional background

2.1 Fertility policies for ethnic minorities

There are 56 ethnic groups in China: the dominant Han and 55 minorities. According to the census in 1990, the aggregate population of minorities was 91.20 million, accounting for 8.01% of the total population. Except for the Zhuang ethnic minority, all other minority groups included less than 10 million people each. With a population of over 9.85 million, the Manchu was the largest in 1990, while the Lhoha, with only 2,322, were the smallest. Given their relatively small population share and political considerations, the government has introduced various policy favours and benefits to ethnic minorities (Scharping, 2013). For example, minorities enjoy extra points in the National College Entrance Exam; priority employment and promotions in state-owned enterprises and government sectors; and poverty alleviation funds tilted towards minority gathering regions.

This preferential treatment is also reflected in fertility policies. When family planning policies were introduced in the 1960s, minorities were more favourably treated than their

Han counterparts (Jia & Persson, 2020).[32] In the 1970s, the government began its launch of a stricter fertility policy called the 'Later, Longer, Fewer' (LLF) campaign, to promote later marriage, longer birth intervals and fewer children. Although the Han people were fully restricted from having three children, reducing the fertility rate substantially, exemptions were fully granted to all minority groups (Chen & Fang, 2021). That is, all minorities remained able to have any number of children at will (Scharping, 2013).[33]

However, beginning in the 1980s, Chinese leaders decided to turn to a much more stringent family planning policy due to increasing fear of a Malthusian catastrophe. To achieve the goal of reducing the population and birth rate as soon as possible, an open letter from the central committee in 1980 formally stated that all the minority groups would be required to follow the family planning policy, marking the first time in history that minority groups had been included in China's fertility programme. However, minorities continued to experience advantages compared to the Han majority.

Based on differences in birth quota requirements, the application of family planning policy in China can be divided into three categories. First, minority groups with a population of less than 100,000 or with strong ethnic awareness followed a 2+ child policy and were allowed to have more than two children per couple. Second, minorities with a population between 100,000 and 10 million had to adapt to the TCP and were allowed to have two children. Third, the Zhuang minority, with over 10 million people, had to obey the strict OCP, similar to the Han majority. With this principle as a foundation, provinces have made certain adjustments according to the composition of ethnic minorities in their regions.[34] For example, in areas such as Guizhou, Yunnan and Qinghai, having over two children was entirely based on having an ethnic population line of 100,000. However, in Inner Mongolia, Mongolians with rural hukou, Daurs, Ewenkis and Oroqen could have three children, while other minority groups in the province were only allowed to have two children. Our study thus makes a distinction at the ethnic level by region. According to our estimations, over 60 million minority people were subject to

[32] China's earliest family planning policy began in December 1962, when the central government issued the Instructions on Seriously Advocating Family Planning. However, it was not effectively promoted until 1970.

[33] Rossi and Xiao (2020) argue that there may have been a spatial spillover effect on fertility decisions between the Han majority and minority groups. However, the LLF never encouraged the notion of birth control within minority groups *per se*.

[34] Minority people tend to locate in gathering regions, such as in ethnic minority autonomous cities and counties. Thus, most provinces only contain parts of ethnic minority groups, and it is unnecessary to implement a policy for all minority groups. We attached the map in the Appendix to illustrate the geographic distribution of ethnic minorities. It suggests that minority groups are highly spatially agglomerated.

the TCP. At the same time, 10 million minority people were allowed have more than two children.

It needs to be highlighted that Han–minority mixed marriages are prevalent in many parts of China, and this has been considered in the policy design of family planning programmes since 1980. In the case of unions between Han males and minority females, the females usually regard such unions as an avenue to social advancement, while the males can save a large amount in terms of the bride price. Regional governments are likely to promote such forms of marriage as they are generally viewed as instrumental in alleviating ethnic tensions. Han–minority mixed marriages are mainly promoted in autonomous areas and places with solid national antagonism. To further promote ethnic fusion, many provinces allow Han–minority couples to follow the fertility policy of the minority side and thus to have two or more children (Scharping, 2013).[35]

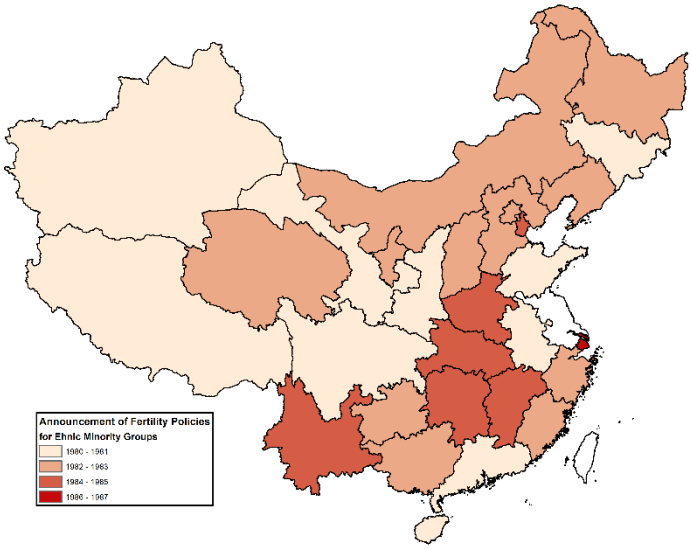
While family planning policy is formulated by central government agencies, such as the National Population and Family Planning Commission, it is always marked with localised features and implemented differently in different provinces (Gu et al., 2007). Regional governments are given substantial discretionary power in the policy's implementation, and there are thus clear spatial and temporal variations at the regional level. This feature is consistent regarding the fertility policies applied to minority groups. For instance, the provinces of Guangdong and Tibet issued government suggestions on birth quota for minorities in 1980, while the promulgation of similar regulations in Jiangxi and Henan occurred in 1985 (Figure 1).[36]

The intensity of family planning policy has also varied over time. It was relatively mild when first enacted, and government documents were usually published as regulations, recommendations or suggestions that were not legally binding. Under these circumstances, even when people had more children than the policies suggested, punishment, such as financial penalty, was unlikely. The central government later noted the laxity and loss of control during policy enforcement and called for strict adherence to population targets. In 1986, the No. 13 central document provided much more stringent control than the previous stipulation (Peng, 1996; Scharping, 2013). Strict monetary punishment

[35] Note that the Liaoning, Shandong and Henan provinces have not provided such immunities. Han–minority mixed couples living in these provinces are required to follow the OCP and thus can only have one child. We removed these families from the sample of those following the TCP in the following analysis.
[36] As Figures 1 and 2 show, the province of Zhejiang has never formulated fertility policies for ethnic minorities. As only 0.85% of the overall population (around 40,000 people) in Zhejiang are from minority groups, the provincial government asked all minorities to follow the OCP, similar to the Han majority, in the region. We thus excluded Zhejiang from our analysis.

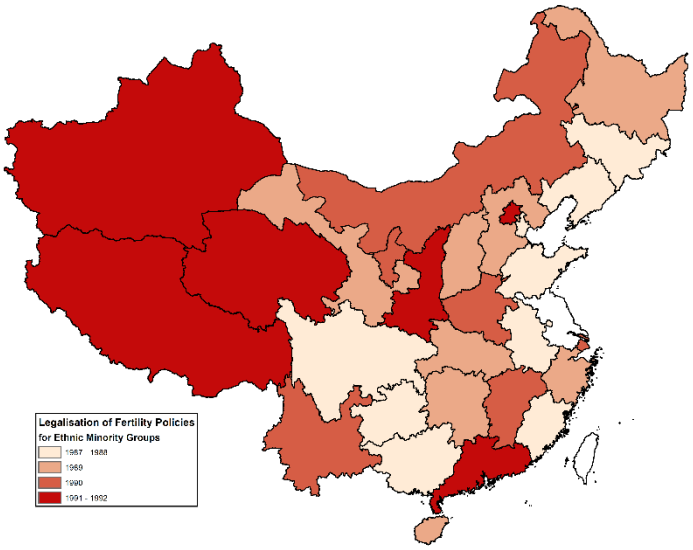
for excess fertility was introduced and would take effect after that.[37] At the same time, the performance appraisal and promotion of officials became linked with the total fertility rate, further contributing to the policy's strength. Provincial governments began to gradually implement the formal legalisation of such policies in 1987 (Ebenstein, 2010). Figure 2 shows the spatial variations regarding the legalisation of fertility policies at the provincial level.

Figure 1: Geographic distribution of announcement of fertility policies for minority groups



[37] If minorities following the TCP violate the rules and have a third birth, they are asked to pay 50% of the provincial fine rate.

Figure 2: Geographic distribution of legalisation of fertility policies for minority groups



2.2 Culture of son preference

Considered alone, the fertility reduction induced by the TCP should not result in gender discrimination. The literature has long taken account of the fact that sex-biased activities are the result of a combination of sex preferences and fertility factors (Becker, 1991; Ben-Porath & Welch, 1976; Li et al., 2010). By limiting the birth quota, family planning policy can catalyse implicit cultural preference for sons, leading to responses that bring about explicit gender inequality.

In the Chinese context, similar to that in many other developing countries, son preference is traditionally rooted in society (Das Gupta et al. 2003). It has been argued that China's first Neolithic Revolution and the introduction of long-term intensive agriculture made land valuable and led to a rigid class structure. Originally, sons had to remain near their parents to defend them and inherit valuable land (Ebenstein, 2021; Putterman, 2008). Land values still play a crucial role even in modern China. Almond et al. (2019) show that land reform has significantly increased family income and the sex ratio at birth based on the adopted land responsibility system. In this manner, patrilocality, the cultural norm according to which sons provide care for their elderly parents, and daughters leave home following marriage to provide care for their in-laws, prevails in China. The prevalence of patrilocality has been verified by the Ethnographic Atlas and Global Data Lab, as shown in Figures A1 (Kirby et al., 2016; Murdock, 1967). Tightly linked to patrilocality, Chinese sons traditionally provide old-age support for their parents

(Jayachandran, 2015).[38] Ebenstein and Leung (2010) show that this norm regarding old-age support is another key factor in the desire to have sons when social welfare has not been fully established. Families without sons are also more likely to participate in the pension program. In summary, sons are generally valued because they inherit family property, contribute to field labour and household income and shoulder the responsibility of caring for parents in their old age.

It is worth noting that there is significant heterogeneity in the degree of sex preference between the Han majority and minority groups. Compared to other ethnicities, Han people are more deeply influenced by Confucianism. Generally, reinforcing the patriarchy to maintain social order and morality is the core of Confucianism.[39] Following the values of Confucianism, the status of men is always higher than that of women in traditional Han Chinese families, and wives are required to be completely subordinated to their husbands. Further, it is the wives' responsibility to take care of other family members and do all the housework (Zhang & Liu, 2012). To further rationalise and sanctify the idea of male superiority, sons play an essential part in ancestor worship rituals under Confucianism (Jayachandran, 2017). Further, only the name of a male can be recorded in an official family genealogy. In comparison, other ethnicities less influenced by Confucian ideology tend to subscribe to different norms from the majority Han in terms of son preference (Attané, 2009; Poston et al., 2006). For example, the patrilineal system within the Yi ethnic minority is shaped by highly hierarchical clan organisations, while for the ethnic Hui, the religion of Islam plays a vital role in shaping patrilineality (Tilt et al., 2019). Thus, although the patrilineal system is also found in ethnic minority groups, their degree of son preference is usually much lower than that of the Han majority.

