



The Cost and Benefit of Regional Cultural Diversity on the Income of Rural Workers: Evidence from China

Ling Wang¹ · Jianqing Ruan²

Accepted: 18 November 2022 / Published online: 5 December 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

This paper studies the impact of regional cultural diversity of working cities on the income of rural workers. Using a nationally representative micro-dataset in China, and adopting dialect diversity at the city level as the indicator of regional cultural diversity, we find that regional cultural diversity has a highly significant hump-shaped effect on rural workers' incomes from the wage employment and self-employment, reflecting a trade-off between the beneficial and the detrimental effects of regional cultural diversity on rural workers' incomes. The results remain robust to the inclusion of a set of covariates, alternative estimation technique, alternative measure of regional cultural diversity, and a placebo test. We also find that the detrimental effect of diversity is weaker for high-educated rural workers, and the impact of regional cultural diversity remains stable in the short term. To address potential endogeneity, we employ an instrumental variable approach by using the river length of cities to instrument the dialect diversity, and the instrumented results confirm the conclusions. In addition, we provide empirical evidence that the cost and benefit of regional cultural diversity on the income of rural workers can be attributed to reduced social capital and increased creativity.

Keywords Cultural diversity · Off-farm income · Rural workers · China

JEL Classification J31 · Z13

1 Introduction

Since the livelihood of farmers increasingly depends on off-farm employment, understanding the determinants of off-farm incomes of rural workers is a critical issue in the analysis of rural poverty (Corral & Reardon, 2001; Haggblade et al., 2010; Otsuka & Yamano, 2006; Sumner, 1982; Thiede et al., 2018). Previous studies have shown that off-farm incomes of rural workers are influenced by a number of demographic and economic factors

✉ Jianqing Ruan
ruanjq@zju.edu.cn

¹ Institute of Food and Strategic Reserves, Nanjing University of Finance and Economics, Nanjing 210003, Jiangsu, China

² China Academy for Rural Development, Zhejiang University, Hangzhou 310058, Zhejiang, China

(Cherdchuchai & Otsuka, 2006; Corral & Reardon, 2001). However, little is known about how some intangible factors affect off-farm incomes, such as regional cultural diversity.

The purpose of this paper is to test the impact of regional cultural diversity of working cities on the off-farm incomes of rural workers. Regional cultural environment may profoundly shape farmers' behavior. Farmers get used to a uniform and closed cultural environment because agricultural production mainly requires small-scale cooperation and interaction among relatives and neighbors in the same village (Fei & Malinowski, 2013; Olsson & Paik, 2016). When farmers participate in off-farm employment, they start to interact with unfamiliar people and need to integrate into a relatively diverse cultural environment, then how does cultural diversity in the working regions affect these rural workers' performance? Do rural workers be challenged or benefited by the regional cultural diversity? While there is no shortage of studies on the impact of cultural diversity in labor markets, there is comparably little work on understanding how regional cultural diversity affects rural workers, especially in developing countries.

In this paper, we hypothesize that regional cultural diversity has a hump-shaped effect on rural workers' incomes. Rural workers are challenged and benefited simultaneously by the regional cultural diversity. Rural workers' labor performance is decreased on the one hand by reduced social capital, while simultaneously enhanced on the other by increased creativity. Hence, the overall effect of regional cultural diversity would be hump-shaped, rather than monotonically negative or positive. Considering the diminishing marginal returns to the both cost and benefit of regional cultural diversity, we will observe that the beneficial effect dominates at lower levels of diversity and the detrimental effect dominates at higher levels, that is, an inverted-U relationship between regional cultural diversity and the incomes of rural workers.

We focus on rural workers, that is, farmers that have participated in off-farm employment and obtained off-farm incomes. We use dialect diversity to measure regional cultural diversity of cities, considering dialects are highly correlated with regional cultures. We take China as an example because it is a large developing country with a rapidly increasing number of rural workers. Additionally, China is a vast country with diverse regional cultures, and despite the large cultural diversity in China, as a politically unified nation for decades, China is far more homogeneous in religious, political, and certain socioeconomic factors.

We conduct our analysis with data on dialect diversity at the city level and a nationally representative micro-dataset from the China Labor-force Dynamic Survey 2014 (the CLDS 2014). We empirically test the impact of regional cultural diversity on rural workers' incomes from the two primary types of off-farm employment: wage employment and self-employment. We find that regional cultural diversity has a highly significant hump-shaped effect on rural workers' incomes from these two types of employment. Furthermore, to address the potential endogeneity issue caused by omitted variables and measurement error, we employ the river length of cities to instrument the dialect diversity, and the instrumented results confirm the conclusions. Finally, we test the possible mechanisms of the hump-shaped effect of regional cultural diversity on rural workers' incomes. We provide empirical evidence for the cost and benefit of regional cultural diversity on rural workers. We find regional cultural diversity has a significantly negative impact on rural workers' social capital, while has a significantly positive effect on rural workers' creativity.

By examining the impact of regional cultural diversity on the off-farm incomes of rural workers, our study contributes to the rich body of literature exploring the determinants of incomes of farmers. Previous studies have mostly focused on demographic and economic factors, such as age, gender, education, training, working experience, mobile internet use,

social capital, and institution restrictions (Cherdchuchai & Otsuka, 2006; Corral & Rardon, 2001; Ibekwe et al., 2010; Knight & Yueh, 2004; Liu, 2005; Luan et al., 2015; Sun et al., 2021; Wang et al., 2015; Yang et al., 2021). We contribute to this literature by proposing an intangible and deep-rooted factor—regional cultural diversity as an important determinant of the off-farm incomes of rural workers.

Our paper is also related to the literature concerning the impact of cultural diversity in labor markets. Previous studies show that cultural diversity in the workplace can be associated with both positive and negative outcomes for workers. On the one hand, cultural diversity creates potential benefits by increasing the variety of ideas and skills. For instance, Ottaviano and Peri (2006) find that cultural diversity measured by the share of foreign-born residents has a positive effect on the wage of US-born individuals. Suedekum et al. (2014), Lee (2015), and Bellini et al. (2013) find results similar to Ottaviano and Peri (2006) for German, UK, and European countries. On the other hand, cultural diversity may increase intercultural frictions and reduce social cohesion, and thus lower productivity (Dale-Olsen & Finseraas, 2020; Hjort, 2014; Koopmans & Schaeffer, 2016; Lazear, 1999; Parrotta et al., 2014). One incremental contribution of this study is that we focus on the rural workers separately. Rural workers require special attention as they are usually disadvantaged groups in the labor market, and they lived in a uniform cultural environment before participating in off-farm employment. It remains an open issue how does cultural diversity affect this special group. The second contribution of our study is that we focus on a developing country rather than developed countries. Developed countries may benefit more from cultural diversity and educated foreign immigrants (Alesina & Ferrara, 2005; Kemeny, 2012; Lee, 2015). It remains unclear how does regional cultural diversity affect productivity within a developing country. The third contribution is that we focus on intra-national cultural diversity rather than cross-country cultural diversity, which helps mitigate the confounding influences due to the omission of unobserved factors of cross-country differences.

The remainder of this paper is structured as follows: Sect. 2 provides a theoretical framework that relates regional cultural diversity to rural workers' incomes. Section 3 explains the definitions and sources of the main variables, and the econometric specification. Section 4 reports the empirical results. Section 5 further investigates the possible channels through which regional cultural diversity affects rural workers' incomes. Section 6 offers a conclusion.

2 Theoretical Framework

We hypothesize that regional cultural diversity has a hump-shaped effect on rural workers' incomes. To illustrate the mechanisms underlying this hypothesis, we review and discuss the literature on the cost and benefit of cultural diversity on productivity, and finally, we use a stylized model to illustrate the inverted-U relationship between regional cultural diversity and the income of rural workers.

2.1 The Cost of Cultural Diversity

The cost of cultural diversity mainly arises from reduced social capital. According to social identity theory, individuals naturally tend to interact with others who are similar to themselves, which gives them a feeling of identity and belonging. People with similar cultural backgrounds therefore form a cohesive social group, and are more likely to trust and

cooperate with in-group members than out-group members (Alesina & La Ferrara, 2000; DiMaggio, 1997; Tajfel, 1982). Cultural diversity negatively affects social capital because the probability of interacting with people with similar cultural backgrounds is low in culturally diverse regions. A large body of evidence shows that cultural diversity and social capital are negatively correlated in many different settings (Koopmans & Schaeffer, 2016; Putnam, 2007).

The negative impact of cultural diversity on social capital further affects productivity, as social capital is the foundation of social and economic activities. For instance, social capital promotes greater coordination and cooperation among individuals and between departments to solve problems. Social capital also promotes economic efficiency by facilitating information exchange and lowering transaction costs (Coleman, 1988; Josten, 2013; Narayan & Pritchett, 1999).

Based on the above premises, a body of literature suggests that cultural diversity generates costs on productivity by increased intercultural frictions and reduced social capital (Dale-Olsen & Finseraas, 2020; Hjort, 2014; Koopmans & Schaeffer, 2016; Lazear, 1999; Parrotta et al., 2014). This may especially apply to rural workers, as this disadvantaged group with less education and training may rely more on social capital to struggle for competence in the labor market.

