Income Inequality in Chinese Provinces: The Role of Human Capital

Kristina Butaeva[a], Albert F. Park[b,c]

[a]Harris School of Public Policy Stone Center for Research on Wealth Inequality and Mobility, University of Chicago, 1307 E 60th St, Chicago, 60637, IL, USA

[b]Asian Development Bank (ADB), 6 ADB Avenue, Mandaluyong City, Metro Manila, Philippines

[c]Hong Kong University of Science and Technology (HKUST), Clear Water Bay, Hong Kong, China

ARTICLE INFO

Keywords:
Income Inequality
Provinces
China
Human Capital

ABSTRACT

We conduct the first systematic empirical analysis of income inequality in China at the provincial level. Using data from the China Household Finance Survey (CHFS) and a semiparametric distribution model, we estimate Gini indices for Chinese provinces in 2012, 2014, 2016, and 2018. We find that differences in the "prices" and "quantities" of human capital are strongly associated with differences in inequality across provinces. Our findings suggest that poor provinces are highly disadvantaged compared to rich provinces, as they face higher income and educational inequality, as well as a higher premium for completing high school, while at the same time exhibiting lower average educational attainment. We conclude that the reduction of existing interprovincial human capital gaps and the acceleration of labor market integration through appropriate government policies could be associated with lower spatial disparities in inequality levels across regions and lower overall income inequality in China.

1. Introduction

Over the past four decades, the Chinese economy has experienced remarkable economic growth. At the same time, the level of spatial and individual income inequality has risen dramatically, and it seems that the fruits of economic success are still unevenly distributed.[1] Individual and household income inequality in China, as well as different types of spatial economic inequalities (rural-urban, regional, provincial, etc.), are well-studied in the existing literature.[2] However, to date there has been no systematic study of the nature of within-province income inequality and its variation across Chinese provinces.

Why is an analysis of provincial inequality important? First, intraprovincial income inequality substantially contributes to national-level inequality. The decomposition analysis from our study suggests that around 90% of the nationwide variation in annual household income per capita in 2012-2018 can be explained by intraprovincial income variation. Therefore, the analysis of factors associated with intraprovincial income inequality, coupled with the estimation of each province's contribution to nationwide inequality, has the potential to generate insights that can help inform policies aimed at reducing nationwide income inequality. Second, income inequality is an important dimension of cross-regional economic disparities, though prior analyses have focused on the comparisons of GRP per capita or mean disposable income. Third, provincial income inequality has the capacity to impede provincial and national economic growth and influence social attitudes that underpin political legitimacy. The negative impact of high income inequality on growth is well-documented in the literature. For example, high income inequality can create inequality

⋆We are grateful to Grigory Kosenok, Shlomo Weber, Steven N. Durlauf, Li Gan, Jin Wang, Li Han, and Sujata Visaria for their insightful suggestions and valuable feedback. We also thank participants of the Applied Micro Seminar at the Hong Kong University of Science and Technology (HKUST) for their helpful comments. Butaeva gratefully acknowledges financial support from the James M. and Cathleen D. Stone Foundation and the Greater Eurasia Research Center (GEAR) at HKUST. She also thanks Grace Kolavo and Joseph P. Quinn for their invaluable assistance in preparing the manuscript.

*Corresponding author

kbutaeva@uchicago.edu (K. Butaeva)

ORCID(s): 0000-0001-5865-577X (K. Butaeva); 0009-0007-3323-120X (A.F. Park)

[1]See, e.g., Naughton (2006), Ravallion and Chen (2009), Kanbur, Wang and Zhang (2021), Jain-Chandra, Khor, Mano, Schauer, Wingender and Zhuang (2018), Gradín and Wu (2020).

[2]Zhang (2021) and Kanbur et al. (2021) provide recent comprehensive reviews of the literature on China's inequality trends. A comprehensive overview of the trends in the Gini index in China is presented in Xie and Zhou (2014). A review of existing studies on inequality in China employing top-income corrections is presented in Li, Li and Wan (2020), whereas spatial inequality studies are described in Wang, Wan and Yang (2014).

of opportunities which limits the human capital development of disadvantaged social groups (Perotti, 1996; Galor and Moav, 2006). The literature has also found that high income inequality reduces social cohesion which can reduce political stability and adversely impact investment and growth.[3] Political instability creates uncertainty, which could adversely impact investments and economic growth (Alesina and Perotti, 1996; Esteban and Ray, 1999). Moreover, high income inequality may facilitate lobbying on behalf of financial elites. This could hurt political institutions, create inefficiency, and impede sustainable economic growth (Savoia, Easaw and McKay, 2010; Chong and Gradstein, 2017). Finally, studying the factors associated with provincial inequality will produce insights into the efficiency of educational investments and the performance of local labor markets.

In this paper, we aim to fill the gap in the existing literature on subnational-level income inequality in China. Our purpose is four-fold. First, we provide Gini index estimates for 29 mainland Chinese provinces (excluding Tibet and Xinjiang), applying a new robust measurement approach developed in Butaeva (2025) to the data from four waves of the China Household Finance Survey (CHFS 2013, 2015, 2017, and 2019). Second, we identify the main factors that account for the variation in income inequality across Chinese provinces, by examining the Mincer (1974) model. The Mincer model posits that education and the returns to education are key determinants of earnings, and therefore we assume that higher inequality in educational attainment and higher returns to education result in higher income inequality. These two human capital factors are policy-relevant because they reflect local educational and labor market institutions. Many economic studies have noted substantial increases in educational attainment, including a remarkable expansion of higher education in China over the last several decades.[4] At the same time, previous studies have found that under a decentralized fiscal system, significant progress in educational development has advanced in the richer coastal provinces and large cities more than other regions (e.g., Xiang, Stillwell, Burns and Heppenstall, 2020).

Many scholars have also documented the positive effects of Chinese institutional market reforms which encourage labor market liberalization. For example, Cai, Park and Zhao (2008) emphasizes that internal migration in China has increased substantially over time, helping to promote local labor market integration, which has already led to progressive convergence of returns to education across regions over time. Significant progress in labor migration is also outlined by Cai et al. (2008). However, despite the improvements, some serious challenges to the full integration of the labor market in China remain (Fleisher and Yang, 2008; Heckman, 2005; Heckman and Yi, 2012; Tombe and Zhu, 2019). In particular, the *hukou*[5] system still limits urbanization and industrialization in China (Ngai, Pissarides and Wang, 2019). Therefore, our third goal is to contribute to the discussion of the unequal rewards of additional schooling across Chinese provinces. Unrestricted mobility implies free flows of skilled and unskilled labor among different locations which should equalize the returns to education across provinces. However, we continue to observe differences in returns to education across provinces, which suggests that despite the significant progress in labor market integration in China there still exist barriers to labor migration. Finally, we evaluate the potential effects of narrowing gaps associated with educational disparities and incomplete labor market integration. To that end, we conduct a simulation analysis that examines how provincial and national income inequality would change in response to: (a) equalization of educational structure across provinces; (b) equalization of returns to education; and (c) both an equalization of educational structure and returns to education.

Previewing the main findings, in this study we document substantial disparities in levels of provincial income inequality across China. In 2018, the Gini index in the most unequal province, Sichuan, was 1.6 times larger than in the least unequal, Shanghai (0.665 versus 0.421, respectively). We also find that poor provinces – those with low GRP per capita – are severely disadvantaged compared to rich provinces. They not only have lower average levels of education and higher inequality in education but also face greater income inequality. Existing human capital disparities and differences in returns to education are both strongly associated with the variation in income inequality across Chinese provinces. In turn, the factors associated with inequities in the returns to education include variation in the relative supply of more skilled workers, the size of the state sector, and the degree of industrial structure and trade openness. We also find that current levels of interprovincial migration are not large enough to shift regional variation in human capital structure to alter the extent of interprovincial differences in income inequality given plausible estimates of the returns to education. Empirically, these findings are consistent with Heckman (2005)'s thesis that human capital is closely linked to regional economic development and inequality in China. Our simulation analysis reveals that keeping

[3] Previous studies find that a notion of fairness is positively related to social tolerance towards income inequality (Lü, 2012). Florida, Mellander and Rentfrow (2013) point out that human capital plays a key role in the happiness of a city's residents.

[4] See, e.g., Chan and Ngok (2011), Lei and Shen (2015), Mendoza (2018).

[5] *Hukou* is a household registration system in mainland China introduced in the 1950s, which imposes formal restrictions on changes of permanent residence.

all other factors constant, equalization of educational structure across provinces is associated with a 10.8% decrease in nationwide income inequality. In contrast, equalization of returns to high school alone is associated with a small 1.7% increase in nationwide income inequality. Combining equalization of educational structure with equalization of returns to high school lowers nationwide income inequality by 8.9%. These findings highlight the potential relevance of regional educational policy and labor market integration in understanding interprovincial income disparities in China.