3 Data sources and hypothesis testing

Our analysis combines family responses on childbirth from micro-censuses with information we assembled regarding the timing of the implementation of family planning policies for minorities at the province level. Data on ultrasound use was also collected to investigate possible mechanisms of gender selection.

[38] A common Chinese saying also verifies the notion of bringing up sons to support parents in their old age.

[39] Confucianism encourages the patrilineal system not only in Han China but also in other Asian countries, such as Vietnam, Japan and South Korea (Jayachandran, 2015).

3.1 Micro-level population censuses

The primary data adopted in this article were taken from the 1% sample of the 1990 and 2000 Chinese population censuses, which cover the period of family planning policy implementation for ethnic minorities. The census data report individuals' ethnicity, relationship to household head, gender, education level, year and month of birth, type of hukou (urban/rural), province, marital status and number of siblings, among other measures. Noting that China's census does not explicitly mention the birth order of children within families, we followed Almond et al. (2019) in using information on the relationship to household head and date of birth to identify birth order. We also required the number of children we matched to be equal to the number of reported surviving births in the family. To eliminate issues resulting from underreporting and children leaving original families, we only focussed on children aged 4–16 in the census year.[40]

3.2 Rollout of fertility policies for ethnic minority groups

We collected information on the family planning policies for minority groups from several sources. Differential treatment across ethnic groups has been embodied in birth planning policy documents and was even officially written into the Law on Regional National Autonomy (Peng, 1996). We used several law and regulation databases (i.e., PkuLaw and Xihu Law Library) to collect original official documents at the provincial level. In cases in which there were no records on certain provinces in the legal database, we also consulted local gazetteers. As archives to record local history, gazetteers generally have a separate chapter detailing family planning policies.[41] We read through all the relevant documents and identified the specific fertility policies applying to different ethnic groups in each province. Finally, we used the information on fertility policies collected by Gu et al. (2007) and Scharping (2013) to further verify the accuracy of our rollout data.

The summary statistics are reported in Table 1. Among first births, the fraction of males remained stable at 0.51 both pre- and post-TCP. For groups asked to follow the TCP, the fraction of males remained at approximately 0.5 among firstborn-boy families before and after the reform. The most striking change can be observed among second births in firstborn-girl families. Here, the fraction of males increased from 0.53 pre-TCP to 0.59 post-TCP. However, we did not find any noticeable change in groups that were allowed

[40] Zhang and Zhao (2006) show that the underreporting of children is quite common when they are younger than four years old. The earliest age for marriage in Chinese society is 16 years.

[41] The accuracy of information provided gazetteers has repeatedly been examined in empirical research, such as by Almond et al. (2019) and Walder (2014).

to have more than two children; before and after TCP implementation, the fraction of males in this case remained consistent in both firstborn-girl and firstborn-boy families.

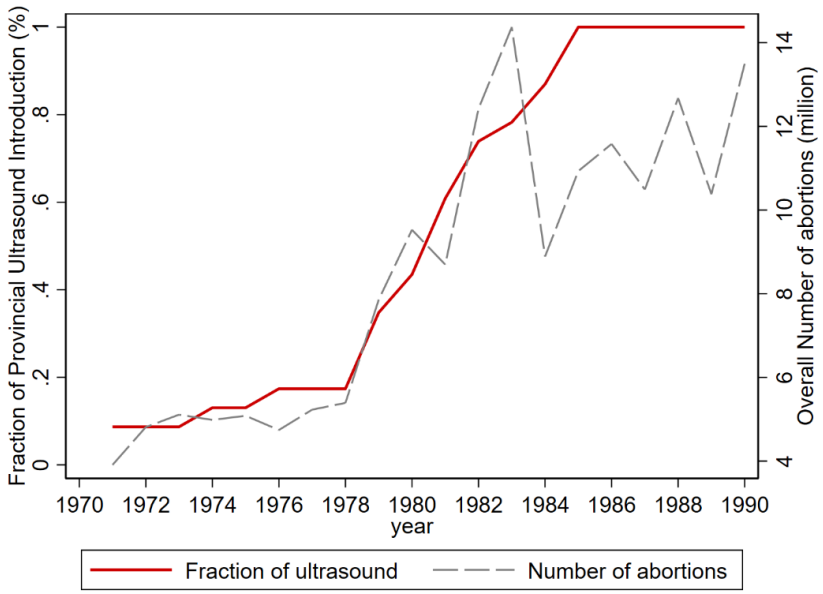
Table 1: Summary statistics

	Pre-TCP		Post-TCP	
First births				
	Mean	Obs.	Mean	Obs.
Male	0.5105	95063	0.511	188554
Second births of groups following TCP				
	Mean	Obs.	Mean	Obs.
Male following a firstborn girl	0.5345	31999	0.5863	49087
Male following a firstborn boy	0.5053	31446	0.5039	45780
Second births of groups following 2+ Child Policy				
	Mean	Obs.	Mean	Obs.
Male following a firstborn girl	0.5411	3855	0.5323	9232
Male following a firstborn boy	0.5089	4123	0.4985	9279

3.3 Introduction of ultrasound scans

Before the introduction of ultrasound technology, parents manipulated the ratio of surviving girls to boys by the practice of son-biased fertility stopping (Arnold et al., 1998; Bhalotra & van Soest, 2008) along with multiple forms of neglect involving breastfeeding duration, immunisation and nutrition (Jayachandran & Kuziemko, 2011; Oster, 2009). However, with the support of prenatal sex detection technology, families became able to select the sex of the foetus at a lower cost (Anukriti et al., 2020; Chen et al., 2013). The main source of the diffusion of ultrasound technology at the province level is provided by Almond et al. (2019). We added the diffusion coefficient regarding ultrasound in other provinces into our dataset. As discussed by Almond et al. (2019), after the introduction of the ultrasound machine in Shaanxi in 1965, other provinces gradually adopted the technology, beginning in the 1970s. Figure 3 shows the positive correlation between the rollout of the ultrasound and the overall number of abortions.

Figure 3: Introduction of ultrasound and abortion rate



3.4 Hypothesis testing

The hypothesis we tested in this paper states that given the introduction of the TCP, male offspring are favoured by human intervention during second births, while the gender of firstborn children is left to providence. The urge to secure the birth of a son grows if it has not been satisfied during preceding births, thereby driving up the distortion for following births. To meet this assumption, we first tested the effectiveness of the TCP on the fertility and exogeneity of the sex of the firstborn to satisfy the premise. Then, we discussed our regression specifications and precisely formulated our empirical specification.

To ensure the validity of TCP, we had to prove that the TCP effectively changed households' decisions on family size in practice. We thus regressed the implementation of the TCP on family size, as shown in Table 2. Columns 1–4 present the estimations regarding the groups subject to the TCP. Column 5 shows the results regarding the groups allowed to have more than two children, playing the role of a placebo. Panel A shows the probability of families having three children after the TCP was introduced, revealing the TCP significantly decreased the number of families with three children by 2.89 to 4.682,

conditional on the controls we applied. In contrast, groups following the three-child policy or even more liberal fertility policies were found to be immune to the implementation of the TCP, demonstrating the effectiveness of the TCP. Panel B shows the relationship between the TCP and the probability of families having two children, revealing that the TCP was unlikely to lead to the desired fertility rate of fewer than two children.

Table 2: Validity of TCP on family size, household-level analysis

	(1)	(2)	(3)	(4)	(5)
Panel A:		2+ kid-family (0/1)			Placebo test
TCP	-4.682***	-4.766***	-4.251***	-2.890***	1.148
	(1.089)	(1.061)	(0.997)	(0.899)	(1.98)
Observations	249,283	249,283	249,281	249,281	34,327
R-squared	0.169	0.176	0.232	0.251	0.329
Panel B:		2 kid-family (0/1)			Placebo test
TCP	0.93	0.815	0.108	0.36	3.738
	(2.09)	(2.061)	(1.784)	(0.467)	(2.871)
Observations	204,142	204,142	204,140	204,140	21,898
R-squared	0.184	0.192	0.315	0.325	0.27
Province FE	Y	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y	Y
Census FE	N	Y	Y	Y	Y
Other controls	N	N	Y	Y	Y
Province linear trend	N	N	N	Y	Y

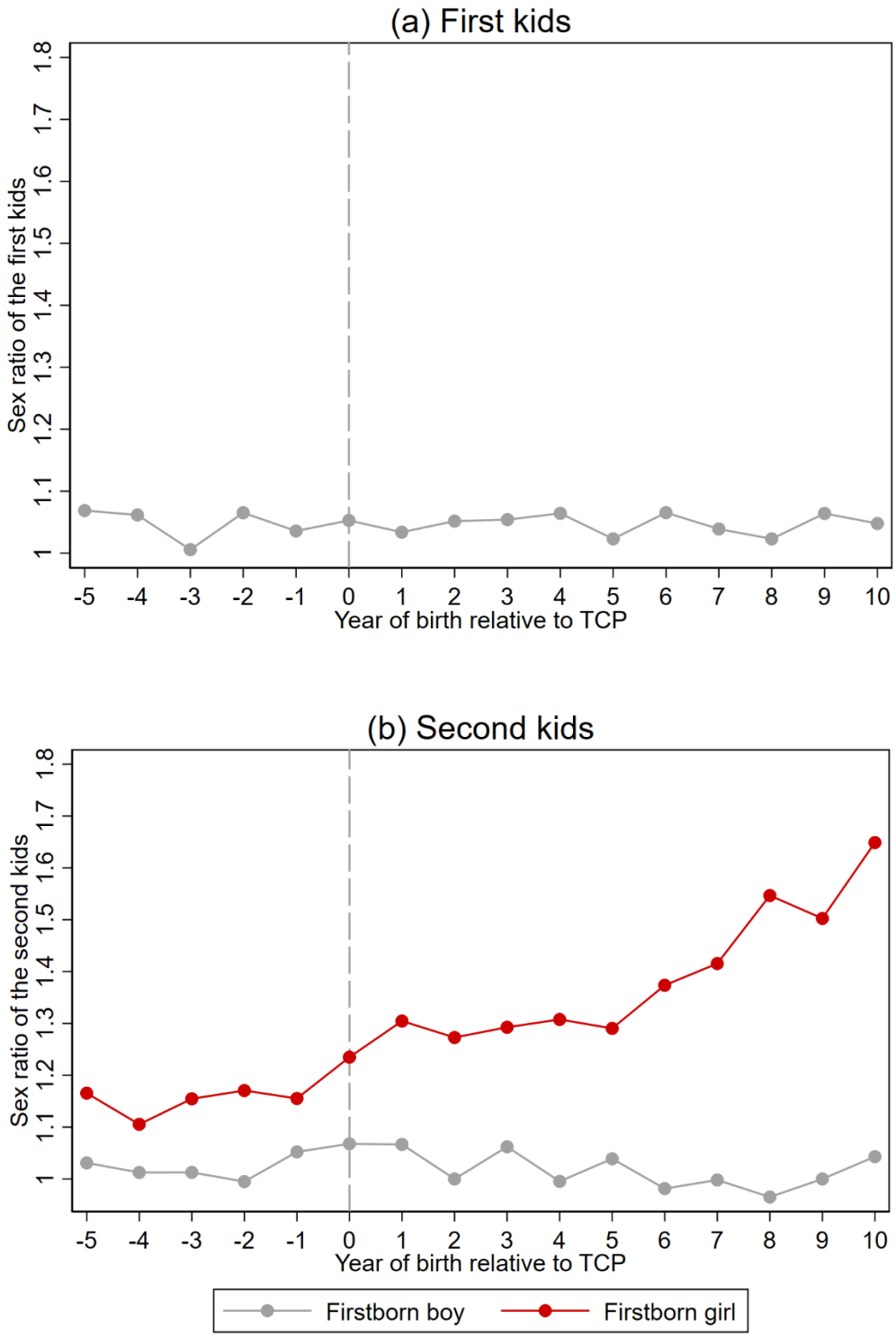
Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to the second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

A key assumption of our strategy for determining the effect of the TCP is that the sex of the first child is randomly determined. Figure 3 presents the sex ratio before and after the TCP at different birth orders. Figure 4(a) shows the sex ratio of firstborn children by year relative to the introduction of the TCP, revealing the sex ratio remained stable pre- and post-TCP. We also performed a regression between TCP and sex ratio, as shown in Table 1A, further indicating the irrelevance of the TCP on the sex ratio of firstborn children. Both results support the assumption that the gender of firstborn children is quasi-random.

