Freie Universität Berlin

Chinese Dialects, Revolutionary War & Economic Performance

Junbing Zhu
Theocharis Grigoriadis

School of Business & Economics
Discussion Paper

Economics

2020/7

Chinese Dialects, Revolutionary War & Economic Performance[1]

Junbing Zhu[2] Theocharis Grigoriadis[3]
FU Berlin FU Berlin

Abstract: In this paper, we explore the effects of dialectal diversity on economic performance by drawing evidence from Chinese prefecture-level cities. Our dataset is a panel of 5-year average data over the period from 2001 to 2015 including 274 cities. We compute five indices of Chinese dialectal diversity: 1. Dialectal fractionalization; 2. Adjusted dialectal fractionalization; 3. Dialectal polarization; 4. Adjusted dialectal polarization and 5. Periphery heterogeneity. We find that dialectal fractionalization and dialectal polarization as well as periphery heterogeneity have a positive effect on both income per capita and economic growth. Adjusted dialectal fractionalization exhibits a positive effect only on the change in economic growth over time. However, adjusted dialectal polarization does not show any robust effects. Furthermore, the experience of being governed by the Chinese Communist Party during the revolutionary war inhibits the negative effects of dialectal diversity in eastern China, while it has persistent negative effects in central and north-eastern regions of the country.

Key words: dialectal diversity, local economic performance, communist governance

JEL: O10, O40, P51, Z19

I. Introduction

In this paper, the relationship between Chinese dialectal diversity and economic performance is explored empirically at the level of Chinese prefecture-level cities[4]. Currently, China has ten major dialectal supergroups, including about 100 dialectal subgroups (see the list of dialects in Appendix A). Since the division of administrative areas is not based on dialects and there have historically been several waves of migration, it is common that citizens of one prefecture-level city belong to different dialect groups, which makes it possible to explore the effect of dialectal diversity on economic performance. Thus, in this study, dialectal diversity is taken as the index of cultural diversity. In

---

[1] Junbing Zhu acknowledges gratefully the financial support of the Chinese Scholarship Council and the DAAD for this research. We are grateful to Klaus Desmet, Steve Nafziger, Max Steinhardt and Akos Dombi for insightful comments and suggestions. All remaining errors are ours.
[2] Freie Universität Berlin, School of Business & Economics & Institute of East European Studies, Garystr. 55, 14195, Berlin, Germany, zhujunbing@zedat.fu-berlin.de, T: +49-30-838-54008, F: +49-30-838-4-50925.
[3] Freie Universität Berlin, School of Business & Economics & Institute of East European Studies, Garystr. 55, 14195, Berlin, Germany, theocharis.grigoriadis@fu-berlin.de, T: +49-30-838-57037, F: +49-30-838-4-50925.
[4] There are four subnational levels in the Chinese administrative system: the provincial level, the prefecture level, the county level and the town level. The prefecture level includes prefecture-level city, league or autonomous prefecture and prefecture-level district. This study focuses on the dialectal diversity in prefecture-level cities.

economic studies, cultural diversity may hinder economic development by inducing communication difficulties, more social conflicts, distorted policies and inefficiency in governance. However, cultural diversity may benefit the economy by increasing innovation and market specialization. Given its long history of diversity, Chinese society is very inclusive of people from different dialect groups and there are few obstacles to their communication, which may undermine the negative effect of cultural diversity on economic development. Furthermore, the writing system is common for all dialects and the official language, Putonghua, has been promoted since the 1950s, providing more pathways for different dialect groups to understand each other.[5] Hence, at the local level in China, the negative effect of dialectal diversity is reduced, and we expect a positive influence on economic growth.

However, studies have found negative effects of both ethnic diversity at the provincial level (Dincer and Wang, 2011) and dialect diversity at the prefectural level (Xu et al., 2015) in China. But the discussion can be improved. Firstly, cultural diversity is not well measured. On the one hand, ethnic diversity is not a good index of cultural diversity in China. Since many ethnicities have been assimilated by the Han culture, they use Han dialects as their only language or the main language. Thus, ethnic diversity only captures a small part of cultural diversity. On the other hand, the number of Han dialects used in each city (Xu et al., 2015) can reflect neither the fractionalization nor the polarization of the population. If one dialect is used only by a small fraction of the population, the equal treatment of all dialects will result in biased results. Although dialectal fractionalization is used in the robustness test by Xu et al. (2015), dialectal distances are not examined in their paper. Secondly, in the research by Xu et al. (2015), only data of the year 2010 is used. This cannot capture the actual effects of dialectal diversity because of unobserved factors. Thirdly, in the analysis of endogeneity of dialectal diversity, Xu et al. (2015) use the railway index in the period of the Republic of China as the instrumental variable because the railway index can be explained as an indicator of land quality. However, the index may affect economic development in other channels, such as trade and freight traffic.

In this paper, we perform an improved empirical analysis of the relationship between dialectal diversity and economic development. Contrary to conventional wisdom, we argue that linguistic

[5] It is true that some people do not master the writing or Putonghua. They may also have difficulty in understanding other dialects or being understood themselves. But these are mainly old people and they account for a very small part of the population in prefecture-level cities. Their economic activities are primarily in local neighbourhoods and they encounter few communication difficulties.

fractionalization has a positive association with growth and development. We also make three main contributions to the research on dialectal diversity in China. First, five indices of Chinese dialectal diversity are calculated to measure dialectal diversity of prefecture-level cities: dialectal fractionalization, adjusted dialectal fractionalization, dialectal polarization, adjusted dialectal polarization and peripheral heterogeneity. Dialectal fractionalization represents the probability that two randomly selected persons are from two different dialect groups and it increases with the number of groups and the balance of population distribution. Dialectal polarization is used as the index reflecting the tension between the two largest groups. The polarization index mainly depicts how much the population distribution across groups deviates from a bimodal distribution and reaches its maximum when there are only two groups of equal size. Adjusted dialectal fractionalization and polarization refer to indices adjusted by dialectal distances, but the adjusted dialectal polarization puts a larger weight on the dialectal distances between the two largest groups. Periphery heterogeneity considers the interaction between the largest group and other groups and the dialectal distance between them. By comparing the effect of these, we can find whether dialectal distances have a role in explaining differences in economic development and the distance between which groups is more relevant.

Second, a panel sample covering the period 2001-2015 is constructed and 5-year average is used in the estimation. Therefore, we have the second contribution that a fixed-effect model is used and the effect of unobserved factors is reduced. Third, the difference in the effect resulting from exposure to the governance of the Chinese Communist Party during the revolutionary war and resources for economic development is determined. Prefecture-level cities with a longer exposure to the governance of the Chinese Communist Party are more deeply affected by the communist value system. Therefore, citizens' values and beliefs are affected by the difference in treatment by communism. On the other hand, the long exposure to the Party's governance leaves a higher proportion of cadres from the native population in local government and this leads to different outcomes in dealing with the interest conflicts between different groups. Furthermore, we consider the effect on the efficiency of governance at the local level of the economic environment given resources and support from the central government to develop the economy.

We find that in China dialectal diversity is conducive to higher levels of growth and development.[6] Analysis of the whole sample indicates that dialectal fractionalization and polarization as well as periphery heterogeneity have a positive impact on economic growth. Dialectal fractionalization adjusted by dialectal distances shows a positive effect only on the change in economic growth over time. But dialectal polarization adjusted by dialectal distances does not show any robust effect. Thus, dialectal distance between two polarized groups is not relevant for economic performance. Furthermore, exposure to the governance of the Chinese Communist Party during the revolutionary war causes a difference in the effect of dialectal diversity. In eastern China, communist experience tends to inhibit the negative impact of dialectal diversity while inducing negative influences in central and north-eastern part of China.

The organization of the paper is the following. The second section covers the literature review of the effect of cultural diversity on economic development and the experience of communism. In the third section, we discuss the relationship between dialectal and cultural diversity. The fourth section reports the data description and empirical strategy. A baseline fixed-effect regression and IV (instrumental variable) analysis are in the fifth section. The sixth section analyzes the differential effects from the longer exposure to the governance of Chinese Communist Party. The seventh section concludes.

II. Literature

Cultural diversity & economic development

As early as in 1967, the effects of cultural diversity on economic development attracted attention. Adelman and Morris (1967) conclude that economic growth rates tend to be higher in less heterogeneous countries, based on the data of 72 less developed countries from 1957 to 1962 and their linguistic diversity. Through re-analysis of data on 114 world polities from A Cross Polity Survey, Haug (1967) also finds that high cultural diversity is related to lower per capita GNP. The first economic study using modern econometric methods is by Easterly and Levine (1997), who adopt three measures of ethnic diversity. The results of a cross-country analysis indicate that high ethnic diversity induces low schooling, political instability, underdeveloped financial systems, distorted foreign exchange markets, high government deficits, insufficient infrastructure, low income and low

[6] See also Table B50 of Desmet et al. (2017) on the effects of diversity on log per capita income to corroborate our argument.

growth rates. The direct effects of ethnic diversity can also explain significant differences in economic growth across African countries and the Asian miracle. Thus, this paper initiates the study of both transmission channels and the direct effects of cultural diversity on economic development. The channel of government consumption is analyzed by La Porta et al. (1999). Cultural diversity tends to increase government consumption, but its effect depends on the utilization of the consumption. More recently, Alesina and La Ferrara (2004) find direct negative effects of both ethnic and linguistic diversity on economic growth, but the negative effect is mitigated by a higher initial income level. Goeren (2014) examines the direct and indirect effects of both ethnic fractionalization and polarization on economic growth through eight transmission channels: investment, civil war, human capital, government consumption, political instability, market distortion, trade openness and fertility. The dataset used is the updated version of the Barro–Lee data set on educational attainment and consists of 100 countries with 651 observations over the period 1960–1999. It does not only confirm that ethnic diversity has a strong direct negative effect on economic growth, but also establishes the indirect negative effect of ethnic polarization. Garcia-Montalvo and Reynal-Querol (2005a) also analyze the indirect effects of ethnic fractionalization and polarization through the channels of investment, government consumption and civil war. Taking child mortality, fertility, education and wealth as the outcomes of human development, Gerring et al. (2015) find that the negative effects of cultural diversity exist at national levels, while not at subnational levels.

However, Lian and Oneal (1997) argue that cultural diversity does not have significant effects on either economic growth or political instability. The reason for the difference may be that political institutions are not controlled in the research above. Collier (2000) develops a theoretical model with respect to government decisions under the influence of ethnic diversity. In this model, there is a tradeoff between economic growth and redistribution and the result depends on the political context. He finds that ethnic diversity leads to decisions reducing the growth rate in dictatorship, while ethnic diversity has no effect in democracy. Empirical evidence is also provided based on the data of 94 countries over the period 1960-1990 and World Bank projects in 89 countries and shows that the political environment exerts influences on the effects of ethnic diversity. Easterly (2001) holds a similar opinion and finds that good institutions reduce the negative effects of ethnic diversity on economic growth by adding the interaction term of institutions and ethnic diversity to the regression model. Furthermore, he tests the effects of institutions on the effect of ethnic diversity on policy factors and obtains results consistent with Easterly and Levine (1997), showing that good institutions significantly mitigate the negative effects of ethnic diversity. Furthermore, the relationship is affected by the level of development.

With respect to effects of cultural diversity within a specific country, Ottaviano and Peri (2006) demonstrate that the productivity of US-born citizens living in metropolitan areas is positively and significantly affected by a rise in the share of foreign-born citizens between 1970 and 1990. Alesina et al. (2000) employ ethnic diversity in a Dixit-Stiglitz production structure and find that diversity can increase total output because of more variety of" intermediate inputs", which can be interpreted as more diversity in individual skills. Diversity in skills may also increase overall productivity even when the cost of diversity is considered (Lazear, 1999 a, b). In addition, Ager and Brueckner (2014) examine the effects of immigrants to the US over the period 1870-1920 on economic growth. They construct measures of fractionalization and polarization and find that fractionalization has a positive effect on output while population polarization decreases output. Based on the data covering the NUTS3 regions of 12 countries in Europe, the same relationship between diversity, in terms of the share of foreigners, and productivity is revealed (Bellini et al., 2008). But the problem here is that they use the percentage of foreign-born citizens as the measure of cultural diversity, but this may not capture the exact cultural differences. Moreover, immigrants may have some common characteristics that affect productivity. Nevertheless, Sparber (2010) takes racial diversity as the measure of cultural diversity and a fixed-effects analysis shows that racial fractionalization of employment creates gains in the productivity of US cities, but the effect at the state level is ambiguous because it is only significant in random-effects specifications. Above all, although cultural diversity is shown to have a negative effect on economic development across countries, the effect is not significant when controlling for the influence of political institutions.

Cultural diversity & endogeneity

There are two possibilities that induce the endogenous problem in analyzing the relationship between cultural diversity and economic development. Firstly, better economic development may decrease cultural diversity. Secondly, people may be attracted by the better economic development and thus the fractionalization of the society increases. These causal effects may result in over- or underestimation of the effects of cultural diversity on economic development. This is a critical problem in exploring the effects of cultural diversity, but there are merely a few papers taking it into consideration. Ahlerup (2009) finds that the underestimation of negative effects exists in the empirical analysis based on OLS estimation. The study is conducted at the national level and four instruments are chosen: the duration of human settlements, the diversity of vegetation types, the number of years since the date of independence and the migratory distance in kilometers from Ethiopia to the centroid of each country. Apart from these factors that affect diversity, Ahlerup and Olsson (2007) also explore how local pathogen loads may affect ethnic diversity. Leigh (2006) instruments neighborhood diversity with regional diversity based on the assumption that population mobility is constrained within the region. In the Chinese city study of Xu et al. (2015), dialectal diversity is instrumented by the railway index in the period of the Republic of China to identify its effect on income growth. When exploring the influence of diversity on openness and urbanization of Chinese cities, the mountain index (Li et al., 2017) and migration in history (Shao et al., 2017) are used as instrumental variables. Michalopoulos (2012) finds that geographical variation, captured by the variation in regional land quality and elevation, fundamentally determines the contemporary ethnolinguistic diversity. Geographical factors may, however, affect economic development through other ways than cultural diversity. Hence, taking both geographical and historical factors into consideration, migration in history and geographical factors with regard to altitude and slope are used as instrumental variables of dialectal diversity in this study.

The effect of Communism

The effect of exposure to communism can be found in two strands of literature. One strand suggests that exposure to communism has a significant effect on the values and attitudes of citizens. Eleches and Tucker (2017) conduct research about how communism influences citizens' attitudes and behavior and find that more exposure to communism leads to more opposition to democracy and capitalism, less civic participation, less support for markets and more support for social welfare provided by the government. Through the analysis of East Germany and West Germany, Alesina and Schündeln (2007) also suggest that the effect of communism on the preference of citizens for government intervention in the economy is positive and significant. Reasons for such effects are path dependence and the communist ideology that individual fortunes are largely determined by the social condition as the responsibility of the government. Similarly, in post-communist countries, the development is associated with less movement towards democracy and less market reform (Treisman, 2014). Therefore, after exposure to communism, citizens are more supportive of collectivism than individualism.

The other strand of literature indicates that conflicts due to heterogeneity of groups are less in regions with longer exposure to communism. On the other hand, the benefits from diversity are also limited because of groups' preference for unification. However, there are also researchers who put forward the idea that exposure to communism has no significant long-run effect on culture and development. Roland (2010) suggests that institutional evolutions, values and beliefs in current transition countries are more affected by the long-run historical past than the experience of communism. In Germany, regardless of drastic political and economic changes, regional entrepreneurship culture tends to have had long-lasting effects over the period 1925-2005 (Fritsch and Wyrwich, 2014). Therefore, from the perspective of individual preference for collectivism, the experience of longer exposure to communism may depress individual market and entrepreneurial activities and reduce the benefits of dialectal diversity or have no effect because of the lack of impact on cultural traits.

However, from a different perspective, Li et al. (2014) show that provinces with longer exposure to the governance of the Chinese Communist Party (CCP) during the revolution war have a higher proportion of native cadres and suggest that local cadres contribute to higher economic growth because they have a better knowledge of local conditions and a better reputation among residents. Thus, we suppose that the coordination cost should be lower in the areas facing conflicts between different dialect groups when there is longer governance by the CCP. Accordingly, such experience should inhibit the negative effect of dialectal diversity on economic development resulting from interest conflicts. Hence, the longer exposure to the governance of CCP may reduce both the benefits and the loss of dialectal diversity at the same time and the combined influence may not be significant.

III. Dialectal vs. Cultural Diversity

While it is a convention that dialectal identity is an important component of cultural identity, there is no direct evidence showing that dialect is a cultural trait of people in China. Desmet et al. (2017) argue that cultural diversity, as measured by the probability of answering a random question of the WVS differently, is positively associated with good policy outcomes (less conflict, more public goods and higher income per capita). We use the China Family Panel Studies (CFPS) to examine the relationship between dialectal and cultural diversity.[7] The study is conducted at both individual level and family level, thus providing individual-level data on answers to questions on norms, values, and preferences as well as observable and individual characteristics. The dialectical areas of individuals involved in the survey are determined according to the individual's county. We ask whether there is a joint significant effect of dialects and how much variation in cultural attitudes can be explained by dialectal identities. For each question, the following specification is estimated:

$$Q_i = \alpha_0 + \sum_{d=1}^{D} \alpha_d X_i^d + \beta' C_i + \varepsilon_i ,$$

where i denotes a respondent, Q_i is the answer of the respondent to the question under consideration, $d = 1,...,D$ proxies dialect groups and X_i^d equals 1 if respondent i belongs to dialect group d and zero otherwise. C_i is a vector of control variables, including the gender dummy, age, the education level, ethnicity identity, the education level of the respondent's parents and household income.

The data used is the first wave of the China Family Panel Studies in 2010, which is the most comprehensive of all waves we have. The survey was conducted in 117 prefecture-level cities in which at most 3 counties were covered. Of all the questions studied, we confine our attention to questions identified as views to norms, values and attitudes, which leaves us with twentyfour questions in the end. Some questions have binary responses, some have an ordered response and the rest are the actual value of deposits, financial assets and total assets of the respondent's family. Binary and ordered responses are readily used as dependent variables and we also compute the ratio of deposits in family assets and the ratio of financial assets in family assets. All dependent variables and their meanings are shown in Table 1. Furthermore, each respondent is matched with one dialect based on his county name; a total of fifty-three dialect dummies are included in the dataset. The matching between dialects and counties follows the Coding Scheme of the Language Atlas of China.[8] To show whether dialects have predicting power for individual values, attitudes and behavior, we run the following regressions. Firstly, regressions are run in the whole sample for each question while

[7] The data is from China Family Panel Studies (CFPS), funded by the 985 Program of Peking University and carried out by the Institute of Social Science Survey of Peking University.
[8] Lavely, William; Berman, Lex, 2012, "Language Atlas of China", https://hdl.handle.net/190 2.1/19004, Harvard Dataverse, V1.

controlling provincial dummies. Secondly, regressions are done in each province for each dependent variable.[9] All regressions are done through OLS. The p-value of joint significance test of dialects and the goodness of fit, R^2, are also reported. We also compute the increase in R^2 by controlling for dummies of dialects.

Table 1. Variables for cultural identity

Variables	Meaning
Status-Achievement	The importance of social status in making achievements
Wealth-Achievement	The importance of wealth in making achievements
Education-Achievement	The importance of education level in making achievements
Talent-Achievement	The importance of talent in making achievements
Effort-Achievement	The importance of effort in making achievements
Luck-Achievement	The importance of luck in making achievements
Social network-Achievement	The importance of social network in making achievements
Social network vs. Ability	View about the statement: Social network is more important than personal ability.
Wealth as achievement	View about taking wealth as achievement
Importance of money	View about the importance of money
Effort-Reward	View about the statement: More effort, more reward.
Smart-Reward	View about the statement: Smarter, more reward.
Attention-Society	Attention to social problems
Attention-Anti-corruption	Attention to news about anti-corruption
Attention-Law and regulation	Attention to news about law and regulation
Attention-Economy	Attention to economic news
Attention-Environment	Attention to environmental problems
Social sympathy	Whether the respondent donated anything last year
Fairness vs. efficiency	The attitude about fairness and efficiency
Attitude about competition	View about the statement: Fair competition is necessary for good interpersonal relationship.
Trust	Willingness to trust the majority
Ratio of financial assets	The ratio of financial assets in family assets
Ratio of deposit	The ratio of deposit in family assets

The results of all regressions are collected in Table 2 and Table 3. In Table 2, all regressions are done in the whole sample and the average observations are 24,006 when dialect dummies are

[9] It would be more useful to examine the relationship between dialect and culture in every Chinese city. However, the CFPS study selects only one county for each city and there is no variance in dialects in the subsample at the city level. Thus, we opt for regressions at the province level.

controlled and 24,386 when they are not. We observe that all dialects are jointly significant at the 1 percent level. Furthermore, by including dialect dummies, R^2 rises in all regressions. For nine out of all the regressions, the addition of the dialect dummies increases the explanatory power of the estimation by more than 50 percent. In these regressions, dependent variables are Education-Achievement, Effort-Achievement, Social network-Achievement, Social network vs. Ability, Effort-Reward, Smart-Reward, Competition, Ratio of financial assets and Ratio of deposits. Therefore, dialect is an important determinant of responses to questions regarding cultural values and behavior.