2.2 The Benefit of Cultural Diversity

The benefit of cultural diversity mainly arises from increased creativity. There is a vast literature on the link between cultural diversity and creativity. People with different cultural backgrounds have diverse ideas and values, and thus may embody complementary skills and abilities. For instance, cultural diversity may help bring new perspectives to capture new opportunities or create new products. Diverse perspectives and skills may also help create better solutions to solve problems (Hong & Page, 2004; Niebuhr, 2010). In culturally diverse regions, people are more likely to interact with and be inspired by those of different cultural backgrounds than in culturally homogeneous regions, and thus may be more creative.

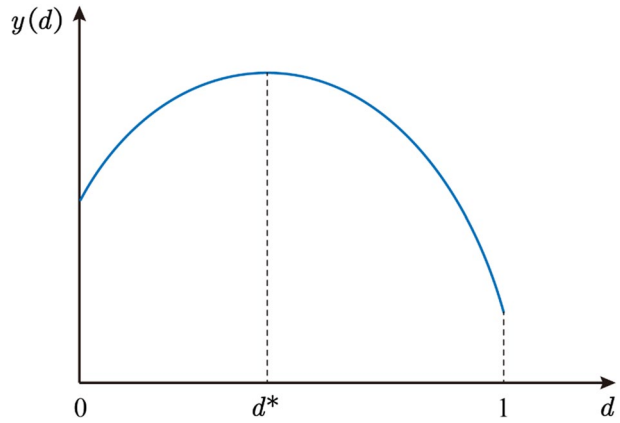
The positive impact of cultural diversity on creativity naturally leads to higher productivity. A wide literature supports this claim that cultural diversity generates benefits on productivity through complementary skills and increased creativity (Bellini et al., 2013; Lee, 2015; Ottaviano & Peri, 2006; Suedekum et al., 2014). Although most of these studies focused on high-skilled workers in developed countries, we assume that this beneficial side of cultural diversity may also apply to rural workers in developing countries.

2.3 A simple Model

As discussed above, regional cultural diversity may confer conflicting effects on rural workers' performance in the labor market, through reducing social capital and fostering creativity. We then use a stylized model to illustrate the relationship between regional cultural diversity and the income of rural workers, which is built on Ashraf and Galor's (2013b) model of diversity and productivity.

Suppose that the degree of regional cultural diversity, $d \in [0, 1]$, has a positive but diminishing effect on the income of rural workers by fostering creativity. The income of a rural worker, y , is determined by a vector of socioeconomic factors, x , such as age, gender,

Fig. 1 The hump-shaped effect of regional cultural diversity on the income of rural workers



education, and GDP, as well as by the degree of regional cultural diversity, d . Specifically, the income of a rural worker is given by:

$$y = f(x, d) \tag{1}$$

where $f(x, d) > 0$, $f_d(x, d) > 0$, and $f_{dd}(x, d) < 0$ for all $d \in [0, 1]$. $\lim_{d \rightarrow 0} f_d(x, d) = \infty$ and $\lim_{d \rightarrow 1} f_d(x, d) = 0$.

Suppose further that the degree of regional cultural diversity also negatively affects the income of rural workers by reducing social capital. In particular, a fraction of the income $f(x, d)$ is lost due to the lack of trust and cooperation at work, which is represented as αd , $\alpha \in (0, 1)$. Hence, the income of a rural worker becomes:

$$y = f(x, d)(1 - \alpha d) \equiv y(d) \tag{2}$$

The first- and second-order conditions of y with respect to d are:

$$y'(d) = (1 - \alpha d)f_d(x, d) - \alpha f(x, d) \tag{3}$$

$$y''(d) = (1 - \alpha d)f_{dd}(x, d) - 2\alpha f_d(x, d) \tag{4}$$

where $\lim_{d \rightarrow 0} y'(d) > 0$, $\lim_{d \rightarrow 1} y'(d) < 0$, and $y''(d) < 0$, implying $y(d)$ is an inverted-U function of d , as depicted in Fig. 1. The intermediate level of regional cultural diversity d^* that maximizes the income of rural workers satisfies:

$$(1 - \alpha d^*)f_d(x, d^*) = \alpha f(x, d^*) \tag{5}$$

3 Data and Econometric Specification

3.1 Measures of Rural Workers' Incomes

The relevant variables that reflect rural workers' productivity in the labor market are their incomes from off-farm employment. Specifically, we focus on incomes from the

Table 1 Summary statistics

Variable	Obs	Mean	SD	Min	Max
Dialect diversity	105	0.29	0.23	0.00	0.76
Dialect diversity squared	105	0.13	0.15	0.00	0.57
<i>Wage employment sample</i>					
Wage employment income	4949	9.81	1.02	2.30	14.00
Age	4949	41.19	12.76	16.00	114.00
Age squared	4949	1859.67	1115.50	256.00	12,996.00
Gender	4949	0.60	0.49	0.00	1.00
Marital status	4949	0.85	0.36	0.00	1.00
Education	4949	3.50	2.01	1.00	11.00
<i>Self-employment sample</i>					
Self-employment income	3205	9.28	1.39	2.08	14.91
Age	3205	47.70	12.05	16.00	81.00
Age squared	3205	2420.53	1154.07	256.00	6561.00
Gender	3205	0.56	0.50	0.00	1.00
Marital status	3205	0.93	0.26	0.00	1.00
Education	3205	2.70	1.33	1.00	9.00
<i>Regional controls</i>					
Topographic slope	105	2.40	2.02	0.05	9.40
Topographic relief	105	1362.00	943.05	37.00	4520.00
North	105	0.44	0.50	0.00	1.00
Coastal	105	0.24	0.43	0.00	1.00
GDP	105	16.91	0.97	14.73	19.21
Population density	105	6.08	0.79	3.43	8.25

two primary types of off-farm employment: wage employment income and self-employment income (both are in logarithmic form).

The data on the two types of incomes of rural workers are derived from the China Labor-force Dynamic Survey 2014 (CLDS 2014) by the Center for Social Survey of Sun Yat-Sen University, China. The CLDS 2014 is a large-scale, nationally representative survey. It adopts a stratified multistage cluster sampling method and covers large geographic areas, including 29 provinces and 124 cities in China.

The CLDS 2014 targets labors aged 15–64 years and contains rich information about the respondents' sociodemographic characteristics. The original data of the CLDS 2014 includes 23,594 individuals. To identify farmers that had participated in off-farm employment, we only retain the respondents with agricultural household registration, having incomes from wage employment or self-employment. Respondents below the legal working age (16 years old) and the student respondents are also excluded. After restricting the sample to rural workers, the sample includes 5,475 observations for wage employment and 3,670 observations for self-employment. Additionally, we only include individuals who have valid information for all variables. The final sample includes 4,949 observations for wage employment, 3,205 observations for self-employment, and covers 105 cities. Table 1 provides a summary of the statistics.

3.2 Measures of Regional Cultural Diversity

The key explanatory variable is the regional cultural diversity of rural workers' working cities. Regional cultural diversity is measured by dialect diversity. As has been well documented, culture and language are inseparable, because language simultaneously reflects culture, and is influenced and shaped by it, which makes dialect diversity a widely used measure of regional cultural diversity (Brown, 1980; Jiang, 2000; Spolaore & Wacziarg, 2016). The dialect diversity index captures the probability that any two random individuals in a given region speak different dialects. The greater the dialect diversity in a region, the greater the probability that any two individuals speak different dialects and thus have different regional cultural backgrounds, that is, the greater the degree of regional cultural diversity of this region.

According to Greenberg (1956), the formula for the dialect diversity in a rural worker's working city i is as follows:

$$\text{Dialect diversity}_i = 1 - \sum_{j=1}^N S_{ij}^2 \quad (6)$$

where S_{ij} is the proportion of speakers of dialect j to the total population in the city i , and N is the total number of dialects in the city i . Data on dialect diversity of cities in China can be obtained from Xu et al. (2015). Figure 2 shows the regional pattern of dialect diversity across cities in China. Darker shades indicate higher values of dialect diversity, while lighter shades indicate lower values. White shades represent cities not in our sample or data not available.

3.3 Controls

In examining the impact of regional cultural diversity on rural workers' incomes, it is necessary to control for potential confounding factors and covariates that affect the outcomes, including personal sociodemographic characteristics and regional characteristics.

Personal sociodemographic characteristics include the rural workers' age, age squared term, gender, marital status, and education level, all of which may affect their incomes. The data on personal sociodemographic characteristics can be obtained in the CLDS 2014.

Regional characteristics include geographic factors and sociodemographic factors. Geographic factors are possible confounding factors that may be correlated with both dialect diversity and rural workers' incomes. We thus control for a series of geographic factors, which include topographic slope, topographic relief, a dummy variable indicating the northern city, and a dummy variable indicating whether the city is coastal.