The paper is organized as follows. In Section 2, we review the main results of previous studies on income and human capital disparities in China. In Section 3, we describe the data. In Section 4, we estimate provincial Gini indices in China, applying the semiparametric method introduced in Butaeva (2025) to the data from CHFS. In Section 5, we test different patterns of nationwide income inequality in China, applying decomposition methods: (a) we decompose the country-level Theil T income inequality index by province, analyzing the relative contribution of inter- and intraprovincial variation in household income per capita to the nationwide variation in income; and (b) we perform a regression-based decomposition by applying the classical Mincer model. In Section 6, we estimate provincial educational Gini indices and provincial returns to education. In Section 7, we study the key factors associated with interprovincial variation in income inequality. We regress the provincial Gini indices on provincial educational Gini indices, provincial returns to education, and their interaction term to evaluate the role of human capital in the variation of income inequality across Chinese provinces. We also regress our provincial returns to education on different provincial labor supply and labor demand factors to identify the key factors associated with existing interprovincial disparities in returns to education. Finally, in Section 8, we conduct the simulation analysis described above. In Section 9, we conclude.

2. Related literature

Provincial income inequality in China

The uneven impact of economic development since the beginning of market reforms in China has attracted the attention of many scholars. Current interprovincial economic inequality is considerable. In 2018, the richest province (Beijing) had 4.7 times higher GRP per capita than the poorest province (Gansu).[6] The coefficient of variation in GRP per capita among provinces is around 0.44-0.46. For comparison, the ratio of rich to poor GRP per capita among the US states in 2016 was 2.8, and the coefficient of variation was 0.18 (Fraumeni, He, Li and Liu, 2019). Nevertheless, Hao, Sun, Tombe and Zhu (2020) show that between 2000 and 2015, the variance in the cross-province (log) GDP per worker fell by one-third, which they claim was mostly explained by the increase in the internal migration.

Due to data limitations, studies focusing specifically on provincial-level inequality remain scarce (existing estimates of provincial-level Gini indices are summarized in Table A1, Appendix A). The earliest studies used survey data from the Chinese Academy of Social Sciences (CASS) (Hussain, Lanjouw and Stern, 1994; Khan and Riskin, 2005). Hussain et al. (1994) estimate Gini indices for most Chinese provinces in 1986 separately for rural and urban populations. They find that intraprovincial income inequality was the main contributing factor to nationwide income inequality (explaining 94% of (log) income variance for urban households and 85% for rural households). They also find that, unlike most other developing countries, income inequality in China's urban areas appeared to be lower than in rural areas. Analyzing provincial income inequality in 1995 and 2002, Khan and Riskin (2005) find a decline in both urban and rural income inequality in most provinces and an increase in the rural-urban gap. Employing the Survey on Urban Household Income, Xu and Zou (2000) study the potential factors explaining provincial income inequality dynamics between 1985 and 1995. They find urban inequality increases with a reduction in the share of state-owned enterprises (SOEs) in GDP, with an increase in inflation and regional growth, and (less significantly) with greater exposure to foreign trade. At the same time, no statistically significant impact of schooling or urbanization was found. Using Urban Household Survey data for 1990-2009, Nie and Xing (2019) find the rise of urban income inequality is mainly attributed to the increase in returns to education. Nonetheless, the fragmented evidence from the existing studies on provincial-level income inequality in terms of their timing and geographic scope does not substitute for a comprehensive analysis of provincial income inequality in China. Our study aims to fill that gap by providing robust Gini index estimates for 29 mainland provinces (excluding Tibet and Xinjiang) for the years 2012, 2014, 2016, and 2018, and, furthermore, by identifying the key factors that are related to inter- and intraprovincial income inequality in China.

[6] Provincial GRP per capita for 2012, 2014, 2016, and 2018 is reported in Table F1, Appendix F.

Educational inequality in China

In this study, we find educational inequality and returns to education to be the main factors statistically associated with the current variation in provincial income inequality levels in China. According to human capital theory, education is a fundamental determinant of earnings. Therefore, disparities in education could considerably widen disparities in income. In a multi-country study, Coady and Dizioli (2018) found a large, positive, and statistically significant relationship between inequality of schooling and income inequality, especially in emerging and developing countries.[7] Educational disparities in China have been studied quite extensively. One must note the remarkable progress made in furthering educational attainment since the launch of the government's higher education expansion policy in 1999. According to the Educational Statistical Yearbook of China, enrollment in higher education (undergraduate and graduate study, excluding adult higher education) increased eight times from 3.6 million individuals in 1998 to 28.9 million in 2016. However, recent studies continue to provide new evidence on educational inequalities.[8] Based on the 2010 China Family Panel Studies (CFPS) data, Lei and Shen (2015) underscore the widening gap in years of schooling between rural and urban residents. Introducing a new multidimensional index, Xiang et al. (2020) show that the northeastern regions of China provide better education compared to the southwest part of the country. Yang, Huang and Liu (2014) estimate the educational Gini index in each province for several years from 1996 to 2008 and find that intraprovincial educational inequality has declined but regional variation remains large. Estimating provincial-level educational Gini from the 2010-2012 CFPS data, Golley and Kong (2018) find the lack of equality in educational opportunities is primarily due to the individual's *hukou* status. Yu, Yu, de Jong and Storm (2015) and Peng, Qi, Wan, Li and Hu (2020) contribute to the provincial educational Gini discussion by considering the regional economic development levels. Yu et al. (2015) show that the relationship between inequality in educational attainment and growth in China depends on the level of economic development of a particular area. More specifically, they claim that educational inequality has a larger impact on growth than educational attainment, and this effect is stronger in less developed western provinces than in more prosperous eastern provinces. Peng et al. (2020) conclude that poor provinces not only encounter lower educational resources but also experience a greater demand for education because they have more children to educate. Fleisher, Li and Zhao (2010) demonstrate that provincial human capital, particularly in the form of higher education, has a positive effect on Total Factor Productivity (TFP) growth in China. Their findings support policies aimed at reducing regional educational inequality to mitigate economic disparities. In summary, there exists an extensive literature linking sub-national educational inequality and economic disparities, as well as studies linking educational and income inequality in China (e.g., Sicular, Yue, Gustafsson and Li, 2007; Chan and Ngok, 2011; Gustafsson and Wan, 2020). However, there are no provincial-level analyses of the combined impact of interrelated economic, educational, and income disparities. Our paper contributes to this literature by providing evidence that in China poor provinces exhibit higher levels of both income and educational inequality as compared to rich provinces.

Returns to education in China

Finally, our study provides an analysis of provincial differences in returns to education in China. We are not the first to find that returns to education per se, as well as in combination with educational inequality, are significant explanatory factors for provincial income inequality variation in China. However, we also seek to explain provincial differences with respect to returns to education itself. To evaluate provincial returns to education we start with the classical Mincer regression model, which has been used in many studies exploring differences in returns to education in China.[9] A recent paper by Asadullah and Xiao (2020) uses two waves (2010 and 2015) of the China General Social Survey (CGSS) to confirm the importance of the human capital factor in the Mincerian earnings equation in China. They also document the dynamics of greater declines in returns to schooling in urban locations, coastal regions, and among women. At the same time, Li, Whalley and Xing (2014) document sharp increases in unemployment among young college graduates in response to China's educational expansion policy predominantly in non-coastal regions. The authors suggest that encouraging interregional mobility in China could help to better match local labor supply and demand. Overall, most existing studies attribute differences in returns to schooling across China to remaining labor market segmentation and labor mobility restrictions. Many scholars claim that the notorious *hukou* system, first introduced in 1951, which has become significantly less impactful in recent decades, still affects people's educational and employment opportunities, economic and social mobility, and labor market returns to education. Employing the data from the 2005 (1%) National

[7] The paper also provides a theoretical framework and comprehensive literature review of the topic.

[8] For e.g., Heckman and Yi (2012) emphasize a substantial inequality in access to higher education in China.

[9] Zhang, Zhao, Park and Song (2005) provide a comprehensive review of existing estimates of returns to schooling in China. Churchill and Mishra (2018) also present a broad literature review on the topic.

Population Sample Survey, Fu and Ren (2010) demonstrate that *hukou* status still considerably shapes people's labor-market returns: an individual's human capital is largely determined by their birthplace (rural versus urban). A recent study by Ngai et al. (2019) also finds that the *hukou* system alters urbanization and industrialization in China. In addition to formal residential mobility restrictions, researchers also outline the presence of informal discrimination against labor migrants (Park, Song, Zhang and Zhao, 2003).

3. Data

We use household data from 29 mainland Chinese provinces (excluding Tibet and Xinjiang) collected in four waves of the China Household Finance Survey (CHFS) conducted in 2013, 2015, 2017, and 2019 by the Research Institute of Economics and Management at Southwestern University of Finance and Economics. The CHFS employs stratified multi-stage random sampling and contains micro-level information on household demographics and finance. The CHFS sample sizes are reported in Table C1, Appendix C.[10] All four aforementioned survey waves are nationally and provincially representative.[11]

In this paper, we use annual household disposable income per capita to measure income inequality. Household disposable income in the CHFS includes income generated from labor, business, assets, transfers, and other sources received during the last twelve months.[12] Labor income includes total after-tax monetary wages and salaries, bonuses, and subsidies from employers (both monetary and in-kind). Business income is the sum of net income from agricultural production and business operations. Asset income includes net after-tax income from non-financial assets (rental income from lands, properties, and vehicles) and income from financial assets (interest on deposits, stocks, bonds, dividends from commercial life insurances, etc.). Transfer and other income include retirement benefits and pensions; unemployment insurance; other government compensation and subsidies; private transfers; net after-tax income from all other sources including lotteries, sale of properties, vehicles, intellectual property rights, etc.).