Table 2. Joint significance of dialect dummies in questions from CFPS-the whole sample

Variables	p-value of joint significance test	R^2 with dialect dummies	R^2 without dialect dummies	ΔR^2	The ratio of rise in R^2
Status-Achievement	0.000	5.8	4.5	1.3	0.289
Wealth-Achievement	0.000	7.3	5.7	1.6	0.281
Education-Achievement	0.000	3.6	1.7	1.9	1.118
Talent-Achievement	0.000	8.3	6.2	2.1	0.339
Effort-Achievement	0.000	3.6	1.6	2.0	1.250
Luck-Achievement	0.000	6.0	4.3	1.7	0.395
Social network-Achievement	0.000	4.3	2.6	1.7	0.654
Social network vs. Ability	0.000	2.5	1.5	1.0	0.667
Wealth as achievement	0.000	4.8	3.6	1.2	0.333
Importance of money	0.000	4.9	3.5	1.4	0.400
Effort-Reward	0.000	5.0	3.0	2.0	0.667
Smart-Reward	0.000	4.8	2.8	2.0	0.714
Attention-Society	0.000	8.9	7.4	1.5	0.203
Attention-Anti-corruption	0.000	9.8	8.4	1.4	0.167
Attention-Law and regulation	0.000	7.5	6.3	1.2	0.190
Attention-Economy	0.000	9.9	8.6	1.3	0.151
Attention-Environment	0.000	9.9	7.8	2.1	0.269
Social sympathy	0.000	11.6	9.1	2.5	0.275
Fairness vs. efficiency	0.000	8.4	6.3	2.1	0.333
Competition	0.000	2.7	1.5	1.2	0.800
Trust	0.000	5.6	4.2	1.4	0.333
Ratio of financial assets	0.000	4.7	3.0	1.7	0.567
Ratio of deposit	0.000	4.3	2.4	1.9	0.792

Notes: p-value shows the joint significance of dialect dummies of each regression. R^2 is expressed in percentage terms. ΔR^2 is the rise in R^2 when dialect dummies are added in the regression. The ratio of rise in R^2 is obtained by the percentage of ΔR^2 in R^2 of regressions without dummies and it reflects the power of dialects in explaining the variation in values compared to control variables.

Table 3 displays the share of joint significant regressions and average R^2 in each province. There are, in total, 21 provinces and 23 regressions for each province. In terms of the joint significance of dialect dummies, more than 50 percent of 23 regressions have significant dialect dummies in 12 provinces, which account for more than half of all provinces. In Shanxi, Henan, Guangdong and Gansu, the share is much higher (more than 80 percent). Except for Liaoning, Shandong and Henan, the average R^2 of regressions is higher than 0.05 when dialect dummies are controlled for. The increase in R^2 is also significant for most provinces. Thus, in most provinces, the explanatory power of dialects for variations in cultural values and attitudes and behavior persists as it does in the whole sample. Hence, according to regressions in the whole sample and selected provinces in the CFPS sample, dialects can explain cultural values and attitudes to a significant extent. Therefore, it is reasonable to proxy cultural diversity by dialectal diversity in China.

Table 3. Joint significance of dialect dummies in questions from CFPS-by province

	Number of regressions	Share of regressions with jointly significant dialect dummies	R^2 with dialect dummies	R^2 without dialect dummies	ΔR^2	The ratio of rise in R^2
Hebei	23	0.696	5.748	3.548	2.200	0.620
Shanxi	23	0.826	6.222	3.274	2.948	0.900
Liaoning	23	0.652	4.552	3.752	0.800	0.213
Jilin	23	0.565	10.874	9.357	1.517	0.162
Heilongjiang	23	0.522	5.004	3.491	1.513	0.433
Jiangsu	23	0.435	7.239	6.304	0.935	0.148
Zhejiang	23	0.217	8.935	7.587	1.348	0.178
Anhui	23	0.348	6.226	5.017	1.209	0.241
Fujian	23	0.304	10.930	8.878	2.052	0.231
Jiangxi	23	0.565	5.941	3.532	2.409	0.682
Shandong	23	0.522	4.857	3.474	1.383	0.398
Henan	23	0.870	4.926	3.874	1.052	0.272
Hubei	23	0.348	7.787	5.843	1.943	0.333
Hunan	23	0.261	5.378	5.039	0.339	0.067
Guangdong	23	1.000	7.539	4.278	3.261	0.762
Guangxi	23	0.391	6.726	5.057	1.670	0.330
Sichuan	23	0.609	5.956	3.439	2.517	0.732
Guizhou	23	0.783	10.287	7.917	2.370	0.299
Yunnan	23	0.478	7.265	5.835	1.430	0.245
Shaanxi	23	0.435	5.387	4.330	1.057	0.244
Gansu	23	0.913	6.748	5.252	1.496	0.285

Notes: R^2 is the average of all regressions in each province and expressed in percentage terms. The ratio of rise in R^2 is obtained by the percentage of ΔR^2 in R^2 of regressions without dummies and it reflects the power of dialects in explaining the variation in values compared to control variables.

IV. Data & Empirical Strategy

Data

The data used in this paper is from four main sources: the population census data, the Chinese Dialects Dictionary together with the Coding Scheme of the Language Atlas of China, the China City Statistical Yearbook and study reports and government documents.[10] To establish the data sample, single-year data of prefecture-level cities is collected firstly over the period 2001-2015.[11] Since changes occurred in jurisdiction areas and units of prefectural cities very often in the 1990s, the panel data is only meaningful when focused on statistics after 1999. Even if there is a change in the administration area from 2001-2015, it is controlled by the respective land area. Furthermore, there is a limitation in accessing official population census data of counties before 2000 and much data on economic development of the same period is missing. In order to reduce endogeneity between economic development and dialectal diversity, 2001 is chosen as the starting year of the sample. To avoid the influence of business cycles, typically 5-year average data is analyzed in the literature. Although shorter period average data can extend the time dimension, 5-year average data is also more appropriate for Chinese economic practices, which is consistent with the 5-year plan regarding social and economic development in China, both at the national level and local levels. Thus, 5-year average data performs better. When data in some year is missing, data of the corresponding period is also treated as missing.

Independent variables. We have five indices for dialectal diversity: dialectal fractionalization (ELF), adjusted dialectal fractionalization (GI), dialectal polarization (RQ), adjusted dialectal polarization (ER) and periphery heterogeneity (PH) (Desmet et al., 2009; Ginsburg and Weber, 2011). ELF is a Herfindahl-based metric measuring the probability that two randomly selected people come from different linguistic groups without considering linguistic distances (Goeren, 2014). We use this in the computation of dialectal fractionalization. It increases with the number of dialect groups and the balance of population distribution among groups. We also consider the other four indices as in the study of linguistic diversity by Desmet et al. (2009). The index taking dialectal distances into consideration based on ELF is called GI, which was proposed by Greenberg (1956). Since dialectal

10 Department of urban social economic investigation, National Bureau of Statistics, China City Statistical Yearbook, 1996-2016, China Statistical Press.

11 There are also cities at the county-level which are under the jurisdiction of prefecture-level cities and same as counties. In this study, we focus on prefecture-level cities.

distances are smaller than 1, GI has smaller values than ELF. RQ was proposed by Reynal-Querol (2002) and is determined by the population distribution between the two largest dialect groups. It is maximized when there are two equally sized groups and decreases with an increasing number of equally sized groups. Thus, fractionalization is positively associated with polarization at low levels, not associated with polarization at intermediate levels and negatively associated with polarization at high levels (Goeren, 2014; Ager and Bruekner, 2013). ER is the polarization measure adjusted by dialectal distances and was proposed by Esteban and Ray (1994). Thus, ER is affected most by the population shares of the two largest groups and the dialectal distance between them. The largest index, PH, was proposed by Desmet et al. (2005) and takes dialectal distances into consideration. It reflects the alienation between peripheral groups and the largest group.

There are three steps to calculate dialectal diversity at the prefectural level. Firstly, since people in each county use one dialect, each county is matched with a dialect code referred to in the Coding Scheme of the Language Atlas of China. The code is designed at the dialect subgroup level, providing information on both low and high levels of dialect groups. For counties in which more than one dialect is used, only the code of the dominant dialect is taken. The matching is conducted through the names of counties directly and 2625 counties are matched. Counties whose names have changed are also considered in the matching. Furthermore, 51 counties not covered in the coding scheme are added in the Chinese Dialects Dictionary, whose codes are added by comparing with other counties with the same dialects.[12]

Secondly, dialectal distances are assigned to each pair of languages used in each city according to the method proposed by Fearon (2003). There are 6 levels in the tree of Chinese dialects, as shown in Figure 1. Levels 2 to 6 are made up of phylum, stock, supergroup, group and dialectal subgroup, respectively. Based on the data available, the analysis is focused on the diversity of Chinese dialects belonging to the Sino-Tibetan phylum. These dialects are divided into 8 supergroups – a Mandarin supergroup and 11 non-Mandarin supergroups. The Mandarin supergroup includes 8 groups while the non-Mandarin supergroups include more than 40 groups. The dialectal distances are assigned to each pair of dialects according to the codes of dialects, (See details in Appendix B).

[12] Xu, Baohua; and Ichiro Miya, Chinese Dialects Dictionary (p. fl156-fl224), 1999, Zhong Hua Book Company.

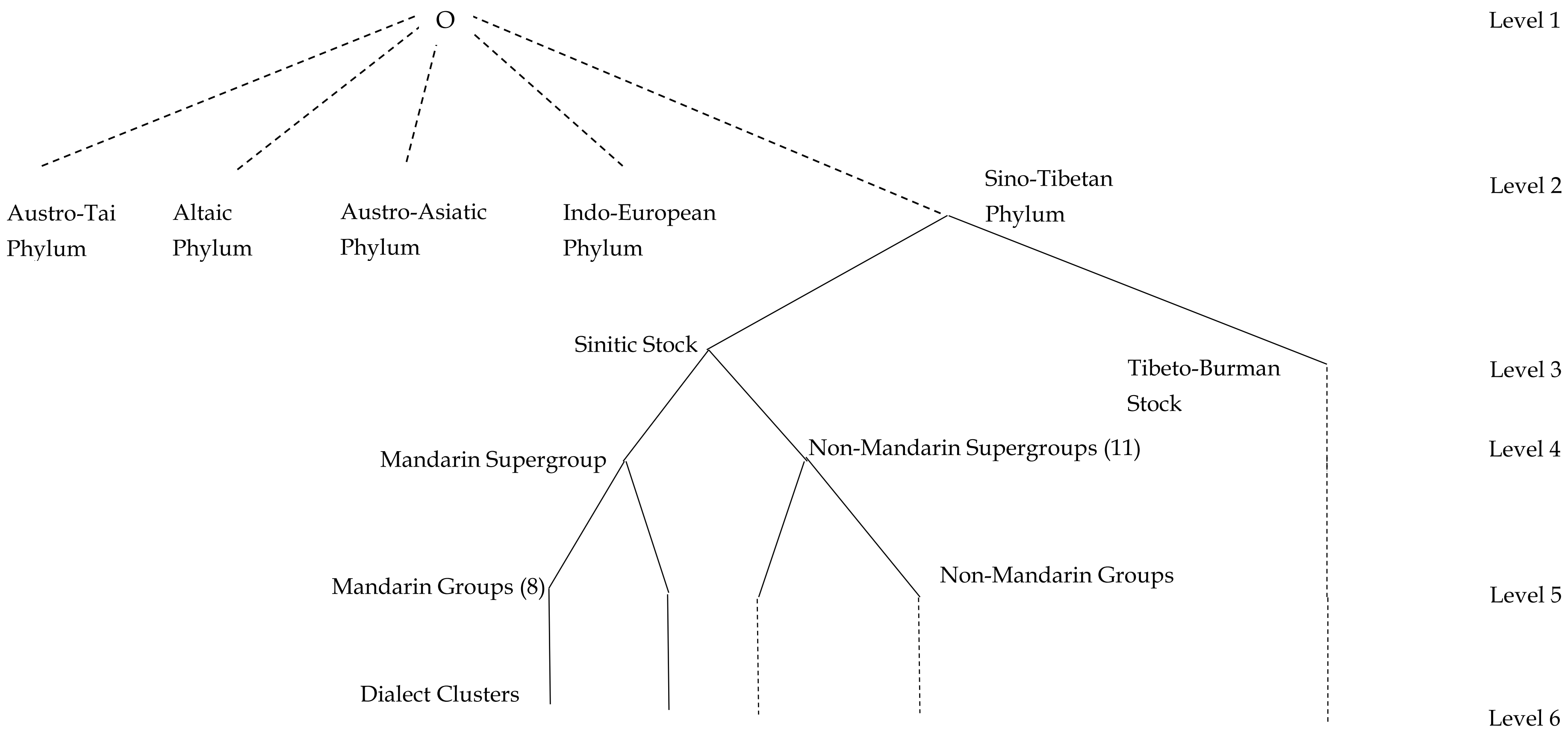


Figure 1. Language Tree of Chinese Dialects

Notes: As in Desmet et al. (2012), we assume there is an original language, O, of all language types that is at Level 1 of the game tree. Since Han dialects are in the Sino-Tibetan phylum, Sinitic stock more specifically, branches of other phyla and Tibeto-Burman are not drawn. For groups who have no subgroups, a dashed line is drawn and we assume that the subgroup is the same as its lower level group. From Level 4, due to limits of space, the specific name of each supergroup, group or cluster is not present.

Thirdly, the population share of each dialectal subgroup in each city is calculated given the population data from the population census (2000, 2010) and the China Population Statistics Yearbook (2006), which provide population information at the county level.[13] Then given the population share of dialect groups and dialectal distances, the five indices are computed according to the method used by Desmet et al. (2009) (see details in Appendix B). Given the limited data on population at the county level, only data on the dialectal diversity of 274 cities in the year 2000 and 2010, and 275 cities in year 2005 is reserved in the sample. Since dialectal diversity does not change in a short period, values of dialectal diversity in the years 2000, 2005 and 2010 are assigned to observations over the period 2001-2005, the period 2006-2010 and the period 2011-2015, respectively, which is one way to reduce the potential problem of endogeneity. Table 4 shows the descriptive statistics of five diversity indices in years 2000, 2005 and 2010 separately.

Table 4. Measurements of dialectal diversity

			Year=2000		
Stats	ELF	GI	RQ	ER	PH
Obs	274	274	274	274	274
Mean	0.2208	0.0847	0.0974	0.0144	0.0760
Std. Dev.	0.2333	0.1050	0.0978	0.0324	0.0899
Min	0.0000	0.0000	0.0000	0.0000	0.0000
Max	0.7802	0.4275	0.2496	0.1334	0.2988
			Year=2005		
Stats	ELF	GI	RQ	ER	PH
Obs	275	275	275	275	275
Mean	0.2219	0.0850	0.0979	0.0369	0.0761
Std. Dev.	0.2339	0.1058	0.0981	0.0429	0.0901
Min	0.0000	0.0000	0.0000	0.0000	0.0000
Max	0.7915	0.4272	0.2498	0.1497	0.2993
			Year=2010		
Stats	ELF	GI	RQ	ER	PH
Obs	0.2219	0.0853	0.0979	0.0369	0.0766
Mean	0.2340	0.1059	0.0981	0.0430	0.0901
Std. Dev.	0.0000	0.0000	0.0000	0.0000	0.0000
Min	0.7915	0.4272	0.2498	0.1497	0.2993
Max	0.2219	0.0853	0.0979	0.0369	0.0766

In the whole sample, there are 123 observations showing no dialectal diversity, accounting for

[13] Tabulation on the 2000 Population Census Data of China and Tabulation on the 2010 Population Census Data of China, China Statistics Yearbook. Department of Population and Employment statistics, National Bureau of Statistics, the China Population Statistics Yearbook, 2006), China Statistics Yearbook.

44.9% of 823 observed prefecture-level cities. Table 4 shows the descriptive statistics of five diversity indices in years 2000, 2005 and 2010 separately. Taking 0 as the minimum value for each index, ELF has the highest maximum value, close to 0.8, and ER has the lowest, which is less than 0.15. Comparing the mean of each index in each year, all the indices have higher values in 2005 than in 2000 and in 2010. For ELF, GI, RQ and PH, the values in 2000 are higher than in 2010, while ER has a higher value in 2010 than in 2000.

The distribution of dialectal diversity among all the observed prefecture-level cities in 2000 can be seen in the maps in Figure 1A – Figure 1E. For all the indices, all the prefectures are divided into five groups: homogeneous, low diversity, middle low diversity, middle high diversity and high diversity. Firstly, there is no significant change in the distribution across all prefectures of all indices. Secondly, cities with a diversity level are not concentrated in one area. Thirdly, when diversity is measured by GI and ER, the proportion of cities with high diversity increases, although there are a few cities that become less diverse compared to the case when diversity is measured by ELF and RQ. Fourthly, by comparing Figure 1A and Figure 1C, cities with middle high and high ELF tend to be located in South China, while cities with a middle high and a high RQ are more evenly distributed. Moreover, cities with a middle high and a high GI, ER and PH are more likely to be located in South China, which can be seen in Figures 1B, 1D and 1E. Furthermore, the distributions of dialectal diversity across cities are similar in the other two periods, which are shown in Figures C1-C5 in Appendix C.

Dependent variable. The dependent variable is income per capita proxied by the gross regional product per capita. We have data on the gross regional product (GRP) per capita at current year's prices in the China City Statistical Yearbook (2001-2016), which is adjusted to the price level in 1995.[14] Due to missing data in the statistical yearbook, data on income per capita is only available for 801 observations in the 5-year average dataset. We report the distribution of ELF and income of each period in Figures 2A- 2C. We also have more observations of average income for the period 2006-2010 and the period 2011-2015 and find that there is no explicit relationship between the distribution of ELF and the distribution of income in each period. High income can be observed in cities with low ELF as well as in cities with high ELF and the same holds for cities with relatively low income. Hence,

[14] The data of GDP inflator and investment price index is obtained from the data in the China City Statistical Yearbook (1995-2015).

there is no clear pattern regarding the relationship between dialectal diversity and economic development. In addition, we use the logarithm value of income per capita in the estimations.

Control variables. In the baseline regression, we have five groups of control variables. The first group includes the public expenditure per capita and the fixed asset investment per capita. The data on these two variables is mainly from the China City Statistical Yearbook (2001-2016), but the data on public expenditure per capita in the years 2001 and 2002 is from provincial statistical yearbooks for each year. Furthermore, public expenditure is adjusted at the price of the year 1995 by the GDP deflator and fixed asset investment is adjusted by the investment price index of the respective province. The logarithm values of these are put in the regression. The second group is the industry structure reflected by the ratio of the primary industry and the ratio of the second industry in the economy of prefecture-level cities. The third group reflects the financial development, including the ratio of loans in the GRP and the ratio of residential deposit in the GRP. The data of these two groups of controls is from the China City Statistical Yearbook (2001-2016). The fourth group is deals with labour and human capital, including the logarithm of population, employment rate, the average years of education per capita, the logarithm of enrolment of students in regular secondary schools and the number of key universities. The data on population, employment rate and enrolment of students is obtained from the China City Statistical Yearbook (2001-2016). The average years of education per capita is abstracted from the population census data in 2000 and 2010. The value in 2000 is matched with the periods 2001-2005 and 2006-2010 and the value in 2010 is matched with the period 2011-2015. To capture the capacity in promoting education development, the number of key universities in each city is obtained from the Ministry of Education.[15] Other control variables include the logarithm of highway freight traffic per capita, total land area, market institutions and intermediate organizations, the number of high technology zones to control the effect of transportation conditions, the constraint of land and related resources, market environment and development in technology. The data on highway freight traffic and total land area is from the China City Statistical Yearbook (2001-2016), with highway freight traffic divided by population to get its per capita level. The index of market institutions and intermediate organizations is from the Marketization Indexes Report of China Provinces (2011, 2016).[16] The data on the number of high technology zones is gained from

[15] http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/moe_648/200506/10003.html

[16] Fan, Gang, Xiaolu Wang, Hengpeng Zhu, China's marketization index: the relative process of regional

government policy documents and the number in each year is adjusted based on the policy of the year before. In addition to the control variables above, period dummies are also included to control for time trends. Furthermore, in order to identify the effect of dialectal diversity on economic growth, income per capita in lagging periods is also taken as a control variable. We also collect data on gross regional product per capita over the period 1996-2000. Table 5 presents the basic information of all the variables apart from the diversity indices.

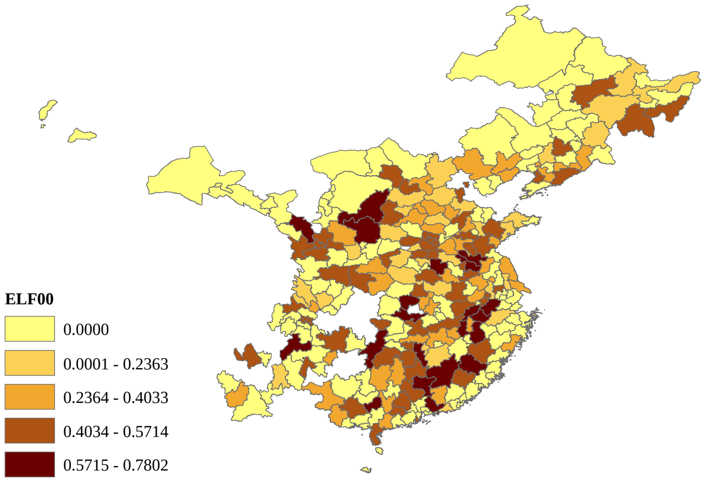


Figure 1A. ELF in the year 2000

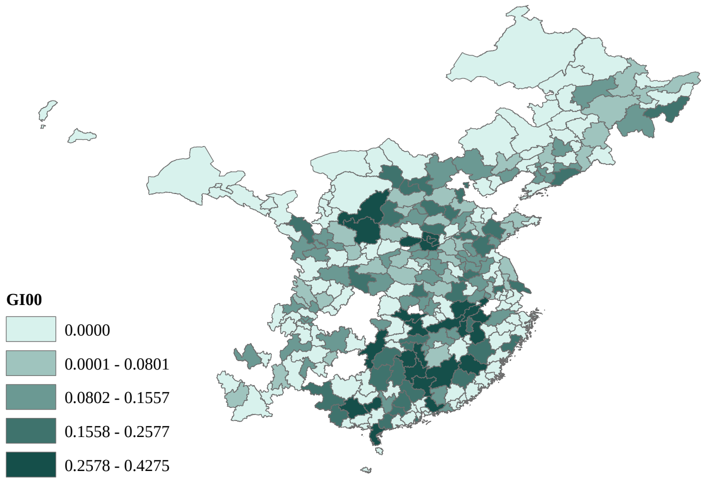


Figure 1B. GI in the year 2000

marketization, 2011, Economic Science Press. Wang, Xiaolu, Gang Fan, Jingwen Yu, China's provincial marketization index report, 2017, Social Sciences Academic Press (China).