Topographic slope and relief. Topographic slope and relief may shape regional cultural diversity because of geographic isolation, whereas topographic slope and relief may also affect regional development and thus rural workers' incomes. The data on topographic slope and relief are calculated by using the SRTM DEM (90 m).

North dummy. There is a large regional variation in cultural diversity between northern and southern China, as cultures are similar in northern China and vary greatly even between neighboring villages in southern China (Kurpaska, 2010; Tang, 2009). Moreover, the average incomes are relatively higher in the south compared to the northern regions in China. Regions in China are divided into those in the south and north, according to

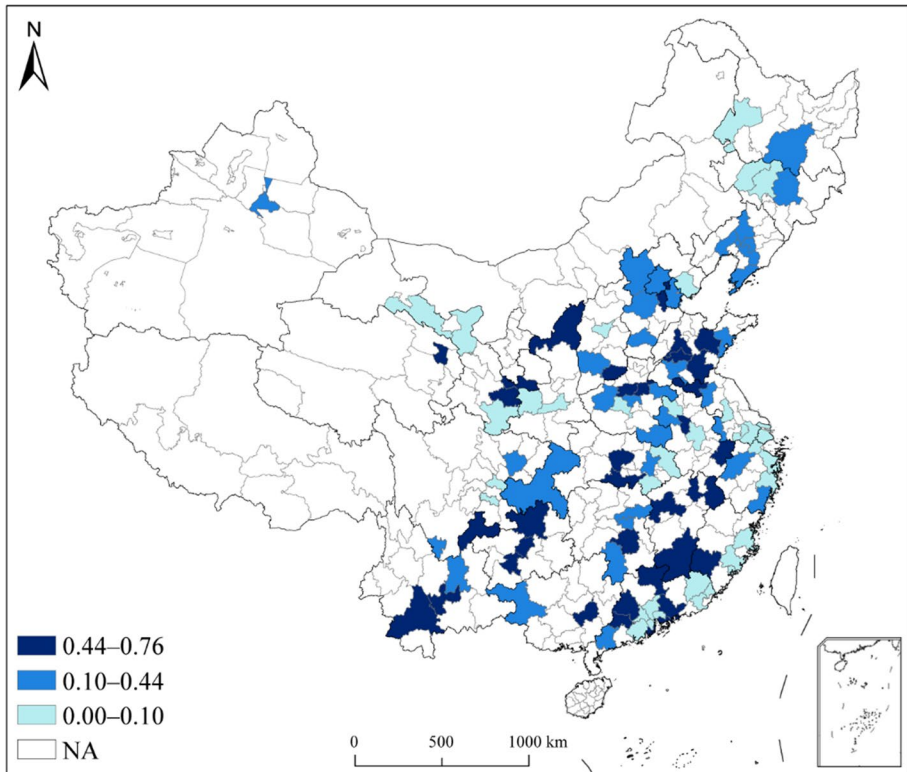


Fig. 2 Regional pattern of regional cultural diversity in China. *Note:* The figure shows the regional pattern of dialect diversity across cities in China. Darker shades indicate higher values of dialect diversity. White shades represent cities not in our sample or data not available. The data on dialect diversity are taken from Xu et al. (2015)

Qinling–Huaihe line, a geomorphological boundary. If the working city of a rural worker is north of the Qinling–Huaihe line, the dummy variable North equals 1; otherwise, it equals 0.

Coastal dummy. Coastal cities may be exposed to diverse cultures, while proximity to the sea may also facilitate trade and promote regional development. To control for this confounding factor, we include a dummy variable indicating whether the working city of a rural worker is coastal. The data of the north and coastal dummy can be derived from Google Earth.

Regional sociodemographic characteristics include GDP and population density (both are in logarithmic form). Both of them are likely correlated with regional cultural diversity, and also affect rural workers' incomes. The data on GDP and population density can be obtained from the China Statistical Yearbook 2014.

3.4 Econometric Specification

To examine the effect of regional cultural diversity on rural workers' incomes, we employ the following OLS specification:

$$y_{ij} = \beta_0 + \beta_1 D_j + \beta_2 D_j^2 + \beta_3' X + \varepsilon_i \quad (7)$$

where y_{ij} is the income of rural worker i in city j , D_j and D_j^2 denote the dialect diversity and its squared term of the city j . X is a set of the aforementioned control variables. The parameters to be estimated are β_0 , β_1 , β_2 , and β_3' ; and ε_i is the stochastic error term. Specifically, according to the aforementioned hypothesis, the impact of regional cultural diversity on the incomes of rural workers should be hump-shaped—i.e., that $\beta_1 > 0$ and $\beta_2 < 0$.

4 Empirical Results

4.1 Baseline Results

In Table 2, we examine the impact of regional cultural diversity on rural workers' incomes. Columns 1 and 2 report the results on wage employment income, whereas columns 3 and 4 report the results on self-employment income. To provide a baseline, we only include the dialect diversity and its squared term in the regressions initially (columns 1 and 3). For both types of off-farm work, the results reveal an inverted-U relationship between regional dialect diversity and the rural workers' incomes. In columns 2 and 4, we further add control variables introduced in Sect. 3, including personal demographic characteristics, a series of geographic confounding factors, and regional sociodemographic characteristics. The results consistently show that, for both types of off-farm work, regional cultural diversity has a significantly hump-shaped effect on rural workers' incomes, and the estimated linear and quadratic coefficients are both statically significant at the 1% level, which lends credence to the assertion that these effects are indeed due to dialect diversity as opposed to potential confounding factors.

Specifically, the optimal level of diversity is 0.27 for both types off-farm incomes conditional on the covariates (columns 2 and 4), suggesting that the regional cultural diversity positively affects rural workers' incomes when dialect diversity is smaller than 0.27, while the negative effect dominates when the diversity is larger than the optimal level. Figure 3 shows the degree of proximity to the optimal diversity of cities in China. Darker shades indicate higher degrees of proximity to the optimal level, while lighter shades indicate lower degrees of proximity.

In terms of magnitude, for wage employment, the estimated linear and quadratic coefficients associated with dialect diversity imply that a one percentage point increase in diversity for the least diverse city in the regression sample would raise rural workers' income by 0.66%, whereas a one percentage point decrease in diversity for the most diverse city would raise rural workers' income by 1.22%; for self-employment, a one percentage point increase in diversity for the least diverse city in the regression sample would raise rural workers' income by 1.25%, whereas a one percentage point decrease in diversity for the most diverse city would raise rural workers' income by 2.30%.¹

¹ The magnitude of the effect of dialect diversity at the specified level \bar{D} is given by: $\Delta y/y = \exp\left\{\Delta D\left(\beta_1 + 2\beta_2\bar{D} + \bar{D}\Delta D\right)\right\} - 1$.

Table 2 Impact of regional cultural diversity on rural workers' incomes

	(1)	(2)	(3)	(4)
	Wage employment income		Self-employment income	
Dialect diversity	0.50** (0.24)	0.67*** (0.22)	1.33*** (0.36)	1.27*** (0.34)
Dialect diversity squared	-1.42*** (0.39)	-1.24*** (0.38)	-2.65*** (0.54)	-2.34*** (0.51)
Age		0.05*** (0.02)		0.08*** (0.01)
Age squared		-0.00*** (0.00)		-0.00*** (0.00)
Gender		0.46*** (0.03)		0.40*** (0.04)
Marital status		0.24*** (0.05)		0.41*** (0.09)
Education		0.10*** (0.01)		0.20*** (0.02)
Topographic slope		-0.03 (0.02)		-0.03* (0.02)
Topographic relief		-0.00 (0.00)		0.00 (0.00)
North		-0.24*** (0.03)		-0.27*** (0.05)
Coastal		0.26*** (0.03)		0.26*** (0.07)
GDP		0.24*** (0.02)		0.29*** (0.03)
Population density		-0.12*** (0.03)		-0.21*** (0.04)
Constant	9.84*** (0.02)	5.21*** (0.35)	9.24*** (0.05)	3.59*** (0.59)
Optimal diversity	0.18	0.27	0.25	0.27
Observations	4949	4949	3205	3205
R-squared	0.008	0.279	0.010	0.272

OLS coefficients are reported

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

4.2 Robustness Checks

4.2.1 Alternative Estimation Technique

In examining the impact of cultural diversity at the city level on rural workers' incomes at the individual level, we use the OLS model, a traditionally used estimation technique (e.g., Suedekum et al., 2014; Thomas & Zhang, 2005). However, the standard OLS model requires the random errors to be independent. One potential violation of this independence