Figure 4(b) compares the sex ratio of second children among firstborn-girl and firstborn-boy families before and after the TCP. It reveals that both firstborn-girl and firstborn-boy families exhibited a trend towards gender change before the fertility policy was implemented. After the TCP was introduced, firstborn-girl families exhibited a sharp

increase in the probability of having a male second birth, from 1.1 to nearly 1.7. As a reference group, firstborn-boy families were unaffected by the implementation of the TCP, and the ratio remained around 1.

Figure 4: Sex ratio trends among TCP groups



4 Empirical results: effects on the sex ratio

4.1 Baseline results

We employed a triple difference strategy in our empirical specification to examine the link between the sex ratio of second children and the sex of firstborn children before and after the TCP. We only modelled the sex ratio for the second birth to mitigate concerns regarding endogeneity. To make our estimations accurate, we also excluded cases of multiple births, including twins, triplets, quadruplets, etc. We estimated the following specification:

$$Y_{i,p,s,t} = \alpha + \beta_1 TCP_{p,t} * Firstborn\ girl_{i,p,t} + \beta_2 Firstborn\ girl_{i,p,t} + \beta_3 TCP_{p,t} + X_{i,p,t} + \gamma_p + \delta_t + \theta_s + \mu_{pt} + \epsilon_{ipst} \quad (1)$$

Here, the subscripts i, p, s and t denote the families, the province, the census year and the year of birth, respectively. The dependent variable $Y_{i,p,s,t}$ is the outcome if family i has a child in year t in province p, as measured in census year s. The outcome variable is a binary variable equal to 1 if the child is a boy and 0 otherwise. For the analysis, the outcome variable was multiplied by 100 so that the coefficient of regression could be interpreted as a change in percentage points regarding the sex ratio at birth. $Firstborn\ girl_{i,p,t}$ is a dummy variable that identifies the sex of the first child. $TCP_{p,t}$ is equal to 1 if the child is born after the TCP and 0 otherwise. The variable of interest is $TCP_{p,t} * Firstborn\ girl_{i,p,t}$, while the coefficient β_1 measures the average probability of a family having a boy after the implementation of the TCP compared between firstborn-boy and firstborn-girl families. $X_{i,p,t}$ is a set of control variables including hukou status, ethnicity, mother's education and age. γ_p, δ_t and θ_s are province, year of birth and census year dummy variables, respectively. It is worth considering that there could be an omitted province-specific time-varying variable correlated with the TCP. We thus introduce the province-by-year of birth fixed effects with the goal of including time-varying province characteristics. This also helps us soak up variation in other fertility-relevant policies at the province level that may have coincided with the introduction of the TCP, such as land reform (Almond et al., 2019). All of our results are clustered at the provincial level.

The results are reported in Table 3. Panel A concentrates on the minority groups subject to the TCP. Column 1 of Panel A, which controls for the province and year of birth dummies, shows that the coefficient of interest $TCP_{p,t} * Firstborn\ girl_{i,p,t}$ (5.284) is

statistically significant at the 1% level and very large. The robustness of the results was checked by including census year and household level controls in columns 2 and 3, respectively. To further deal with possible omitted variables, we included province-year fixed effects to flexibly control any omitted province-year variation that could be correlated with the passage of the TCP. The results are reported in column 4 of Panel A. Overall, the results were consistent, revealing that the probability of the second child being male sharply increased in firstborn-daughter families. Almond et al. (2019) found that the sex ratio increased by 3% over time, led by the land reform in the early 1980s. Our results indicate that the TCP has also had an enormous impact on gender discrimination at birth.

To ensure that other factors were not driving our results, we ran placebo tests, as shown in Panel B. Given that fertility policies for minorities were always issued at the same time, we were able to examine whether the TCP was influential among groups allowed to have more than two children. The results in Panel B confirm that the development of son preference was mitigated by a higher birth quota allowing for additional childbirth. As predicted, columns 1–4 in Panel B reveal that even after TCP implementation, the sex ratio of second births remained in balance when additional childbirths were allowed.

As shown in Table B2 (Appendix), our estimations were found to be robust when including all births at the second parity or above. We also tested the robustness of our assessments by including children under the age of four along with conditions in which families reported a different number of surviving children than that observed (Tables B3 and B4). Overall, the results support the findings presented in Table 3.

Table 3: Sex ratio at the second birth

	(1)	(2)	(3)	(4)
	Child is male (0/1)			
Panel A: Minorities following TCP				
Firstborn girl*TCP	5.284***	5.283***	5.283***	5.251***
	(0.76)	(0.757)	(0.752)	(0.773)
Observations	158,312	158,312	158,309	158,307
R-squared	0.006	0.006	0.007	0.012
Panel B: Minorities following 2+ child policy (placebo test)				
Firstborn girl*TCP	-0.176	-0.203	-0.148	-0.104
	(1.632)	(1.629)	(1.708)	(1.759)
Observations	26,489	26,489	26,487	26,487
R-squared	0.002	0.002	0.006	0.012
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province-by-year of birth FE	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported. The full specification can be found in Appendix Table A1.

After estimating the effect of the TCP on sex imbalance, we examined whether the results of our regression were influenced by son-biased fertility stopping. If families decided not to birth second children when they already had a boy as a first child, our results would face a sample selection issue. It was therefore necessary to examine whether the TCP affected the fertility desires of firstborn-girl and firstborn-boy families. To test this, we regressed the dummy variable on having a younger sibling at the second birth on the TCP, as shown in Table 4. This regression helped us to detect the difference in fertility stopping at the second birth between firstborn-girl and firstborn-boy families. As shown in Table 4, all the coefficients were small and insignificant, indicating that firstborn-boy families still expected to have additional children when the policy allowed for it. This concern regarding selected second births was thus alleviated, and this issue did not bias our results. This finding also matched the situation regarding the total fertility rate before the year 2000 and the traditional notion of 'more children, more happiness' in China (Greenhalgh, 1988; Lu & Zhai, 2009).

Table 4: Examining sample selection: son-biased fertility stopping

	(1)	(2)	(3)	(4)
	Having younger sibling at the second birth (0/1)			
Firstborn girl*TCP	0.0196	0.0204	0.0131	0.0156
	(0.0122)	(0.0122)	(0.0113)	(0.011)
Observations	204,142	204,142	204,140	204,137
R-squared	0.192	0.2	0.325	0.345
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province-by-year of birth FE	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported. The full specification can be found in Appendix Table A2.

4.2 Different phases of policy implementation

In addition to examining the general effects of the TCP, we investigated whether there might be differential effects among the different phases of policy implementation. After 1987, provinces further legalised the TCP by promulgating specific laws, and the enforcement of the fertility policy was much stricter during this phase. Since disobeying the TCP began to entail paying hefty fines and losing employment, fewer families were willing to violate it during the legalisation phase (Gu et al., 2007). Thus, the influence of the TCP during the legalisation phase should have been much greater than that during the announcement phase. To investigate this empirically, we established the following specification:

$$Y_{i,p,s,t} = \alpha + \beta_1 TCP\ Announcement_{p,t} * Firstborn\ girl_{i,p,t} + \beta_2 TCP\ Legalisation_{p,t} * Firstborn\ girl_{i,p,t} + \beta_3 Firstborn\ girl_{i,p,t} + \beta_4 TCP\ Announcement_{p,t} + \beta_5 TCP\ Legalisation_{p,t} + X_{i,p,t} + \gamma_p + \delta_t + \theta_s + \mu_{pt} + \epsilon_{ipst} \quad (2)$$

Here, $TCP\ Announcement_{p,t}$ and $TCP\ Legalisation_{p,t}$ are dummy variables indicating whether a child was born after the TCP legalisation phase or after the announcement phase. By combining $Firstborn\ girl_{i,p,t}$ with $TCP\ Announcement_{p,t}$ and $TCP\ Legalisation_{p,t}$, these two coefficients were able to specify the effect of the TCP during these phases. Table 5 shows that both the announcement and legalisation periods had a significant influence on the sex ratio. However, the male-biased sex ratio was

significantly larger during the legalisation stage. The legalisation of the policy doubled the effect compared to that during the announcement phase, increasing to 8.293% when including the strictest controls (column 5 in Panel A). Similar to Table 3, we performed additional checks and ran placebo tests to ensure the validity of these results regarding the TCP. The results given in Panel B further confirm the accuracy of the main results.

Table 5: Sex ratio at the second birth: phases of policy implementation

	(1)	(2)	(3)	(4)
	Second Birth are Twins (0/1)			
Panel A: Minorities following TCP				
Firstborn girl	3.215***	3.212***	3.213***	3.157***
* TCP Announcement	(0.928)	(0.924)	(0.918)	(0.919)
Firstborn girl	8.276***	8.277***	8.276***	8.293***
* TCP Legalisation	(0.909)	(0.91)	(0.912)	(0.946)
Observations	158,312	158,312	158,309	158,307
R-squared	0.007	0.007	0.008	0.012
Panel B: Minorities following 2+ child policy (placebo test)				
Firstborn girl	-0.128	-0.167	-0.106	-0.125
* TCP Announcement	(1.80)	(1.805)	(1.84)	(1.878)
Firstborn girl	-0.27	-0.273	-0.232	-0.064
* TCP Legalization	(2.456)	(2.456)	(2.595)	(2.636)
Observations	26,489	26,489	26,487	26,487
R-squared	0.002	0.002	0.006	0.012
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province-by-year of birth FE	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported. The full specification can be found in Appendix Table A3.

4.3 Heterogeneous effect

The facts that Han-Minority mixed marriages following the minorities' fertility allows us to explore whether the skewed sex ratio results vary by the type of marriage. It helps us to understand the sensitivities of cultures during the TCP regulations. In Panel A of Table 6, we estimate the aggregate effect of TCP and find that if one of the spouses is

from Han ethnic, the sex imbalance effects would be doubled, rising from 3.2 to 7.0 percentage points. This may reflect the fact that although the culture of son preference is conventional in agricultural or nomadic civilisation, as the cornerstone of Confucianism, the notion of male superiority such as 'The Three Cardinal Guides and The Five Constant Virtues' and 'Women Inferior to Men' in the Han majority has further depreciated the status of females and exacerbate favouritism toward males. Thus, compared to pure minority families, the desire of having a boy will be even stronger in Han-Minority mixed families.