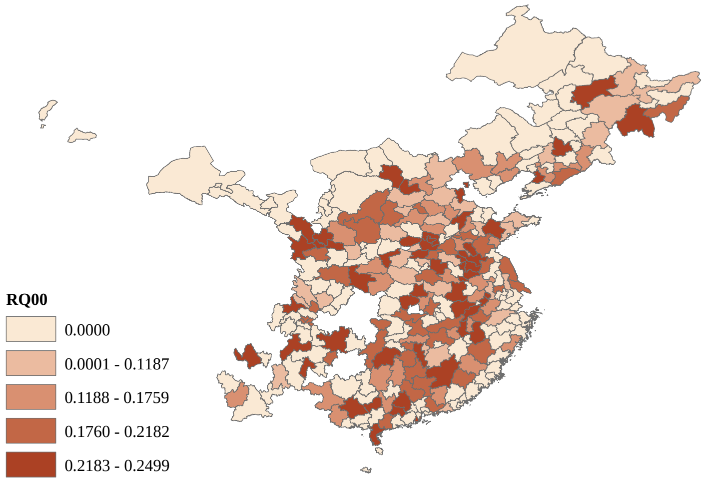


Figure 1C. RQ in the year 2000

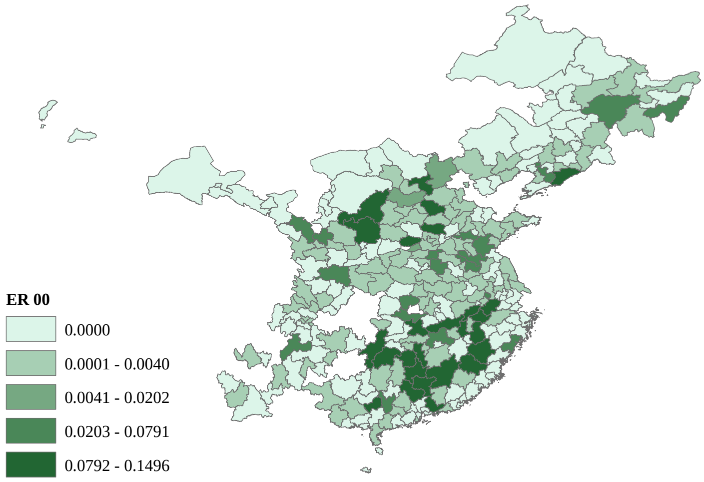


Figure 1D. ER in the year 2000

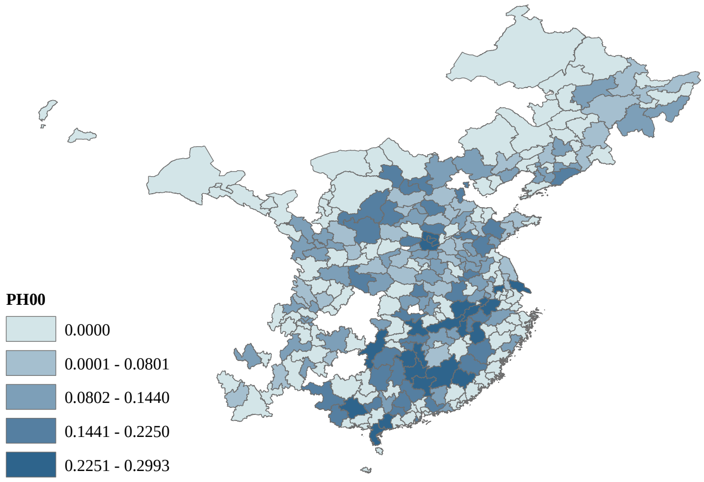


Figure 1E. PH in the year 2000

Figure 1. Distribution of dialect diversity in the year 2000 in the observed cities

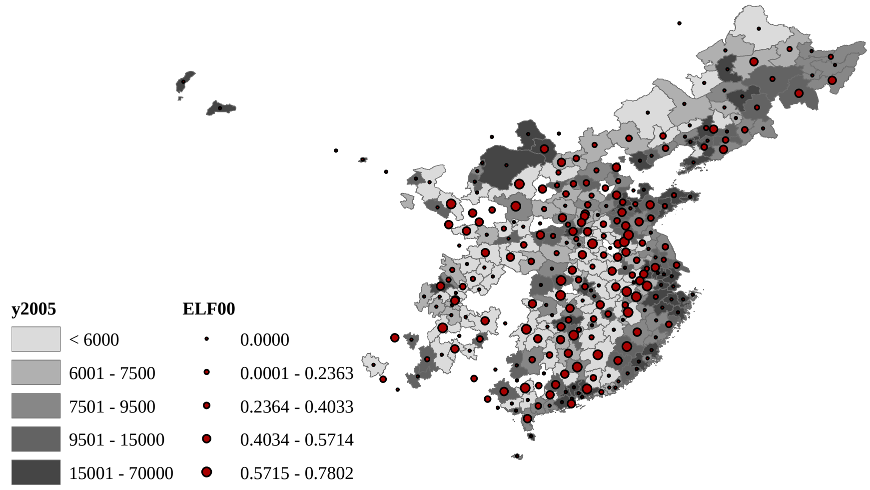


Figure 2A. Distribution of dialectal diversity and average income of 2001-2005

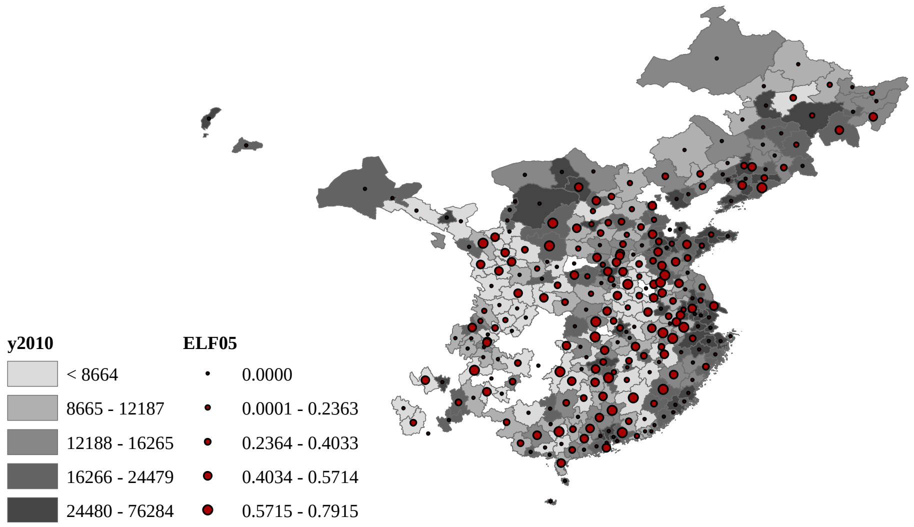


Figure 2B. Distribution of dialectal diversity and average income of 2006-2010

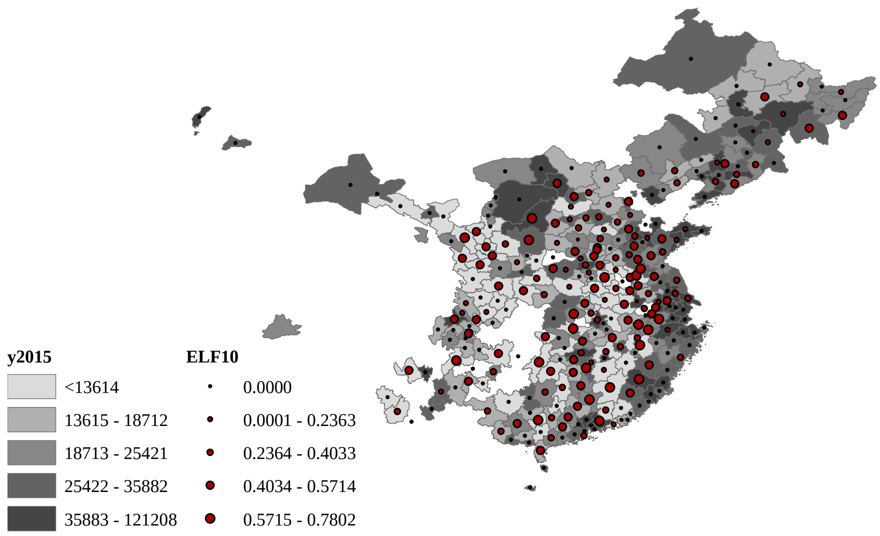


Figure 2C. Distribution of average income of 2011-2015

Figure 2. Distribution of dialectal diversity and average income in 2001-2015

Table 5. Variable description and sources

VARIABLES	Notation	Unit	Source
Income	lny	Yuan	Calculation based on data of China City Statistical Yearbook (2001-2016)
Public expenditure	lnpe	Yuan	Calculation based on data of China City Statistical Yearbook (2001-2016)
Fixed asset investment	lnfai	Yuan	Calculation based on data of China City Statistical Yearbook (2001-2016)
Ratio of primary industry	ppg	%	Calculation based on data of China City Statistical Yearbook (2001-2016)
Ratio of secondary industry	spg	%	Calculation based on data of China City Statistical Yearbook (2001-2016)
Ratio of loans	rlnb	%	Calculation based on data of China City Statistical Yearbook (2001-2016)
Ratio of residential deposit	rdnb	%	Calculation based on data of China City Statistical Yearbook (2001-2016)
Population	lnapop		Calculation based on data of China City Statistical Yearbook (2001-2016)
Employment rate	empr	%	Calculation based on data of China City Statistical Yearbook (2001-2016)
Education level	hc	Year	Population Census Data (2000, 2010)
Enrolment of students	lnnrss		Population Census Data (2000, 2010)
Number of key universities	n29	Integral	Ministry of Education of the People's Republic China
Highway freight traffic	lnhft	Ton	Calculation based on data of China City Statistical Yearbook (2001-2016)
Total land area	lntlaar	km^2	Calculation based on data of China City Statistical Yearbook (2001-2016)
Market institutions and organization	mio	Index	Marketization Indexes Report of China Provinces (2011, 2016)
High technology zones	htdz	Integral	Government policy documents

Table 6 provides the basic descriptive statistics for all the main variables in addition to the dialectal diversity indices. There is no outlier for any variables. Although there are large differences between the minimum values and maximum values for the ratio of loans, ratio of resident deposits and highway freight traffic, the standard deviation is smaller than the mean. The Pearson correlation coefficients between the independent variables and their significance level can be seen in Table A2. Firstly, there are no correlation coefficients between dialectal diversity indices and other variables larger than 0.8. Thus, we believe that there is no collinearity problem in the regression analysis.

Secondly, although high correlation appears between public expenditure and fixed asset investment and between the ratio of loans and the ratio of resident deposits, the regression result shows that they are all significant and the correlation has no potential problem. Besides, they are also controlled simultaneously in the literature. Hence, controlling these variables will not cause a collinearity problem in the regression.

Table 6. Descriptive analysis of main variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Income	801	9.5627	0.7253	7.7098	11.7053
Public expenditure	802	7.5018	0.8069	5.6925	10.6270
Fixed asset investment	733	8.9635	1.0116	6.5705	11.3043
Ratio of the primary industry	800	15.2599	9.1648	0.0440	48.5700
Ratio of the secondary industry	800	48.6292	10.7448	16.3040	86.1200
Ratio of loans	816	80.8407	43.9151	15.0000	302.0000
Ratio of residential deposit	817	116.4468	51.3686	52.0000	576.0000
Population	818	5.8345	0.6621	2.7732	7.1001
Employment rate	818	0.9631	0.0228	0.7580	0.9930
Education level	823	8.0574	0.9984	5.0800	11.1200
Enrolment of students	818	2.9937	0.6881	-0.1625	4.4951
Number of key universities	823	0.3096	1.1739	0.0000	9.0000
Highway freight traffic	812	2.9505	0.7299	0.7326	5.5700
Total land area	819	9.2886	0.7469	6.9694	11.2132
Market institutions and organization	823	5.1957	2.9087	1.0800	14.5100
High technology zones	823	0.2423	0.4216	0.0000	2.0000

Note: Because of data missing for some prefecture-level cities, the number of observations for most variables is smaller than 823 in the sample of 5-year average data.

Empirical Strategy

The analysis starts with the estimation of the basic specification of the two-way fixed-effects model with panel data:

$$Outcome_{it} = \beta_0 + \beta_1 Diversity_{it} + \left(control\ \text{var}iables\right)_{it} \gamma^T + \alpha_i + \eta_t + \varepsilon_{it}$$

The dependent variable is represented by the logarithm of income per capita, with the coefficients of the independent variables reflecting their effect on economic growth. $Diversity_{it}$ represents the dialectal diversity of city *i* in period *t* and regressions regarding ELF, GI, RQ, ER and PH will be run separately. The control variables include public expenditure, fixed asset investments, industrial structure, financial development and variables of labor and human capital, as well as other variables shown in Table 2. These variables will be included in the regression step by step; α and η are included to control individual and time effects. To avoid the impact of the persistence of economic development in lagging periods, we add income per capita in lagging period 1 as a control variable in the estimation. Furthermore, to identify the effect of dialectal diversity on changes in economic growth over periods, we also run regressions controlling income per capita in lagging period 2.

To solve the potential endogeneity problem between dialectal diversity and economic development, pooled-2SLS, FE-2SLS and IV-GMM are applied after the baseline regression. We instrument dialectal diversity with historical migration, the average altitude and the share of land area with an altitude under 500 meters. There are five large-scale migration waves within China at different periods in history – the Yong Jia Rebellion in the Western Jin dynasty, migration in the Sui, Tang and Five dynasties, migration because of the shame of Jing Kang of the Song dynasty, migration at the beginning and middle of the Ming dynasty and "Hu-Guang people fill Sichuan" in the Ming and Qing dynasties. As in Shao et al. (2017), all five migration waves are considered in constructing the instrumental variable of historical migration. Firstly, five dummies of each migration wave are constructed. If a city received immigration in the given migration wave, the corresponding dummy takes the value 1, and 0 if otherwise. Based on the migration map and records for all cities in the area where immigrants were densely populated, the dummy takes the value 1.[17] If cities are in the area where immigrants are dispersed, only cities that accepted immigrants are specified. Secondly, we take the sum of these dummies.[18] Of all the dummies for each migration wave and summations of these dummies, we find that migration at the beginning and middle of the Ming dynasty works best as an instrumental variable, thus choosing the dummy for this as the proxy of historical migration. The data of the other two instrumental variables, the average altitude and the share of land area with an altitude under 500 meters, is abstracted from the DEM data by ArcGIS. If the altitude is above 500 meters, it is difficult for people historically to communicate and mobilize and languages are kept isolated from each other, thus resulting in higher dialectal distances. But it may also be positively related to dialectal diversity because regions with better geographical conditions are easier for

[17] Migration maps and records can be found in Ge, Jianxiong (chief editor), Shuji Cao, Songdi Wu, 1997, Migration history of China, Vol. 1-Vol. 6, Fujian People's Publishing House.

[18] We sum the dummies of all migration waves, the latest four waves, the latest three waves and the latest two waves, respectively.

population mobility and the formation of a higher number of dialect groups. Since historical migration, the average altitude and the share of land with an altitude below 500 meters are time-invariant, we first perform the IV regression with a pooled 2SLS model. Then we apply 2SLS and IV-GMM regressions in the fixed-effects model, in which instrumental variables are represented by interaction terms of each variable and period dummies according to the method proposed by Acemoglu et.al (2005). Furthermore, the influence of exposure to governance by the Chinese Communist Party (CCP) during the revolutionary war is obtained by estimating the effect of the experience of being in an area controlled by the CCP from the 1920s to the 1940s. A dummy variable, revolutionary area, is constructed indicating whether a city was governed by the CCP during the revolutionary area.[19] If more than 50% of the counties of a city have revolutionary towns accounting for over 59% of all towns, the variable takes a value of 1.[20] Otherwise, it takes 0. All the cities in the sample are divided into two groups, which are shown in Figure C6 (see Appendix C). The regression is run through the fixed-effects model by including the interaction term between revolutionary area and dialectal diversity as well as control variables:

$$\begin{aligned} Outcome_{it} = {} & \beta_0 + \beta_1 Diversity_{it} + \beta_2 Diversity_{it} * \mathrm{Re}volutionary\ Area \\ & + \left(control\ \ \mathrm{var}iables * \mathrm{Re}volutionary\ Area\right)_{it} \gamma^T + \alpha_i + \eta_t + \varepsilon_{it}\ . \end{aligned}$$[21]

V. Results

Baseline Results

Taking economic growth as the outcome variable, the results of baseline estimations regressed on ELF, RQ and PH are shown in Table 7A to Table 7C. In each table, column (1) is the result of the regression when only dialectal diversity is included in the model. Column (2) is the result of the regression when public expenditure and fixed asset investment are added as control variables and column (3) is the result when the industrial structure is also controlled. Based on the estimated model of column (3), financial development, population and human capital, and other control variables are

---

[19] Based on the Soviet area map as well as the map of revolutionary bases, http://dangshi.people.com.cn/GB /151935/164962/

[20] Other dummies, indicating more than 30% and 75% of counties of each prefecture-level city, are also constructed and used in the regression analysis. And we select the one performing the best.

[21] Interaction terms between control variables and revolutionary area are also regressed to control the potential effect on the effect of control variables of the experience governed by the CCP during the revolutionary war.

included in estimations of columns (4), (5) and (6). Column (7) and column (8) show the regression results when income in lagging period 1 and income in lagging period 2, respectively, are controlled.

In Table 7A where ELF is the independent variable, the coefficient of ELF is positive and significant at the level of 0.01 from column (1) to column (6) when income in lagging periods is not considered. In column (7) and (8) when income in lagging periods is controlled, coefficients of ELF are not significant, but still positive. In addition, income in lagging period 1 shows a positive effect on current income and the coefficients are less than 1, which is consistent with the growth theory that economies with a higher initial income level have slower economic growth. We also notice that income levels in lagging period 1 and 2 are not significant, but we find that this is the result of collinearity between them and other control variables. Table 7B shows the result of the estimation when RQ is taken as the independent variable. The coefficient is higher than the coefficient of ELF, but the significance level is same as that in Table 7A. In both estimations, we find that there is a large decrease in the coefficients of ELF and RQ in column (2) compared to those in column (1). This may be because dialectal diversity has a significant impact on public expenditure and fixed asset investment, and the indirect effect of dialectal diversity is separated from the direct effect on economic growth. Furthermore, the magnitude and significance of the coefficients of RQ become lower when income in the lagging periods is controlled, but they are still positive although they are not significant. In contrast, Table 7C shows that PH has no significant effect if economic development in the lagging periods is not controlled. The coefficients are positive and significant at the level of 0.01 and 0.05, however, when income in the lagging periods is controlled.

We also run an estimation when dialectal diversity is proxied by the adjusted dialectal fractionalization, GI, and adjusted dialectal polarization, ER, respectively. The results are shown in Table A3 and Table A4 (see Appendix A). When the dialectal distances between all the groups are equally considered, column (2)–column (6) of Table A3 indicate that GI is not significant in explaining differences in economic growth. Moreover, it is significant only when income in both lagging period 1 and 2 is included as a regressor. However, when income in lagging period 2 is considered, GI becomes significant at the level of 0.05. However, in Table A4, we observe that coefficients are not significant whether income in lagging periods is considered or not.

Therefore, based on the baseline result, ELF and RQ have a significant and positive effect on economic growth. GI and PH show a significant and positive effect on economic growth when lagged economic development is controlled, while ER has no significant effect in any cases. Therefore, the effect of GI, ER and PH on economic growth is related to dialectal distances and also the way in which indices are adjusted by dialectal distances. This may be because the potential benefits of the difference have not been exploited completely. Furthermore, the significant effect of PH on economic growth also suggests that the effect is determined by how indices are adjusted by dialectal distances, but the dialectal distance between polarized groups has no significant influence. But the result may also suffer from reverse causality between income and dialectal diversity.

In the literature, the endogeneity problem is that economic development tends to reduce linguistic diversity because people tend to be assimilated by the mainstream culture and languages. But cultural evolution is a long-term process. On the other hand, economic development may promote population diversity by promoting population mobility. Along with economic development, the population of smaller dialect groups grows faster and thus the distribution of population among dialect groups becomes more balanced. Then, ELF and RQ increase. On the other hand, the positive effect on dialectal diversity may be smaller for indices adjusted by dialectal distances precisely because they are also determined by dialectal distances. Since it is easier for dialect groups with less distant dialects to benefit from economic development, their population share may grow faster than others. For example, small dialect groups who have less dialectal distance from the central group may grow faster than other groups. In this case, GI, ER and PH face a less positive effect from economic development and the effect might be negative if groups closer to each other in dialect become large enough. Therefore, when economic development has a positive effect on ELF as well as on RQ, ELF and RQ increase with increasing economic growth, with the result that the significant positive effect of ELF and RQ may be overestimated. When GI, ER and PH are affected by income per capita negatively, they decline along with economic growth. If the true effect of these on economic growth is positive, the insignificance of their effect in the baseline estimation should be the result of underestimation. Therefore, to verify whether the true effect of dialectal diversity is identified, IV analysis will be conducted as well.