In the CHFS, some missing income values were recorded as zeros. Therefore, in our study, we exclude households with zero income to avoid potential measurement bias. Negative incomes are also excluded from the analysis. The restriction to positive income values reduces the sample size across waves by 1-2.5%. All provincial-level estimates of our key variables, as well as the individual-level regression results reported in Table 3, are derived using survey sampling weights. It should be noted that CHFS contains a larger sample size of those living in high-income neighborhoods compared to other existing Chinese household surveys. In Tables C2 and C3 in the Appendix C we report summary statistics for all household- and provincial-level variables used in the current study. Figure C1 illustrates the provincial-level sample sizes.

4. Provincial-level income inequality

We begin by reporting provincial Gini index estimates derived using a new methodological approach introduced in Butaeva (2025). The method is robust to a well-known income inequality measurement issue – the high sensitivity of the Gini index to the presence of top income outliers.[13] On the one hand, sparse and incomplete top income data yield a high variance of Gini estimates across different years, because any random change in several top incomes may cause a sizable shift in inequality estimates for the whole geographic area under consideration. On the other hand, we can't just ignore or winsorize high income observations because they represent an essential element of income inequality analysis. The problem is especially concerning for small samples, which is relevant for the current study because the CHFS provincial samples range from 400 to 3,000 observations in different survey years. The method we apply yields a semiparametric probability density function (PDF):

[10] For more details on the CHFS dataset, see `https://chfs.swufe.edu.cn/sjzx/sjsq.htm`. A comprehensive review of the survey and key statistics of the data are presented in Gan, Yin, Jia, Xu, Ma, Zheng et al. (2014).

[11] Other household surveys with income information are available for China. Appendix B provides a comparative overview of available years, sample sizes, representativeness, sampling design, and income definitions for the Chinese General Social Survey (CGSS), China Household Income Project (CHIP), China Family Panel Studies (CFPS), and CHFS. As reflected in the Table B1, none of the other existing surveys is designed to be representative at the provincial level. In contrast, CHFS was explicitly constructed to achieve provincial representativeness, offering a unique opportunity for provincial-level analysis of income inequality in China.

[12] Therefore, the inequality estimates are for the year preceding each survey year.

[13] See, e.g., Cowell and Flachaire (2007).

$$f(x) = \begin{cases} p_\tau \times \left[\frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x_i - x}{h} \right) \right], & \text{if } x \leq \tau \\ (1 - p_\tau) \times \left[\frac{1}{\sigma} \left(1 + \xi \frac{x - \tau}{\sigma} \right)^{-\frac{1}{\xi} - 1} \right], & \text{if } x > \tau. \end{cases} \tag{1}$$

where τ is a threshold income, and p_τ is a corresponding threshold income percentile. The bottom p_τ% of incomes are represented by the nonparametric kernel density, where N is the number of observations, h is a smoothing parameter (bandwidth), and $K(\cdot)$ is a kernel function. The upper $(1 - p_\tau)$% of incomes are approximated by a Generalized Pareto Distribution (GPD) model where τ is a location parameter, ξ is a shape, and σ is a scale. The weights p_τ and $(1 - p_\tau)$ are introduced to guarantee the integral of the semiparametric PDF is equal to one.

The parameters of the distribution in (1) are estimated as follows. Common rules of thumb are applied to the choice of bandwidth h and kernel function $K(\cdot)$. Namely, for $K(\cdot)$ we use a Gaussian kernel and h is estimated by cross-validation (CV). A threshold percentile p_τ was arbitrarily chosen at 90^{th} percentile, and this choice was governed by balancing a standard bias-variance trade-off: the fewer observations in the tail considered to be a part of a GPD density, the more volatile the GPD parameters estimates are but the lower the potential bias, and vice versa (Charpentier and Flachaire, 2022). The parameters of the Generalized Pareto top income tail are estimated by applying a maximum penalized likelihood estimator (MPLE) originating from the extreme values literature (Coles and Dixon, 1999; Martins and Stedinger, 2000, 2001; Pauli and Coles, 2001; Huang, Lin and Ren, 2012). Following Butaeva (2025), a beta distribution with parameters $\hat{\alpha} = 8$ and $\hat{\beta} = 8$ ($beta(8, 8)$) has been chosen as a penalty function. Butaeva (2025) shows that unlike other estimators commonly used in the income inequality literature (e.g. MLE, OBRE[14]), a MPLE exhibits high efficiency and robustness in the presence of sparse high-income observations as well as potential top-income data contaminations. We report bootstrap bias-corrected estimates (Cameron and Trivedi, 2005, p. 365) of Gini index and 95% basic (reverse percentile) bootstrap confidence intervals (Davison and Hinkley, 1997, p. 194). The robustness of the baseline results was checked against the threshold, bandwidth, and penalty function choices. For more details on the estimation procedures, see Butaeva (2025).

Figures 1 and 2 show the estimated provincial Gini index and mean household income per capita, respectively.[15] The spatial distribution of provincial inequality estimates is presented in maps in Figures D1 to D4, Appendix D. It can be seen from Figure 1 that there is considerable variation in point estimates of income inequality among Chinese provinces. In 2018, provincial Gini in China ranged from 0.42-0.43 in Shanghai and Tianjin to 0.62-0.67 in Guangdong and Sichuan. The largest increase (6-10%) in income inequality from 2012 to 2018 occurred in Heilongjiang and Sichuan. The largest decrease in income inequality, ranging from 13% to 17%, occurred in Tianjin, Qinghai, and Zhejiang. Notably, during the study period, Sichuan - one of the provinces with the highest levels of inequality - experienced a significant increase in inequality of 6%. In contrast, Tianjin and Shanghai, which are among the least unequal provinces, experienced sizable declines in inequality of 13% and 9%, respectively. These trends created a widening gap in income inequality across Chinese provinces from 2012 to 2018, with the ratio of maximum Gini to minimum Gini increasing from 1.48 in 2012 to 1.58 in 2018.

Economic disparities across provinces are also substantial. Mean household income per capita in 2018 ranges from around 17K CN¥ in Guangxi and Hebei to 63K CN¥ in Shanghai and 73K CN¥ in Beijing (Figure 2). The ratio of rich to poor GRP per capita (Table F1, Appendix F) among Chinese provinces in 2018 (4.7 times) is even larger than the ratio of rich to poor mean household income per capita (4.2 times). From Figure 3, we also observe that poor provinces (e.g., Guizhou, Guangxi, Yunnan) are also more unequal, and richer provinces (e.g., Beijing, Shanghai, Jiangsu, Tianjin) are less unequal. We find a negative correlation between log GRP per capita and estimated provincial Gini: −24% in 2012, −29% in 2014, −22% in 2016, and −34% in 2018. The exceptions are Guangdong, which is both rich (by GRP per capita) and highly unequal, and Jilin, which is relatively poor and less unequal.

Several points should be added regarding the benefits of the estimation method used. First, while estimating Gini coefficients, we noticed that several observations of top incomes in each province have a large effect on provincial

[14] OBRE was introduced by Victoria-Feser and Ronchetti (1994) and later used in many other inequality studies (e.g., Victoria-Feser, 2000; Cowell and Victoria-Feser, 2007; Van Kerm, 2007; Jenkins, 2017).

[15] The corresponding point estimates are reported in Table D1, Appendix D. Provincial abbreviations presented in Table D1 are used throughout the paper. The provincial order in all tables is similar: descending by average GRP per capita (for the years 2012, 2014, 2016, and 2018).

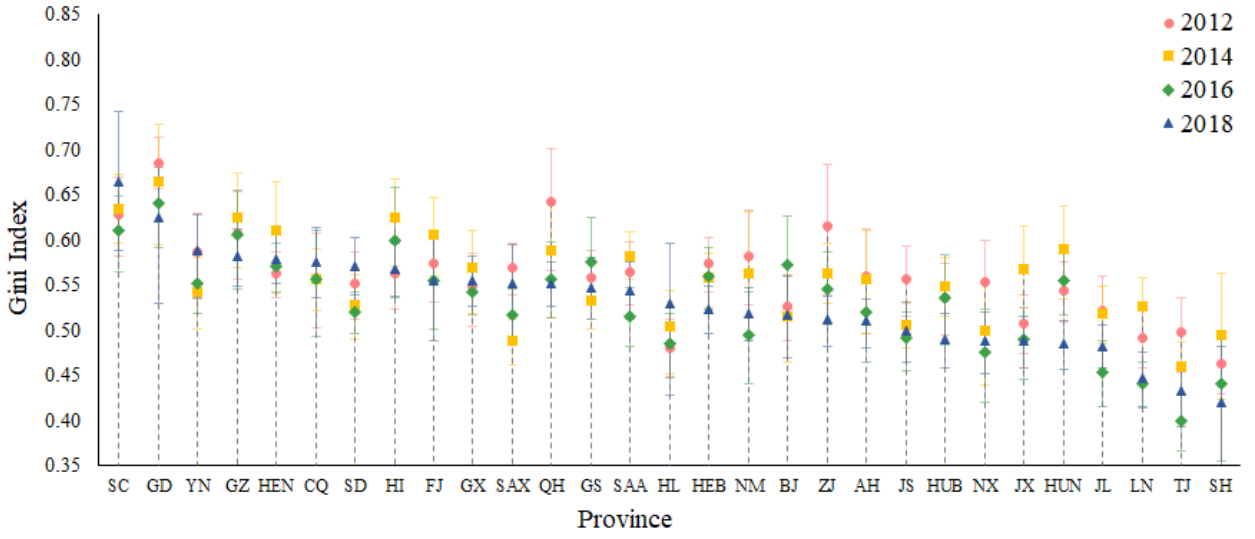


Figure 1: Provincial Gini index estimates for China. *Notes:* This figure presents our baseline estimates of the provincial income Gini index in China, based on data from the CHFS. Provinces on the x-axis are ordered in descending Gini values for 2018. Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 bootstrap samples.