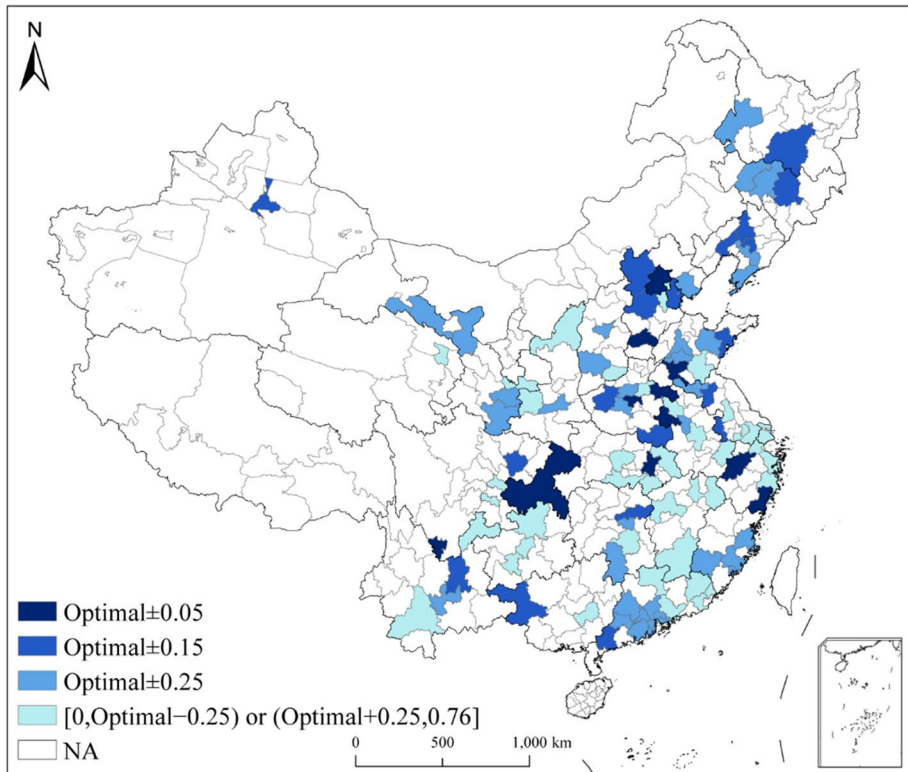


Fig. 3 The degree of proximity to the optimal diversity of cities in China. *Note:* The figure shows the degree of proximity to the optimal dialect diversity of cities in China. Darker shades indicate higher degrees of proximity to the optimal level. White shades represent cities not in our sample or data not available. The data are obtained from the authors' calculations

assumption is that the errors may be interdependent across individuals within each city if we omit city-level controls, thus causing the OLS estimates to be biased. Our baseline estimation attempts to alleviate this concern by controlling for a set of city-level factors. To further address this concern, we employ an alternative estimation technique—the hierarchical linear model (HLM), which is a complex form of OLS that is used to analyze variance in the outcome variables when the predictor variables are at varying hierarchical levels (Raudenbush & Bryk, 2002; Woltman et al., 2012).

Table 3 reports the results using HLM. Columns 1 and 2 report the results concerning wage employment and self-employment income, respectively. The results consistently show that regional cultural diversity has a significantly hump-shaped effect on rural workers' incomes of the two types of off-farm employment.

4.2.2 Alternative Measure of Regional Cultural Diversity

The baseline analysis uses dialect diversity within a city as the measure of regional cultural diversity. For robustness, we employ genetic diversity as an alternative measure, as it is commonly accepted that genetic diversity is positively correlated with cultural diversity

Table 3 Impact of regional cultural diversity on rural workers' incomes: HLM estimates

	(1) HLM Wage employment income	(2) HLM Self- employment income
Dialect diversity	0.83** (0.34)	1.20** (0.56)
Dialect diversity squared	-1.24** (0.55)	-2.53*** (0.88)
Age	0.05*** (0.01)	0.08*** (0.01)
Age squared	-0.00*** (0.00)	-0.00*** (0.00)
Gender	0.47*** (0.03)	0.41*** (0.04)
Marital status	0.23*** (0.04)	0.43*** (0.08)
Education	0.08*** (0.01)	0.17*** (0.02)
Topographic slope	-0.05*** (0.02)	-0.03 (0.03)
Topographic relief	-0.00 (0.00)	0.00 (0.00)
North	-0.25*** (0.05)	-0.12 (0.08)
Coastal	0.32*** (0.05)	0.31*** (0.10)
GDP	0.28*** (0.03)	0.34*** (0.05)
Population density	-0.15*** (0.04)	-0.23*** (0.07)
Constant	4.72*** (0.44)	3.00*** (0.84)
Observations	4949	3205
Number of groups	104	105

Coefficients are reported

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

(Ashraf & Galor, 2013a; Cavalli-Sforza, 1997; Cavalli-Sforza et al., 1994; Desmet et al., 2017). Genetic diversity is typically measured by an index called expected heterozygosity. Like most other measures of diversity, this index may be interpreted simply as the probability that any two random individuals in a given population are genetically different from one another. The expected heterozygosity of a given city is constructed by geneticists using sample data on allelic frequencies (Ashraf & Galor, 2013b; Cavalli-Sforza, 2005).

Specifically, consider a single gene with k observed alleles in the population in city j , and p_i denotes the frequency of the i -th allele. Then, the genetic diversity of the city j measured by expected heterozygosity is:

$$\text{Genetic diversity}_j = 1 - \sum_{i=1}^k p_i^2 \quad (8)$$

Genetic diversity is calculated with data on the allelic frequencies of Gm (a neutral genetic marker) of cities in China, which are obtained from Zhao and Lee (1989). The genetic diversity data covers 33 cities in our sample, reducing the wage employment sample to 1,290 individuals, and the self-employment sample to 928 individuals.

Table 4 reports the results using genetic diversity as the alternative measure of cultural diversity. Column 1 reports the result concerning wage employment income, whereas column 2 reports the result concerning self-employment income. Despite the constrained sample sizes, however, the empirical findings once again reveal a highly statistically significant hump-shaped relationship between regional cultural diversity, predicted by genetic diversity, and rural workers' incomes of the two types of off-farm employment.

4.3 Heterogeneity

The differences among certain characteristics of individuals may lead to heterogeneity in the impact of regional cultural diversity on rural workers' incomes. First, considering the possible gender differences in the labor market, it is interesting to examine whether the hump-shaped impact of regional cultural diversity varies by gender. We run regressions for male and female workers separately and report the results in Table 5. The results show that, for wage employment, the hump-shaped effect of diversity is stronger for females. For self-employment, there is not much difference in the effects of regional cultural diversity on the incomes by gender.

4.4 The Time Pattern of the Impact of Regional Cultural Diversity

Additionally, we examine the possible heterogeneous impacts by rural workers' educational levels. While there is a trade-off between the beneficial and the detrimental effects of diversity on incomes, we assume that the detrimental effects of diversity may be weaker for high-educated rural workers, because they rely less on social support in the labor market. We therefore divide the rural workers into low-educated and high-educated groups. The low-educated group comprises rural workers whose education level is lower than high school, whereas the high-educated group comprises those who have an education level beyond high school. We run regressions for the two groups separately and report the results in Table 6. The results show that hump-shaped effects of regional cultural diversity on incomes are significant for low-educated rural workers while non-significant for high-educated ones. However, the results should be interpreted with caution since the entire sample is relatively low educated, so the sample size of the high-educated rural workers is small. Furthermore, we examine the linear impact of diversity for high-educated rural workers and report the results in columns 3 and 6 of Table 6. Despite the more constrained sample size, regional cultural diversity has a significantly

Table 4 Impact of regional cultural diversity on rural workers' incomes: genetic diversity as an alternative measure of regional cultural diversity

	(1) Wage employment income	(2) Self- employment income
Genetic diversity	9.92** (4.77)	19.29** (9.80)
Genetic diversity squared	-8.35** (4.13)	-15.18* (8.22)
Age	0.02 (0.02)	0.08*** (0.02)
Age squared	-0.00 (0.00)	-0.00*** (0.00)
Gender	0.36*** (0.05)	0.44*** (0.08)
Marital status	0.31*** (0.10)	0.39** (0.17)
Education	0.12*** (0.01)	0.16*** (0.04)
Topographic slope	-0.04 (0.03)	-0.06** (0.03)
Topographic relief	-0.00 (0.00)	-0.00 (0.00)
North	-0.12* (0.06)	-0.35*** (0.13)
Coastal	0.33*** (0.08)	0.58*** (0.15)
GDP	0.17*** (0.04)	0.21*** (0.06)
Population density	-0.11* (0.06)	-0.25*** (0.09)
Constant	3.92** (1.69)	-0.37 (3.02)
Observations	1290	928
R-squared	0.221	0.229

OLS coefficients are reported

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

positive effect on the self-employment income of high-educated rural workers, which partially supports our assumption that the negative side of regional cultural diversity may be mitigated by education level.