Table 7A. Baseline results of the relationship between ELF and economic performance

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ELF	1.571***	1.174***	1.068***	1.059***	1.128***	1.063***	0.537	0.523
	(0.335)	(0.338)	(0.320)	(0.322)	(0.344)	(0.352)	(0.379)	(0.406)
Public expenditure		0.209***	0.138**	0.138**	0.122**	0.130**	0.107	0.074
		(0.063)	(0.054)	(0.054)	(0.055)	(0.058)	(0.066)	(0.083)
Fixed asset investment		0.166***	0.083***	0.086***	0.082***	0.084**	0.114***	0.049
		(0.031)	(0.031)	(0.031)	(0.031)	(0.033)	(0.033)	(0.056)
Ratio of primary industry			-0.005*	-0.005*	-0.006**	-0.006*	-0.003	-0.007*
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Ratio of secondary			0.009***	0.009***	0.009***	0.009***	0.009***	0.011***
industry			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Ratio of loans				-0.000	-0.000	-0.000	0.000	0.001**
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ratio of residents' deposit				-0.000	-0.000	-0.000	-0.000	-0.000*
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Population					-0.074	-0.077	-0.074	-0.434***
					(0.130)	(0.164)	(0.154)	(0.138)
Employment rate					0.503**	0.481*	0.115	-0.148
					(0.247)	(0.256)	(0.247)	(0.219)
Education level					0.014	0.010	-0.043	-0.009
					(0.041)	(0.042)	(0.027)	(0.029)
Number of key					-0.007	-0.006	0.004	-0.175
universities					(0.016)	(0.016)	(0.014)	(0.136)
Enrolment of students					-0.054	-0.054	-0.005	-0.047
					(0.039)	(0.039)	(0.039)	(0.049)
Market institutions and						0.002	0.001	0.002
organizations						(0.004)	(0.003)	(0.008)
Land area						0.023	-0.015	0.027
						(0.098)	(0.096)	(0.073)
High technology zones						0.001	-0.023	-0.015
						(0.029)	(0.019)	(0.022)
Highway freight traffic						-0.027	-0.030	-0.066**
						(0.024)	(0.025)	(0.027)
Income per capita in							0.286***	0.070
lagging period 1							(0.039)	(0.060)
Income per capita in								0.018
lagging period 2								(0.067)
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	9.671***	6.387***	7.427***	7.394***	7.544***	7.403***	5.699***	10.463***
	(0.072)	(0.473)	(0.456)	(0.459)	(1.185)	(1.202)	(1.303)	(1.587)
Observations	801	729	728	723	723	712	656	412
R-squared	0.933	0.956	0.964	0.964	0.964	0.964	0.972	0.954
Number of cities	275	253	253	253	253	252	250	229

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 7B. Baseline results of the relationship between RQ and economic performance

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RQ	3.779***	2.530***	2.275***	2.253***	2.370***	2.274***	1.117	0.998
	(0.758)	(0.736)	(0.716)	(0.720)	(0.788)	(0.802)	(0.837)	(0.898)
Public expenditure		0.208***	0.137**	0.137**	0.122**	0.129**	0.107	0.076
		(0.063)	(0.054)	(0.054)	(0.055)	(0.058)	(0.066)	(0.083)
Fixed asset investment		0.166***	0.083***	0.086***	0.082***	0.084**	0.114***	0.048
		(0.031)	(0.031)	(0.031)	(0.031)	(0.033)	(0.033)	(0.056)
Ratio of primary industry			-0.005*	-0.005*	-0.006*	-0.006*	-0.003	-0.006
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Ratio of secondary industry			0.009***	0.009***	0.009***	0.009***	0.009***	0.012***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Ratio of loans				-0.000	-0.000	-0.000	0.000	0.001**
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ratio of residents' deposit				-0.000	-0.000	-0.000	-0.000	-0.000*
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Population					-0.075	-0.078	-0.074	-0.434***
					(0.130)	(0.164)	(0.154)	(0.138)
Employment rate					0.500**	0.480*	0.113	-0.152
					(0.247)	(0.256)	(0.247)	(0.222)
Education level					0.014	0.009	-0.043	-0.010
					(0.041)	(0.042)	(0.027)	(0.029)
Number of key universities					-0.007	-0.006	0.004	-0.172
					(0.016)	(0.016)	(0.014)	(0.136)
Enrolment of students					-0.053	-0.052	-0.004	-0.047
					(0.039)	(0.039)	(0.039)	(0.050)
Market institutions and						0.002	0.001	0.002
organizations						(0.004)	(0.003)	(0.008)
Land area						0.025	-0.014	0.028
						(0.098)	(0.096)	(0.074)
High technology zones						0.000	-0.024	-0.016
						(0.029)	(0.019)	(0.022)
Highway freight traffic						-0.027	-0.030	-0.067**
						(0.024)	(0.025)	(0.027)
Income per capita in							0.285***	0.070
lagging period 1							(0.039)	(0.061)
Income per capita in								0.019
lagging period 2								(0.067)
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	9.650***	6.405***	7.443***	7.410***	7.568***	7.407***	5.710***	10.464***
	(0.072)	(0.476)	(0.459)	(0.462)	(1.191)	(1.209)	(1.303)	(1.587)
Observations	801	729	728	723	723	712	656	412
R-squared	0.934	0.956	0.964	0.964	0.964	0.964	0.972	0.954
Number of cities	275	253	253	253	253	252	250	229

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 7C. Baseline results of the relationship between PH and economic performance

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PH	1.749	1.004	0.965	0.989	1.016	1.002	1.243***	1.269**
	(1.083)	(0.983)	(0.848)	(0.829)	(0.832)	(0.811)	(0.416)	(0.545)
Public expenditure		0.215***	0.143**	0.143**	0.128**	0.134**	0.107	0.075
		(0.065)	(0.056)	(0.056)	(0.057)	(0.059)	(0.066)	(0.083)
Fixed asset investment		0.169***	0.085***	0.088***	0.085***	0.088***	0.118***	0.053
		(0.031)	(0.031)	(0.031)	(0.031)	(0.033)	(0.033)	(0.055)
Ratio of primary industry			-0.005*	-0.005*	-0.006**	-0.006*	-0.003	-0.006
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Ratio of secondary industry			0.009***	0.009***	0.009***	0.009***	0.009***	0.011***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Ratio of loans				-0.000	-0.000	-0.000	0.000	0.001**
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ratio of residents' deposit				-0.000	-0.000	-0.000	-0.000	-0.001**
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Population					-0.070	-0.073	-0.063	-0.405***
					(0.129)	(0.164)	(0.153)	(0.143)
Employment rate					0.461*	0.438*	0.143	-0.128
					(0.250)	(0.257)	(0.249)	(0.210)
Education level					0.003	-0.000	-0.048*	-0.011
					(0.038)	(0.038)	(0.026)	(0.028)
Number of key universities					-0.011	-0.010	0.003	-0.182
					(0.017)	(0.017)	(0.014)	(0.135)
Enrolment of students					-0.055	-0.053	-0.005	-0.053
					(0.039)	(0.039)	(0.039)	(0.049)
Market institutions and						0.002	0.001	0.003
organizations						(0.004)	(0.003)	(0.008)
Land area						0.026	-0.011	0.028
						(0.099)	(0.096)	(0.075)
High technology zones						0.001	-0.024	-0.013
						(0.028)	(0.019)	(0.022)
Highway freight traffic						-0.030	-0.032	-0.066**
						(0.025)	(0.025)	(0.027)
Income per capita in lagging							0.281***	0.064
period 1							(0.039)	(0.061)
Income per capita in lagging								0.022
period 2								(0.068)
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	9.880***	6.491***	7.526***	7.487***	7.762***	7.582***	5.667***	10.299***
	(0.083)	(0.470)	(0.454)	(0.457)	(1.111)	(1.159)	(1.308)	(1.613)
Observations	801	729	728	723	723	712	656	412
R-squared	0.932	0.956	0.963	0.963	0.964	0.964	0.972	0.954
Number of cities	275	253	253	253	253	252	250	229

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Results of IV estimation

All three potential instrumental variables are examined for dialectal diversity separately and the F-statistic of the exclusion test and the values of Prob>F in the first-stage regression are reported in Tables A5-A14 (see Appendix A). On the basis of pooled 2SLS regression results, we observe that taking historical migration or the average altitude as instruments leads to higher F-statistics in first stage regressions when ELF, GI, ER and PH are used as independent variables (the F-statistics of the instruments in the first stage are higher than 10 in most cases). When RQ is the independent variable, regressions with historical migration and the share of land with altitude below 500m have higher F-statistics in the first stage. Thus, we include historical migration and one geographical factor of the above as instruments in IV regressions through pooled 2SLS. In regressions using FE-2SLS and IV-GMM, we can see that all F-statistics in the first stage are smaller than 10, but the coefficients of the average altitude or the share of land with altitude below 500m are significant for all dialectal indices. In regressions using FE-2SLS and IV-GMM, we select one geographical factor with a higher F-statistic in the first stage as the instrument. Thus, we use the average altitude as the instrument for ELF, GI, RQ and PH and the share of land with altitude below 500m as instrument for ER. We estimate robust standard errors in all the regressions.

The results of IV regressions on ELF, RQ and PH with selected instrumental variables are reported in Tables 8A-8C, with each table containing one index of diversity as the independent variable. Columns (1)-(3) display the results of the pooled 2SLS regression. Columns (4)-(6) display the results of the FE 2SLS regression and columns (7)-(8) show the results of IV-GMM regression. Table 8A shows that ELF has a positive and significant effect on economic growth no matter whether the economic development in the lagging period 1 is controlled or not. But the effect does not persist when income per capita in lagging period 2 is controlled. Thus, ELF has an effect on the level of economic growth, but no effect on the increase in growth over periods. We can see the result of regressions on RQ in Table 8B, which shows that RQ has a positive and robust effect on economic growth, but no significant effect on the change in economic growth when income in lagging period 2 is included as a control variable. Furthermore, we can see a similar significant, positive and robust effect of PH on economic growth in Table 8C in all regressions. The results of the regression on GI are reported in Table A15, where we find no robust effect of GI on economic growth, but it may affect change in economic growth positively (see Appendix A). However, as the polarization index adjusted by dialectal distances, ER only shows a positive effect on economic growth and an increase over periods in the pooled 2SLS estimation, but the effect is not robust (see Table A16 in Appendix A). Furthermore, in the first stage of pooled 2SLS regressions, most of the F-statistics are higher than 10, with others close to 10, and the coefficients of the instrumental variables are significant. In the first stage of FE-2SLS and IV-GMM regressions, the coefficients of instrumental variables are significant although the F-statistics are small. As the average altitude and the share of land with altitude below 500m are indeed exogenous and the results of the second stage regressions through FE-2SLS are very different from that of the baseline regression, we think the two geographical factors are effective instruments which tackle the endogeneity problem efficiently.

Table 8A. Results of IV regression on ELF

	Pooled 2SLS			FE 2SLS			IV-GMM	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ELF	0.949***	0.290***	0.040	12.762**	11.878**	5.809	12.793**	11.996**
	(0.244)	(0.101)	(0.105)	(5.674)	(6.023)	(4.026)	(5.670)	(5.968)
Income in lagging		0.691***	0.787***		0.169**	-0.022		0.167**
period 1		(0.027)	(0.052)		(0.084)	(0.110)		(0.083)
Income in lagging			-0.078*			-0.091		
period 2			(0.045)			(0.100)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	656	412	704	631	366	704	631
R-squared	0.833	0.964	0.966	0.902	0.906	0.903	0.901	0.905
Number of cities				245	225	183	245	225
F-statistic of the 1st stage	14.55	17.18	11.36	4.38	3.27	3.8	4.38	3.27
Historical migration	0.081***	0.090***	0.10***					
	(0.020)	(0.021)	(0.026)					
Altitude*t1	-0.032***	-0.037***	-0.036**	-0.003***	-0.003**		-0.003***	-0.003**
	(0.011)	(0.011)	(0.016)	(0.001)	(0.001)		(0.001)	(0.001)
Altitude*t2				-0.002*	-0.002	-0.004*	-0.002*	-0.002
				(0.001)	(0.001)	-0.002	(0.001)	(0.001)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Note: Additional control variables include all those in column (7) of Table 7A. In the regression on instrumental variables, t1 means the period 2001-2005 and t2 means the period 2006-2010. However, in the first stage of the pooled 2SLS regression, instrumental variables are regressed without interacting with period dummies.

Table 8B. Results of IV regression on RQ

	Pooled 2SLS			FE 2SLS			IV-GMM	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RQ	1.328***	0.248	-0.244	27.088**	24.388**	12.784	27.187**	24.618**
	(0.477)	(0.230)	(0.290)	(12.206)	(12.321)	(9.119)	(12.179)	(12.197)
Income in lagging period 1		0.686***	0.794***		0.168**	-0.038		0.166**
		(0.026)	(0.053)		(0.084)	(0.124)		(0.083)
Income in lagging period 2			-0.093**			-0.100		
			(0.046)			(0.110)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	656	412	704	631	366	704	631
R-squared	0.885	0.971	0.965	0.902	0.911	0.899	0.901	0.910
Number of cities				245	225	183	245	225
F-statistic of the 1st stage	13.51	14.48	8.83	4.01	3.32	3.37	4.01	3.32
Historical migration	0.031***	0.035***	0.036***					
	(0.008)	(0.009)	(0.010)					
Share of land with altitude below 500m	0.033***	-0.032***	0.032**					
	(0.010)	(0.011)	(0.014)					
Altitude*t1				-0.001***	-0.001***		-0.001***	-0.001***
				(0.000)	(0.001)		(0.000)	(0.001)
Altitude*t2				-0.001*	-0.001	-0.002*	-0.001*	-0.001
				(0.000)	-0.001	(0.001)	(0.000)	-0.001

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Note: Additional control variables include all those in column (7) of Table 7A. In the regression on instrumental variables, t1 means the period 2001-2005 and t2 means the period 2006-2010. However, in the first stage of the pooled 2SLS regression, instrumental variables are regressed without interacting with period dummies.

Table 8C. Results of IV regression on PH

	Pooled 2SLS			FE 2SLS			IV-GMM	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PH	2.218***	0.716***	0.170	16.738**	14.568**	9.712	16.028**	14.403**
	(0.505)	(0.226)	(0.241)	(7.075)	(6.806)	(6.084)	(7.056)	(6.805)
Income in lagging		0.689***	0.790***		0.174**	-0.034		0.170**
period 1		(0.026)	(0.053)		(0.072)	(0.096)		(0.072)
Income in lagging			-0.079*			-0.021		
period 2			(0.045)			(0.063)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	656	412	704	631	366	704	631
R-squared	0.846	0.964	0.966	0.926	0.943	0.916	0.929	0.944
Number of cities				245	225	183	245	225
F-statistic of the 1st stage	21.01	23.13	14.51	4.5	4.74	5.87	4.5	4.74
Historical migration	0.034***	0.037***	0.041***					
	(0.008)	(0.008)	(0.010)					
Altitude*t1	-0.018***	-0.019***	-0.020***	-0.001*	-0.002***		-0.001*	-0.002***
	(0.004)	(0.004)	(0.006)	(0.001)	(0.001)		(0.001)	(0.001)
Altitude*t2				-0.001***	-0.002***	-0.002**	-0.001***	-0.002***
				-0.001	(0.001)	(0.001)	-0.001	(0.001)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Note: Additional control variables include all those in column (7) of Table 7A. In the regression on instrumental variables, t1 means the period 2001-2005 and t2 means the period 2006-2010. However, in the first stage of the pooled 2SLS regression, instrumental variables are regressed without interacting with period dummies.

Through IV analysis, we find ELF, RQ and PH each have a robust and positive effect on economic growth across all regressions**.** However, GI and ER have no robust effect on economic growth. Compared with the significant effect of PH, we can conclude that dialectal distance between the central group and other groups plays a greater role than dialectal distances between other groups in explaining differences in economic development. In addition, compared with baseline results, the increase in the magnitudes of the effects of ELF and RQ on economic growth indicates that the positive effects of ELF and RQ in the baseline results are overestimated, while the fact that the effect of PH on economic growth becomes significant provides evidence that peripheral heterogeneity is negatively associated with economic development. Thus, economic development contributes more to the balance of population distribution across dialect groups, but average dialectal distances between the central group and other groups become smaller.

Furthermore, to affirm the positive effect of dialectal diversity, we introduce a random-effects model using the dialectal diversity in the year 2000 and single-year data of economic development from 2011 to 2015. Compared with the dialectal diversity and income of the same periods, the dialectal diversity in the lagging period is less likely to influence the economic development of the current period. Thus, in the following analysis, the dialectal diversity in 2000 is regressed and the indices are represented by ELF00, GI00, RQ00, ER00 and PH00. The results of this analysis are displayed in Table A17 (see Appendix A). We can observe that the coefficients of ELF00, GI00, RQ00 and PH00 are positive and significant. Therefore, dialectal fractionalization and polarization as well as periphery heterogeneity each show a robust and positive effect on income and economic growth, as shown in the IV analysis. The positive effect of dialectal fractionalization adjusted by dialectal distances also gains more supportive evidence. But we still have no evidence for the significant effect of ER.

According to our analysis above, ELF, GI, RQ and PH each have a positive effect on local economic growth in China, which is different from the conclusion in the literature. In the literature, on the one hand, diversity should have a positive effect on the innovation capacity, which is true in China, as established in the research of Pan et al. (2017), which suggests that private high-tech firms have more innovative output in more diverse cities. On the other hand, dialectal diversity has a negative impact because of the difficulty in the communication between different dialect groups, which results in less economic interaction. Nevertheless, we suggest that this kind of negative effect does not exist in China.[22] For one thing, the coexistence of dialects has a long history and people speaking different dialects can understand each other to some extent. In addition, Putonghua has been the official language for more than 50 years and most people are able to communicate with each other using Putonghua. For another thing, the same writing system is common to all Chinese dialects and thus speaking different dialects does not affect the ability to communicate in writing. We also find empirical evidence showing that larger dialectal distance has no influence on the effect of dialectal diversity on economic growth (see Appendix D). Furthermore, there are also studies showing that cultural diversity has a negative impact on the level of public spending due to heterogeneous preferences and interest conflicts. But this does not affect our result because we include both public expenditure and fixed asset investments as controls in the regression. However,

[22] On the impact of linguistic factor on the ease of communication and therefore bilateral trade, see Melitz and Toubal (2014).

the study of Liu et al. (2017) states that dialectal diversity is detrimental to the efficiency of resource allocation. We conjecture that the efficiency loss is not common across all cities and it is also affected by the capacity of local government and the economic environment. We will discuss this more in the next section.

VI. The Effect of CCP Governance during the Revolutionary War

During the period of the Agrarian Revolutionary War and the anti-Japanese national revolutionary war, some districts functioned as revolutionary bases controlled by the CCP. The governance of these districts was primarily through congress and democratic government composed of local people. In the process of war, government members as well as the government itself formed a close relationship with local residents. After the war ended, these people continued to be government members, and this has had the lasting effect on the contemporary government structure that the government of this area has a higher proportion of cadres selected from local residents. The closer relationship between the local government and citizens has resulted in higher government capacity in the coordination and efficiency of resources allocation. But the effect of the government capacity may also be affected by the economic environment and resource support in developing the economy. Conflicts and deficient allocation of resources may only appear when resources are sufficient such that the governments of counties have choices concerning the availability of resources.

Since the beginning of economic reforms starting in 1978 when the unbalanced development strategy was first implemented, the eastern part of China has been the pioneer in economic development, receiving more support and resources through preferential policies in relation to investments, fiscal decentralization, tax, credit, investment and the introduction of new technologies by establishing special economic zones and economic-technological development zones.[23] Facing relatively high amounts of resources and policy support for economic development, it is more difficult for dialect groups to reach agreement regarding resource allocation among local county

[23] There are, in total, four economic regions in mainland China, the eastern part, the central part, the north-eastern part and the western part. The Western Development Strategy started in year 2000 and then the central government has been providing support to develop economy in the western part. There is also a western development office in the central government making related policies and decisions. Thus, the local government in the western part plays a smaller role in developing economy than other parts of China. Hence, as we want to examine the effect from different roles of local government, we focus on the eastern, central and north-eastern regions in China.

governments. Thus, the government at the prefecture level with a higher proportion of local cadres can allocate resources more efficiently by coordinating the conflicts between counties arising from different dialect groups. Hence, in the East of China, dialect diversity may have a negative effect because of conflicts in resource allocation, but the experience of being governed by the CCP during the revolutionary war should contribute to the reduction in the negative role of diversity.

Since the economic reforms started in the 1980s, local governments in the central and northeastern part of China have received less motivation from the central government to promote economic development. Although the situation has improved since the beginning of the 21st century, the advantage in the East of China has persisted. For example, the ratio of special economic zones and high technology development zones to cities in the East is 0.869, while the ratio in the other two regions is 0.21. Hence, in these regions, there are only limited resources to be allocated to sectors and departments which have a critical need for local economic development. Thus, dialectal diversity may not result in conflicts in resource allocation. But the higher proportion of local cadres may play a negative role in the efficiency of resource allocation because of their preference for counties having close relations with native cadres in the upper level government. Therefore, dialectal diversity should have a less positive effect in cities in central and northeastern regions of China that experienced governance by the CCP during the revolutionary war. Taking the dummy revolutionary area as the proxy for governance by the CCP during the revolutionary war, our hypothesis is verified by empirical estimation.