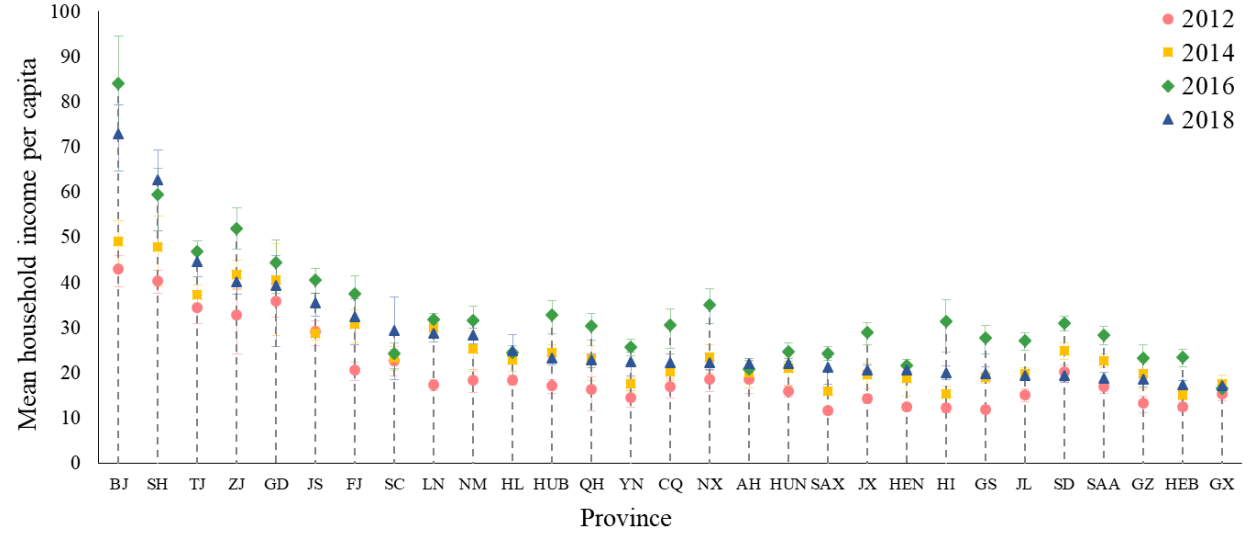


Figure 2: Provincial mean household income per capita (in 1000 CN¥). *Notes:* This figure presents our provincial mean household disposable income per capita estimates, based on data from the CHFS. Provinces on the x-axis are ordered by descending values of the indicator in 2018. Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 bootstrap samples.

Gini index estimates. There could be different reasons for yearly shifts in the composition of top income outliers: for example, an increase in the sample size could add a new top income household or, conversely, some of the top income participants may quit the survey. Moreover, households may experience large swings in their household finance. All these non-systematic factors could contribute to large changes in the resulting Gini index estimates from year to year. The application of the estimation method from Butaeva (2025), which is based on the semiparametric density function (1), helps address this issue by treating top incomes on the distributional tail as a continuous Generalized Pareto function in which parameters are estimated with robust and efficient MPLE. This substantially improves the stability

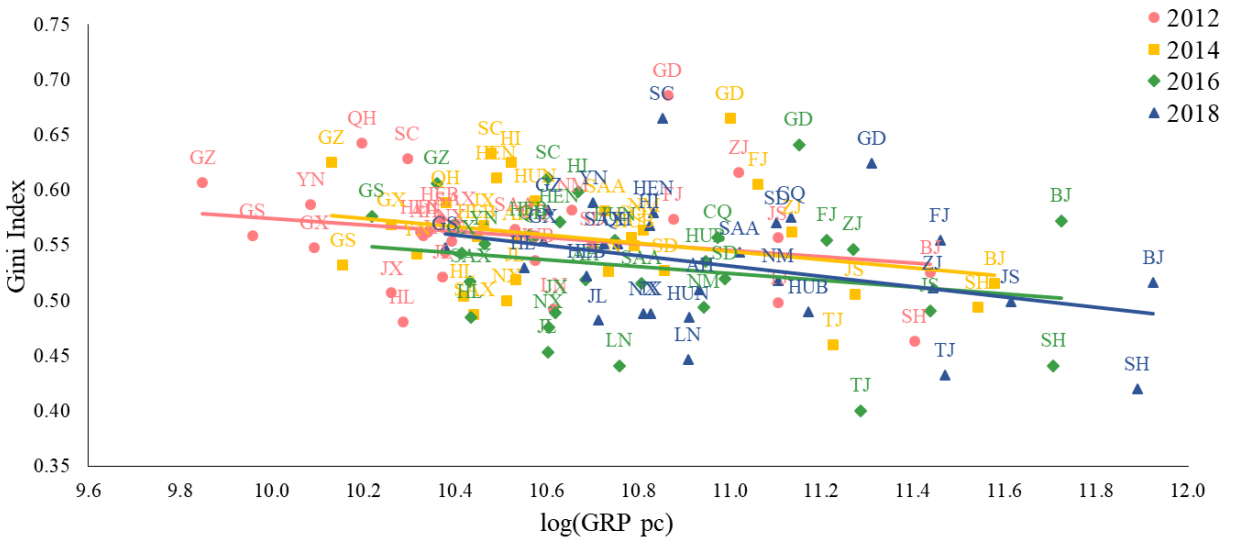


Figure 3: Comparison of provincial Gini index estimates and log GRP per capita. *Notes:* This figure presents our baseline provincial Gini index estimates obtained from the CHFS plotted versus the log GRP per capita from China Statistical Yearbook. Correlations are: -24% in 2012, -29% in 2014, -22% in 2016, and -34% in 2018.

of the obtained income inequality estimates. It can be seen from Figure E1, Appendix E that the resulting provincial Gini's are much more stable over time than the Gini estimates from the raw data (the average variation in Gini index estimates for the same province across years drops by 42%). Second, this substantial improvement in the stability of Gini estimates generates more stable correlations with key provincial economic statistics reported by NBS. For example, Gini estimates from the raw data are correlated with the log GRP per capita by -31% in 2012, -34% in 2014, -19% in 2016, and -26% in 2018, which is less stable compared to the correlation between log GRP per capita and our robust Gini estimates reported in the previous paragraph. Finally, the estimation method doesn't upset the provincial income rankings. Our mean household income per capita estimates are correlated with the raw data mean incomes by 98% in 2014, and 99% in other years. Our estimates are also strongly correlated with per capita disposable income reported in the China Statistical Yearbook (CSY) based on household surveys conducted by the National Bureau of Statistics of China. The relevant correlations are: 89% in 2012 and 2014, 91% in 2016, and 95% in 2018.

Figures K4-K32 in Appendix K present robustness checks with respect to the choice of threshold, bandwidth, and the likelihood penalty function. Together with the baseline threshold of 90^{th} percentile, we consider lower (85^{th} percentile) and higher (95^{th} percentile) thresholds for separating the nonparametric kernel density approximating the bulk of the distribution from the Generalized Pareto density representing the top income tail. Together with the default cross-validated bandwidths ($h = CV$), we consider $h = 0.5 \times CV$ and $h = 1.5 \times CV$ bandwidths. We also tried more, $beta(11, 11)$, and less, $beta(5, 5)$, restrictive penalties in the likelihood function, compared to the baseline $beta(8, 8)$. Based on the confidence intervals, none of the alternative approaches produce changes in Gini estimates compared to our baseline results.

5. Inequality decomposition

5.1. Theil T index decomposition by population subgroups

To analyze the relative contribution of inter- and intraprovincial variation in incomes to nationwide income inequality in China, we perform a subgroup decomposition of the Theil T inequality index by province. The Theil T index formula is:

$$T = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} ln\left(\frac{y_i}{\bar{y}}\right), \tag{2}$$

where y_i is an annual income per capita of a household i, $\bar{y}$ is mean household income per capita, and N is the number of households in the sample.

Table 1
Decomposition of nationwide income inequality by province.