Regional cultural diversity is inherently stable while the income of rural workers is relatively dynamic. This naturally raises an interesting question: what is the time pattern of the impact of regional cultural diversity? We expect that the impact of regional cultural diversity may remain similar in the short term. The impact could only change in the long term if

Table 5 Impact of regional cultural diversity on rural workers' incomes by gender

	(1)	(2)	(3)	(4)
	Wage employment income		Self-employment income	
	Male	Female	Male	Female
Dialect diversity	0.41 (0.29)	0.97*** (0.36)	1.17** (0.46)	1.42*** (0.51)
Dialect diversity squared	-0.95* (0.48)	-1.59*** (0.61)	-2.30*** (0.71)	-2.43*** (0.75)
Age	0.08*** (0.01)	0.02 (0.02)	0.08*** (0.02)	0.08*** (0.02)
Age squared	-0.00*** (0.00)	-0.00* (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
Marital status	0.26*** (0.06)	0.16** (0.08)	0.47*** (0.12)	0.32** (0.15)
Education	0.08*** (0.01)	0.11*** (0.01)	0.18*** (0.02)	0.23*** (0.03)
Topographic slope	-0.02 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.03)
Topographic relief	-0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
North	-0.24*** (0.04)	-0.25*** (0.06)	-0.29*** (0.07)	-0.24*** (0.08)
Coastal	0.25*** (0.04)	0.28*** (0.04)	0.29*** (0.09)	0.24** (0.11)
GDP	0.21*** (0.02)	0.27*** (0.03)	0.30*** (0.04)	0.28*** (0.05)
Population density	-0.07** (0.04)	-0.19*** (0.05)	-0.21*** (0.05)	-0.22*** (0.06)
Constant	5.21*** (0.37)	5.59*** (0.49)	4.03*** (0.77)	3.75*** (0.94)
Observations	2953	1996	1783	1422
R-squared	0.270	0.258	0.266	0.246

OLS coefficients are reported

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

the structure or the importance of cultural environment has gradually changed over time, or if there is a major shift in socioeconomic factors that condition the use of social capital in the labor market.

Ideally, to observe the time pattern of the impact of regional cultural diversity, we would like to know the incomes and related characteristics of rural workers in every year over a long period. However, so far, the CLDS has been carried out and released only once every two years over a short period (in 2012, 2014, and 2016). Moreover, the sample size of the CLDS 2012 is much smaller compared with the CLDS 2014 and

Table 6 Impact of regional cultural diversity on rural workers' incomes by education level

	(1)	(2)	(3)	(4)	(5)	(6)
	Wage employment income			Self-employment income		
	Low-educated	High-educated	High-educated	Low-educated	High-educated	High-educated
Dialect diversity	0.69*** (0.25)	0.26 (0.50)	-0.19 (0.12)	1.25*** (0.35)	-1.19 (1.68)	0.88** (0.42)
Dialect diversity squared	-1.31*** (0.41)	-0.79 (0.88)		-2.38*** (0.52)	3.78 (2.75)	
Age	0.07*** (0.01)	0.03*** (0.01)	0.03*** (0.01)	0.07*** (0.01)	0.21*** (0.06)	0.22*** (0.06)
Age squared	-0.00*** (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
Gender	0.49*** (0.03)	0.22*** (0.05)	0.22*** (0.05)	0.39*** (0.05)	0.10 (0.18)	0.09 (0.18)
Marital status	0.20*** (0.06)	0.14** (0.06)	0.14** (0.06)	0.42*** (0.10)	0.03 (0.27)	0.03 (0.27)
Education	0.15*** (0.02)	0.07*** (0.02)	0.06*** (0.02)	0.26*** (0.03)	0.17** (0.08)	0.18** (0.09)
Topographic slope	-0.03 (0.02)	-0.00 (0.03)	0.00 (0.03)	-0.03* (0.02)	0.09 (0.07)	0.07 (0.07)
Topographic relief	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
North	-0.25*** (0.04)	-0.23*** (0.07)	-0.22*** (0.07)	-0.26*** (0.05)	-0.36 (0.22)	-0.38* (0.22)
Coastal	0.26*** (0.03)	0.19*** (0.06)	0.19*** (0.05)	0.28*** (0.07)	-0.06 (0.27)	-0.02 (0.27)
GDP	0.24*** (0.02)	0.18*** (0.03)	0.19*** (0.03)	0.30*** (0.03)	0.18 (0.15)	0.11 (0.15)
Population density	-0.13*** (0.03)	-0.05 (0.06)	-0.05 (0.06)	-0.23*** (0.04)	0.04 (0.18)	0.10 (0.19)
Constant	4.90*** (0.33)	6.09*** (0.51)	5.94*** (0.50)	3.76*** (0.61)	1.94 (2.47)	2.56 (2.50)
Observations	4136	813	813	3052	153	153
R-squared	0.281	0.189	0.188	0.258	0.223	0.216

OLS coefficients are reported

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

2016.² Given these constraints, we focus on the common sample of rural workers in the CLDS 2014 and 2016 to observe the time pattern, at least in the short term.

² The number of the common sample of rural workers between the CLDS 2014 and 2016 is 2,609 for wage employment and 2,307 for self-employment. After merging data from the CLDS 2012, the number of the common sample of rural workers drops dramatically to 877 for wage employment and 755 for self-employment.

Table 7 Impact of regional cultural diversity on rural workers' incomes in 2014 and 2016

	(1)	(2)	(3)	(4)
	Wage employment income		Self-employment income	
	2014	2016	2014	2016
Dialect diversity	0.70* (0.37)	0.69* (0.37)	1.47*** (0.42)	1.04** (0.43)
Dialect diversity squared	-1.27** (0.61)	-1.24** (0.62)	-2.47*** (0.65)	-1.77*** (0.67)
Age	0.05*** (0.01)	0.06*** (0.01)	0.05*** (0.02)	0.04** (0.02)
Age squared	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
Gender	0.53*** (0.04)	0.47*** (0.04)	0.39*** (0.05)	0.40*** (0.05)
Marital status	0.33*** (0.08)	0.25*** (0.08)	0.50*** (0.11)	0.52*** (0.11)
Education	0.10*** (0.01)	0.12*** (0.01)	0.13*** (0.03)	0.16*** (0.02)
Topographic slope	-0.04 (0.02)	-0.04* (0.02)	-0.02 (0.02)	-0.04 (0.02)
Topographic relief	-0.00 (0.00)	-0.00 (0.00)	-0.00* (0.00)	-0.00 (0.00)
North	-0.39*** (0.05)	-0.41*** (0.05)	-0.25*** (0.06)	-0.25*** (0.06)
Coastal	0.27*** (0.05)	0.26*** (0.05)	0.19** (0.08)	0.22*** (0.08)
GDP	0.27*** (0.03)	0.25*** (0.03)	0.23*** (0.04)	0.30*** (0.04)
Population density	-0.19*** (0.05)	-0.17*** (0.05)	-0.15*** (0.05)	-0.20*** (0.05)
Constant	4.89*** (0.49)	4.96*** (0.50)	5.19*** (0.72)	4.18*** (0.76)
Optimal diversity	0.28	0.28	0.30	0.29
Observations	2609	2609	2307	2307
R-squared	0.291	0.302	0.217	0.241

Coefficients are reported

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

Table 7 reports the results. We do find that the effects of regional cultural diversity on rural workers' incomes are similar in the short term. Specifically, both in 2014 and 2016, regional cultural diversity has a significantly hump-shaped effect on rural workers' incomes. The optimal levels of diversity are also similar between 2014 and 2016. Moreover, the magnitudes of the cultural diversity effect change little over these two years: for wage employment, increasing the diversity of the least diverse city by one percentage point would raise rural workers' income by 0.69% in 2014 and by 0.68% in 2016, whereas

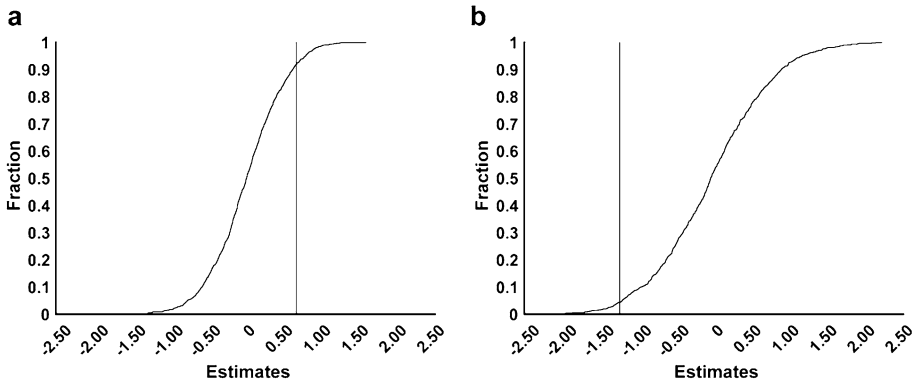


Fig. 4 Placebo test for the impact of regional cultural diversity on rural workers' wage employment income. *Note:* To estimate placebo effects, we first randomly assign the dialect diversity to the cities in our sample, and second regress the rural workers' wage employment income on the randomly assigned dialect diversity, dialect diversity squared, and relevant controls. We repeat these two steps 1000 times. The figure shows the CDF of 1000 placebo estimates of the dialect diversity and dialect diversity squared. The vertical line marks our true estimate. For details, see Sect. 4.5

decreasing the diversity of the most diverse city by one percentage point would raise the rural workers' income by 1.22% in 2014 and by 1.19% in 2016; for self-employment, increasing the diversity of the least diverse city by one percentage point would raise rural workers' income by 1.45% in 2014 and by 1.02% in 2016, whereas decreasing the diversity of the most diverse city by one percentage point would raise the rural workers' income by 2.28% in 2014 and by 1.64% in 2016.