We firstly run the regression by propensity score matching on the dummy of revolutionary area while controlling different indices of dialectal diversity. We find that there is no difference in economic growth between the revolutionary area and the non-revolutionary area (see Table A18 in Appendix A). Therefore, the effect of dialectal diversity on economic growth will not be disturbed by the experience of being governed by the CCP during the revolutionary war. Furthermore, by introducing the interaction terms of dialectal diversity and the dummy for the experience in the fixed effect model, we observe that the experience of being governed by the CCP during the revolutionary war might contribute to the positive effect of dialect diversity (See Table A19 in Appendix A). Hence, there is no difference in the effect of dialectal diversity between the revolutionary area and the non-revolutionary area.

Tables 9A-9D report the results of regressions in the subsample of the East and other regions regarding the effect on economic growth. Firstly, we observe that, in the East, the coefficients of the interaction term between ELF and the revolutionary area are positive and significant and have higher absolute values than the negative coefficients of ELF (Table 9A). In other regions, in contrast, ELF shows a positive and significant effect on economic growth in the non-revolutionary area. Moreover, coefficient of interaction terms in columns (5) is significant and negative. Therefore, consistent with our hypothesis, the negative role of the conflicts over resource allocation is reduced by the experience of being governed by the CCP during the revolutionary war in the East, but such experience in central and north-eastern regions leads to more negative impacts. Similarly, regarding the effect of GI (Table 9B), we find that GI also shows a negative effect on economic growth in the revolutionary area of the East. Thus, the conflicts over resource allocation are common among different dialect groups and deeper as dialectal distance increases. Furthermore, in the East, RQ has a negative effect on income and economic growth in cities of the revolutionary area and a positive effect in cities that were not in the revolutionary area. In other regions, RQ may also induce more negative effect on economic growth in the revolutionary area. We also observe similar results regarding the influence of the experience of being governed by the CCP during the revolutionary war on the effect of PH in the East and other regions (Tables 9D). We additionally run regressions in which ER is included as the independent variable and the results are shown in Table A20. However, we observe that ER has no significant effect in most cases, as was the case in the results of the analysis in the sections above.

In brief, the experience of being governed by the CCP during the revolutionary war inhibits the negative impact of dialectal diversity and contributes to its positive effect in the eastern part of China. In central and north-eastern regions, the experience tends to promote the negative influence of dialectal diversity. Furthermore, the results regarding the effect of GI suggest that dialectal distances also play some role in determining economic development. But the difference in the significance of ER and PH predicts that different distance has different roles. The effect of ER is only significant in the East when income in lagging period 1 and 2 is controlled, while the effect of PH is significant and robust. Hence, dialectal distances between the central group and other dialect groups has a larger role in influencing economic outcomes.

Table 9A. The effect of ELF: revolutionary area vs. non-revolutionary area

VARIABLES	The East			Other regions		
	(1)	(2)	(3)	(4)	(5)	(6)
ELF	-1.961*	-1.892*	-2.726***	1.262**	0.865	0.693
	(1.027)	(0.991)	(0.864)	(0.637)	(0.653)	(0.757)
ELF*Revolutionary area	2.301*	2.261**	3.489***	-1.447	-2.157**	-0.279
	(1.198)	(1.111)	(0.939)	(0.896)	(1.003)	(0.980)
Income in lagging period 1		0.347***	0.208*		0.279***	0.186*
		(0.066)	(0.115)		(0.071)	(0.102)
Income in lagging period 2			0.052			-0.024
			(0.154)			(0.072)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	13.922***	8.147*	4.663	13.057**	8.635**	9.994**
	(4.430)	(4.223)	(7.231)	(5.351)	(4.245)	(4.202)
Observations	177	172	112	330	310	201
R-squared	0.986	0.989	0.989	0.979	0.982	0.980
Number of cities	60	60	59	115	115	113

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Time variant control variables shown in column (7) of Table 7A are also included in the regression.

Table 9B. The effect of GI: revolutionary area vs. non-revolutionary area

VARIABLES	The East			Other regions		
	(1)	(2)	(3)	(4)	(5)	(6)
GI	-0.650	-3.271*	-4.611***	1.954	1.175	1.156
	(0.525)	(1.682)	(1.544)	(2.292)	(1.910)	(1.546)
GI*Revolutionary area	1.744	4.252**	6.287***	-2.569	-1.671	0.350
	(1.489)	(1.942)	(1.568)	(2.345)	(2.259)	(2.052)
Income in lagging period 1		0.345***	0.199*		0.277***	0.181*
		(0.065)	(0.115)		(0.073)	(0.104)
Income in lagging period 2			0.042			-0.011
			(0.153)			(0.075)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	14.660***	8.096*	4.577	13.030**	7.144	9.663**
	(4.625)	(4.306)	(6.985)	(5.250)	(4.465)	(4.153)
Observations	177	172	112	330	310	201
R-squared	0.986	0.990	0.990	0.979	0.982	0.980
Number of cities	60	60	59	115	115	113

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Time variant control variables shown in column (7) of Table 7A are also included in the regression.

Table 9C. The effect of RQ: revolutionary area vs. non-revolutionary area

VARIABLES	The East			Other regions		
	(1)	(2)	(3)	(4)	(5)	(6)
RQ	-4.027*	-3.866*	-5.826***	2.805**	2.060	1.839
	(2.037)	(1.961)	(1.709)	(1.301)	(1.393)	(1.788)
RQ*Revolutionary area	4.129	4.370*	7.276***	-3.221	-3.945*	-1.135
	(2.716)	(2.534)	(2.228)	(2.010)	(2.099)	(2.150)
Income in lagging period 1		0.349***	0.196*		0.280***	0.187*
		(0.068)	(0.117)		(0.071)	(0.102)
Income in lagging period 2			0.088			-0.028
			(0.159)			(0.071)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	14.242***	8.275*	4.466	13.086**	8.155*	10.084**
	(4.368)	(4.260)	(7.298)	(5.398)	(4.399)	(4.221)
Observations	177	172	112	330	310	201
R-squared	0.986	0.989	0.989	0.979	0.982	0.980
Number of cities	60	60	59	115	115	113

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Time variant control variables shown in column (7) of Table 7A are also included in the regression.

Table 9D. The effect of PH: revolutionary area vs. non-revolutionary area

VARIABLES	The East			Other regions		
	(1)	(2)	(3)	(4)	(5)	(6)
PH	-3.431*	-3.343*	-5.096***	1.871	1.176	0.929
	(1.721)	(1.672)	(1.512)	(1.578)	(1.364)	(1.102)
PH*Revolutionary area	4.684**	4.586**	6.775***	-1.375	-0.160	0.355
	(1.994)	(1.943)	(1.724)	(1.950)	(2.222)	(1.593)
Income in lagging period 1		0.344***	0.152		0.277***	0.180*
		(0.064)	(0.116)		(0.073)	(0.106)
Income in lagging period 2			0.103			-0.012
			(0.154)			(0.075)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	13.782***	8.048*	2.939	12.485**	6.362	9.768**
	(4.587)	(4.313)	(7.011)	(5.354)	(4.425)	(4.216)
Observations	177	172	112	330	310	201
R-squared	0.987	0.990	0.990	0.979	0.982	0.980
Number of cities	60	60	59	115	115	113

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Time variant control variables shown in column (7) of Table 7A are also included in the regression.

VII. Conclusions

In this paper, the effect of dialectal diversity on economic performance at the prefectural level in China is re-examined. Firstly, five indices of dialectal diversity are analyzed to identify the effect of diversity resulting from a variety of languages and diversity resulting from both the variety and differences of languages. Of these five indices, ELF and RQ are determined only by the population distribution among the dialect groups. GI and ER also consider the dialectal distances of each pair of dialects. PH is determined by the population distribution and dialectal distance between the central group other groups. Secondly, dialectal diversity in the years 2000, 2005 and 2010 are calculated and a fixed-effects model is implemented. Thirdly, a panel sample of 5-year average data covering the period 2001-2015 is used in the estimation. There are 274 prefecture-level cities in the sample. To solve the problem of endogeneity, instrumental variable analysis is applied using the approaches pooled 2SLS, FE-2SLS and IV-GMM. Furthermore, differences in the effect between cities that were governed by the CCP during the revolutionary war and those that were not are also explored.

We find that ELF and RQ each have a significant and robust effect on economic growth. Of the indices considering dialectal distances, only PH has a robust and positive effect on economic growth. GI, however, shows a significant effect on the change in economic growth over time, whereas ER does not show a robust effect in any cases. Hence, the effect of GI and PH imply that dialectal distances also play a role in explaining variation in income, but the insignificant effect of ER indicates that dialectal distances between the central group and other groups, instead of that between the two largest groups, are more relevant. Furthermore, the significant effect of GI also suggests that dialectal diversity may be related to the variation in economic growth over time. In addition, the influence of the experience of being governed by the CCP during the revolutionary war on the effect of dialectal diversity is different between the East and other parts (including the central and north-eastern regions) of China. In the East, exposure to communist governance tends to inhibit the economic loss from dialectal diversity and promotes its positive effect, while it is the opposite in the Center and Northeast other regions. The discussion of the experience of being in the revolutionary base also shows that the actual effect of dialectal diversity is determined by the relative scale of benefits compared to economic loss caused by diversity.

On the basis of this research, there is still much potential for further study. Firstly, due to the limitation of data access, a longer period panel will be better for the analysis. It is necessary to do further research in the future to identify the long-run effect of dialectal diversity on economic development, especially on the time trend of economic growth. Secondly, more potential channels need to be investigated. But this requires better data on efficiency in resource allocation, productivity and technological improvement at the prefecture level. Thirdly, conclusions of this study are limited to the sample we have, and it is still meaningful to conduct this research in future when more data is available. Above all, this study contributes to a better understanding of the effect of dialectal diversity on economic development in China under the same cultural and institutional environment.

References

Alesina, Alberto, Eliana La Ferrara. 2004, Ethnic Diversity and Economic Performance. NBER Working Paper 10313.

Annett, Anthony, 2001, Social Fractionalization, Political Instability, and the Size of Government. IMF Staff Papers 48, 3.

Alesina, Alberto, Caterina Gennaioli, Stefani Lovo. 2015, Public Goods and Ethnic Diversity: Evidence from Deforestation in Indonesia. Working paper, Department of Economics, Harvard University.

Alesina, Alberto, Nicola Fuchs-Schündeln, 2007, Good-Bye Lenin (or Not?): The Effect of Communism on People's Preferences. The American Economic Review, 97(4): 1507-1528

Alesina, Alberto, Enrico Spolaore, Romain Wacziarg, 2000, Economic Integration and Political Disintegration. American Economic Review, 90.

Ager, Philipp, Markus Brückner, 2013, Cultural diversity and economic growth: Evidence from the US during the age of mass migration. European Economic Review 64: 76-97.

Ahlerup, P. and O. Olsson, 2007, The Roots of Ethnic Diversity. Working Papers in Economics No. 281, University of Gothenburg.

Ahlerup, Pelle, 2009, The Causal Effects of Ethnic Diversity: An Instrumental Variables Approach. Working Papers in Economics, 386:1403-2465, Department of Economics School of Business, Economics and Law at University of Gothenburg.

Anderson, Lisa R., Jennifer M. Mellor, Jeffrey Milyo, 2006, Induced heterogeneity in trust experiments. Exp Econ 9: 223–235.

Adelman, Irma, Cynthia Taft Morris, Society, Politics, and Economic Development (Baltimore: Johns Hopkins University Press, 1967); Marie R. Haug, ''Social and Cultural Pluralism as a Concept in Social System Analysis,'' American Journal of Sociology 73: 294–304.

Acemoglu, Daron; Simon Johnson and James Robinson, 2005, The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth, American Economic Review, 95(3), pp.546-579.

Bellini, Elena; Gianmarco I. P. Ottaviano; Dino Pinelli; Giovanni Prarolo,,2008: Cultural diversity and economic performance: Evidence from European regions. HWWI Research Paper, 3-14.

Bovea, Vincenzo, Leandro Elia. 2017, Migration, Diversity, and Economic Growth, World Development 89: 227–239.

Bahry, Donna, Mikhail Kosolapov, Polina Kozyreva, Rick K. Wilson, 2005, Ethnicity and Trust: Evidence from Russia, American Political Science Review, 99(4).

Collier, P. 2000, Ethnicity, Politics and Economic Performance. Economics & Politics, 12: 225–45.

Dincer, O.C., Wang, F., 2011. Ethnic diversity and economic growth in China. Journal of Economic Policy Reform 14:1–10.

Desmet, Klaus, Shlomo Weber, Ignacio Ortuño-Ortín, 2009, Linguistic Diversity and Redistribution. Journal of the European Economic Association December 7(6): 1291–1318.

Desmet, Klaus, Ignacio Ortuño-Ortín, Romain Wacziarg, 2012, The political economy of linguistic cleavages. Journal of Development Economics 97 (2): 322-338.

Desmet, Klaus, Ignacio Ortuño-Ortín, Romain Wacziarg, 2017, Culture, ethnicity and diversity. American Economic Review 107 (9): 2479-2513.

Desmet, Klaus, Ignacio Ortuño-Ortín and Shlomo Weber, 2009, Linguistic Diversity and Redistribution. Journal of the European Economic Association, 7(6), 1291-1318.

Easterly, W., Levine, R., 1997. Africa's growth tragedy: policies and ethnic divisions. Quarterly Journal of Economics 112: 1203–1250.

Eleches, Grigore Pop, Joshua A. Tucker, 2017, Communism's Shadow: Post-Communist Legacies, Values and Behavior. Princeton Studies in Political Behavior (Book 3), Princeton University Press.

Foley, R. A., M. Mirazo´n Lahr, 2011, The evolution of the diversity of cultures. Phil. Trans. R. Soc. B 366: 1080–1089.

Falck, Oliver, Stephan Heblich, Alfred Lameli, Jens Südekum, 2012, Dialects, Cultural Identity, and Economic Exchange. Journal of Urban Economics 72(2-3):225-239.

Goeren, Erkan, 2014, How Ethnic Diversity Affects Economic Growth. World Development Vol. 59: 275–297.

Gerring, John, Strom C. Thackera, Yuan Lu and Wei Huang, 2015, Does Diversity Impair Human Development? A Multi-Level Test of the Diversity Debit Hypothesis. World Development Vol. 66: 166–188.

Ginsburg Viktor and Shlomo Weber, 2011, How Many Language Do We Need? The Economics of Linguistic Diversity. Princeton University Press.

Lazear, E., 1999a. Globalization and the market for team-mates. Economic Journal 109, 15-40.

Lazear, E., 1999b. Culture and Language. Journal of Political Economy, Supplement, 95-125.

LaPorta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1999. The Quality of Government. Journal of Law, Economics and Organization 15 (1): 222-279.

Leigh, Andrew, 2006, Trust, Inequality and Ethnic Heterogeneity. The Economic Record, 82(258): 268–280.

Lian, Brad, John R. Oneal, 1997, Cultural Diversity and Economic Development: A Cross-National Study of 98 Countries, 1960–1985. Economic Development and Cultural Change. The University of Chicago.

Li, Guangqin, Jianhua Cao, Shuai Sha, 2017, Linguistic Diversity and Difference in Openness. The World Economy 3.

Li, Feiyue, Dong Zhang, Mingxing Liu, 2014, Real Political Power Structure and Local Economic Growth: Long-term Impact of China's Revolutionary War, Economic Research Journal, 12.

Lin, Shuanglin, Shunfeng Song, 2002, Urban Economic Growth in China: Theory and Evidence. Urban Studies 39 (12): 2251–2266.

Liu, Yuyun, Tianshi Dia, Xianxiang Xu, 2017, Chinese Dialects, Market Segmentation and Resource Misallocation, China Economic Quarterly, 16(4): 1583-1600.

Montalvoa, Jose G., Marta Reynal-Querol, 2014, Cultural Diversity, Conflict, and Economic Development. Handbook of the Economics of Art and Culture, Vol 2: 485-506.

Michalopoulos, Stelios. 2012. The Origins of Ethnolinguistic Diversity. American Economic Review, 102(4): 1508-39.

Melitz Jacques and Farid Toubal. 2014. Native language, spoken language, translation and trade. Journal of International Economics, 93(2): 351-363.

Nathan, Max, 2011, The Economics of Cultural Diversity: Lessons from British cities. A thesis submitted to the Department of Geography and Environment of the London School of Economics for the degree of Doctor of Philosophy.

Ottaviano, Gianmarco, Giovanni Peri, 2006, The economic value of cultural diversity: evidence from US cities. Journal of Economic Geography 6: 9–44.

Posner, Daniel N., 2004, Measuring Ethnic Fractionalization in Africa, American Journal of Political

Science 48 (4): 849–863.
Pan, Yue, Jinli Xiao, Yiyi Dai, 2017, Cultural Diversity and Innovation of Enterprises: From the Perspective of Chinese Dialects, Journal of Financial Research 44 (General N.448): 147-161.
Quigley, John M.,1998, Urban Diversity and Economic Growth, Journal of Economic Perspective 12(2): 127–138.
Shao, Shuai, Guanqin Li, Jianhua Cao, 2017, Is Cultural Diversity Detrimental to Urbanization? - from the Perspective of Dialects, Journal of Southeast University (Philosophy and Social Science), 19 (5).
Sparber, C., 2010. Racial diversity and macroeconomic productivity across US states and cities. Regional Studies 44: 71–85.
Treisman, Daniel, 2014, The political economy of change after communism, Forthcoming in Anders Åslund and Simeon Djankov, eds., The Great Rebirth: Lessons from the Victory of Capitalism over Communism, Washington, DC: Peterson Institute for International Economics
Wang, Zhenyu, Xin Gu, 2018, Ethnic diversity and local economic growth: evidence from cities of China[J]. Journal of Finance and Economics, 44(2): 127-140.
Xu, Xianxiang, Yuyun Liu, Zekai Xiao, 2015, Dialects and Economic Growth, China Journal of Economics, 2.
Yogo, U.T., 2015. Ethnic Diversity and the Efficiency of Public Spending in Developing Countries. Etudes et Documents, N13, CERDI.