In Panel B of Table 6, we further investigate the effect of TCP in different implementation phases. Our estimations are consistent with findings in Table 5 which shows that in both types of families, reactions in the stage of legalisation are always much stronger than the stage of the policy announcement. Interestingly, our results show that if both parents belong to minor ethnic, our coefficients are small and insignificant in the announcement phase, which indicates that groups with less son preference are less motivated to intervene in children's gender when the fertility policy is not mandatory. In contrast, Han-Minority couples heavily affected by solid son-biased notions would further intensify their interference when legalising the procedure. The actual effect has been raised to 10.95 percentage points.

Table 6: Sex ratio at the second birth: heterogeneous effect

	(1)	(2)	(3)	(4)
	Second Birth are Twins (0/1)			
Panel A:	Both Parents are minorities		Either Parents is Han ethnic	
Firstborn girl*TCP	3.125***	3.162***	6.975***	7.010***
	(1.038)	(1.035)	(0.835)	(0.837)
Observations	59,402	59,397	98,873	98,868
R-squared	0.013	0.015	0.015	0.016
Panel B:	Both Parents are minorities		Either Parents is Han ethnic	
Firstborn girl	1.033	1.082	4.624***	4.649***
* TCP Announcement	(1.013)	(1.002)	(1.055)	(1.06)
Firstborn girl	5.697***	5.720***	10.90***	10.95***
* TCP Legalization	(1.327)	(1.328)	(0.931)	(0.939)
Observations	59,402	59,397	98,873	98,868
R-squared	0.013	0.015	0.015	0.016
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	Y	Y	Y	Y
Other controls	Y	Y	Y	Y
Province-by-year of birth FE	Y	Y	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported. The full specification can be found in Appendix Table A4.

4.4 Birth spacing

The influence of fertility policies on families' behaviour is also reflected by the float of birth spacing. Two possible forces push and pull the birth gap between first and second children within families. The first is an effect associated with prenatal and postnatal sex selection behaviours, which are unobserved in the fertility history data. In particular, the observed gap between the first and second births will be enlarged among firstborn-girl families if these sex-selection behaviours are engaged in. However, this widening of the birth interval may be diluted by a second force, that is, the TCP may also increase firstborn-girl families' desire to conceive a son earlier, even if it means shortening the breastfeeding of their first daughter. For this reason, the birth spacing is also likely to be shortened to some extent as a result of the TCP (Jayachandran & Kuziemko, 2011; Milazzo, 2018). We adopted the birth interval between first and second children as an outcome variable to examine these effects. The direction of the coefficient of interest was the result of the combination of these two forces. However, when activities such as sex-selective abortion, female infanticide and abandonment are widespread, the effect should still be positive.

The results shown in Table 7 agree with this judgment and show that the TCP widened the birth interval for firstborn-girl families. In line with our estimations in Tables 5 and 6, the results regarding the different stages in Panel B further confirm that sex selection behaviours occurred most prominently during the TCP legalisation phase. Interestingly, the adverse effects shown in our placebo tests indicate that there may have been spillover effects in response to the TCP in terms of birth spacing. In particular, spouses allowed to have three children or more may have delayed the birth of their second children. However, they were not interested in intervening in their children's sex due to their additional permitted fertility.

Table 7: Birth spacing at the second birth

	(1)	(2)	(3)	(4)	(5)
	Birth spacing (logged)				
Panel A:					Placebo
Firstborn girl*TCP	0.0281***	0.0272***	0.0212***	0.0100**	-0.0382**
	(0.00948)	(0.00946)	(0.00637)	(0.0047)	(0.0115)
Observations	158,310	158,310	158,304	158,302	26,486
R-squared	0.081	0.084	0.31	0.329	0.259
Panel B:					Placebo
Firstborn girl	0.00503	0.00318	0.00116	-0.00477	-0.0399**
* TCP Announcement	(0.00842)	(0.00833)	(0.00597)	(0.00498)	(0.0153)
Firstborn girl	0.0614***	0.0617***	0.0496***	0.0315***	-0.0351***
* TCP Legalisation	(0.0146)	(0.0146)	(0.0097)	(0.0068)	(0.00568)
Observations	158,310	158,310	158,304	158,302	26,486
R-squared	0.082	0.084	0.31	0.329	0.259
Province FE	Y	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y	Y
Census FE	N	Y	Y	Y	Y
Other controls	N	N	Y	Y	Y
Province-by-year of birth FE	N	N	N	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported. All the other terms are included but not reported. The full specification can be found in Appendix Table A5.

4.5 Ultrasound mechanism

Our estimations have confirmed that the probability of having a second-born son sharply increased among firstborn-girl families. It is thus natural to inquire as to the ways in which families distorted the gender of their second child. Theoretically, both prenatal and postnatal activities could lead to a biased sex ratio. Without detailed relevant surveys, such as the National Family Health Survey in India, direct investigation of the possibility of postnatal activities, such as infanticide or abandonment, is unlikely within the Chinese context. However, the introduction of ultrasound data makes it possible to test the feasibility of selective abortion. We followed Chen et al. (2013) in measuring this based on the year in which ultrasound technology was introduced. In particular, our specification is as follows:

$$Y_{i,p,s,t} = \alpha + \beta_1 TCP_{p,t} * Firstborn\ girl_{i,p,t} * Ultra_{p,t} + \beta_2 Firstborn\ girl_{i,p,t} * Ultra_{p,t} + \beta_3 TCP_{p,t} * Ultras_{p,t} + \beta_4 Ultra_{p,t} + \beta_5 TCP_{p,t} * Firstborn\ girl_{i,p,t} + \beta_6 Firstborn\ girl_{i,p,t} + \beta_7 TCP_{p,t} + X_{i,p,t} + \gamma_p + \delta_t + \theta_s + \mu_{pt} + \epsilon_{ipst} \quad (3)$$

We emphasised the interaction term $Firstborn\ girl_{i,p,t} * Ultra_{p,t}$, which captures the impact of the TCP on the probability of families with firstborn girls having sons relative to families with firstborn boys in the post-ultrasound period, examining its effect on the sex ratio. Table 8 reports the findings. Panel A shows that the triple interaction term$TCP_{p,t} * Firstborn\ girl_{i,p,t} * Ultras_{p,t}$ has a positive coefficient, although it is not significant. We went on to separate the effect of the TCP into different implementation stages (Panel B), and the results again stress the salience of policy enforcement. In particular, it is unlikely that the prenatal approach was universal in the initial TCP announcement phase. The differences between the coefficients in Tables 8 and 4 (–1.304 and 3.157, respectively) indicate that postnatal methods were the standard means of sex selection in the case of weak TCP enforcement. However, sex-selective abortion became routine when TCP was legalised. In the latter case, the gender-biased results could then be explained almost entirely by prenatal ultrasound.

Table 8: Mechanism of ultrasound on sex ratio at the second birth

	(1)	(2)	(3)	(4)	(5)
			Child is male (0/1)		
Panel A:					Placebo
Firstborn girl	0.961	0.964	0.975	0.996	7.122
*TCP*Ultra	(1.794)	(1.792)	(1.785)	(1.823)	(4.949)
Observations	158,312	158,312	158,309	158,307	26,487
R-squared	0.006	0.006	0.007	0.012	0.012
Panel B:					Placebo
Firstborn girl	-1.314	-1.311	-1.300	-1.304	7.161
*TCP Announcement*Ultra	(1.746)	(1.746)	(1.735)	(1.769)	(4.562)
Firstborn girl	8.246***	8.248***	8.264***	8.306***	3.568
*TCP Legalisation*Ultra	(1.243)	(1.244)	(1.242)	(1.287)	(3.376)
Observations	158,312	158,312	158,309	158,307	26,487
R-squared	0.007	0.007	0.008	0.012	0.012
Province FE	Y	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y	Y
Census FE	N	Y	Y	Y	Y
Other controls	N	N	Y	Y	Y
Province-by-year of birth FE	N	N	N	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported. The full specification can be found in Appendix Table A6.

We further verified the effectiveness of ultrasound technology by replacing our outcome variable with birth spacing. When abortion driven by ultrasound is prevalent, the interval between first and second birth should widen because abortion goes unobserved in the census survey. Our results in Table 9 are in line with those in Table 8. By combining our findings on the sex ratio and birth gap, our results show that when fines for over birth were much higher than the cost of ultrasound scans, sex selection by abortion became a conventional choice among firstborn-girl families. Taken together, the effects on the sex ratio and birth spacing strengthen our finding that the TCP exacerbated sex selection at the second birth. We also highlight that the impact of the TCP was intensified during the legalisation stage and confirm that ultrasound was used as a method of sex selection.

Table 9: Mechanism of ultrasound on birth spacing at the second birth

	(1)	(2)	(3)	(4)	(5)
	Birth Spacing (logged)				
Panel A:					Placebo
Firstborn girl*TCP*Ultra	0.0342	0.0370*	0.0406**	0.0303	-0.0474
	(0.0202)	(0.0194)	(0.0175)	(0.0181)	(0.0432)
Observations	158,310	158,310	158,307	158,305	26,487
R-squared	0.083	0.085	0.311	0.329	0.259
Panel B:					Placebo
Firstborn girl	0.00918	0.0111	0.019	0.014	-0.0496
*TCP Announcement*Ultra	(0.0207)	(0.0198)	(0.0183)	(0.0191)	(0.0461)
Firstborn girl	0.0757***	0.0767***	0.0656***	0.0468***	-0.0474*
*TCP Legalisation*Ultra	(0.0143)	(0.0143)	(0.00898)	(0.0068)	(0.0198)
Observations	158,310	158,310	158,307	158,305	26,487
R-squared	0.083	0.086	0.311	0.329	0.259
Province FE	Y	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y	Y
Census FE	N	Y	Y	Y	Y
Other controls	N	N	Y	Y	Y
Province-by-year of birth FE	N	N	N	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported. The full specification can be found in Appendix Table A7.

5 Empirical results: effects of artificial twinning

5.1 Baseline results

In the previous section, we showed that families with firstborn girls tended to manipulate the gender of their second children to fulfil the demand of son preference given the

constraints of fertility policies. However, sacrificing girls' lives is not the only option for families to respond to fertility quotas. A covert way to resist fertility policies is to have 'artificial twins' during the second birth. Under natural conditions, twinning during a second birth is a random event with a small probability and has thus been exempted from any form of penalty among families following the TCP. Given this pardon, couples can possibly bypass TCP restrictions by producing twins through human intervention. Anecdotes and news reports have revealed the prevalence of artificial twins in Chinese society.[42] For example, it has been reported that the number of twins has doubled within the past decade in Nanjing.[43] Huang et al. (2016) have also shown that twinning in the first birth significantly increased during implementation of the OCP. We here explore the possibility of second-birth artificial twinning as a reaction to the TCP.