	2012		2014		2016		2018	
	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Beijing	0.374	5%	0.343	3%	0.440	5%	0.342	4%
Shanghai	0.281	3%	0.251	2%	0.232	2%	0.197	2%
Jiangsu	0.424	5%	0.377	6%	0.323	6%	0.353	7%
Tianjin	0.328	2%	0.288	1%	0.215	1%	0.256	1%
Zhejiang	0.473	8%	0.449	7%	0.394	6%	0.376	6%
Fujian	0.491	3%	0.502	3%	0.408	3%	0.393	3%
Guangdong	0.752	11%	0.558	11%	0.557	11%	0.510	9%
Shandong	0.458	4%	0.395	7%	0.387	5%	0.458	6%
Inner Mongolia	0.504	1%	0.506	2%	0.350	1%	0.385	2%
Hubei	0.372	3%	0.414	3%	0.376	4%	0.328	3%
Chongqing	0.429	3%	0.449	2%	0.425	2%	0.454	2%
Shaanxi	0.467	2%	0.543	3%	0.375	2%	0.460	2%
Liaoning	0.333	2%	0.388	4%	0.265	3%	0.290	3%
Hunan	0.441	3%	0.477	4%	0.406	4%	0.329	3%
Anhui	0.438	3%	0.434	3%	0.367	2%	0.378	3%
Hainan	0.502	1%	0.580	0%	0.477	1%	0.493	0%
Ningxia	0.455	1%	0.345	0%	0.305	0%	0.350	0%
Henan	0.475	2%	0.525	5%	0.477	4%	0.533	6%
Sichuan	0.553	6%	0.653	8%	0.512	5%	0.506	6%
Jilin	0.369	2%	0.384	1%	0.278	1%	0.341	1%
Jiangxi	0.364	1%	0.461	2%	0.325	2%	0.330	2%
Hebei	0.476	2%	0.484	3%	0.422	4%	0.394	3%
Shanxi	0.488	2%	0.350	1%	0.378	2%	0.430	2%
Qinghai	0.618	2%	0.570	0%	0.465	0%	0.477	0%
Heilongjiang	0.310	2%	0.344	2%	0.318	2%	0.361	2%
Yunnan	0.476	2%	0.437	2%	0.454	3%	0.491	3%
Guangxi	0.453	2%	0.495	2%	0.485	2%	0.470	2%
Guizhou	0.530	2%	0.586	2%	0.535	2%	0.497	2%
Gansu	0.477	2%	0.422	1%	0.465	2%	0.457	1%
Within	0.453	86%	0.449	90%	0.395	87%	0.397	86%
Between	0.073	14%	0.048	10%	0.059	13%	0.065	14%
Total	0.526	100%	0.498	100%	0.453	100%	0.462	100%

Notes: This table reports the results of the Theil T index (2) decomposition by province, as formulated in eq. (3).

We choose the Theil T index because, in contrast to many other inequality measures, it can be effectively decomposed into within and between population sub-group components without an overlap component (which is difficult to interpret). We decompose the Theil T index into between-province and within-province shares applying the following formula:[16]

$$T = \sum_{g=1}^{h} s_g T_g + \sum_{g=1}^{h} s_g ln\left(\frac{s_g}{p_g}\right), \tag{3}$$

where the first term is the "within" component (T_g is each subgroup's Theil index, weighted by the subgroup's share in total income $s_g = p_g \bar{y}_g / \bar{y}$), and the second term is the "between" component. Note that the Theil T index is calculated

[16]This formula goes back to (Bourguignon, 1979; Cowell, 1980; Shorrocks, 1984). More details can also be found in Cowell (2000) and Jain-Chandra et al. (2018).

Table 2
Decomposition of nationwide income inequality by educational level of working-age (16-60) household members.

	2012		2014		2016		2018	
	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary school and below	0.541	7%	0.562	7%	0.573	7%	0.484	8%
Middle school	0.473	20%	0.499	20%	0.437	18%	0.387	20%
High school	0.403	22%	0.415	24%	0.375	26%	0.344	23%
College and above	0.350	28%	0.325	29%	0.288	29%	0.272	24%
Within	0.406	77%	0.400	80%	0.359	80%	0.336	76%
Between	0.123	23%	0.101	20%	0.092	20%	0.108	24%
Total	0.529	100%	0.501	100%	0.451	100%	0.444	100%

Notes: This table reports the results of the Theil T index (2) decomposition by the education level of working-age (16-60) household members, as formulated in eq. (3).

by using the mean income of each subgroup, where $p_g = N_g/N$ is the share of each subgroup's population N_g in the total population N.

Table 1 presents the results of the decomposition of nationwide income inequality in China by province. We can see that within-provincial inequality accounts for around 86-90% of nationwide income inequality measured by the Theil T index, and the remaining 10-14% is attributable to the between-provincial income variation. It can be also observed that the largest share of nationwide inequality, around 9-11%, is accounted for by the income inequality in Guangdong province (due to the large share of the national population in Guangdong). The next most important contributors to nationwide inequality are Sichuan and Zhejiang provinces (6-8%). The correlation coefficient between the provincial Theil T indices and our baseline Gini indices are: 96% in 2012, 93% in 2014, 97% in 2016, and 93% in 2018.

We also apply eq. (3) to study the variation of incomes within and between educational groups. In Table 2, we summarize the results of the inequality decomposition by educational level of working-age[17] household members. We can see that the dominant part, 76-80%, of nationwide income inequality in China is accounted for by variation within educational groups, and the remaining 20-24% is accounted for by between-group variation.

5.2. Regression-based decomposition of income inequality

To analyze relative importance of factors associated with the levels of income inequality in China, we turn to regression-based decomposition. We apply the following variant of the Mincer model (Mincer, 1974) to each year of data:

$$\ln Y_i = \delta_0 + \delta_1 S_i + \delta_2 E_i + \delta_3 E_i^2 + \delta_4 \ln N_i + \sum_{l=1}^{3} \kappa_l K_{li} + \sum_{k=1}^{2} \eta_k X_{ki} + \sum_{q=1}^{Q-1} \gamma_q G_{qi} + \varepsilon_i, \tag{4}$$

where $\ln Y_i$ is the logarithm of annual household income per capita i; S_{ijt} is years of schooling per worker defined as average years of schooling among working-age (16-60) household members; E_i is average years of potential work experience of working-age household members (we follow the existing literature and use potential experience defined as age minus years of schooling minus six); and ε_i is an error term which captures other unobserved household-specific factors and measurement error. According to human capital theory, more education enhances productivity and thus is associated with higher earnings. Consequently, we expect to the δ_1 coefficient estimate ($\hat{\delta_1} > 0$) to be positive. The Mincer model uses individual earnings as the dependent variable. However, in the current study, we focus on household income per capita. Therefore, we must transform the regression equation using per-worker measures of education and experience, and we also control for the log of household size $\ln N_i$. We also control for K_{li}, a set of factors shaping components of household income other than labor income. First, to control for land endowments affecting agricultural business income, we include a *Land dummy* variable, which indicates whether the household owns agricultural land. We also include *Ln*(*land per capita*), which is log of the amount of the household's land (in *mu*[18]) per capita. To

[17] Here and below, we use 16 to 60 years-old as the definition of working-age individuals.
[18] One *mu* is equivalent to 0.06667 hectare.

control for capital endowments supporting other potential business income sources, we add *Ln*(*business assets per capita*), the log of total household asset value for all business projects (in 100K CN¥) divided by household size. We also control for household-level covariates X_{ik} which may also affect household income levels in China (namely, gender and Communist Party membership of the household head). To capture geographic factors, we also include G_{iq} spatial dummies represented in different specifications by rural and provincial dummies or rural and city dummies.

Following Fields (2003), we additionally estimate the share of the log-variance of household income per capita that can be attributed to each regressor in eq. (4):

$$s_z = \frac{\hat{\alpha}_z cov(V_z, \ln Y)}{\sigma^2(\ln Y)}, \tag{5}$$

where s_z is a resulting % of income inequality which can be attributed to the zth regressor in eq. (4), $\hat{\alpha_z}$ is the estimated regression coefficient on each z^{th} independent variable, $cov(V_z, \ln Y)$ is the covariance between each z^{th} independent variable and the explanatory variable, and $\sigma^2(\ln Y)$ is the variance of the explanatory variable. The sign of the resulting s_z indicates whether the z^{th} regressor increases ($s_z > 0$) or decreases ($s_z < 0$) income inequality.

Table 3 reports the regression-based decomposition results. The OLS estimates for eq. (4) controlling for provincial dummies are reported in odd columns from (1) to (7) and those controlling for city dummies are reported in odd columns from (9) to (15). Each even column presents $\hat{s}_z$ estimates from eq. (5) reflecting the relative contribution of regressors from eq. (4) in nationwide income inequality. We find that a one-year increase in the years of schooling per worker is associated in different years with a 9-13% increase in household income per capita. Variation in the years of schooling per worker accounts for the largest share of the variation in household income per capita (9-17%). The positive sign of $\hat{s}_z$ estimates for schooling suggests that greater variation in the years of schooling per worker among households is associated with higher income inequality. Our results confirm positive but diminishing returns to experience per worker for the years 2014 to 2018. In 2012, we observed increasing returns to experience, but these results are not statistically significant. The inflection point for experience per worker was around 18 years in 2014 and between 30 to 35 years in 2016 and 2018. Experience accounts for less than 1% of the variation in household income per capita.

Household size is negatively correlated with (log) household income per capita, and it accounts for 1-4% of its variation. Households that own agricultural land tend to have lower average incomes compared to those without land. A 1% increase in *mu* of owned agricultural land is associated with a 0.10-0.27% increase in the household income per capita. Variation in household land accounts for less than 2% in the variation of (log) household income per capita. A 1% increase in the value of household business assets is associated with an increase in income of about 0.05-0.16%, and its variation accounts for up to 2.5% of the variation in income. Households whose heads are Communist Party members have 16-25% higher incomes per capita. Households with female heads receive 5-13% lower income than those with male heads. However, both party membership and gender of the household head accounts for less than 1% of the variation in (log) household income per capita. Rural households have 20-53% lower incomes than urban households. The rural dummy accounts for 3-8% of income inequality in China. Provincial and city dummies accounts for 4-6% and 8-14% of inequality in China, respectively.