4.5 Placebo Test

One concern is that the statistical significance of the hump-shaped effect of regional cultural diversity on rural workers' incomes may simply arise from some random factors. One way to address this concern is to examine the impact of placebo regional cultural diversity. We should not likely find estimates that are similar to our "true estimates" for the effect of regional cultural diversity when assigning random cultural diversity to the cities. Thus, we perform a Monte Carlo falsification test: First, we randomly assign the values of dialect diversity to the cities in our sample. Second, we estimate the effects of the randomly assigned dialect diversity with the regression specification used in columns 2 and 4 of Table 2. We repeat these two steps 1,000 times and plot the distribution of placebo estimates for dialect diversity and its squared term in Figs. 4 and 5, along with a vertical line that marks our true estimates. If there is a significant hump-shaped effect of dialect diversity on rural workers' incomes, we expect the true estimate to fall into the tails of the distribution of placebo estimates, i.e., among the 5–10 percent we are least likely to find. As Figs. 4 and 5 illustrate, our true estimates fall into the tails of the distribution of placebo estimates that we create (among the 5–10 percent most extreme estimates), which gives further credibility to our results.

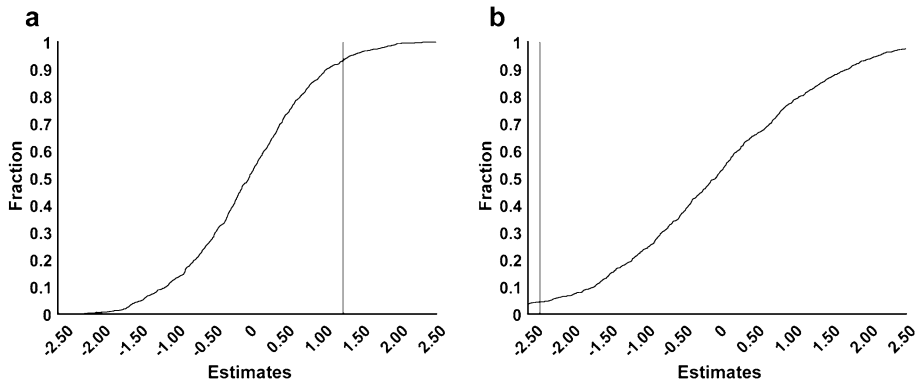


Fig. 5 Placebo test for the impact of regional cultural diversity on rural workers' self-employment income. *Note:* To estimate placebo effects, we first randomly assign the dialect diversity to the cities in our sample, and second regress the rural workers' self-employment income on the randomly assigned dialect diversity, dialect diversity squared, and relevant controls. We repeat these two steps 1000 times. The figure shows the CDF of 1000 placebo estimates of the dialect diversity and dialect diversity squared. The vertical line marks our true estimate. For details, see Sect. 4.5

4.6 Instrumental Results

In estimating the impact of regional cultural diversity on rural workers' incomes, we are confronted with possible endogeneity issues. Regional cultural diversity is likely associated with a complex myriad of factors. Despite our attempt to control for a variety of factors that may be correlated with both regional cultural diversity and rural workers' incomes, we are still unable to rule out the biases due to omitted variables. Another concern is the possible measurement error, because dialect diversity is but one dimension of regional cultural diversity.

To deal with these concerns, we employ an instrumental variable approach. We use the river length (in logarithmic form) of cities to instrument the dialect diversity. River length is likely to be correlated with dialect diversity because the distribution of rivers may shape the geographic pattern of dialects. Historically, on the one hand, rivers may cause geographic barriers that decreased population communication, and thus gradually formed diverse dialects. On the other hand, rivers had played a major role in the pattern of settlement, because people tended to settle along rivers, therefore rivers may create a network for communication and thus shaped dialects (Carver, 1986). In addition, river length may have limited effects on contemporary rural workers' incomes considering upgraded transportation. Therefore, we believe that river length could serve as a valid instrument. The data on river length of cities are derived from the National Geomatics Center of China.

The instrumented results are reported in Table 8 (with the first-stage results reported in the bottom panel).³ For both types of off-farm employment, the instrumented results confirm that dialect diversity has a significantly hump-shaped effect on rural workers' incomes. In comparison to their OLS counterparts in Table 2, the estimated 2SLS coefficients on dialect diversity and its squared term are substantially larger, suggesting that measurement error is likely the main source of endogeneity. The optimal level of regional cultural diversity is 0.26 for

³ The values of Kleibergen–Paap F-statistic are both above Staiger & Stock's (1997) rule-of-thumb value of 10, suggesting that the river length is by no means a weak instrument of dialect diversity.

Table 8 Impact of regional cultural diversity on rural workers' incomes: instrumental results

	(1)	(2)
	2SLS	2SLS
	Wage employment income	Self-employment income
Dialect diversity	3.30*** (1.18)	7.11** (3.44)
Dialect diversity squared	-6.38*** (2.13)	-11.30** (4.78)
Optimal diversity	0.26	0.31
Controls	Yes	Yes
Observations	4949	3205
<i>First stage results</i>		
Dependent variable is dialect diversity		
River length	-1.54*** (0.07)	-0.90*** (0.10)
River length squared	0.10*** (0.00)	0.06*** (0.01)
Dependent variable is dialect diversity squared		
River length	-0.90*** (0.05)	-0.55*** (0.07)
River length squared	0.06*** (0.00)	0.04*** (0.00)
Kleibergen-Paap rk	70.796	21.445
Wald F-stat		

Controls include age, age squared, gender, marital status, education, topographic slope, topographic relief, north, coastal, GDP, and population density of current residence

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

wage employment income and 0.31 for self-employment income, which are similar to base-line results. In terms of magnitude, for wage employment, a one percentage point increase in diversity for the least diverse city in the regression sample would raise rural workers' income by 3.29%, whereas a one percentage point decrease in diversity for the most diverse city would raise rural workers' income by 6.55%; for self-employment, a one percentage point increase in diversity for the least diverse city in the regression sample would raise rural workers' income by 7.25%, whereas a one percentage point decrease in diversity for the most diverse city would raise rural workers' income by 10.47%.

5 Testing Mechanisms

While it has been convincingly shown that regional cultural diversity has a hump-shaped effect on rural workers' incomes, we endeavor to identify the possible underlying mechanisms and provide empirical evidence for the cost and benefit of regional cultural diversity in this section.

5.1 Excluding the Communication Barrier Mechanism

One concern is that the cost of regional cultural diversity on labor market performance may simply arise from communication barriers caused by dialect differences, as regional cultural diversity is measured by dialect diversity in this paper. To address this concern, we first control for rural workers' Putonghua proficiency in the regressions. Putonghua is the official language for reading and writing in China, and Chinese people compulsorily learn Putonghua in elementary education. Therefore, most Chinese people can understand Putonghua, which has largely reduced communication barriers caused by dialect differences.⁴ As reported in columns 1 and 2 of Table 9, after controlling for rural workers' Putonghua proficiency, the hump-shaped effects of regional cultural diversity on rural workers' incomes remain significant.

To further separate the communication effects, we restrict the sample to Mandarin regions. Chinese dialects can be divided into Mandarin and non-Mandarin branches. Dialects within the non-Mandarin branch are heterogeneous and vary by language, but dialects in the Mandarin branch share many internal similarities. Chinese people have little difficulty in communication if their native languages belong to the Mandarin branch (Kurpaska, 2010; Tang, 2009). The results are reported in columns 3 and 4 of Table 9. Since Mandarin regions are also homogeneous in regional cultures, it is not surprising that the significant effect of dialect diversity on wage employment income disappears. However, for self-employment income, the hump-shaped effect of regional cultural diversity remains significant even in Mandarin regions. This further provides evidence that communication barriers cannot account for the effect of regional cultural diversity on rural workers' incomes.

5.2 The Cost and Benefit of Regional Cultural Diversity

As proposed, there exists a trade-off between the beneficial and the detrimental effects of regional cultural diversity on rural workers' incomes. According to previous studies, regional cultural diversity may confer cost on productivity, in the form of lower social capital. The benefit of regional cultural diversity, on the other hand, is that the diverse cultural values may expand ideas and improve creativity, thereby increasing rural workers' performance. The cost and benefit of regional cultural diversity will be tested in the rest of the section.

We first examine the impact of regional cultural diversity on the social capital of rural workers. We use three questions in the CLDS 2014 to measure social capital, including "the degree of trust to people in the community", "the degree of familiarity with people in the community", and "the frequency of mutual help in the community", all of which are scored on a five-point scale. Based on these three questions, we

⁴ Data on Putonghua proficiency can be obtained from the CLDS 2014.