To test the interplay between the TCP and the probability of artificial twinning, we included in our sample families who had twins in their second birth. We used Specification (1) and replaced the outcome variable with a binary variable representing whether twinning took place in the second birth to compare the probability of twinning between firstborn-girl and firstborn-boy families. In Table 10, column 1 of Panel A shows that application of the TCP did not increase the probability of having second-birth twins among firstborn-girl families at the aggregate level. We additionally split the TCP into two stages (Panel B). The results regarding the TCP announcement phase indicate people were reluctant to take the risk of creating artificial twins when policy enforcement was weak. However, after legalisation of the TCP, a 0.198% increase was found in second-birth twinning when comparing firstborn-girl and firstborn-boy families, indicating a positive relationship between the cost of breaking the law and the probability of distorting behaviour. In contrast, groups following the 2+ child policy were unlikely to engage in second-birth artificial twinning during any TCP stage (column 4).

Similar to Table 6, we explored the heterogeneity of the effect of the TCP on the probability of having second-birth twins by dividing the families into two groups: Han–minority mixed families and ethnic minority families. Columns 2 and 3 in Table 11 show that the positive and significant results in Table 10 were entirely driven by Han–minority mixed families. There is no evidence indicating that creating artificial twins was a conventional approach for pure minority couples at either TCP stage, while having one Han spouse in the family appears to have increased the probability of second-birth artificial twins. This result indicates that the previously identified positive correlation between TCP legalisation and having second-birth twins is not accidental. When the TCP became

[42] http://tech.sina.com.cn/d/2009-10-13/18473503428.shtml

[43] https://www.gmw.cn/01gmrb/2007-03/02/content_561792.htm

stricter, the positive coefficients increased and became more significant, at a 1% level, within Han–minority mixed families.

Table 10: Second birth are twins

	(1)	(2)	(3)	(4)
		Second birth are twins (0/1)		
Panel A:		Both minorities	Either Han	Placebo
Firstborn girl*TCP	0.108	-0.0711	0.225**	-0.132
	(0.08)	(0.112)	(0.104)	(0.181)
Observations	159,251	59,649	99,560	26,592
R-squared	0.008	0.016	0.011	0.009
Panel B:		Both minorities	Either Han	Placebo
Firstborn girl*TCP Announcement	0.0456	-0.0679	0.096	-0.218
	(0.0932)	(0.165)	(0.107)	(0.138)
Firstborn girl*TCP Legalisation	0.198*	-0.0751	0.440***	0.0277
	(0.111)	(0.133)	(0.157)	(0.292)
Observations	159,251	59,649	99,560	26,592
R-squared	0.008	0.016	0.011	0.009
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	Y	Y	Y	Y
Other controls	Y	Y	Y	Y
Province-by-year of birth FE	Y	Y	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported. The full specification can be found in Appendix Table A8.

5.2 Mechanism

It remains to be asked how people could succeed in having additional children in the first place. There are two plausible answers: People may take a fertility drug *ex-ante* or misreport single children as twins intentionally *ex-post*, increasing the apparent probability of twinning (Huang et al., 2016). We first examine whether the former was typical.

To create the illusion that their second and third children are twins, parents need to wait until the birth of the third child to register the second and third children together. Thus, the reported birthdate of the twins tends to be based on the younger child. Thus, if the mechanism of birth record falsification has a significant effect, the observed interval between first and second births should increase. Otherwise, the birth spacing should not be altered. Table 11 examines this influence by focusing on the relationship between the triple interaction term $TCP_{p,t} * Firstborn\ girl_{i,p,t} * Twinning\ in\ the\ second\ birth_{p,t}$ and the birth spacing between the first and second reported births. The results show

that among pure minority families, who were shown not to engage in creating artificial twins (Table 11), the birth spacing between first and second births decreased because the TCP increased the sense of urgency among firstborn-girl families. The gap in delivery was reversed and turned positive in Han–minority mixed families due to the influence of Confucianism, although the effect was not significant. These results indicate that the mechanism of falsifying birth records has an effect in practice.

Using several pieces of circumstantial evidence, we also investigated the possibility that couples increased the probability of having twins by taking fertility drugs.[44] First, we found that the first two pharmaceutical factories in China that produced or imported clomiphene were established after 1996, which is later than the study period of our research. Therefore, fertility drugs are very unlikely to have been available to families in the 1980s and 1990s.[45] Second, in birth records collected by hospitals, the mechanism of taking fertility drugs could be confirmed if associated with a surge in second-birth twinning among firstborn-girl families post-TCP. Although the information from the birth records of hospitals is inaccessible, multiple news outlets in different cities reported that hospitals began to report a twinning boom after the year 2000. Third, since fertility drugs cannot control the number of fertilised eggs, the possibility of triplets or quadruplets should also be increased when the drugs are taken. However, in our sample, no families had multiple second births other than twins, indicating that taking fertility drugs was very rare. To summarise, our results indicate that the use of fertility drugs is highly unlikely to have contributed to the phenomenon of artificial twinning.

[44] One possible approach is to compare height or weight between twins. However, such survey data (i.e., China Health and Nutrition Survey) contain too few samples of second-birth ethnic minority twins, making it impossible to formulate any regression estimations.

[45] These first two pharmaceutical factories in China that produced or imported clomiphene were Pioneer Pharmaceutical Ltd and Guangzhou Kanghe Pharmaceutical Ltd. They were established in 1996 and 2001, respectively. http://www.pioneer-pharma.com/; http://www.gkh.com.cn/aboutus/

Table 11: Birth spacing of twins at the second birth

	(1)	(2)	(3)	(4)
		Birth Spacing (logged)		
Panel A:		Both minorities	Either Han	Placebo
Firstborn girl*TCP	0.0411	-0.0332	0.0397	-0.0939
*Twin at the second birth	(0.0594)	(0.138)	(0.0698)	(0.211)
Observations	159,249	59,648	99,559	26,592
R-squared	0.33	0.344	0.33	0.26
Panel B:		Both minorities	Either Han	Placebo
Firstborn girl*TCP Announcement	0.033	-0.0244	0.0321	-0.0872
*Twin at the second birth	(0.0749)	(0.154)	(0.0842)	(0.258)
Firstborn girl*TCP Legalisation	0.0483	-0.0451	0.0463	-0.103
*Twin at the second birth	(0.0529)	(0.128)	(0.0771)	(0.188)
Observations	159,249	59,648	99,559	26,592
R-squared	0.33	0.344	0.33	0.26
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	Y	Y	Y	Y
Other controls	Y	Y	Y	Y
Province-by-year of birth FE	Y	Y	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported. The full specification can be found in Appendix Table A9.

6 Conclusion

A high fertility rate is always regarded as a hindrance to economic take-off in underdeveloped countries. In particular, booming populations can threaten less-developed countries' food security and social protection systems. Accordingly, family planning programmes have been initiated and gradually popularised since the 1970s.

Drawing on the context within China in the late 20th century, this study estimates the sex-bias costs of the most commonly used fertility policy worldwide – the TCP. Our findings indicate that the application of the TCP greatly exacerbated sex discrimination at birth. We first find that to improve their likelihood of having sons and avoid the fertility quota, parents tended to manipulate the sex of their children. Among the second births of firstborn-girl families, the boy-to-girl ratio rose from 1.1 pre-TCP to 1.6 post-TCP within a decade. Legalizing this fertility policy further escalated family responses, and the male-biased sex imbalance became even more pronounced. Our analysis also confirmed the mechanism behind this behaviour. In particular, sex-selective abortion fuelled by ultrasound technology was the main form of sex-selection behaviour in the

TCP legalisation phase. We also found evidence indicating faking twins was another a possible approach to circumventing strict birth policies.

It needs to be highlighted that the sex-discrimination impacts of the TCP in China are highly heterogenous. Although almost all ethnic groups in China are characterised by patriarchy, the extent of son preference still differs according to the Han–minority division. Influenced by Confucianism, the cultural preference for sons is much more pronounced in the Han majority. Our estimations also indicate that sex-selection behaviour was much more significant in Han–minority mixed families than in pure ethnic minority families, suggesting that cultural preference is a key factor affecting the degree of gender bias.

Our research also provides insights for policymaking. Facing population pressure, many developing countries and regions are considering applying birth restriction policies.[46] It is worth noting that the effects of such policies heavily depend on cultural factors, such as son preferences and dowries. Our research shows that failing to carefully consider local cultural factors and the degree of enforcement in policy designs can lead to unintended social consequences.

[46] For example, in 2021, Indian states, including Uttar Pradesh and Assam, were considering drafting a fertility policy stating that couples with over two children would be stripped of benefits, subsidies and opportunities to work for the government. https://www.theguardian.com/world/2021/jul/14/india-states-considering-two-child-policy-and-incentives-for-sterilisation

References

Almond, D., Li, H., & Zhang, S. (2019). Land reform and sex selection in China. *Journal of Political Economy, 127*(2), 560-585.

Anukriti, S., Bhalotra, S., & Tam, E. H. (2020). On the quantity and quality of girls: Fertility, parental investments, and mortality. *The Economic Journal*.

Arnold, F., Kishor, S., & Roy, T. K. (2002). Sex-selective abortions in India. *Population and development review, 28*(4), 759-785.

Ashraf, N., Bau, N., Nunn, N., & Voena, A. (2020). Bride price and female education. *Journal of Political Economy, 128*(2), 591-641.

Attané, I. (2009). The determinants of discrimination against daughters in China: evidence from a provincial-level analysis. *Population Studies, 63*(1), 87-102.

Banister, J. (2004). Shortage of girls in China today. *Journal of Population Research, 21*(1), 19-45.

Bau, N. (2021). Can policy change culture? Government pension plans and traditional kinship practices. *American Economic Review, 111*(6), 1880-1917.

Becker, G. S. (1991). *A treatise on the family: Enlarged edition*. Harvard university press.

Ben-Porath, Y., & Welch, F. (1976). Do sex preferences really matter?. *The Quarterly Journal of Economics, 90*(2), 285-307.

Bhalotra, S., Brulé, R., & Roy, S. (2020). Women's inheritance rights reform and the preference for sons in India. *Journal of Development Economics*, 146, 102275.

Bhalotra, S., & Van Soest, A. (2008). Birth-spacing, fertility and neonatal mortality in India: Dynamics, frailty, and fecundity. *Journal of Econometrics, 143*(2), 274-290.

Chao, F., Gerland, P., Cook, A. R., & Alkema, L. (2019). Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels. *Proceedings of the National Academy of Sciences, 116*(19), 9303-9311.

Chen, Y., & Fang, H. (2021). The long-term consequences of China's "later, longer, fewer" campaign in old age. *Journal of Development Economics, 151*, 102664.

Chen, Y., Li, H., & Meng, L. (2013). Prenatal sex selection and missing girls in China: Evidence from the diffusion of diagnostic ultrasound. *Journal of Human Resources, 48*(1), 36-70.

Ebenstein, A. (2010). The "missing girls" of China and the unintended consequences of the one child policy. *Journal of Human resources, 45*(1), 87-115.

Ebenstein, A., & Leung, S. (2010). Son preference and access to social insurance: evidence from China's rural pension program. *Population and Development Review, 36*(1), 47-70.

Ebenstein, A. (2014). Patrilocality and missing women. Available at SSRN 2422090.