To sum up, Table 3 shows that the factor most strongly associated with the variation in household income is the educational attainment of working-age household members which accounts for around 9-17%. The second-largest contributors to income inequality are geographic factors: city dummies (8-14%), provincial dummies (4-6%), and the rural dummy (3-8%). The dominant role of education in accounting for household income in China is also found by Goh, Xubei and Nong (2009), who apply the Mincer regression decomposition framework to 1989–2004 data from the China Health and Nutrition Survey (CHNS). However, in contrast to Wan and Zhou (2005)[19], we haven't found any evidence on the importance of capital input in accounting for variation in household income in China. It could be that the relative importance of human capital versus physical capital has increased over time in China.

[19] Wan and Zhou (2005) who uses a similar Mincer regression decomposition framework to study rural household incomes in China in 2002 using data from the Research Centre for Rural Economy (RCRE, Ministry of Agriculture).

Table 3
Mincer regression: accounting for household income per capita variation in China.

	Dependent variable: Ln(household income per capita)															
	(2012)	($\hat{s}_z$)	(2014)	($\hat{s}_z$)	(2016)	($\hat{s}_z$)	(2018)	($\hat{s}_z$)	(2012)	($\hat{s}_z$)	(2014)	($\hat{s}_z$)	(2016)	($\hat{s}_z$)	(2018)	($\hat{s}_z$)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Schooling per worker	0.124***	13.6%	0.096***	10.1%	0.138***	17.2%	0.122***	14.7%	0.112***	12.1%	0.088***	9.3%	0.128***	15.9%	0.110***	13.3%
	(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)	
Experience per worker	0.002	-0.6%	0.012***	-2.0%	0.028***	-3.9%	0.022***	-3.8%	0.003	-0.7%	0.012***	-2.0%	0.030***	-4.1%	0.022***	-3.8%
	(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)	
Experience per worker^2 : 100	0.006	-0.1%	−0.034***	2.4%	−0.041***	3.1%	−0.037***	3.5%	0.001	0.1%	−0.035***	2.4%	−0.043***	3.3%	−0.037***	3.5%
	(0.008)		(0.009)		(0.007)		(0.008)		(0.008)		(0.009)		(0.007)		(0.008)	
Ln(household size)	−0.384***	2.8%	−0.499***	4.1%	−0.271***	1.4%	−0.323***	1.4%	−0.359***	2.6%	−0.445***	3.6%	−0.228***	1.2%	−0.294***	1.3%
	(0.025)		(0.030)		(0.027)		(0.027)		(0.024)		(0.030)		(0.027)		(0.027)	
Land dummy	−0.084***	0.7%	−0.280***	3.6%	−0.192***	2.0%	−0.045*	0.4%	−0.063***	0.6%	−0.270***	3.5%	−0.135***	1.4%	−0.029	0.2%
	(0.022)		(0.036)		(0.032)		(0.025)		(0.023)		(0.036)		(0.032)		(0.026)	
Ln(land per capita)	0.168**	-0.0%	0.200***	-1.9%	0.125***	-1.2%	0.264***	-0.1%	0.109	-0.0%	0.193***	-1.9%	0.091***	-0.9%	0.203***	-0.1%
	(0.079)		(0.029)		(0.026)		(0.067)		(0.082)		(0.030)		(0.027)		(0.066)	
Ln(business assets per capita)	0.158***	2.3%	0.164***	2.5%	0.100***	1.2%	0.051***	0.4%	0.150***	2.2%	0.160***	2.4%	0.095***	1.1%	0.050***	0.4%
	(0.007)		(0.008)		(0.008)		(0.010)		(0.007)		(0.008)		(0.008)		(0.009)	
Party member	0.232***	1.0%	0.190***	0.8%	0.162***	0.6%	0.179***	0.8%	0.246***	1.1%	0.210***	0.9%	0.168***	0.6%	0.190***	0.8%
	(0.025)		(0.025)		(0.024)		(0.024)		(0.024)		(0.025)		(0.024)		(0.024)	
Female	−0.045**	-0.1%	−0.073***	-0.2%	−0.126***	-0.2%	0.011	0.0%	−0.051**	-0.1%	−0.077***	-0.2%	−0.134***	-0.2%	0.002	0.0%
	(0.021)		(0.022)		(0.022)		(0.023)		(0.021)		(0.022)		(0.022)		(0.022)	
Rural	−0.269***	3.5%	−0.527***	8.2%	−0.304***	3.9%	−0.336***	4.8%	−0.197***	2.5%	−0.485***	7.7%	−0.239***	3.1%	−0.230***	3.3%
	(0.023)		(0.031)		(0.024)		(0.024)		(0.024)		(0.033)		(0.025)		(0.025)	
Constant	8.468***		9.102***		8.407***		8.740***		8.150***		8.985***		8.355***		8.920***	
	(0.090)		(0.094)		(0.090)		(0.093)		(0.139)		(0.136)		(0.140)		(0.137)	
Provincial dummies	Yes	5.6%	Yes	3.7%	Yes	6.0%	Yes	6.1%	No		No		No		No	
City dummies	No	No	No	No	No	No	No	No	Yes	10.7%	Yes	7.5%	Yes	11.5%	Yes	13.6%
Residual		71.4%		68.8%		70.0%		71.9%		68.9%		66.8%		67.2%		67.5%
Observations	21,880		27,821		26,642		23,243		21,880		27,821		26,642		23,243	
R^2	0.315		0.339		0.324		0.309		0.340		0.358		0.351		0.353	

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

Notes: This table reports the results of the Mincer-type regression formulated in eq. (4), controlling for provincial dummies in columns (1) to (7) and controlling for city dummies in columns (9) to (15). We employ heteroskedasticity-robust standard errors, and apply survey weights.

6. Educational inequality and returns to education

To measure educational (human capital) inequality, we estimate educational Gini indices from two different data sources: CHFS for our baseline results and the China Population and Employment Statistics Yearbook (CPESY) as a robustness check. To calculate provincial-level educational Gini coefficients, we map years of schooling to educational attainment levels as follows: zero years for no schooling; six years for primary school; nine years for junior school; twelve years for regular or technical high school; fifteen years for vocational college; sixteen years for regular college and bachelor's; nineteen years for a master's degree; and twenty-two years for a doctoral degree.

Figure 4 reports our estimates of the provincial-level educational Gini indices using the CHFS data.[20] Point estimates of the educational Gini index range from around 0.16 in Shanghai and Beijing to 0.39 in Qinghai. One notable observation is that provincial educational inequality has a strong negative correlation with average years of schooling (Figure 5): −83% in 2012, −86% in 2014, −83% in 2016, and −81% in 2018. The average number of years of schooling also varies considerably across provinces, from 7.7 in Qinghai to 12.3 in Beijing. Provincial educational patterns are also highly correlated with the level of economic development. Richer provinces, like Beijing, Shanghai, and Tianjin, exhibit lower levels of educational inequality (the correlation between (log) GRP per capita and the educational Gini is −45% in 2012, −43% in 2014, −42% in 2016, and −44% in 2018), and their citizens have higher average levels of education (the correlation between (log) GRP per capita and average years of schooling is 73% in 2012, 72% in 2014, 67% in 2016, and 69% in 2018). These patterns are associated with provincial differences in educational attainment structure. Based on the data from the CPESY, in Figure 6, we summarize the provincial-level distributions of the employed population by their educational level averaged across four waves and grouped into four categories: primary school and below, middle school, high school, or "college and above".[21] We can see that the richer provinces, like Beijing, Shanghai, and Tianjin, have a significantly lower share of the employed population with primary school and below education and a higher share of employed individuals with college and above degrees. In poorer provinces like Yunnan, Guizhou, and Gansu, the distribution of education levels is reversed: the largest share of labor is unskilled and the smallest share of labor is high-skilled compared to other provinces. It can also be noticed that except in Beijing, Shanghai, and Tianjin, high school and college and above categories are more equally distributed across provinces compared to primary and middle school categories. Therefore, it seems that differences in educational inequality across provinces in China are most closely associated with the variation in the share of less skilled labor.

Next, we evaluate differences in returns to human capital across provinces. To do so, we estimate provincial returns to schooling and returns to high school and above for each province and year.[22] For returns to schooling, we apply the following provincial-level version of eq. (4)[23] to each year of the CHFS household-level data:

$$\ln Y_{ijt} = \delta_{0jt} + \delta_{1jt} Z_{ijt} + \delta_{2jt} E_{ijt} + \delta_{3jt} E_{ijt}^2 + \delta_{4jt} \ln N_{ijt} + \varepsilon_{ijt}, \tag{6}$$

where for each household i in province j and year t, $Z_{ijt} = \{S_{ijt}; H_{ijt}\}$ denotes either years of schooling per worker S_{ijt} or a dummy variable indicating whether working-age household members obtained, on average, high school or above educational level ($H_{ijt} = 1$), or not ($H_{ijt} = 0$). In case of $Z_{ijt} = S_{ijt}$, we call the estimate of $\hat{\delta}_{1jt}$ *"returns to schooling"*, which defines an increase in household income per capita in response to an increase in average years of schooling among working-age household members.[24] In case of $Z_{ijt} = H_{ijt}$ the estimate of $\hat{\delta}_{1jt}$ is called *"returns to high school"*.

Figures 7 and 8 report estimated provincial returns to schooling and returns to high school (respectively). On average during 2012-2018, point estimates of the effect of an additional year of schooling of working-age household members on household income per capita varied from 10.3% (Jilin) to 18.7% (Hainan). Point estimates of the returns to high school vary from 50.4% in Shanghai to 130.9% in Qinghai. The substantially higher returns to high school

[20] The corresponding point estimates of all educational variables presented in the current Section are reported in Table D2, Appendix D.