Appendix A

Table A1: List of Chinese Dialects

Dialect supergroup	Dialect group	Dialect subgroup	Dialect code[24]	Dialect supergroup	Dialect group	Dialect subgroup	Dialect code
Mandarin	Dongbei	Jishen	1101	Mandarin	Southwest	Changhe	1712
		Hafu	1102		Jianghui	Hongchao	1801
		Heisong	1103			Tairu	1802
	Beijing	Jingshi	1201			Huangxiao	1803
		Huaicheng	1202	Jin	Bingzhou	Bingzhou	2100
		Chaofeng	1203		Luliang	Luliang	2200
		Shike	1204		Shangdang	Shangdang	2300
	Jilu	Baotang	1301		Wutai	Wutai	2400
		Shiji	1302		Dabao	Dabao	2500
		Canghui	1303		Zhanghu	Zhanghu	2600
	Jiaoliao	Qingzhou	1401		Hanxin	Hanxin	2700
		Denglian	1402		Zhiyan	Zhiyan	2800
		Gaihuan	1403	Wu	Taihu	Taihu	3100
	Zhongyuan	Zhengcao	1501		Taizhou	Taizhou	3200
		Cailu	1502		Oujiang	Oujiang	3300
		Luoxu	1503		Wuzhou	Wuzhou	3400
		Xinbeng	1504		Chuqu	Chuqu	3500
		Fenhe	1505		Xuanzhou	Xuanzhou	3600
		Guanzhong	1506	Jiangxi	Changjing	Changjing	4100
		Qinlong	1507		Yiliu	Yiliu	4200
		Longzhong	1508		Jicha	Jicha	4300
		Nanjiang	1509		Fuguang	Fuguang	4400
	Lanyin	Jincheng	1601		Yingyi	Yingyi	4500
		Yinwu	1602		Leizi	Leizi	4700
		Hexi	1603		Dongsui	Dongsui	4800
		Beijiang	1605		Huaiyue	Huaiyue	4900
	Southwest	Chengyu	1701	Hunan	Changyi	Changyi	5100
		Dianxi	1702		Loushao	Loushao	5200
		Qianbei	1703		Jixu	Jixu	5300
		Kungui	1704	Fukienese	Minnan	Minnan	6100
		Guanchi	1705		Puxian	Puxian	6200
		Ebei	1706		Mindong	Mindong	6300
		Wutian	1707		Minbei	Minbei	6400
		Cenjiang	1708		Minzhong	Minzhong	6500
		Qiannan	1709		Qiongwen	Qiongwen	6600
		Xiangnan	1710		Leizhou	Leizhou	6700
		Guiliu	1711		Shaojiang	Shaojiang	6800

[24] Lavely, William; Berman, Lex, 2012, "Language Atlas of China", https://hdl.handle.net/1902.1/19004 , Harvard Dataverse, V1

Continued Table A: List of Chinese Dialects

Dialect supergroup	Dialect group	Dialect subgroup	Dialect code	Dialect supergroup	Dialect group	Dialect subgroup	Dialect code
Cantonese	Guangfu	Guangfu	7100	Hakka	Ninglong	Ninglong	8600
	Siyi	Siyi	7200		Yugui	Yugui	8700
	Gaoyang	Gaoyang	7300		Tonggu	Tonggu	8800
	Goulou	Goulou	7400	Hui	Jingzhan	Jingzhan	9100
	Yongxun	Yongxun	7600		Jishe	Jishe	9200
	Qinlian	Qinlian	7700		Xiuyi	Xiuyi	9300
Hakka	Yuetai	Yuetai	8100		Qide	Qide	9400
	Yuezhong	Yuezhong	8200		Yanzhou	Yanzhou	9500
	Huizhou	Huizhou	8300	Pinghua	Pinghua	Pinghua	100
	Yuebei	Yuebei	8400	Other	Shaozhou Tuhua	Shaozhou Tuhua	400
	Tingzhou	Tingzhou	8500				

Table A2: Pearson Correlation Matrix (5-year average data)

	Income	ELF	GI	RQ	ER	PH	Public expenditure
ELF	-0.199**						
GI	-0.131**	0.888**	1				
RQ	-0.194**	0.957**	0.813**	1			
ER	-0.090*	0.726**	0.816**	0.642**	1		
PH	-0.128**	0.867**	0.970**	0.853**	0.784**	1	
Public expenditure	0.822**	-0.154**	-0.126**	-0.151**	-0.055	-0.129**	1
Fixed asset investment	0.895**	-0.124**	-0.070*	-0.110**	-0.013	-0.064*	0.915**
Ratio of the primary industry	-0.719**	0.116**	0.053	0.099**	0.028	0.041	-0.549**
Ratio of the secondary industry	0.526**	-0.139**	-0.083*	-0.130**	-0.043	-0.079*	0.322**
Ratio of loans	-0.007	-0.024	-0.044	-0.017	-0.091**	-0.039	0.053
Ratio of residential deposit	0.093**	-0.017	-0.025	-0.011	-0.074*	-0.02	0.191**
Population	-0.141**	0.190**	0.153**	0.206**	0.160**	0.177**	-0.223**
Employment rate	0.214**	0.007	0.013	0.014	0.025	0.016	0.218**
Education level	0.720**	-0.146**	-0.082*	-0.133**	-0.111**	-0.068*	0.676**
Number of key universities	0.274**	-0.065*	-0.066*	-0.044	-0.066*	-0.052	0.205**
Enrolment of students	-0.218**	0.192**	0.154**	0.210**	0.160**	0.179**	-0.335**
Market institution and intermediary organization	0.385**	-0.097**	-0.043	-0.100**	0.045	-0.042	0.296**
Land area	-0.280**	0.200**	0.182**	0.163**	0.183**	0.159**	-0.156**
Number of special zones	0.503**	-0.117**	-0.101**	-0.107**	-0.091**	-0.095**	0.385**
Highway freight traffic	0.061*	0.046	0.078*	0.051	0.091**	0.094**	-0.035

Continued Table A2. Correlation matrix (5-year average data)

	Fixed asset investment	Ratio of the primary industry	Ratio of the secondary industry	Ratio of loans	Ratio of residential deposit	Population	Employment rate
Ratio of the primary industry	-0.654**	1					
Ratio of the secondary industry	0.484**	-0.718**	1				
Ratio of loans	0.012	-0.106**	0.032	1			
Ratio of residential deposit	0.152**	-0.186**	0.080*	0.838**	1		
Population	-0.115**	0.160**	-0.230**	-0.130**	-0.101**	1	
Employment rate	0.271**	-0.254**	0.197**	-0.102**	-0.003	0.029	1
Education level	0.678**	-0.520**	0.266**	0.115**	0.213**	-0.081*	0.152**
Number of key universities	0.239**	-0.263**	-0.023	0.095**	0.062*	0.234**	-0.039
Enrolment of students	-0.241**	0.158**	-0.210**	-0.132**	-0.107**	0.944**	0.054
Market institution and intermediary organization	0.349**	-0.254**	0.106**	-0.135**	-0.074*	0.142**	0.155**
Land area	-0.219**	0.409**	-0.387**	-0.054	-0.05	0.352**	-0.110**
Number of special zones	0.413**	-0.432**	0.148**	0.089*	0.145**	0.209**	0.131**
Highway freight traffic	0.034	0.014	-0.007	-0.202**	-0.176**	0.154**	0.062*

	Education level	Number of key universities	Enrolment of students	Market institution and intermediary	Land area	Number of special zones	Highway freight traffic
Education level	1						
Number of key universities	0.341**	1					
Enrolment of students	-0.164**	0.206**	1				
Market institution and intermediary organization	0.056	0.052	0.118**	1			
Total land area	-0.208**	-0.014	0.290**	-0.248**	1		
Number of special zones	0.507**	0.437**	0.176**	0.133**	-0.073*	1	
Highway freight traffic	0.067*	-0.067*	0.133**	0.189**	-0.116**	-0.001	1

Notes: ** *significant at 0.01, ** significant at 0.05, * significant at 0.1.

Table A3. Baseline results of the relationship between GI and economic development

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GI	1.572**	0.771	0.657	0.631	0.670	0.599	0.756	1.111**
	(0.754)	(0.754)	(0.648)	(0.662)	(0.702)	(0.713)	(0.509)	(0.556)
Public expenditure		0.216***	0.144***	0.145***	0.129**	0.135**	0.110*	0.076
		(0.064)	(0.055)	(0.055)	(0.056)	(0.059)	(0.066)	(0.083)
Fixed asset investment		0.167***	0.084***	0.088***	0.083***	0.087***	0.116***	0.052
		(0.031)	(0.031)	(0.031)	(0.031)	(0.033)	(0.033)	(0.056)
Ratio of primary industry			-0.005*	-0.005*	-0.006**	-0.006*	-0.003	-0.007*
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Ratio of secondary industry			0.009***	0.009***	0.009***	0.009***	0.009***	0.011***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Ratio of loans				-0.000	-0.000	-0.000	0.000	0.001**
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ratio of residents' deposit				-0.000	-0.000	-0.000	-0.000	-0.000**
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Population					-0.076	-0.080	-0.067	-0.415***
					(0.131)	(0.166)	(0.154)	(0.140)
Employment rate					0.448*	0.421	0.112	-0.138
					(0.254)	(0.263)	(0.251)	(0.214)
Education level					0.004	0.000	-0.047*	-0.011
					(0.041)	(0.042)	(0.026)	(0.028)
Number of key universities					-0.011	-0.010	0.002	-0.181
					(0.017)	(0.017)	(0.014)	(0.136)
Enrolment of students					-0.055	-0.053	-0.006	-0.050
					(0.039)	(0.039)	(0.039)	(0.049)
Market institutions and organizations						0.002	0.001	0.003
						(0.004)	(0.003)	(0.008)
Land area						0.025	-0.015	0.028
						(0.099)	(0.096)	(0.074)
High technology zones						0.001	-0.024	-0.014
						(0.029)	(0.019)	(0.022)
Highway freight traffic						-0.029	-0.031	-0.066**
						(0.025)	(0.025)	(0.027)
Income per capita in lagging period 1							0.288***	0.068
							(0.039)	(0.060)
Income per capita in lagging period 2								0.021
								(0.069)
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	9.880***	6.506***	7.545***	7.511***	7.833***	7.664***	5.692***	10.331***
	(0.063)	(0.486)	(0.468)	(0.471)	(1.201)	(1.220)	(1.327)	(1.609)
Observations	801	729	728	723	723	712	656	412
R-squared	0.932	0.956	0.963	0.963	0.964	0.964	0.972	0.954
Number of cities	275	253	253	253	253	252	250	229

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table A4. Baseline results of the relationship between ER and economic development

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ER	0.070	-0.017	0.034	0.025	0.048	0.061	0.092	-0.229
	(0.209)	(0.166)	(0.146)	(0.145)	(0.146)	(0.147)	(0.164)	(0.224)
Public expenditure		0.218***	0.147***	0.147***	0.130**	0.137**	0.111*	0.084
		(0.064)	(0.055)	(0.055)	(0.056)	(0.059)	(0.067)	(0.084)
Fixed asset investment		0.168***	0.084***	0.088***	0.084***	0.087***	0.116***	0.055
		(0.031)	(0.031)	(0.031)	(0.031)	(0.033)	(0.033)	(0.057)
Ratio of primary industry			-0.005*	-0.005*	-0.006**	-0.006**	-0.003	-0.007
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Ratio of secondary industry			0.009***	0.009***	0.009***	0.009***	0.009***	0.011***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Ratio of loans				-0.000	-0.000	-0.000	0.000	0.001**
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ratio of residents' deposit				-0.000	-0.000	-0.000	-0.000	-0.000*
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Population					-0.083	-0.085	-0.075	-0.439***
					(0.130)	(0.167)	(0.155)	(0.131)
Employment rate					0.414	0.391	0.070	-0.217
					(0.256)	(0.265)	(0.251)	(0.223)
Education level					0.001	-0.003	-0.051*	-0.023
					(0.041)	(0.041)	(0.026)	(0.028)
Number of key universities					-0.013	-0.011	0.001	-0.147
					(0.017)	(0.017)	(0.015)	(0.137)
Enrolment of students					-0.055	-0.054	-0.006	-0.049
					(0.039)	(0.039)	(0.039)	(0.049)
Market institutions and						0.002	0.001	0.002
organizations						(0.004)	(0.003)	(0.008)
Land area						0.024	-0.016	0.030
						(0.099)	(0.096)	(0.074)
High technology zones						0.000	-0.024	-0.016
						(0.029)	(0.019)	(0.023)
Highway freight traffic						-0.030	-0.032	-0.066**
						(0.025)	(0.025)	(0.029)
Income per capita in							0.291***	0.076
lagging period 1							(0.039)	(0.059)
Income per capita in								0.035
lagging period 2								(0.072)
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	10.010***	6.542***	7.580***	7.543***	7.985***	7.806***	5.857***	10.415***
	(0.008)	(0.489)	(0.471)	(0.474)	(1.173)	(1.187)	(1.306)	(1.567)
Observations	801	729	728	723	723	712	656	412
R-squared	0.932	0.956	0.963	0.963	0.964	0.964	0.972	0.954
Number of cities	275	253	253	253	253	252	250	229

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table A5. Results of IV regression on ELF-Pooled 2SLS (one instrument in each regression)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ELF	0.464**	1.929***	0.939*	0.079	0.745***	0.227	-0.114	0.521*	-0.060
	(0.212)	(0.602)	(0.561)	(0.098)	(0.237)	(0.254)	(0.116)	(0.276)	(0.324)
Income in lagging				0.683***	0.708***	0.688***	0.795***	0.761***	0.792***
period 1				(0.025)	(0.037)	(0.028)	(0.053)	(0.073)	(0.054)
Income in lagging							-0.093**	-0.030	-0.088
period 2							(0.046)	(0.066)	(0.053)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	712	712	656	656	656	412	412	412
R-squared	0.897	0.575	0.835	0.972	0.918	0.967	0.965	0.930	0.966
F-statistic of the 1st stage	20.93	15.08	4.60	24.43	17.11	3.62	17.02	8.88	1.54
Historical migration	0.092***			0.101***			0.108***		
	(0.020)			(0.020)			(0.026)		
Altitude		-0.042***			-0.048***			-0.047***	
		(0.011)			(0.012)			(0.016)	
Share of land with			0.060**			0.057*			0.050
altitude below 500m			(0.020)			(0.030)			(0.040)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A6. Results of IV regression on GI-Pooled 2SLS (one instrument in each regression)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GI	0.849**	2.827***	1.544*	0.149	1.154***	0.390	-0.217	0.818**	-0.112
	(0.375)	(0.672)	(0.812)	(0.185)	(0.312)	(0.423)	(0.218)	(0.391)	(0.602)
Income in lagging				0.681***	0.693***	0.684***	0.788***	0.791***	0.788***
period 1				(0.025)	(0.029)	(0.026)	(0.054)	(0.061)	(0.053)
Income in lagging							-0.086*	-0.064	-0.084*
period 2							(0.046)	(0.051)	(0.046)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	712	712	656	656	656	412	412	412
R-squared	0.903	0.766	0.871	0.971	0.944	0.969	0.966	0.945	0.967
F-statistic of the 1st stage	27.53	33.5	8.16	30.56	32.56	5.80	21.48	16.14	2.03
Historical migration	0.050***			0.053***			0.057***		
	(0.010)			(0.010)			(0.012)		
Altitude		-0.029***			-0.031***			-0.030***	
		(0.005)			(0.005)			(0.007)	
Share of land with			0.037***			0.033**			0.027
altitude below 500m			(0.013)			(0.014)			(0.019)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A7. Results of IV regression on RQ-Pooled 2SLS (one instrument in each regression)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RQ	1.251**	6.002***	1.437**	0.207	2.121***	0.328	-0.317	1.442*	-0.080
	(0.602)	(2.285)	(0.705)	(0.259)	(0.753)	(0.343)	(0.325)	(0.807)	(0.433)
Income in lagging				0.685***	0.736***	0.688***	0.795***	0.760***	0.790***
period 1				(0.026)	(0.044)	(0.027)	(0.054)	(0.076)	(0.053)
Income in lagging							-0.097**	-0.014	-0.085*
period 2							(0.047)	(0.072)	(0.048)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	712	712	656	656	656	412	412	412
R-squared	0.888	0.321	0.879	0.971	0.894	0.970	0.964	0.919	0.966
F-statistic of the 1st stage	16.55	9.01	13.65	20.12	12.42	12.23	12.71	6.84	6.46
Historical migration	0.034***			0.038***			0.039***		
	(0.008)			(0.009)			(0.011)		
Altitude		-0.014***			-0.017***			-0.017***	
		(0.005)			(0.005)			(0.006)	
Share of land with			0.039***			0.039**			0.037**
altitude below 500m			(0.011)			(0.011)			(0.015)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A8. Results of IV regression on ER-Pooled 2SLS (one instrument in each regression)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ER	2.592**	10.314***	6.734	0.440	3.884***	1.742	-0.602	2.545**	-0.262
	(1.187)	(2.859)	(4.340)	(0.548)	(1.131)	(2.024)	(0.599)	(1.274)	(1.414)
Income in lagging				0.681***	0.695***	0.686***	0.785***	0.806***	0.787***
period 1				(0.025)	(0.031)	(0.028)	(0.054)	(0.064)	(0.055)
Income in lagging							-0.083*	-0.075	-0.082*
period 2							(0.046)	(0.051)	(0.045)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	712	712	656	656	656	412	412	412
R-squared	0.900	0.666	0.808	0.971	0.933	0.964	0.966	0.939	0.967
F-statistic of the 1st stage	22.99	21.10	3.95	27.28	24.79	2.82	20.81	12.54	3.50
Historical migration	0.016***			0.018***			0.020***		
	(0.003)			(0.003)			(0.004)		
Altitude		-0.008***			-0.009***			-0.010***	
		(0.002)			(0.002)			(0.003)	
Share of land with			0.008**			0.007**			0.011*
altitude below 500m			(0.004)			(0.004)			(0.006)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A9. Results of IV regression on PH-Pooled 2SLS (one instrument in each regression)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PH	1.064**	3.697***	1.342**	0.184	1.485***	0.326	-0.270	1.010**	-0.085
	(0.482)	(0.913)	(0.637)	(0.229)	(0.418)	(0.341)	(0.271)	(0.485)	(0.458)
Income in lagging				0.682***	0.699***	0.684***	0.787***	0.794***	0.788***
period 1				(0.025)	(0.031)	(0.025)	(0.054)	(0.061)	(0.053)
Income in lagging							-0.087*	-0.063	-0.083*
period 2							(0.046)	(0.050)	(0.045)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	712	712	656	656	656	412	412	412
R-squared	0.900	0.724	0.890	0.971	0.937	0.970	0.966	0.943	0.967
F-statistic of the 1st stage	24.34	29.81	18.11	27.8	29.91	14.17	19.55	15.42	6.14
Historical migration	0.040***			0.043***			0.045***		
	(0.008)			(0.008)			(0.010)		
Altitude		-0.022***			-0.024***			-0.024***	
		(0.004)			(0.004)			(0.006)	
Share of land with			0.042***			0.040***			0.035**
altitude below 500m			(0.010)			(0.011)			(0.014)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A10. Results of IV regression on ELF: FE-2SLS &IV-GMM (one instrument in each regression)

	FE-2SLS									IV-GMM					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
ELF	11.553*	12.762**	16.185*	10.586	11.878**	11.550*	7.327	5.809	5.315	11.177*	12.793**	16.511*	11.425	11.996**	11.791*
	(6.412)	(5.674)	(8.663)	(7.003)	(6.023)	(6.818)	(5.655)	(4.026)	(4.684)	(6.407)	(5.670)	(8.570)	(6.978)	(5.968)	(6.759)
Income in lagging				0.182**	0.169**	0.172*	-0.048	-0.022	-0.013				0.168*	0.167**	0.169*
period 1				(0.093)	(0.084)	(0.094)	(0.141)	(0.110)	(0.114)				(0.092)	(0.083)	(0.093)
Income in lagging							-0.122	-0.091	-0.080						
period 2							(0.147)	(0.100)	(0.121)						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	704	706	706	629	631	631	366	366	366	704	706	706	629	631	631
R-squared	0.766	0.901	0.858	0.639	0.906	0.910	0.870	0.903	0.912	0.760	0.900	0.854	0.639	0.905	0.907
Number of cities	245	246	246	224	225	225	183	183	183	245	246	246	224	225	225
F-statistic in 1st stage	2.31	4.35	2.38	1.52	3.27	2.18	1.21	3.8	2.14	2.31	4.35	2.38	1.52	3.27	2.18
Historical migration*t1	0.004*			0.004						0.004*			0.004		
	(0.002)			(0.002)						(0.002)			(0.002)		
Historical migration*t2	0.005**			0.004			0.004			0.005**			0.004		
	(0.002)			(0.002)			(0.003)			(0.002)			(0.002)		
Altitude*t1		-0.003***			-0.003**						-0.003***			-0.003**	
		(0.001)			(0.001)						(0.001)			(0.001)	
Altitude*t2		-0.002*			-0.002			-0.004*			-0.002*			-0.002	
		(0.001)			(0.001)			(0.001)			(0.001)			(0.001)	
Share of land with altitude			0.005**			0.006**						0.005**			0.006**
below 500m*t1			(0.002)			(0.003)						(0.002)			(0.003)
Share of land with altitude			0.004*			0.004			0.006			0.004*			0.004
below 500m*t2			(0.003)			(0.003)			(0.004)			(0.003)			(0.003)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A11. Results of IV regression on GI: FE-2SLS &IV-GMM (one instrument in each regression)

	FE-2SLS									IV-GMM					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
GI	20.174	33.506	34.273	18.308**	37.672	17.434	9.716	36.092	10.841	27.916*	32.715	34.380	37.644	18.589**	17.890*
	(17.421)	(28.424)	(21.483)	(8.386)	(26.155)	(10.849)	(6.034)	(60.697)	(10.052)	(16.653)	(21.585)	(28.416)	(26.154)	(8.382)	(10.842)
Income in lagging				0.219***	0.143	0.222***	-0.015	-0.268	-0.025				0.143	0.217***	0.216***
period 1				(0.057)	(0.137)	(0.064)	(0.087)	(0.633)	(0.117)				(0.137)	(0.057)	(0.063)
Income in lagging							-0.042	-0.234	-0.050						
period 2							(0.072)	(0.506)	(0.097)						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	706	706	706	631	631	631	366	366	366	706	706	706	631	631	631
R-squared	0.893	0.765	0.755	0.927	0.775	0.931	0.916	0.320	0.905	0.826	0.774	0.754	0.775	0.926	0.929
Number of cities	246	246	246	225	225	225	183	183	183	246	246	246	225	225	225
F-statistic in 1st stage	1.30	1.43	0.74	1.22	4.95	2.02	0.32	6.68	1.68	1.30	1.43	0.74	1.22	4.95	2.02
Historical migration*t1	0.001			0.002						0.001			0.002		
	(0.001)			(0.001)						(0.001)			(0.001)		
Historical migration*t2	-0.001			0.001			0.001			-0.001			0.001		
	(0.002)			(0.001)			(0.002)			(0.002)			(0.001)		
Altitude*t1		0.001			-0.002***						0.001			-0.002***	
		(0.001)			(0.001)						(0.001)			(0.001)	
Altitude*t2		0.001			-0.001***			-0.002**			0.001			-0.001***	
		(0.001)			(0.000)			(0.001)			(0.001)			(0.000)	
Share of land with altitude			0.001			0.002*						0.001			0.002*
below 500m*t1			(0.001)			(0.001)						(0.001)			(0.001)
Share of land with altitude			0.002			0.003*			0.003			0.002			0.003*
below 500m*t2			(0.002)			(0.002)			(0.002)			(0.002)			(0.002)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A12. Results of IV regression on RQ: FE-2SLS &IV-GMM (one instrument in each regression)