Ebenstein, A. (2021). The Historical Origins of Son Preference: Patrilocality and Missing Women. *Available at SSRN 3829406.*

Greenhalgh, S. (1988). Fertility as mobility: Sinic transitions. *Population and Development Review*, 629-674.

Greenhalgh, S. (2003). Science, modernity, and the making of China's one-child policy. *Population and Development Review, 29*(2), 163-196.

Gupta, M. D. (2005). Explaining Asia's "missing women": a new look at the data. *Population and development review, 31*(3), 529-535.

Gu, B., Wang, F., Guo, Z., & Zhang, E. (2007). China's local and national fertility policies at the end of the twentieth century. *Population and Development Review*, 129-147.

Huang, W., Lei, X., & Zhao, Y. (2016). One-child policy and the rise of manmanmadens. *Review of Economics and Statistics, 98*(3), 467-476.

Huang, W., & Zhang, C. (2021). The power of social pensions: Evidence from China's new rural pension scheme. *American Economic Journal: Applied Economics, 13*(2), 179-205.

Hvistendahl, M. (2010). Has China outgrown the one-child policy?. *Science*, 1458-1461.

Jayachandran, S., & Kuziemko, I. (2011). Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India. *The Quarterly journal of economics, 126*(3), 1485-1538.

Jayachandran, S. (2017). Fertility decline and missing women. *American Economic Journal: Applied Economics, 9*(1), 118-39.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *economics, 7*(1), 63-88.

Jia, R., & Persson, T. (2021). Choosing ethnicity: The interplay between individual and social motives. *Journal of the European Economic Association, 19*(2), 1203-1248.

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H. J., ... & Gavin, M. C. (2016). D-PLACE: A global database of cultural, linguistic and environmental diversity. *PloS one, 11*(7), e0158391.

Loh, C., & Remick, E. J. (2015). China's skewed sex ratio and the one-child policy. *The China Quarterly*, 222, 295-319.

Lu, Y., & Zhai, Z. (2009). *Sixty Years of New China Population (In Chinese)*. China Population Publishing House.

Milazzo, A. (2018). Why are adult women missing? Son preference and maternal survival in India. *Journal of Development Economics, 134*, 467-484.

Murdock, G. P. (1967). Ethnographic atlas: a summary. *Ethnology, 6*(2), 109-236.

Ngo, A. P. (2020). Effects of Vietnam's two-child policy on fertility, son preference, and female labor supply. *Journal of Population Economics, 33*(3), 751-794.

Osters, E. (2009). Proximate sources of population sex imbalance in India. *Demography, 46*(2), 325-339.

Peng, P. (1996). *Family Planning Encyclopedia of China (In Chinese)*. Beijing: China Population Press.

Poston, D. L., Chang, C. F., & Dan, H. (2006). Fertility differences between the majority and minority nationality groups in China. *Population Research and Policy Review, 25*(1), 67-101.

Putterman, L. (2008). Agriculture, diffusion and development: Ripple effects of the neolithic revolution. *Economica, 75*(300), 729-748.

Rossi, P., & Xiao, Y. (2020). *Spillovers in Childbearing Decisions and Fertility Transitions: Evidence from China* (No. 20-031/V). Tinbergen Institute.

Scharping, T. (2013). *Birth Control in China 1949-2000: Population policy and demographic development*. Routledge.

Tilt, B., Li, X., & Schmitt, E. A. (2019). Fertility trends, sex ratios, and son preference among Han and minority households in rural China. *Asian Anthropology, 18*(2), 110-128.

Walder, A. G. (2014). Rebellion and repression in China, 1966–1971. *Social Science History, 38*(3-4), 513-539.

Wei, Y., & Zhang, L. (2014). Re-examination of the Yicheng two-child program. *The China Journal*, (72), 98-120.

Xu, B., & Pak, M. (2015). Gender ratio under China's two-child policy. *Journal of Economic Behavior & Organization,* 119, 289-307.

Zhang, G., & Zhao, Z. (2006). Reexamining China's fertility puzzle: Data collection and quality over the last two decades. *Population and Development Review, 32*(2), 293-321.

Zhang, J. (2017). The evolution of China's one-child policy and its effects on family outcomes. *Journal of Economic Perspectives, 31*(1), 141-60.

Zhang, J., & Liu, E. Y. (2012). The Confucian Ethic of Female Subordination and Depression Among Young People in Rural China. *Interdisciplinary Journal of Research on Religion*, 8.

Zeng, Y., P. Tu, B. Gu, Y. Xu, B. Li, and Y. Li (1993). Causes and implications of the recent increase in the reported sex ratio at birth in China. *Population and Development Review 19* (2), 283–302.

Appendix A

Figure A1: Marital Residence with Kin, Worldwide

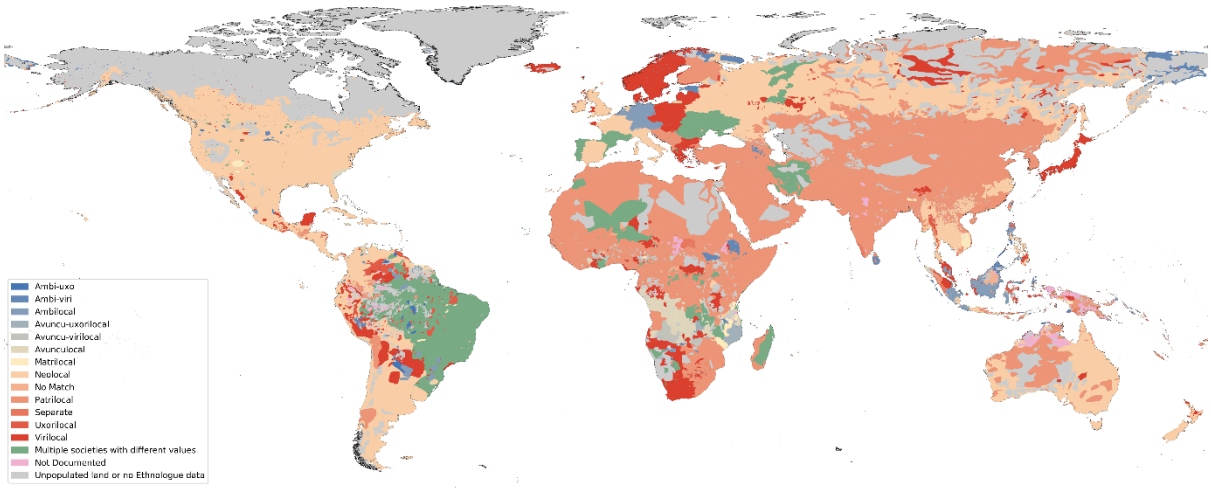


Figure A2: Trends of Sex Ratio in Groups following 3/3+ Child Policy

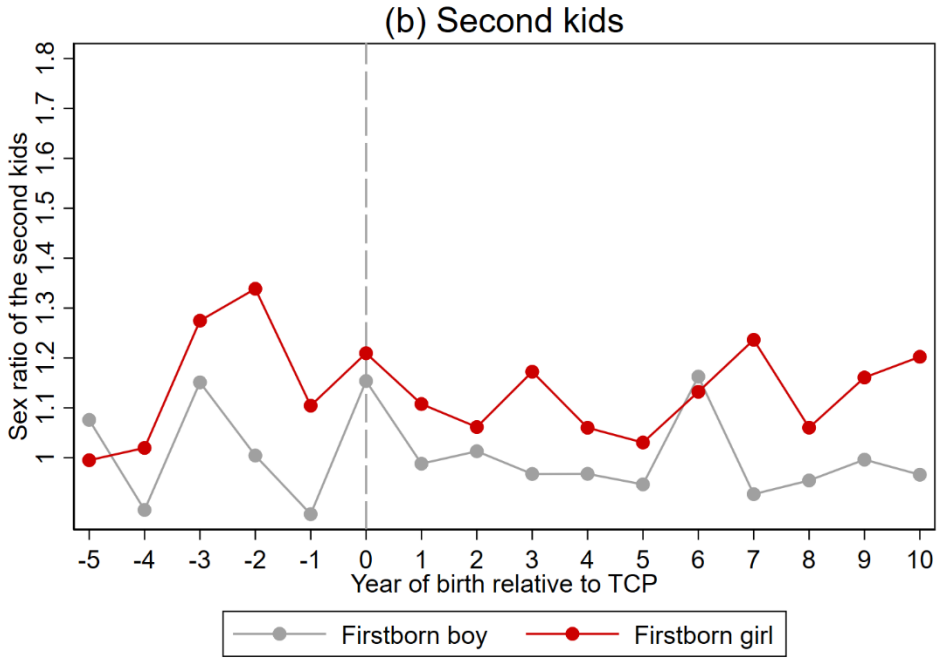


Table A1: Sex Ratio at the Second Birth: Full Specification

	(1)	(2)	(3)	(4)
	Child is male (0/1)			
Panel A: Minorities following TCP				
Firstborn girl*TCP	5.284***	5.283***	5.283***	5.251***
	(0.76)	(0.757)	(0.752)	(0.773)
TCP	-2.149**	-2.148**	-2.170**	
	(0.79)	(0.792)	(0.835)	
Firstborn girl	2.882***	2.881***	2.838***	2.841***
	(0.512)	(0.511)	(0.511)	(0.508)
Constant	51.77***	51.77***	51.80***	50.51***
	(0.541)	(0.543)	(0.563)	(0.268)
Observations	158,312	158,312	158,309	158,307
R-squared	0.006	0.006	0.007	0.012
Panel B: Minorities following 2+ child policy (placebo test)				
Firstborn girl*TCP	-0.176	-0.203	-0.148	-0.104
	(1.632)	(1.629)	(1.708)	(1.759)
TCP	-0.15	-0.0784	-0.0861	
	(1.652)	(1.665)	(1.705)	
Firstborn girl	3.421	3.424	3.324	3.306
	(1.974)	(1.973)	(1.992)	(2.04)
Constant	50.29***	50.24***	50.28***	50.21***
	(1.383)	(1.391)	(1.409)	(0.409)
Observations	26,489	26,489	26,487	26,487
R-squared	0.002	0.002	0.006	0.012
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province-by-year of birth FE	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported.

Table A2: Examining Sample Selection: Son-biased Fertility Stopping: Full Specification

	(1)	(2)	(3)	(4)
	Having younger sibling at the second birth (0/1)			
Firstborn girl*TCP	0.0196	0.0204	0.0131	0.0156
	(0.0122)	(0.0122)	(0.0113)	(0.011)
TCP	-0.00025	-0.00181	-0.0057	
	(0.0214)	(0.0212)	(0.0187)	
Firstborn girl	0.0758***	0.0764***	0.0907***	0.0891***
	(0.016)	(0.0159)	(0.0154)	(0.0151)
Constant	0.500***	0.501***	0.499***	0.495***
	(0.0153)	(0.0148)	(0.0127)	(0.00749)
Observations	204,142	204,142	204,140	204,137
R-squared	0.192	0.2	0.325	0.345
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province-by-year of birth FE	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to the second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported.