[21] In Figure E1, Appendix D, we report the educational structure reported in the China Statistical Yearbook (CSY). Figure E1 and Figure 6 demonstrate similar patterns for larger (six-years of education and above) and smaller (employed) subpopulations in Chinese provinces.

[22] We estimate returns to schooling using separate regressions for each province. This approach is preferred to a pooled model with interaction terms because it allows the coefficients on all covariates to vary across provinces. The validity of this specification is statistically justified: a Chow test formally rejects the null hypothesis of parameter stability (p-value < 0.05). This provides strong evidence of significant coefficient heterogeneity and robustly supports our choice, as it indicates that a single pooled model would be misspecified.

[23] To evaluate provincial returns to schooling and returns to high school, we focus on the importance of human capital in explaining the variation in the household income per capita and ignore non-labor income covariates X_{ik}, household head characteristics K_{li}, and spatial dummies G_{iq} from eq. (4).

[24] Our definition of returns to schooling is close to the "ex-post returns to schooling" from Heckman, Lochner and Todd (2006).

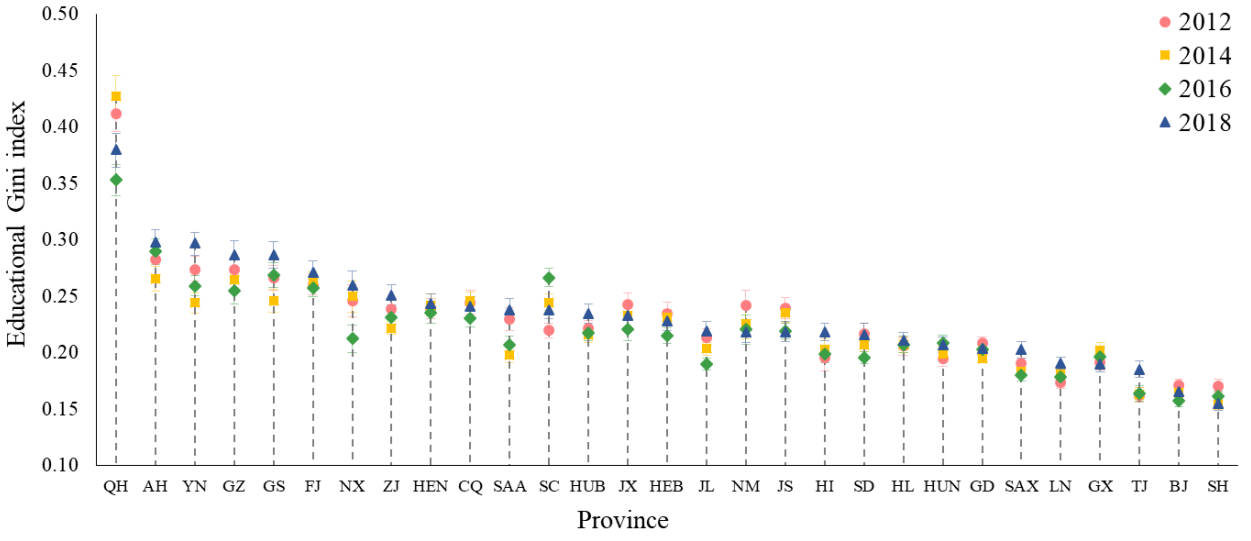


Figure 4: Provincial educational Gini index in China. *Notes:* This figure presents the provincial educational Gini index estimated from the CHFS. Provinces on the x-axis are ordered by descending values of the indicator in 2018. Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 bootstrap samples.

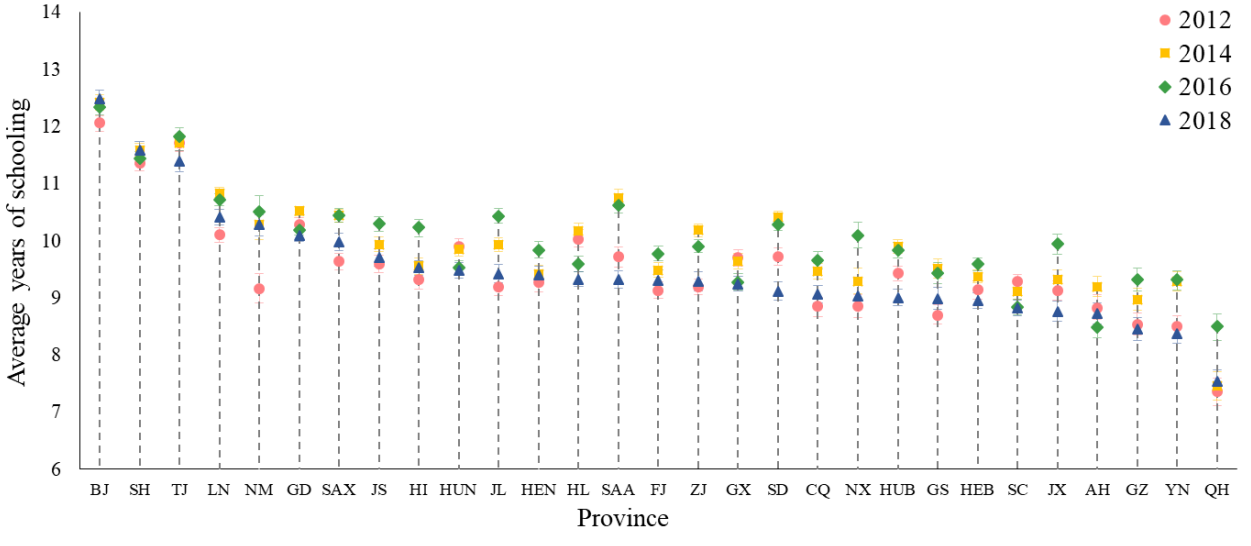


Figure 5: Provincial average years of schooling in China. *Notes:* This figure presents the provincial average years of schooling estimated from the CHFS. Provinces on the x-axis are ordered by descending values of the indicator in 2018. Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 bootstrap samples.

and above, compared to the average return per year of schooling, reflect existing nonlinearities in the education wage premium in China. The economic payoff for completing upper-secondary and tertiary education is significantly larger than the payoff for marginal years of schooling. This pattern is also evident in Table C4 from Appendix C, which shows per capita household income by levels of educational attainment and province.

In general, richer provinces exhibit lower returns to schooling (the correlation between (log) GRP per capita and returns to schooling is −19% in 2012, −40% in 2014, −33% in 2016, and −7% in 2018) and lower returns to high school and above (the correlation between (log) GRP per capita and returns to high school is −40% in 2012, −57% in 2014, −32% in 2016, and −36% in 2018). These negative correlations are consistent with richer provinces having a greater relative supply of skilled labor than poor provinces, and with earlier evidence that returns to education in

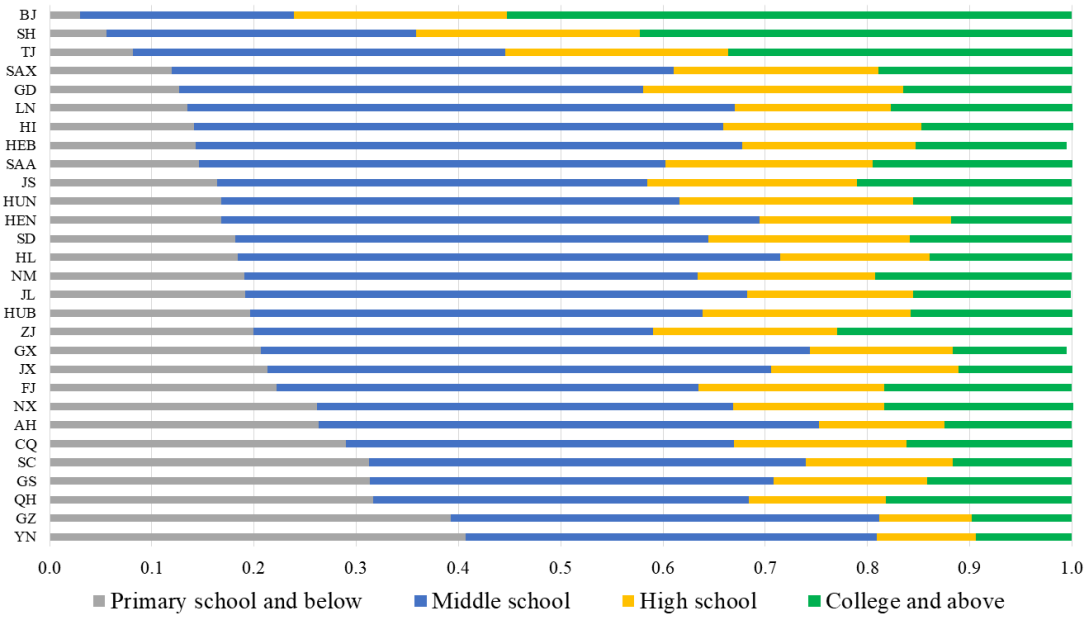


Figure 6: Average educational structure in Chinese provinces, 2012-2018. *Notes:* This figure presents the average educational structure of the employed population in Chinese provinces for 2012-2018 estimated from the China Population and Employment Statistics Yearbook (CPESY). The estimates are ordered in the "Primary school and below" category.