Table 9 Excluding the communication barrier mechanism

	(1)	(2)	(3) Mandarin region	
	Wage employment income	Self-employment income	Wage employment income	Self-employment income
Dialect diversity	0.70*** (0.22)	1.23*** (0.34)	0.53 (0.33)	2.00*** (0.38)
Dialect diversity squared	-1.30*** (0.38)	-2.29*** (0.51)	-0.71 (0.52)	-3.01*** (0.55)
Putonghua proficiency	0.12*** (0.01)	0.14*** (0.02)		
Age	0.05*** (0.02)	0.08*** (0.01)	0.03 (0.02)	0.08*** (0.01)
Age squared	-0.00*** (0.00)	-0.00*** (0.00)	-0.00** (0.00)	-0.00*** (0.00)
Gender	0.45*** (0.03)	0.39*** (0.04)	0.42*** (0.04)	0.34*** (0.05)
Marital status	0.23*** (0.05)	0.37*** (0.09)	0.32*** (0.08)	0.43*** (0.10)
Education	0.08*** (0.01)	0.18*** (0.02)	0.11*** (0.01)	0.21*** (0.02)
Topographic slope	-0.03 (0.02)	-0.03* (0.02)	-0.06*** (0.02)	-0.09*** (0.02)
Topographic relief	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00* (0.00)
North	-0.23*** (0.03)	-0.27*** (0.05)	-0.20*** (0.05)	-0.15** (0.06)
Coastal	0.23*** (0.03)	0.26*** (0.07)	-0.14* (0.07)	-0.36*** (0.11)
GDP	0.20*** (0.02)	0.26*** (0.03)	0.32*** (0.03)	0.29*** (0.04)
Population density	-0.09*** (0.03)	-0.16*** (0.04)	-0.22*** (0.04)	-0.31*** (0.04)
Constant	4.97*** (0.34)	3.33*** (0.59)	4.68*** (0.52)	4.06*** (0.67)
Observations	4906	3191	2396	2275
R-squared	0.291	0.290	0.258	0.262

OLS coefficients are reported

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

construct three dummy variables, taking the value of one if the score is larger than three and zero otherwise.

We regress these three indicators of social capital on regional cultural diversity using Probit and IV-Probit. Table 10 reports the results. Columns 1, 3, and 5 report results of

Table 10 The cost of regional cultural diversity

	(1)	(2)	(3)	(4)	(5)	(6)
	Probit	IV-Probit	Probit	IV-Probit	Probit	IV-Probit
	Trust in the community		Familiarity in the community		Mutual help in the community	
Dialect diversity	-0.43*** (0.10)	-4.70*** (0.04)	-0.42*** (0.10)	-4.69*** (0.04)	-0.30*** (0.08)	-4.69*** (0.04)
Age	0.04*** (0.01)	-0.01 (0.01)	0.05*** (0.01)	-0.01 (0.01)	0.05*** (0.01)	-0.01 (0.01)
Age squared	-0.00** (0.00)	-0.00 (0.00)	-0.00*** (0.00)	-0.00 (0.00)	-0.00*** (0.00)	-0.00 (0.00)
Gender	0.09* (0.05)	0.05** (0.03)	0.08* (0.04)	0.05** (0.02)	0.02 (0.04)	0.05** (0.02)
Marital status	0.16** (0.07)	-0.06 (0.04)	-0.05 (0.07)	-0.07* (0.04)	0.04 (0.06)	-0.06 (0.04)
Education	0.05*** (0.01)	-0.03*** (0.01)	-0.05*** (0.01)	-0.03*** (0.01)	-0.00 (0.01)	-0.03*** (0.01)
Topographic slope	0.05** (0.02)	0.05*** (0.01)	0.09*** (0.02)	0.05*** (0.01)	0.11*** (0.02)	0.05*** (0.02)
Topographic relief	-0.00** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
North	0.37*** (0.06)	0.00 (0.05)	0.39*** (0.05)	-0.00 (0.04)	0.48*** (0.04)	0.00 (0.05)
Coastal	-0.15** (0.06)	-0.48*** (0.03)	-0.19*** (0.05)	-0.47*** (0.03)	-0.17*** (0.04)	-0.48*** (0.03)
GDP	-0.11*** (0.03)	0.00 (0.02)	-0.16*** (0.03)	-0.00 (0.02)	-0.11*** (0.03)	0.00 (0.02)
Population density	-0.01 (0.05)	-0.26*** (0.03)	-0.07 (0.05)	-0.26*** (0.03)	-0.09** (0.04)	-0.26*** (0.03)
Constant	2.06*** (0.51)	3.45*** (0.33)	3.33*** (0.46)	3.47*** (0.35)	2.16*** (0.40)	3.45*** (0.33)
Observations	7389	7389	7389	7389	7389	7389
Wald Chi2 exog test	-	3.19*	-	3.92**	-	3.40*
Log likelihood	-1807.578	-869.957	-2229.060	-1268.34	-3296.077	-2345.455

Coefficients are reported

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

Probit regressions and columns 2, 4, and 6 of IV-Probit regressions with river length used as instrument for the dialect diversity. As reported in Table 10, in both Probit and IV-Probit regressions, dialect diversity has a significantly negative effect on the three indicators of social capital, which is consistent with our hypothesis.

We then try to provide empirical evidence for the beneficial effects of regional cultural diversity. To find out whether regional cultural diversity expands ideas and improves the creativity of rural workers, we exploit the question "whether start business due to good opportunity" in the CLDS 2014, because capturing good business opportunities

Table 11 The benefit of regional cultural diversity

	(1)	(2)	(3)	(4)
	Probit	IV-Probit	Probit	IV-Probit
	Business opportunity		Entrepreneurship	
Dialect diversity	0.39** (0.18)	0.84 (2.23)	0.10* (0.06)	3.46*** (0.46)
Age	0.02 (0.03)	0.02 (0.03)	0.04*** (0.01)	0.04*** (0.01)
Age squared	-0.00 (0.00)	-0.00 (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
Gender	0.05 (0.08)	0.04 (0.08)	0.10*** (0.03)	0.05** (0.03)
Marital status	-0.23 (0.17)	-0.25 (0.18)	0.23*** (0.05)	0.21*** (0.04)
Education	0.10*** (0.03)	0.10*** (0.03)	-0.03*** (0.01)	0.01 (0.01)
Topographic slope	0.15*** (0.04)	0.15*** (0.05)	0.02 (0.01)	-0.02* (0.01)
Topographic relief	-0.00*** (0.00)	-0.00*** (0.00)	-0.00 (0.00)	0.00*** (0.00)
North	-0.06 (0.09)	-0.04 (0.13)	0.16*** (0.03)	0.23*** (0.03)
Coastal	-0.04 (0.10)	-0.00 (0.21)	-0.26*** (0.04)	0.21** (0.09)
GDP	0.17*** (0.06)	0.17*** (0.06)	-0.02 (0.02)	-0.05*** (0.02)
Population density	-0.07 (0.09)	-0.03 (0.22)	0.05** (0.03)	0.23*** (0.03)
Constant	-2.83*** (0.99)	-3.28 (2.38)	-1.86*** (0.31)	-3.44*** (0.29)
Observations	1106	1106	10,746	10,746
Wald Chi2 exog test	-	0.04	-	18.45***
Log likelihood	-735.709	-579.197	-6357.827	-5071.117

Coefficients are reported

*Significant at 10%; **significant at 5%; ***significant at 1%; robust standard error in parentheses

particularly requires an open mind and creativity. We also examine the impact of regional cultural diversity on entrepreneurship, which is defined as a dummy variable indicating whether the farmer is self-employed, as it is closely related to new ideas and creativity.

We regress these two proxies of creativity on regional cultural diversity using Probit and IV-Probit. Table 11 reports the results. Columns 1 and 3 report results of Probit regressions and columns 2 and 4 of IV-Probit regressions with river length used as instrument for the dialect diversity. First, we find that dialect diversity has a significantly positive effect on business opportunity in the Probit specification, while the IV-Probit estimate becomes less precise but remains positive. However, the Wald test of exogeneity for the IV-Probit

model is insignificant, suggesting that we cannot reject the hypothesis of no endogeneity and thus the regular Probit regression may be more appropriate. Likewise, dialect diversity has a significantly positive effect on entrepreneurship in both Probit and IV-Probit regressions. Together, these results provide evidence for the beneficial side of regional cultural diversity.

6 Conclusions

This paper studies the impact of regional cultural diversity of working cities on the incomes of rural workers. Taking China as an example, our results show that regional cultural diversity has a highly significant hump-shaped effect on the two types of off-farm incomes (wage employment income and self-employment income) of rural workers. The results are robust to the inclusion of a set of covariates, alternative estimation technique, alternative measure of regional cultural diversity, and a placebo test. We also find that the detrimental effect of diversity is weaker for high-educated rural workers, and the impact of regional cultural diversity remains stable in the short term. Moreover, we address the potential endogeneity by employing the river length of cities to instrument the dialect diversity, and the instrumented results reaffirm the conclusions. Finally, we explore the underlying mechanisms of the hump-shaped effect of regional cultural diversity, and provide empirical evidence that regional cultural diversity has a significantly negative impact on rural workers' social capital and a significantly positive effect on rural workers' creativity.