Table A3: Sex Ratio at the Second Birth: Different Stages of Implementation: Full Specification

	(1)	(2)	(3)	(4)
		Second Birth are Twins (0/1)		
Panel A: Minorities following TCP				
Firstborn girl	3.215***	3.212***	3.213***	3.157***
* TCP Announcement	(0.928)	(0.924)	(0.918)	(0.919)
Firstborn girl	8.276***	8.277***	8.276***	8.293***
* TCP Legalisation	(0.909)	(0.91)	(0.912)	(0.946)
TCP Announcement	-1.118	-1.115	-1.138	
	(0.84)	(0.843)	(0.88)	
TCP Legalisation	-4.471***	-4.470***	-4.524***	
	(1.097)	(1.098)	(1.125)	
Firstborn girl	2.881***	2.881***	2.837***	2.841***
	(0.512)	(0.512)	(0.512)	(0.508)
Constant	51.95***	51.94***	51.99***	50.48***
	(0.578)	(0.58)	(0.593)	(0.279)
Observations	158,312	158,312	158,309	158,307
R-squared	0.007	0.007	0.008	0.012
Panel B: Minorities following 2+ child policy (placebo test)				
Firstborn girl	-0.128	-0.167	-0.106	-0.125
* TCP Announcement	(1.8)	(1.805)	(1.84)	(1.878)
Firstborn girl	-0.27	-0.273	-0.232	-0.064
* TCP Legalization	(2.456)	(2.456)	(2.595)	(2.636)
TCP Announcement	-0.221	-0.144	-0.15	
	(1.591)	(1.609)	(1.634)	
TCP Legalization	0.918	0.995	0.897	
	(2.577)	(2.598)	(2.702)	
Firstborn girl	3.421	3.424	3.325	3.306
	(1.974)	(1.974)	(1.993)	(2.04)
Constant	50.06***	50.01***	50.07***	50.21***
	(1.514)	(1.525)	(1.551)	(0.408)
Observations	26,489	26,489	26,487	26,487
R-squared	0.002	0.002	0.006	0.012
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province-by-year of birth FE	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to the second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age. All the other terms are included but not reported.

Table A4: Sex Ratio at the Second Birth: Heterogeneous Effect: Full Specification

	(1)	(2)	(3)	(4)
	Second Birth are Twins (0/1)			
Panel A:	Both Parents are minorities		Either Parents is Han ethnic	
Firstborn girl*TCP	3.125***	3.162***	6.975***	7.010***
	(1.038)	(1.035)	(0.835)	(0.837)
Firstborn girl	2.315***	2.271***	3.147***	3.116***
	(0.572)	(0.567)	(0.559)	(0.561)
Constant	50.26***	50.27***	50.58***	50.59***
	(0.243)	(0.236)	(0.301)	(0.301)
Observations	59,402	59,397	98,873	98,868
R-squared	0.013	0.015	0.015	0.016
Panel B:	Both Parents are minorities		Either Parents is Han ethnic	
Firstborn girl	1.033	1.082	4.624***	4.649***
* TCP Announcement	(1.013)	(1.002)	(1.055)	(1.06)
Firstborn girl	5.697***	5.720***	10.90***	10.95***
* TCP Legalization	-1.327	-1.328	-0.931	-0.939
Firstborn girl	2.315***	2.270***	3.147***	3.116***
	(0.572)	(0.567)	(0.559)	(0.561)
Constant	50.26***	50.27***	50.58***	50.59***
	(0.243)	(0.236)	(0.301)	(0.301)
Observations	59,402	59,397	98,873	98,868
R-squared	0.013	0.015	0.015	0.016
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	Y	Y	Y	Y
Other controls	Y	Y	Y	Y
Province-by-year of birth FE	Y	Y	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

Table A5: Birth Spacing at the Second Birth: Full Specification

	(1)	(2)	(3)	(4)	(5)
	Birth spacing (logged)				
Panel A:					*Placebo*
Firstborn girl*TCP	0.0281***	0.0272***	0.0212***	0.0100**	-0.0382**
	(0.00948)	(0.00946)	(0.00637)	(0.0047)	(0.0115)
TCP	-0.0274	-0.0257	-0.0176		
	(0.02)	(0.0205)	(0.0231)		
Firstborn girl	-0.0578***	-0.0581***	-0.0503***	-0.0472***	-0.00848
	(0.00874)	(0.0087)	(0.00694)	(0.00605)	(0.0137)
Constant	3.530***	3.530***	3.523***	3.514***	3.481***
	(0.0122)	(0.0124)	(0.0136)	(0.00347)	(0.00581)
Observations	158,310	158,310	158,304	158,302	26,486
R-squared	0.081	0.084	0.31	0.329	0.259
Panel B:					*Placebo*
Firstborn girl	0.00503	0.00318	0.00116	-0.00477	-0.0399**
* TCP Announcement	(0.00842)	(0.00833)	(0.00597)	(0.00498)	(0.0153)
Firstborn girl	0.0614***	0.0617***	0.0496***	0.0315***	-0.0351***
* TCP Legalisation	(0.0146)	(0.0146)	(0.0097)	(0.0068)	(0.00568)
TCP Announcement	-0.0156	-0.0134	-0.00643		
	(0.0204)	(0.0209)	(0.0229)		
TCP Legalization	-0.0415	-0.0408	-0.00029		
	(0.0358)	(0.0358)	(0.031)		
Firstborn girl	-0.0578***	-0.0581***	-0.0503***	-0.0472***	-0.00848
	(0.00872)	(0.00868)	(0.00692)	(0.00605)	(0.0137)
Constant	3.529***	3.529***	3.514***	3.514***	3.481***
	(0.0139)	(0.014)	(0.0147)	(0.00339)	(0.00578)
Observations	158,310	158,310	158,304	158,302	26,486
R-squared	0.082	0.084	0.31	0.329	0.259
Province FE	Y	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y	Y
Census FE	N	Y	Y	Y	Y
Other controls	N	N	Y	Y	Y
Province-by-year of birth FE	N	N	N	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

Table A6: Mechanism of Ultrasound on Sex Ratio at the Second Birth: Full Specification

	(1)	(2)	(3)	(4)	(5)
			Child is male (0/1)		
Panel A:					Placebo
Firstborn girl*TCP*Ultra	0.961	0.964	0.975	0.996	7.122
	(1.794)	(1.792)	(1.785)	(1.823)	(4.949)
Firstborn girl*Ultra	0.0369	0.036	0.0116	-0.0182	-5.369
	(1.595)	(1.595)	(1.601)	(1.632)	(3.73)
TCP*Ultra	0.738	0.737	0.817		
	(1.12)	(1.12)	(1.12)		
Ultra	-0.642	-0.64	-0.642		
	(0.952)	(0.952)	(0.966)		
Firstborn girl*TCP	4.354***	4.351***	4.356***	4.331***	-3.487
	(1.289)	(1.284)	(1.287)	(1.322)	(3.494)
TCP	-2.511**	-2.509**	-2.587**		
	(1.015)	(1.018)	(1.049)		
Firstborn girl	2.871***	2.871***	2.836***	2.847***	5.044**
	(0.559)	(0.559)	(0.568)	(0.581)	(1.463)
Constant	52.01***	52.01***	52.04***	50.51***	50.21***
	(0.693)	(0.697)	(0.735)	(0.271)	(0.297)
Observations	158,312	158,312	158,309	158,307	26,487
R-squared	0.006	0.006	0.007	0.012	0.012
Panel B:					Placebo
Firstborn girl	-1.314	-1.311	-1.3	-1.304	7.161
*TCP Announcement*Ultra	(1.746)	(1.746)	(1.735)	(1.769)	(4.562)
Firstborn girl	8.246***	8.248***	8.264***	8.306***	3.568
*TCP Legalisation*Ultra	(1.243)	(1.244)	(1.242)	(1.287)	(3.376)
Firstborn girl*Ultra	0.0362	0.0347	0.00971	-0.0189	-5.369
	(1.596)	(1.595)	(1.601)	(1.633)	(3.73)
TCP Announcement*Ultra	1.88	1.88	1.957*		
	(1.147)	(1.148)	(1.15)		
TCP Legalisation*Ultra	-4.006***	-4.005***	-4.036***		
	(1.29)	(1.292)	(1.31)		
Ultra	-0.574	-0.571	-0.572		
	(0.955)	(0.955)	(0.968)		
Firstborn girl	4.357***	4.353***	4.357***	4.329***	-3.487
*TCP Announcement	(1.288)	(1.283)	(1.286)	(1.323)	(3.492)
TCP Announcement	-2.507**	-2.503**	-2.580**		
	(1.017)	(1.021)	(1.052)		
Firstborn girl	2.871***	2.871***	2.836***	2.847***	5.044**
	(0.558)	(0.559)	(0.568)	(0.581)	(1.463)
Constant	52.13***	52.13***	52.16***	50.48***	50.21***
	(0.701)	(0.706)	(0.74)	(0.282)	(0.295)
Observations	158,312	158,312	158,309	158,307	26,487
R-squared	0.007	0.007	0.008	0.012	0.012
Province FE	Y	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y	Y
Census FE	N	Y	Y	Y	Y
Other controls	N	N	Y	Y	Y
Province-by-year of birth FE	N	N	N	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

Table A7: Mechanism of Ultrasound on Birth Spacing at the Second Birth: Full Specification

	(1)	(2)	(3)	(4)	(5)
	Birth Spacing (logged)				
Panel A:					Placebo
Firstborn girl*TCP*Ultra	0.0342	0.0370*	0.0406**	0.0303	-0.0474
	(0.0202)	(0.0194)	(0.0175)	(0.0181)	(0.0432)
Firstborn girl*Ultra	0.00954	0.00658	-0.00218	-0.00476	-0.00391
	(0.0142)	(0.0134)	(0.0141)	(0.0135)	(0.0267)
TCP*Ultra	0.0192	0.0184	-0.022		
	(0.0381)	(0.0372)	(0.0298)		
Ultra	-0.0812**	-0.0793**	-0.0321		
	(0.0333)	(0.0331)	(0.026)		
Firstborn girl*TCP	0.00954	0.00658	-0.00218	-0.00476	-0.00391
	(0.0142)	(0.0134)	(0.0141)	(0.0135)	(0.0267)
TCP	-0.0570*	-0.0545	-0.0137		
	(0.033)	(0.0325)	(0.0368)		
Firstborn girl	-0.0498***	-0.0498***	-0.0421***	-0.0399***	-0.0145
	(0.00953)	(0.00953)	(0.00839)	(0.00781)	(0.015)
Constant	3.594***	3.592***	3.555***	3.514***	3.481***
	(0.0251)	(0.0252)	(0.0244)	(0.00344)	(0.00552)
Observations	158,310	158,310	158,307	158,305	26,487
R-squared	0.083	0.085	0.311	0.329	0.259
Panel B:					Placebo
Firstborn girl	0.00918	0.0111	0.019	0.014	-0.0496
*TCP Announcement*Ultra	(0.0207)	(0.0198)	(0.0183)	(0.0191)	(0.0461)
Firstborn girl	0.0757***	0.0767***	0.0656***	0.0468***	-0.0474*
*TCP Legalisation*Ultra	(0.0143)	(0.0143)	(0.00898)	(0.0068)	(0.0198)
Firstborn girl*Ultra	-0.0235***	-0.0245***	-0.0250***	-0.0224**	0.0179
	(0.00842)	(0.00855)	(0.00853)	(0.00874)	(0.0268)
TCP Announcement*Ultra	0.0326	0.0323	-0.00936		
	(0.0364)	(0.0357)	(0.0283)		
TCP Legalisation*Ultra	-0.0426	-0.0419	-0.012		
	(0.0288)	(0.0289)	(0.0255)		
Ultra	-0.0815**	-0.0797**	-0.034		
	(0.0325)	(0.0324)	(0.0253)		
Firstborn girl	0.00955	0.00657	-0.00222	-0.00477	-0.0039
*TCP Announcement	(0.0142)	(0.0133)	(0.014)	(0.0135)	(0.0267)
TCP Announcement	-0.0571*	-0.0545*	-0.0139		
	(0.0325)	(0.0321)	(0.0353)		
Firstborn girl	-0.0498***	-0.0498***	-0.0421***	-0.0399***	-0.0145
	(0.00951)	(0.00952)	(0.00838)	(0.00781)	(0.015)
Constant	3.591***	3.589***	3.547***	3.514***	3.481***
	(0.027)	(0.0271)	(0.0245)	(0.00336)	(0.00545)
Observations	158,310	158,310	158,307	158,305	26,487
R-squared	0.083	0.086	0.311	0.329	0.259
Province FE	Y	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y	Y