	FE-2SLS									IV-GMM					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
RQ	26.149*	27.244**	30.744*	24.249	24.388**	20.584*	16.803	12.784	9.864	25.474*	27.372**	31.876**	25.954	24.618**	21.217*
	(14.673)	(12.278)	(16.139)	(17.180)	(12.321)	(11.455)	(13.552)	(9.119)	(8.072)	(14.667)	(12.247)	(15.998)	(17.137)	(12.197)	(11.378)
Income in lagging				0.168	0.168**	0.187**	-0.074	-0.038	-0.011				0.155	0.166**	0.181**
period 1				(0.108)	(0.084)	(0.081)	(0.168)	(0.124)	(0.108)				(0.108)	(0.083)	(0.080)
Income in lagging							-0.140	-0.100	-0.071						
period 2							(0.169)	(0.110)	(0.107)						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	706	706	706	631	631	631	366	366	366	706	706	706	631	631	631
R-squared	0.906	0.901	0.882	0.912	0.911	0.929	0.855	0.899	0.923	0.910	0.900	0.875	0.903	0.910	0.927
Number of cities	246	246	246	225	225	225	183	183	183	246	246	246	225	225	225
F-statistic in 1st stage	2.13	3.99	2.46	1.27	3.32	2.69	1.72	3.37	2.95	2.13	3.99	2.46	1.27	3.32	2.69
Historical migration*t1	0.002*			0.002						0.002*			0.002		
	(0.001)			(0.001)						(0.001)			(0.001)		
Historical migration*t2	0.002*			0.002			0.002			0.002*			0.002		
	(0.001)			(0.001)			(0.001)			(0.001)			(0.001)		
Altitude*t1		-0.001***			-0.001***						-0.001***			-0.001***	
		(0.000)			(0.001)						(0.000)			(0.001)	
Altitude*t2		-0.001*			-0.001			-0.002*			-0.001*			-0.001	
		(0.000)			(0.000)			(0.001)			(0.000)			(0.000)	
Share of land with altitude			0.003**			0.003**						0.003**			0.003**
below 500m*t1			(0.001)			(0.001)						(0.001)			(0.001)
Share of land with altitude			0.002*			0.003*			0.003*			0.002*			0.003*
below 500m*t2			(0.001)			(0.001)			(0.002)			(0.001)			(0.001)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A13. Results of IV regression on ER: FE-2SLS &IV-GMM (one instrument in each regression)

	FE-2SLS									IV-GMM					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
ER	-5.948	-4.412	-0.512	-5.788	-0.192	-0.481	-7.923	11.339	2.244	-6.599	-2.347	0.975	-6.615	0.704	-0.030
	(4.338)	(4.067)	(1.438)	(5.059)	(2.234)	(1.225)	(10.134)	(15.520)	(1.765)	(4.324)	(3.999)	(1.343)	(5.032)	(2.214)	(1.213)
Income in lagging				0.290***	0.291***	0.291***	-0.024	0.225	0.108				0.300***	0.270***	0.281***
period 1				(0.057)	(0.040)	(0.041)	(0.186)	(0.281)	(0.074)				(0.057)	(0.040)	(0.041)
Income in lagging							0.244	-0.279	-0.032						
period 2							(0.289)	(0.425)	(0.090)						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	706	706	706	631	631	631	366	366	366	706	706	706	631	631	631
R-squared	0.910	0.934	0.963	0.916	0.971	0.971	0.697	0.373	0.927	0.897	0.955	0.962	0.899	0.971	0.971
Number of cities	246	246	246	225	225	225	183	183	183	246	246	246	225	225	225
F-statistic in 1st stage	1.71	2.26	5.90	0.98	2.82	5.99	0.69	0.67	6.30	1.71	2.26	5.90	0.98	2.82	5.99
Historical migration*t1	-0.006*			-0.004						-0.006*			-0.004		
	(0.003)			(0.004)						(0.003)			(0.004)		
Historical migration*t2	0.002			0.003			-0.004			0.002			0.003		
	(0.004)			(0.004)			(0.005)			(0.004)			(0.004)		
Altitude*t1		0.003*			0.001						0.003*			0.001	
		(0.002)			(0.002)						(0.002)			(0.002)	
Altitude*t2		-0.001			-0.003**			-0.002			-0.001			-0.003**	
		(0.002)			(0.002)			(0.002)			(0.002)			(0.002)	
Share of land with altitude			-0.007*			-0.007*						-0.007*			-0.007*
below 500m*t1			(0.004)			(0.004)						(0.004)			(0.004)
Share of land with altitude			0.011**			0.012***			0.015**			0.011**			0.012***
below 500m*t2			(0.004)			(0.005)			(0.006)			(0.004)			(0.005)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A14. Results IV regression on PH: FE-2SLS &IV-GMM (one instrument in each regression)

	FE-2SLS									IV-GMM					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
PH	43.039	16.680**	9.097*	30.293	14.568**	7.978*	-106.280	9.712	6.242	40.986	15.961**	10.421**	30.140	14.403**	8.083*
	(32.882)	(7.060)	(5.172)	(21.056)	(6.806)	(4.461)	(620.445)	(6.084)	(4.667)	(32.589)	(7.041)	(5.129)	(21.054)	(6.805)	(4.460)
Income in lagging				0.047	0.174**	0.227***	1.316	-0.034	0.006				0.038	0.170**	0.216***
period 1				(0.191)	(0.072)	(0.055)	(7.115)	(0.096)	(0.080)				(0.190)	(0.072)	(0.055)
Income in lagging							0.570	-0.021	-0.003						
period 2							(3.115)	(0.063)	(0.066)						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	706	706	706	631	631	631	366	366	366	706	706	706	631	631	631
R-squared	0.696	0.927	0.954	0.835	0.943	0.964	-5.223	0.916	0.941	0.722	0.930	0.950	0.836	0.944	0.964
Number of cities	246	246	246	225	225	225	183	183	183	246	246	246	225	225	225
F-statistic in 1st stage	0.97	4.51	3.90	1.24	4.74	4.57	0.03	5.87	4.71	0.97	4.51	3.90	1.24	4.74	4.57
Historical migration*t1	0.002			0.002						0.002			0.002		
	(0.001)			(0.001)						(0.001)			(0.001)		
Historical migration*t2	0.000			-0.000			-0.000			0.000			-0.000		
	(0.001)			(0.001)			(0.002)			(0.001)			(0.001)		
Altitude*t1		-0.001*			-0.002***						-0.001*			-0.002***	
		(0.001)			(0.001)						(0.001)			(0.001)	
Altitude*t2		-0.001***			-0.002***			-0.002**			-0.001***			-0.002***	
		(0.001)			(0.001)			(0.001)			(0.001)			(0.001)	
Share of land with altitude			0.002			0.003**						0.002			0.003**
below 500m*t1			(0.002)			(0.002)						(0.002)			(0.002)
Share of land with altitude			0.005***			0.006***			0.005**			0.005***			0.006***
below 500m*t2			(0.002)			(0.002)			(0.002)			(0.002)			(0.002)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Additional control variables include all those in column (7) of Table A3.

Table A15. Results of IV regression on GI (selected instruments)

	Pooled 2sls			FE 2sls			IV-GMM	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GI	1.768***	0.586***	0.134	18.113*	18.034**	9.599	17.657*	18.342**
	(0.381)	(0.175)	(0.193)	(9.380)	(8.299)	(6.012)	(9.290)	(8.247)
Income in lagging		0.686***	0.789***		0.220***	-0.014		0.223***
period 1		(0.026)	(0.053)		(0.056)	(0.086)		(0.056)
Income in lagging			-0.079*			-0.041		
period 2			(0.044)			(0.072)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	656	412	706	631	366	706	631
R-squared	0.857	0.965	0.966	0.907	0.929	0.917	0.910	0.927
Number of cities				246	225	183	246	225
F-statistic of the 1st stage	24.29	26.45	17.05	1.43	4.95	6.68	1.43	4.95
Historical migration	0.042***	0.046***	0.051***					
	(0.009)	(0.010)	(0.012)					
Altitude*t1	-0.024***	-0.025***	-0.024***	0.001	-0.002***		0.001	-0.002***
	(0.005)	(0.005)	(0.007)	(0.001)	(0.001)		(0.001)	(0.001)
Altitude*t2				0.001	-0.001***	-0.002***	0.001	-0.001***
				(0.001)	(0.000)	(0.001)	(0.001)	(0.000)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Note: Additional control variables include all those in column (7) of Table A3. In the regression on instrumental variables, t1 means the period 2001-2005 and t2 means the period 2006-2010. However, in the first stage of the pooled 2SLSsls regression, instrumental variables are regressed without interacting with period dummies.

TableA16. Results of IV regression on ER (selected instruments)

	Pooled 2sls			FE 2sls			IV-GMM	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ER	5.341***	1.671***	0.280	-0.512	-0.481	2.244	0.975	-0.030
	(1.311)	(0.545)	(0.551)	(1.438)	(1.225)	(1.765)	(1.343)	(1.213)
Income in lagging period 1		0.686***	0.791***		0.291***	0.108		0.281***
		(0.026)	(0.054)		(0.041)	(0.074)		(0.041)
Income in lagging period 2			-0.081*			-0.032		
			(0.044)			(0.090)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	712	656	412	706	631	366	706	631
R-squared	0.847	0.965	0.966	0.963	0.971	0.927	0.962	0.971
Number of cities				246	225	183	246	225
F-statistic of the 1st stage	17.48	21.51	14.05	5.9	5.99	6.3	5.9	5.99
Historical migration*t1	0.014***	0.016***	0.019***					
	(0.003)	(0.003)	(0.004)					
Altitude	-0.006***	-0.007***	-0.008***					
	(0.002)	(0.002)	(0.002)					
Share of land with altitude below 500m *t1				-0.007*	-0.008*		-0.007*	-0.008*
				(0.004)	(0.004)		(0.004)	(0.004)
Share of land with altitude below 500m*t2				0.011**	0.012***	-0.015**	0.011**	0.012***
				(0.004)	(0.005)	(0.006)	(0.004)	(0.005)

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Note: Additional control variables include all those in column (7) of Table A3. In the regression on instrumental variables, t1 means the period 2001-2005 and t2 means the period 2006-2010. However, in the first stage of the pooled 2SLSsls regression, instrumental variables are regressed without interacting with period dummies.

Table A17. Results of regressions using dialectal diversity in year 2000 (single-year data)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
VARIABLES	ELF00	ELF00	ELF00	GI00	GI00	GI00	RQ00	RQ00	RQ00	ER00	ER00	ER00	PH00	PH00	PH00
Dialect diversity	1.414***	1.238***	1.387***	5.317***	5.423***	5.592***	2.724***	2.706***	2.673***	-0.063	-0.023	-0.011	5.952***	5.803***	5.972***
in year 2000	(0.496)	(0.161)	(0.487)	(0.593)	(0.591)	(0.589)	(0.956)	(0.953)	(0.939)	(0.321)	(0.317)	(0.315)	(0.664)	(0.757)	(0.741)
Income in lagging		0.002	0.001		0.002	0.001		0.002	0.001		-0.002	-0.002		0.002	0.001
period 1		(0.005)	(0.005)		(0.005)	(0.005)		(0.005)	(0.005)		(0.005)	(0.005)		(0.005)	(0.005)
Income in lagging			0.000			0.000			0.000			-0.003			0.000
period 2			(0.007)			(0.007)			(0.007)			(0.006)			(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,292	1,257	1,245	1,292	1,257	1,245	1,292	1,257	1,245	1,292	1,257	1,245	1,292	1,257	1,245
Number of cities	263	262	262	263	262	262	263	262	262	263	262	262	263	262	262

Robust standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: In addition to control variables in Table A3, time invariant variables, such as geographical factors, including region dummy, provincial dummy and distance to central cities, historical variables, proxies for economic policy as well as individual effect are also considered. Control variables of Table 7 and individual effect are controlled. In the regression of ER00, the individual effect is not controlled because of collinearity between ER00 and the individual effect.

Table A18: The effect of the experience under CCP control by PS-match

Panel A. Income in lagging periods is not included as the covariate					
	Treated	Controls	Difference	S.E.	T-stat
ELF	9.508	9.432	0.075	0.148	0.51
GI	9.494	9.442	0.052	0.154	0.34
RQ	9.492	9.469	0.023	0.144	0.16
ER	9.503	9.457	0.046	0.142	0.32
PH	9.492	9.461	0.031	0.154	0.20
Panel B. Income in lagging period 1 is included as a covariate					
	Treated	Controls	Difference	S.E.	T-stat
ELF	9.556	9.469	0.087	0.152	0.58
GI	9.558	9.472	0.086	0.138	0.62
RQ	9.558	9.475	0.083	0.149	0.55
ER	9.559	9.468	0.091	0.145	0.63
PH	9.559	9.465	0.093	0.138	0.68
Panel C. Income in lagging period 1 and 2 is included as covariates					
	Treated	Controls	Difference	S.E.	T-stat
ELF	9.788	9.687	0.101	0.191	0.53
GI	9.786	9.680	0.106	0.171	0.62
RQ	9.782	9.620	0.163	0.162	0.90
ER	9.788	9.687	0.101	0.190	0.53
PH	9.790	9.614	0.177	0.177	1.00

Note: The results above are obtained by taking different indices of dialect diversity as one covariate. The significance of the difference between the treated group and the control group is decided by values of the T-statistic. Note: In addition to control variables in Table A3, time invariant variables, such as geographical factors, including region dummy, provincial dummy and distance to central cities, and historical variables as well as economic policy are also considered.

Table A19. The effect of dialectal diversity in Revolutionary area vs non-Revolutionary area

Indep. variable	ELF	ELF	ELF	GI	GI	GI	RQ	RQ	RQ	ER	ER	ER	PH	PH	PH
Dialectal diversity	0.466	0.230	-0.270	-0.892*	-0.981	-0.798	1.007	0.533	-0.548	-0.157	-0.064	-0.211	-0.131	-0.718	-0.454
	(0.651)	(0.640)	(0.543)	(0.492)	(1.204)	(0.859)	(1.364)	(1.374)	(1.170)	(0.239)	(0.253)	(0.333)	(1.207)	(1.054)	(0.883)
Dialectal diversity*	0.120	0.056	1.247*	1.224	1.287	2.489**	-0.215	0.007	2.589	0.527	0.527	0.340	0.936	1.724	2.523**
Revolutionary area	(0.810)	(0.817)	(0.713)	(0.892)	(1.423)	(1.105)	(1.817)	(1.812)	(1.607)	(0.348)	(0.350)	(0.445)	(1.366)	(1.269)	(1.010)
Income in lagging		0.279***	0.186***		0.279***	0.187***		0.280***	0.187***		0.282***	0.205***		0.280***	0.180***
period 1		(0.053)	(0.062)		(0.054)	(0.063)		(0.053)	(0.062)		(0.054)	(0.063)		(0.054)	(0.064)
Income in lagging			0.011			0.019			0.012			0.003			0.023
period 2			(0.067)			(0.066)			(0.067)			(0.067)			(0.064)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	11.483***	7.852***	8.190***	12.088***	8.160***	8.211***	11.611***	7.823***	7.990***	11.975***	7.994***	9.540***	11.712***	7.736***	7.683***
	(3.410)	(2.894)	(2.518)	(3.350)	(2.906)	(2.518)	(3.377)	(2.908)	(2.510)	(3.437)	(3.043)	(2.327)	(3.407)	(2.965)	(2.486)
Observations	507	482	313	507	482	313	507	482	313	507	482	313	507	482	313
R-squared	0.977	0.980	0.971	0.977	0.980	0.972	0.977	0.980	0.971	0.977	0.980	0.971	0.977	0.980	0.972
Number of cities	175	175	172	175	175	172	175	175	172	175	175	172	175	175	172

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Note: Time variant control variables shown in column (7) Table A3 are also included in the regression.

Table A20. The effect of ER: revolutionary area vs. non-revolutionary area

VARIABLES	The East			Other regions		
	(1)	(2)	(3)	(4)	(5)	(6)
ER	-0.937**	-1.181***	-2.826***	0.068	0.142	0.333
	(0.362)	(0.405)	(0.467)	(0.243)	(0.217)	(0.315)
ER* Revolutionary area	0.574	1.110**	3.259***	0.452	0.229	-0.725*
	(0.505)	(0.507)	(0.593)	(0.401)	(0.391)	(0.379)
Income in lagging period 1		0.353***	0.213*		0.277***	0.191*
		(0.064)	(0.127)		(0.074)	(0.103)
Income in lagging period 2			0.078			-0.003
			(0.143)			(0.075)
controls	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes
Constant	15.918***	9.838**	11.409***	13.567***	7.432*	10.353***
	(4.381)	(4.320)	(3.846)	(4.812)	(4.421)	(3.775)
Observations	177	172	112	330	310	201
R-squared	0.986	0.989	0.991	0.979	0.982	0.980
Number of cities	60	60	59	115	115	113

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Note: Time variant control variables shown in column (7) Table A3 are also included in the regression.

Appendix B

Indices of dialectal diversity

Taking dialect j and k as an example, the distance between j and k is defined by $\tau_{jk} = 1 - (\frac{l}{m})^{\delta}$, in which $\frac{l}{m}$ is the proportion of shared branches of j and m is the maximum number of branches between dialects at the lowest level and languages at the highest level. The dialectal distance is calculated at the cluster level, with $m = 5$ from Level 1 to Level 6 and the value of $\frac{l}{m}$ for a pair of different dialects ranging from 0.4 to 0.8. The parameter δ decides the declining speed of distance as the number of shared branches increase (Fearon, 2003). The lower the value of δ is, the larger the dialectal distance is. While Fearon (2003) takes 0.5 as the value of δ, Desmet et al. (2009) settle on a value of 0.05 and show that diversity indices perform better when $\delta \in [0.04, 0.10]$ than indices without distances. In this paper, the values 0.05, 0.1, 0.5, 0.8 and 1 are examined, showing that indices with δ equal to 1 or without distances perform better. Thus, in this paper, only the results of these diversity indices are shown in the following analysis.

Then, given the population share of each dialect group in each city and dialectal distances, five indices of dialect diversity can be calculated in the following way[25]:

$$ELF = 1 - \sum_{k=1}^{K} s_k^2, GI = \sum_{k=1}^{K}\sum_{j=1}^{K} s_j s_k \tau_{jk};$$

$$RQ = \sum_{k=1}^{K} s_k^2 (1 - s_k), ER = \sum_{k=1}^{K}\sum_{j=1}^{K} s_j s_k^2 \tau_{jk};$$

$$PH = 2\sum_{k=1}^{K} s_c s_k \tau_{ck}.$$

s_k, s_j and s_c are population share dialect group, k, j and the largest group c . τ_{jk} and τ_{ck} are corresponding dialectal distances of each pair of dialects.

25. Desmet et al., 2009, Linguistic diversity and redistribution, Journal of European Economics Association, p. 1294-1297.

Appendix C

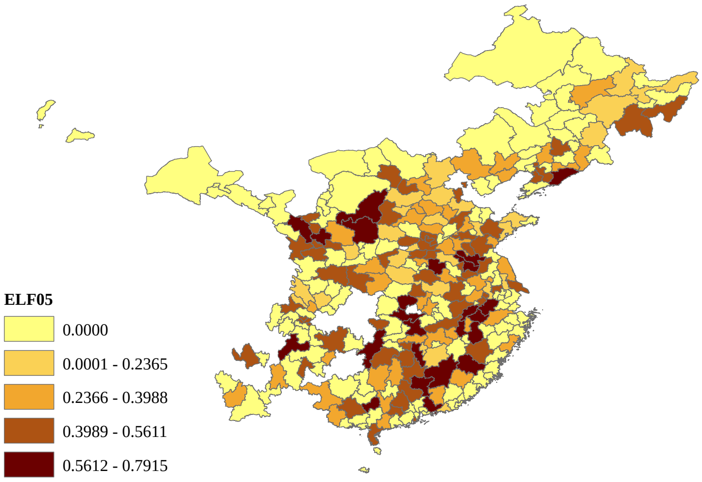
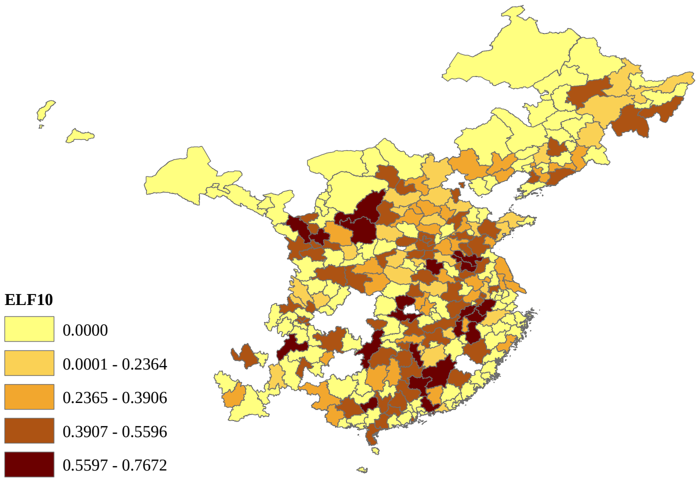


Figure C1.a. ELF in 2005



FigureC1.b. ELF in 2010

Figure C1. Distribution of linguistic fractionalization (ELF)