Our study contributes to the rich body of literature on the determinants of incomes of farmers by emphasizing the role of regional cultural diversity. Our results suggest a trade-off between the beneficial and the detrimental effects of regional cultural diversity on the incomes of rural workers. The results in this paper have significant implications for the understanding of how regional cultural diversity of working cities affects the performance of rural workers, especially within a developing country.

The conclusions in this paper provide possible policy implications. If regional cultural diversity matters for rural workers, policies should strengthen the benefit and mitigate the cost of regional cultural diversity. For instance, efforts could be made to promote interpersonal communication, or improve cultural assimilation in the working cities for rural workers. Efforts could also be made to reduce rural workers' reliance on social capital in the labor market, such as increasing their education level and skill level.

Our study has limitations. First, although we have tried to theoretically and empirically test the mechanisms through which regional cultural diversity works, we construct limited measures of these mechanisms due to the lack of data. Second, we do not explore the time pattern of the impact of regional cultural diversity over a long period because of data limitations. Finally, we do not investigate the heterogeneity of the effects of regional cultural diversity across different skill groups of rural workers, which may be helpful to gain a better understanding of the mechanisms of regional cultural diversity. These limitations can be overcome in further research.

Funding This work was supported by the National Natural Science Foundation of China [Grant number 71873121, 72173060]; the Ministry of Education of China [Grant number 20YJC790138].

Data Availability and Materials The datasets generated or analysed during the current study are available from the corresponding author on reasonable request.

Code Availability The code generated or used during the current study is available from the corresponding author on reasonable request.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

- Alesina, A., & Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3), 762–800.
- Alesina, A., & La Ferrara, E. (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115(3), 847–904.
- Ashraf, Q., & Galor, O. (2013a). Genetic diversity and the origins of cultural fragmentation. *American Economic Review*, 103(3), 528–533.
- Ashraf, Q., & Galor, O. (2013b). The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1), 1–46.
- Bellini, E., Ottaviano, G. I., Pinelli, D., & Prarolo, G. (2013). Cultural diversity and economic performance: Evidence from European regions. *Geography, institutions and regional economic performance* (pp. 121–141). Springer.
- Brown, H. D. (1980). *Principles of language learning and teaching*. Prentice Hall Regents.
- Carver, C. (1986). The influence of the Mississippi River on northern dialect boundaries. *American Speech*, 61(3), 245–261.
- Cavalli-Sforza, L. L. (1997). Genes, peoples, and languages. *Proceedings of the National Academy of Sciences*, 94(15), 7719–7724.
- Cavalli-Sforza, L. L. (2005). The human genome diversity project: Past, present and future. *Nature Reviews Genetics*, 6(4), 333–340.
- Cavalli-Sforza, L. L., Cavalli-Sforza, L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton University Press.
- Cherdchuchai, S., & Otsuka, K. (2006). Rural income dynamics and poverty reduction in Thai villages from 1987 to 2004. *Agricultural Economics*, 35, 409–423.
- China Labor-force Dynamic Survey. (2012). *Center for Social Survey of Sun Yat-Sen University*. <http://css.sysu.edu.cn>.
- China Labor-force Dynamic Survey. (2014). *Center for Social Survey of Sun Yat-Sen University*. <http://css.sysu.edu.cn>.
- China Labor-force Dynamic Survey. (2016). *Center for Social Survey of Sun Yat-Sen University*. <http://css.sysu.edu.cn>.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95–S120.
- Corral, L., & Reardon, T. (2001). Rural nonfarm incomes in Nicaragua. *World Development*, 29(3), 427–442.
- Dale-Olsen, H., & Finseraas, H. (2020). Linguistic diversity and workplace productivity. *Labour Economics*, 64, 101813.
- Desmet, K., Ortuno-Ortín, I., & Wacziarg, R. (2017). Culture, ethnicity, and diversity. *American Economic Review*, 107(9), 2479–2513.
- DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology*, 23(1), 263–287.
- Fei, H.-T., & Malinowski, B. (2013). *Peasant life in China*. Hesperides Press.
- Greenberg, J. H. (1956). The measurement of linguistic diversity. *Language*, 32(1), 109–115.
- Haggblade, S., Hazell, P., & Reardon, T. (2010). The rural non-farm economy: Prospects for growth and poverty reduction. *World Development*, 38(10), 1429–1441.
- Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 129(4), 1899–1946.
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46), 16385–16389.
- Ibekwe, U., Eze, C., Onyemauwa, C., Henri-Ukoha, A., Korie, O., & Nwaiwu, I. (2010). Determinants of farm and off-farm income among farm households In South East Nigeria. *Academia Arena*, 2(10), 58–61.
- Jiang, W. (2000). The relationship between culture and language. *ELT Journal*, 54(4), 328–334.

- Josten, S. D. (2013). Middle-class consensus, social capital and the fundamental causes of economic growth and development. *Journal of Economic Development*, 38(1), 1–26.
- Kemeny, T. (2012). Cultural diversity, institutions, and urban economic performance. *Environment and Planning A*, 44(9), 2134–2152.
- Knight, J., & Yueh, L. (2004). Job mobility of residents and migrants in urban China. *Journal of Comparative Economics*, 32(4), 637–660.
- Koopmans, R., & Schaeffer, M. (2016). Statistical and perceived diversity and their impacts on neighborhood social cohesion in Germany, France and the Netherlands. *Social Indicators Research*, 125(3), 853–883.
- Kurpaska, M. (2010). *Chinese language (s)*. De Gruyter Mouton.
- Lazear, E. P. (1999). Globalisation and the market for team-mates. *The Economic Journal*, 109(454), 15–40.
- Lee, N. (2015). Migrant and ethnic diversity, cities and innovation: Firm effects or city effects? *Journal of Economic Geography*, 15(4), 769–796.
- Liu, Z. (2005). Institution and inequality: The hukou system in China. *Journal of Comparative Economics*, 33(1), 133–157.
- Luan, J., Chen, J.-C., He, Z.-W., Li, Q., & Qiu, H. (2015). The education treatment effect on the non-farm income of Chinese western rural labors. *China Agricultural Economic Review*, 7(1), 122–142.
- Narayan, D., & Pritchett, L. (1999). Cents and sociability: Household income and social capital in rural Tanzania. *Economic Development and Cultural Change*, 47(4), 871–897.
- Niebuhr, A. (2010). Migration and innovation: Does cultural diversity matter for regional R&D activity? *Papers in Regional Science*, 89(3), 563–585.
- Olsson, O., & Paik, C. (2016). Long-run cultural divergence: Evidence from the neolithic revolution. *Journal of Development Economics*, 122, 197–213.
- Otsuka, K., & Yamano, T. (2006). Introduction to the special issue on the role of nonfarm income in poverty reduction: Evidence from Asia and East Africa. *Agricultural Economics*, 35(s3), 393–397.
- Ottaviano, G. I., & Peri, G. (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography*, 6(1), 9–44.
- Parrotta, P., Pozzoli, D., & Pytlíkova, M. (2014). Labor diversity and firm productivity. *European Economic Review*, 66, 144–179.
- Putnam, R. D. (2007). E Pluribus Unum: Diversity and community in the twenty-first century. *Scandinavian Political Studies*, 30(2), 137–174.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.
- Spolaore, E., & Wacziarg, R. (2016). Ancestry, language and culture. *The Palgrave handbook of economics and language* (pp. 174–211). Springer.
- Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3), 557–586.
- Suedekum, J., Wolf, K., & Blien, U. (2014). Cultural diversity and local labour markets. *Regional Studies*, 48(1), 173–191.
- Sumner, D. A. (1982). The off-farm labor supply of farmers. *American Journal of Agricultural Economics*, 64(3), 499–509.
- Sun, R., Mi, J., Cao, S., & Zhang, T. (2021). Trends and determinants of income-oriented livelihood transitions in rural China. *Social Indicators Research*, 155(2), 601–624.
- Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33(1), 1–39.
- Tang, C. (2009). *Mutual intelligibility of Chinese dialects: An experimental approach*. Leiden University.
- Thiede, B. C., Lichter, D. T., & Slack, T. (2018). Working, but poor: The good life in rural America? *Journal of Rural Studies*, 59, 183–193.
- Thomas, S. L., & Zhang, L. (2005). Post-baccalaureate wage growth within four years of graduation: The effects of college quality and college major. *Research in Higher Education*, 46(4), 437–459.
- Wang, C., Zhou, X., & Zhang, C. (2015). The impact of social capital on wages of rural migrants and its gender difference in China. *Journal of Macromarketing*, 35(2), 202–217.
- Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1), 52–69.
- Xu, X., Liu, Y., & Xiao, Z. (2015). Dialect and economic growth. *China Journal of Economics*, 2(2), 1–32.
- Yang, L., Lu, H., Wang, S., & Li, M. (2021). Mobile internet use and multidimensional poverty: Evidence from a household survey in rural China. *Social Indicators Research*, 158(3), 1065–1086.
- Zhao, T., & Lee, T. D. (1989). Gm and Km allotypes in 74 Chinese populations: A hypothesis of the origin of the Chinese nation. *Human Genetics*, 83(2), 101–110.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.