Census FE	N	Y	Y	Y	Y
Other controls	N	N	Y	Y	Y
Province-by-year of birth FE	N	N	N	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

Table A8: Second birth are twins: Full Specifications

	(1)	(2)	(3)	(4)
		Second birth are twins (0/1)		
Panel A:		Both minorities	Either Han	Placebo
Firstborn girl*TCP	0.108	-0.0711	0.225**	-0.132
	(0.08)	(0.112)	(0.104)	(0.181)
Firstborn girl	-0.0398	0.0159	-0.0638	-0.0133
	(0.0487)	(0.0726)	(0.0647)	(0.102)
Constant	0.580***	0.439***	0.662***	0.447***
	(0.0189)	(0.0284)	(0.025)	(0.0441)
Observations	159,251	59,649	99,560	26,592
R-squared	0.008	0.016	0.011	0.009
Panel B:		Both minorities	Either Han	Placebo
Firstborn girl*TCP Announcement	0.0456	-0.0679	0.096	-0.218
	(0.0932)	(0.165)	(0.107)	(0.138)
Firstborn girl*TCP Legalisation	0.198*	-0.0751	0.440***	0.0277
	(0.111)	(0.133)	(0.157)	(0.292)
Firstborn girl	-0.0398	0.0159	-0.0639	-0.0134
	(0.0487)	(0.0725)	(0.0647)	(0.102)
Constant	0.579***	0.439***	0.660***	0.446***
	(0.0191)	(0.0285)	(0.0237)	(0.0466)
Observations	159,251	59,649	99,560	26,592
R-squared	0.008	0.016	0.011	0.009
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	Y	Y	Y	Y
Other controls	Y	Y	Y	Y
Province-by-year of birth FE	Y	Y	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

Table A9: Birth Spacing of Twins at the Second Birth: Full Specification

	(1)	(2)	(3)	(4)
		Birth Spacing (logged)		
Panel A:		Both minorities	Either Han	Placebo
Firstborn girl*TCP	0.0411	-0.0332	0.0397	-0.0939
*Twin at the second birth	(0.0594)	(0.138)	(0.0698)	(0.211)
TCP	-0.012	0.13	-0.0405	0.154
* Twin at the second birth	(0.05)	(0.101)	(0.0542)	(0.125)
Firstborn girl	0.0207	0.103	0.0151	0.181
* Twin at the second birth	(0.0507)	(0.109)	(0.0612)	(0.26)
Twin at the second birth	0.000475	-0.000808	0.000776**	-0.000762
	(0.000355)	(0.00093)	(0.000324)	(0.0014)
TCP	0.00999**	-0.00573	0.0185**	-0.0373**
	(0.00462)	(0.00985)	(0.00756)	(0.0114)
Firstborn girl	-0.0472***	-0.0371***	-0.0522***	-0.00883
	(0.00613)	(0.01)	(0.00556)	(0.0136)
Constant	3.515***	3.535***	3.503***	3.481***
	(0.00352)	(0.00538)	(0.00315)	(0.00585)
Observations	159,249	59,648	99,559	26,592
R-squared	0.33	0.344	0.33	0.26
Panel B:		Both minorities	Either Han	Placebo
Firstborn girl*TCP Announcement	0.033	-0.0244	0.0321	-0.0872
*Twin at the second birth	(0.0749)	(0.154)	(0.0842)	(0.258)
Firstborn girl*TCP Legalisation	0.0483	-0.0451	0.0463	-0.103
*Twin at the second birth	(0.0529)	(0.128)	(0.0771)	(0.188)
TCP Announcement	-0.0152	0.12	-0.0427	0.15
* Twin at the second birth	(0.0562)	(0.127)	(0.0568)	(0.154)
TCP Legalisation	-0.00742	0.143	-0.0381	0.16
* Twin at the second birth	(0.0579)	(0.0873)	(0.0762)	(0.116)
Firstborn girl	0.0207	0.103	0.0151	0.181
* Twin at the second birth	(0.0507)	(0.109)	(0.0611)	(0.26)
Twin at the second birth	0.000475	-0.000808	0.000776**	-0.000762
	(0.000355)	(0.00093)	(0.000324)	(0.0014)
TCP Announcement	-0.00495	-0.0200*	0.00383	-0.0383**
	(0.00492)	(0.0116)	(0.00717)	(0.0151)
TCP Legalisation	0.0317***	0.0118	0.0429***	-0.0352***
	(0.00687)	(0.0126)	(0.0103)	(0.00588)
Firstborn girl	-0.0472***	-0.0371***	-0.0522***	-0.00883
	(0.00613)	(0.01)	(0.00556)	(0.0137)
Constant	3.514***	3.535***	3.502***	3.481***
	(0.00344)	(0.00522)	(0.00315)	(0.00582)
Observations	159,249	59,648	99,559	26,592
R-squared	0.33	0.344	0.33	0.26
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	Y	Y	Y	Y
Other controls	Y	Y	Y	Y

Province-by-year of birth FE	Y	Y	Y	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

Appendix B Robustness Checks

Table B1: Sex Ratio at the First Birth

	(1)	(2)	(3)	(4)
	Child is male (0/1)			
TCP	-0.00295	-0.00296	-0.00338	-0.00331
	(0.00567)	(0.00568)	(0.00565)	(0.00563)
Observations	249,283	249,283	249,281	249,281
R-squared	0	0	0.001	0.001
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province linear trend	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to the first birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

Table B2: Sex Ratio at Second and 2+ Births: Full Specification

	(1)	(2)	(3)	(4)
	Child is male (0/1)			
Panel A: Minorities following TCP				
Firstborn girl*TCP	4.139***	4.148***	4.242***	4.273***
	(0.66)	(0.664)	(0.668)	(0.703)
TCP	-0.854*	-0.853*	-0.854	
	(0.493)	(0.495)	(0.51)	
Firstborn girl	1.595***	1.597***	1.472***	1.495***
	(0.449)	(0.45)	(0.435)	(0.43)
Constant	52.87***	52.87***	52.89***	52.41***
	(0.286)	(0.288)	(0.294)	(0.147)
Observations	249,355	249,355	249,354	249,352
R-squared	0.004	0.004	0.005	0.008
Panel B: Minorities following 2+ child policy (placebo test)				
Firstborn girl*TCP	-1.091	-1.027	-0.908	-1.031
	(1.328)	(1.326)	(1.33)	(1.395)
TCP	-0.261	-0.242	-0.229	
	(1.639)	(1.658)	(1.707)	
Firstborn girl	2.973*	2.978*	2.769*	2.873*
	(1.397)	(1.398)	(1.385)	(1.425)
Constant	51.36***	51.33***	51.36***	51.20***
	(1.164)	(1.176)	(1.203)	(0.119)
Observations	53,443	53,443	53,442	53,442
R-squared	0.001	0.001	0.003	0.006
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province-by-year of birth FE	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

Table B3: Sex Ratio at the Second Birth: Including children who are under age 4

	(1)	(2)	(3)	(4)
	Child is male (0/1)			
Panel A: Minorities following TCP				
Firstborn girl*TCP	6.111***	6.107***	6.094***	6.026***
	(0.778)	(0.776)	(0.773)	(0.792)
TCP	-2.519***	-2.511***	-2.523***	
	(0.821)	(0.823)	(0.862)	
Firstborn girl	2.891***	2.890***	2.852***	2.859***
	(0.513)	(0.513)	(0.512)	(0.505)
Constant	52.08***	52.08***	52.11***	50.43***
	(0.645)	(0.647)	(0.659)	(0.323)
Observations	196,430	196,430	196,428	196,424
R-squared	0.009	0.009	0.009	0.014
Panel B: Minorities following 2+ child policy (placebo test)				
Firstborn girl*TCP	0.151	0.147	0.222	0.251
	(1.874)	(1.865)	(1.915)	(1.977)
TCP	-0.318	-0.309	-0.286	
	(1.625)	(1.627)	(1.657)	
Firstborn girl	3.425	3.426	3.278	3.255
	(1.973)	(1.972)	(1.976)	(2.021)
Constant	50.29***	50.28***	50.31***	50.09***
	(1.403)	(1.401)	(1.41)	(0.314)
Observations	32,936	32,936	32,933	32,933
R-squared	0.003	0.003	0.006	0.011
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province-by-year of birth FE	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.

Table B4: Sex Ratio at the Second Birth: Including Families that Reporting a different number of surviving children than that observed in the family

	(1)	(2)	(3)	(4)
	Child is male (0/1)			
Panel A: Minorities following TCP				
Firstborn girl*TCP	5.453***	5.456***	5.617***	5.557***
	(0.683)	(0.68)	(0.726)	(0.746)
TCP	-2.200***	-2.205***	-2.440***	
	(0.587)	(0.589)	(0.787)	
Firstborn girl	3.293***	3.293***	2.750***	2.772***
	(0.426)	(0.425)	(0.403)	(0.401)
Constant	51.52***	51.52***	51.71***	50.36***
	(0.403)	(0.406)	(0.418)	(0.255)
Observations	226,118	226,118	195,334	195,332
R-squared	0.006	0.006	0.007	0.01
Panel B: Minorities following 2+ child policy (placebo test)				
Firstborn girl*TCP	0.148	0.147	-0.243	-0.213
	(1.196)	(1.178)	(1.433)	(1.45)
TCP	-2.209*	-2.207*	-0.694	
	(1.106)	(1.089)	(1.395)	
Firstborn girl	3.424**	3.424**	3.470*	3.443*
	(1.25)	(1.249)	(1.6)	(1.628)
Constant	51.67***	51.67***	50.64***	50.19***
	(0.89)	(0.882)	(1.097)	(0.341)
Observations	39,353	39,353	35,414	35,414
R-squared	0.002	0.002	0.005	0.009
Province FE	Y	Y	Y	Y
Year of birth FE	Y	Y	Y	Y
Census FE	N	Y	Y	Y
Other controls	N	N	Y	Y
Province-by-year of birth FE	N	N	N	Y

Notes: Robust standard errors clustered at the provincial level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The data are from the 1990 and 2000 micro census. Analysis restricts the sample to second birth of ethnic minorities. Other controls include hukou status, ethnics, mother's education, and age.