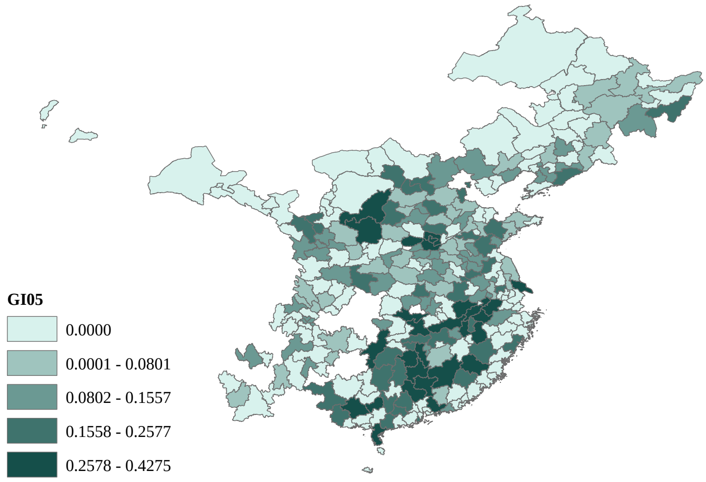


Figure C2.a. GI in 2005

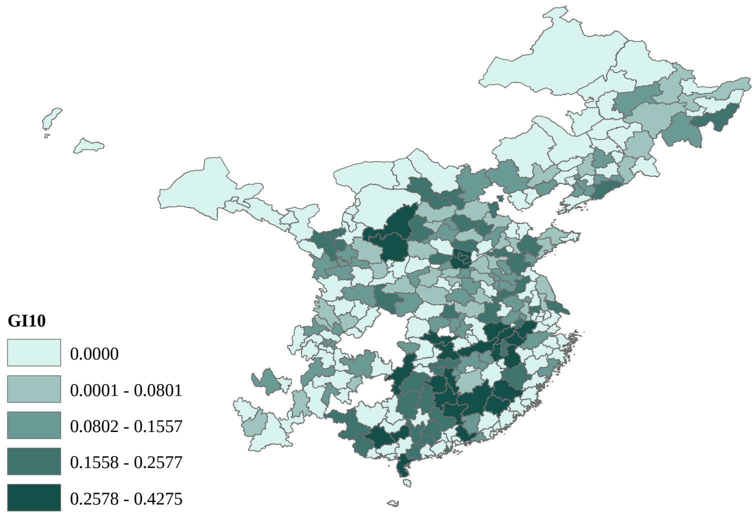


Figure C2.b. GI in 2010

Figure C2. Distribution of adjusted dialect fractionalization (GI)

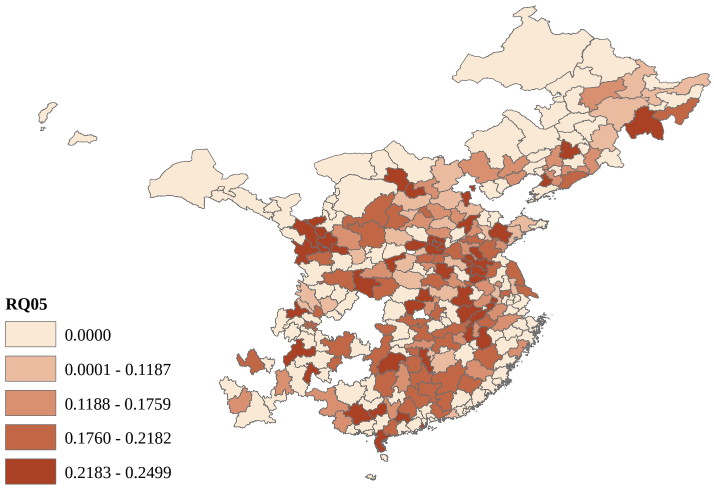


Figure C3.a. RQ in 2005

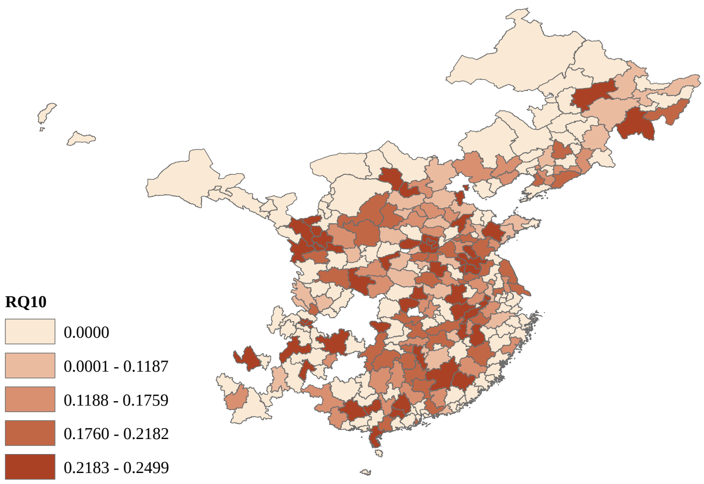


Figure C3.b. RQ in 2010

Figure C3. Distribution of dialect polarization (RQ)

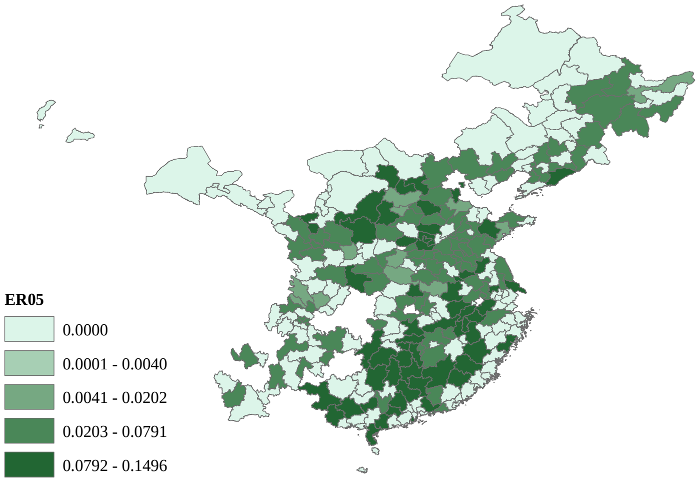


Figure C4.a. ER in 2005

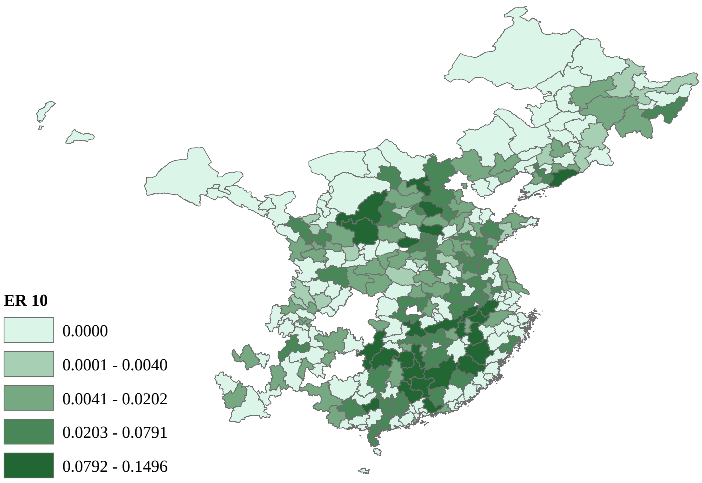


Figure C4.b. ER in 2010

Figure C4. Distribution of adjusted dialect polarization (ER)

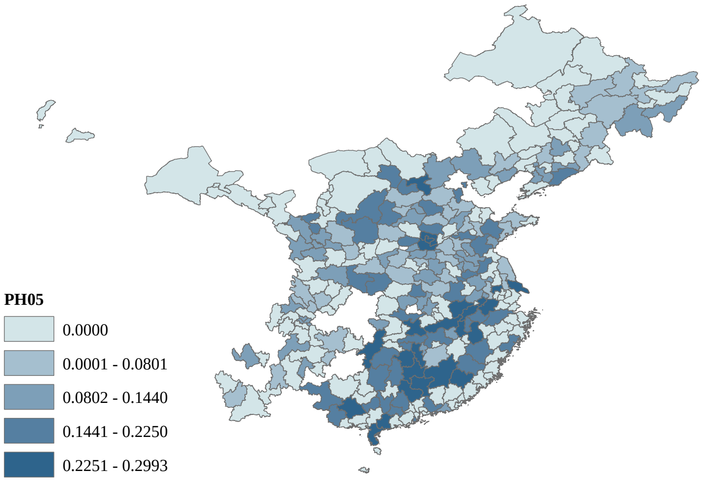


Figure C5.a. PH in 2005

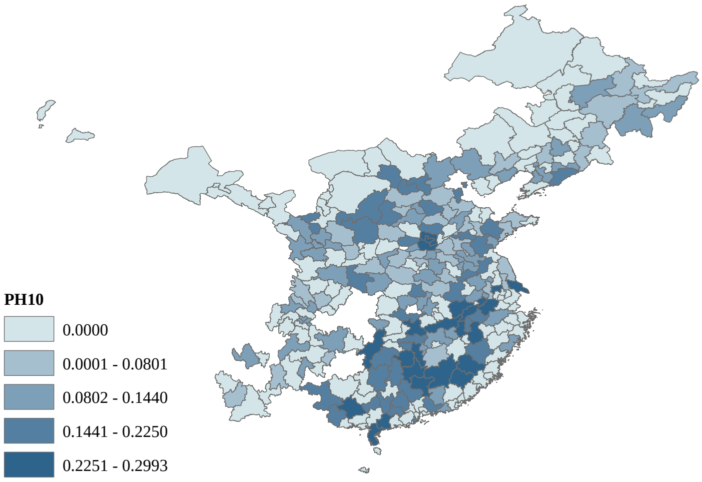


Figure C5.b. PH in 2010

FigureC5. Distribution of periphery heterogeneity (PH)

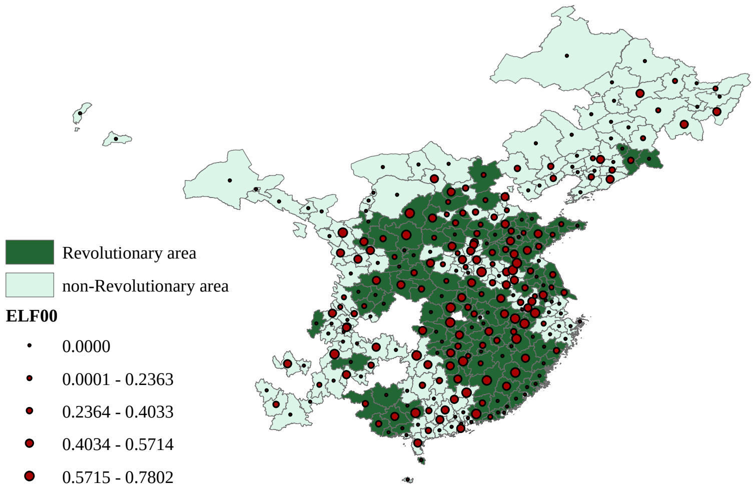


Figure C6.a. The distribution of ELF in revolutionary area vs. non-revolutionary area

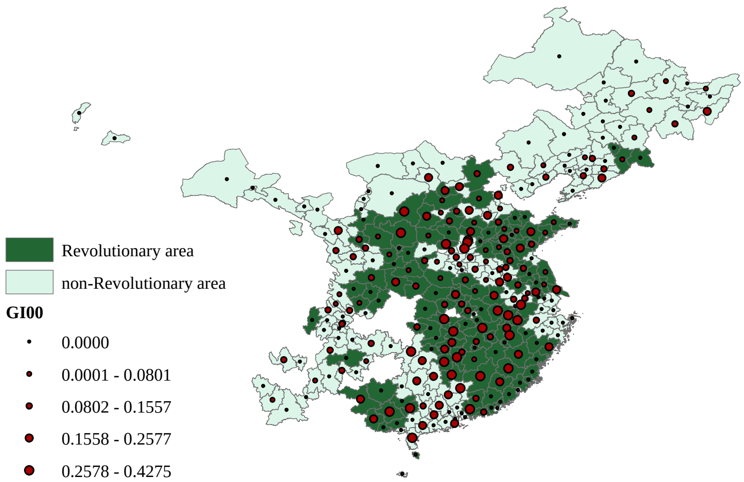


Figure C6.b. The distribution of GI in revolutionary area vs. non-revolutionary area

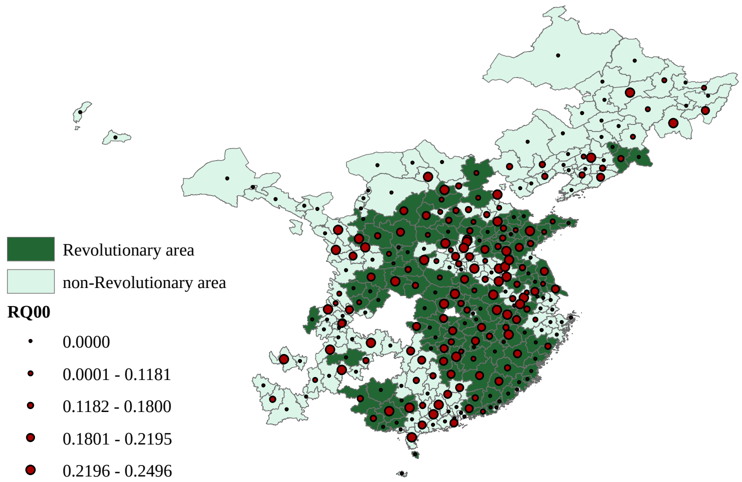


Figure C6.c. The distribution of RQ in revolutionary area vs. non-revolutionary area

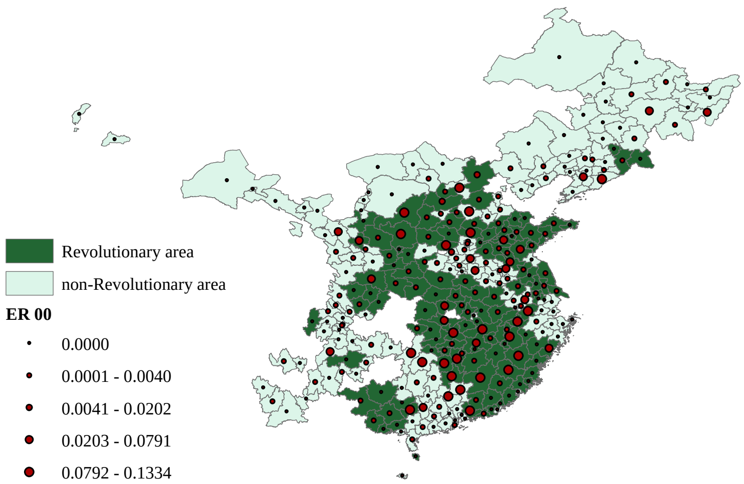


Figure C6.d. The distribution of ER in revolutionary area vs. non-revolutionary area

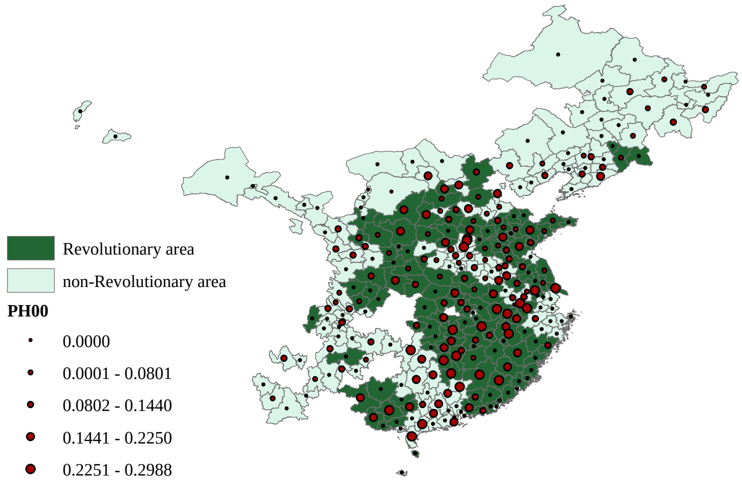


Figure C6.e. The distribution of PH in revolutionary area vs. non-revolutionary area

Figure C6. The distribution of dialect diversity in revolutionary area vs. non-revolutionary area

Appendix D: Mandarin vs. Non-Mandarin Dialects

According to the language tree of Chinese dialects, dialects belong to either the Mandarin supergroup or non-Mandarin supergroup. Correspondingly, based on the category of dialects used by people, prefecture-level cities belong to either the Mandarin area or non-Mandarin area. It should be noted that there is no city where both Mandarin dialects and non-Mandarin dialects are used. Hence, a dummy, Mandarin, is constructed indicating whether a city belongs to the Mandarin area. The estimated result is obtained by the regression of the fixed-effects model introducing the interaction term between the Mandarin dummy and diversity indices. The reason why some dialects in China are called Mandarin dialects is that they have Putonghua as their foundation. Putonghua was formed in the 1920s and promoted in China from 1956 on. Compared with non-Mandarin dialects, Mandarin dialects are closer to the Chinese Putonghua and therefore people speaking Mandarin dialects have advantages in communicating with people in neighborhoods and master Putonghua. Thus, it is generally assumed that it is easier for people speaking Mandarin dialects to engage in market economic activities. Xu et al. (2015) propose that dialectal diversity should have a less negative effect in the Mandarin area. In addition, the distance between non-Mandarin dialects is larger than that between Mandarin dialects. According to our hypothesis that the positive role of diversity increases with the dialectal distance, however, the positive effect in the Mandarin area should also be smaller because of the smaller distance between Mandarin dialects. The opposite will happen in the non-Mandarin area. Therefore, we are not certain about the difference in the effect of diversity between the Mandarin area and the non-Mandarin area.

After introducing an interaction term between dialectal diversity indices and Mandarin in the baseline fixed-effects model, we obtain the estimated results shown in Table D1. The difference in the effect of Mandarin dialect diversity and non-Mandarin dialectal diversity is described by the coefficient of the interaction terms. We can observe that interaction terms based on ELF, RQ and PH are not significant and their effects on economic growth are the same in the Mandarin area and non-Mandarin area. Furthermore, their main effects are consistent with the results of the baseline estimation. For GI, the corresponding interaction term is not significant, but the main effect on income is significant, which is inconsistent with the baseline analysis. We also notice that the coefficient of the interaction term of Mandarin and GI is negative. Although it is not significant, the effect of GI in the whole sample may be disturbed and it is thus not significant in the baseline estimation. In addition, in the estimation of the effect of ER, the main effect is significant and positive, while the coefficients of its interaction term with Mandarin are significant and negative with larger absolute values. Thus, the effects of ER on economic growth are positive in the non-Mandarin area and negative in the Mandarin area. This is also consistent with the insignificant result of ER in the baseline regression. Above all, we do not observe the larger positive role of dialectal diversity in the Mandarin area, and the difference in the effect of ER between the Mandarin area and the non-Mandarin area also implies that larger dialectal distance results in a larger positive role of diversity, which is different from the conclusion of Xu et al. (2015). Hence, we state that there is no significant effect of difficulty in communication due to dialectal diversity.

Table D1. The effect of Mandarin dialectal diversity and non-Mandarin dialectal diversity

Indep. Variable	ELF	ELF	ELF	GI	GI	GI	RQ	RQ	RQ	ER	ER	ER	PH	PH	PH
Dialect diversity	1.455***	0.855**	0.815**	1.126*	0.696	1.102**	3.531***	1.985**	1.603*	0.277*	0.284*	-0.181	0.780	1.338***	1.399**
	(0.407)	(0.400)	(0.408)	(0.585)	(0.497)	(0.547)	(1.057)	(0.904)	(0.928)	(0.148)	(0.145)	(0.216)	(0.967)	(0.393)	(0.551)
Mandarin* Dialect diversity	-0.704	-0.584	-0.567	-1.404	0.523	0.077	-2.084	-1.481	-1.108	-0.662**	-0.692*	-0.186	1.430	-0.597	-0.718
	(0.630)	(0.683)	(0.670)	(1.102)	(2.055)	(1.995)	(1.445)	(1.482)	(1.477)	(0.270)	(0.352)	(0.476)	(1.825)	(1.423)	(1.411)
Income in lagging period 1		0.285***	0.069		0.288***	0.068		0.284***	0.068		0.293***	0.076		0.281***	0.064
		(0.039)	(0.060)		(0.039)	(0.060)		(0.039)	(0.061)		(0.039)	(0.059)		(0.039)	(0.061)
Income in lagging period 2			0.019			0.021			0.020			0.034			0.024
			(0.068)			(0.068)			(0.068)			(0.071)			(0.069)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	7.344***	5.641***	10.341***	7.603***	5.691***	10.336***	7.322***	5.646***	10.352***	7.815***	5.858***	10.407***	7.551***	5.677***	10.267***
	(1.206)	(1.309)	(1.594)	(1.223)	(1.327)	(1.599)	(1.219)	(1.308)	(1.599)	(1.192)	(1.314)	(1.573)	(1.164)	(1.308)	(1.619)
Observations	712	656	412	712	656	412	712	656	412	712	656	412	712	656	412
R-squared	0.964	0.972	0.954	0.964	0.972	0.954	0.964	0.972	0.954	0.964	0.972	0.954	0.964	0.972	0.954
Number of cities	252	250	229	252	250	229	252	250	229	252	250	229	252	250	229

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, p<0.1.

Note: In addition, to control variables in column (7) of Table A3 and individual effects, time invariant variables, such as geographical factors, including region dummy, provincial dummy and distance to central cities, and historical variables as well as economic policy are also considered.

Diskussionsbeiträge - Fachbereich Wirtschaftswissenschaft - Freie Universität Berlin
Discussion Paper - School of Business and Economics - Freie Universität Berlin

2020 erschienen:

2020/1	ASSEYER, Andreas: Wholesale price discrimination with regulatory asymmetry *Economics*
2020/2	JESSEN, Robin und Johannes KÖNIG: Hours Risk and Wage Risk: Repercussions over the Life-Cycle *Economics*
2020/3	ASSEYER, Andreas: Collusion and Delegation under Information Control *Economics*
2020/4	ADAM, Marc Christopher: Liquidating Bankers' Acceptances: International Crisis, Doctrinal Conflict and American Exceptionalism in the Federal Reserve 1913-1932 *Economics*
2020/5	BÖNKE, Timm; Rick GLAUBITZ; Konstantin GÖBLER; Astrid HARNACK; Astrid PAPE und Miriam WETTER: Die Entwicklung und Prognose von Lebenserwerbseinkommen in Deutschland *Economics*
2020/6	HUNDSDOERFER, Jochen und Eva MATTHAEI: Gender Discriminatory Taxes, Fairness Perception, and Labor Supply *FACTS*