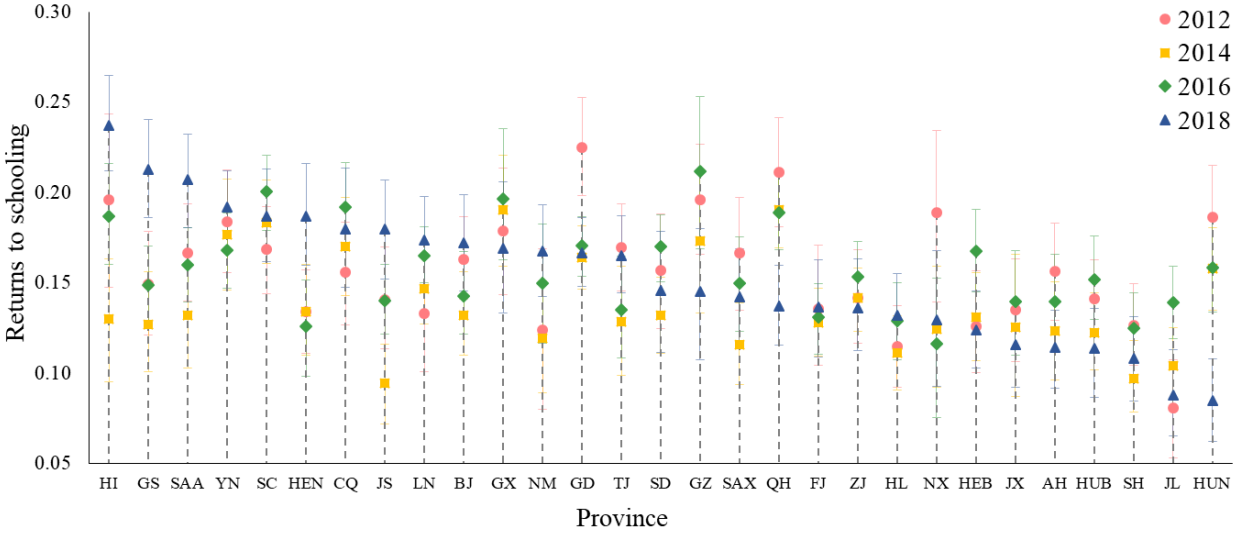


Figure 7: Provincial returns to schooling in China. *Notes:* This figure presents the provincial returns to schooling estimated from the CHFS. Provinces on the x-axis are ordered by descending values of the indicator in 2018. Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 bootstrap samples.

China tend to be lower in more economically developed regions (Xie and Hannum, 1996; Hauser and Xie, 2005; Li, 2003). We use returns to high school rather than the returns to schooling as our preferred measure for examining the persistent component of interprovincial variation in returns to education. The returns to high school exhibits greater explanatory power in our empirical analysis, likely because it captures a key nonlinearity in the returns to schooling. It is also more comparable across provinces because it identifies the returns to a common and important discrete level

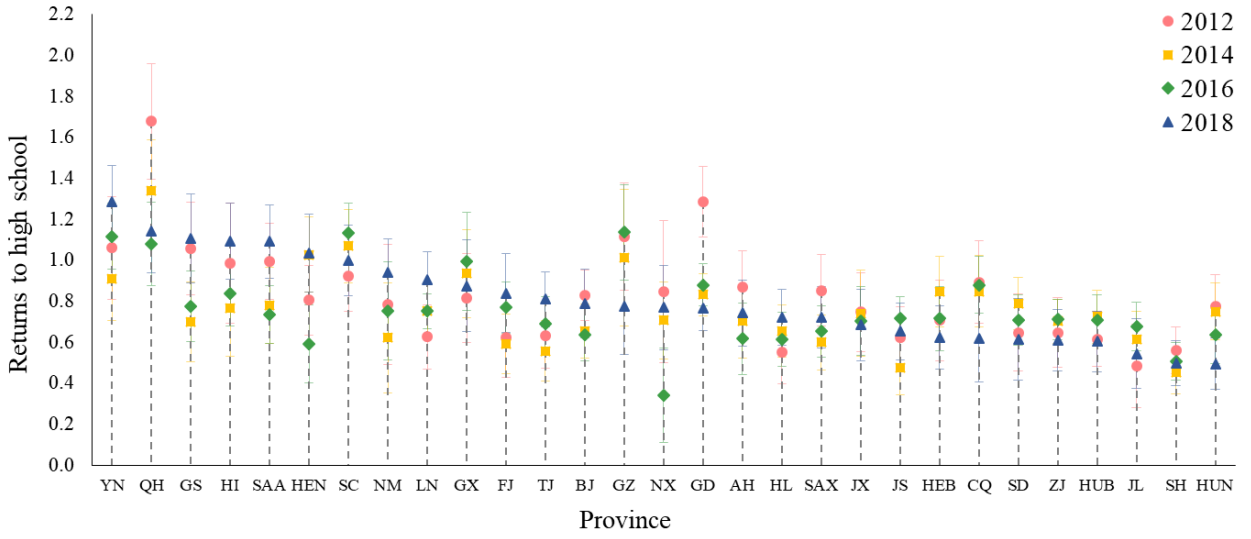


Figure 8: Provincial returns to high school in China. *Notes:* This figure presents the provincial returns to high school estimated from the CHFS. Provinces on the x-axis are ordered by descending values of the indicator in 2018. Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 bootstrap samples.

of schooling whereas the returns to schooling is identified from the levels of schooling with the most variation in each province (which will be at higher levels of schooling in rich provinces than in poor provinces). Our robustness analysis in Appendix K also finds that returns to high school is less subject to bias from influential outliers than returns to schooling.

In Figures D5-D8 (Appendix D), we also show the interrelations between all of the estimated human capital indicators and the provincial-level income Gini estimates introduced in Section 4. Provincial income inequality is positively correlated with provincial educational inequality, provincial returns to schooling, and provincial returns to high school, and negatively correlated with provincial average years of schooling.

7. Interprovincial variation in income inequality

7.1. Explaining variation in provincial income inequality

To evaluate the empirical relationship between the level of human capital development and variation in income inequality across provinces, we regress provincial income Gini indices, introduced in Section 4, on provincial educational Gini indices and returns to schooling from Section 6:

$$I_{jt} = \beta_0 + \beta_1 I^E_{jt} + \beta_2 \hat{\delta}_{1jt} + \beta_3 I^E_{jt} \times \hat{\delta}_{1jt} + \sum_{m=1}^{4} \tau_m T_{mjt} + \nu_t + \varepsilon_{jt}, \tag{7}$$

where I_{jt} is the income inequality Gini index in province j at year t, I^E_{jt} is the educational Gini index, $\hat{\delta}_{1jt}$ is the estimated returns to schooling (returns to high school), ν_t is year fixed effect, and ε_{jt} captures other unobserved factors that can vary across time and provinces.

To motivate our expectations regarding the signs of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, consider a simplified version of the initial Mincer model from eq. (4):

$$\ln Y_i = \delta_0 + \delta_1 S_i + u_i, \tag{8}$$

where Y_i is annual household income per capita, S_i is average years of schooling among working-age household members, and u_i captures other factors that may influence household income per capita. Then the variance of household income per capita can be expressed by the formula:

$$\mathrm{Var}(\ln Y_i) = \delta_1^2 \mathrm{Var}(S_i) + \mathrm{Var}(u_i) + 2\mathrm{Cov}(\delta_1 S_i, u_i), \tag{9}$$

where the variance in household income per capita Var(ln Y) measures income inequality, and Var(S) captures educational inequality. If we believe that the income-generating process is defined by our Mincer model introduced in (4), then from (9) it is clear that an increase in educational inequality or an increase in returns to education, and their interaction, should all increase provincial income inequality among households within the same province. Therefore, we expect to obtain positive coefficients for the educational Gini, returns to education, and their interaction term.[25]

To capture other provincial-specific factors which could be important for explaining variation in provincial income inequality, we add four provincial covariates T_{jtm} to eq. (7) using data from the China Statistical Yearbook. First, we include the $Social\ spending$ variable which measures local governments' expenditures on social insurance per working-age adult (in 1,000 CN¥).[26] We use this variable as a proxy for social public spending that aims to reduce income inequality, so we expect a negative coefficient. Second, we include $SOE\ assets\ share$ - the share of the total current assets of state-owned industrial enterprises in provincial GRP. This variable captures the size of the public sector in the provincial economy. Usually, the public sector exhibits a more egalitarian wage structure than the private sector. Therefore, we expect to find a negative relationship between $SOE\ assets\ share$ and provincial-level income inequality. Third, we include $Foreign\ trade\ share$ - the total volume of foreign trade (imports and exports) as a share of provincial GRP. This variable is used as a proxy for the degree of openness of the provincial economy. We expect to obtain a negative coefficient on this variable because in China greater trade openness is likely associated with increased demand and prices for unskilled labor, negatively impacting income inequality. Finally, we include the $Ln(GRP\ per\ capita)$ to control for differences in levels of economic development across provinces. We documented in Section 4 the negative relationship between log GRP per capita and income inequality in Chinese provinces, and we expect to confirm this correlation with a negative $Ln(GRP\ per\ capita)$ coefficient estimate.

Our key regressors, the educational Gini and returns to schooling and to high school, are estimated from CHFS microdata, and the panel contains 29 provinces observed over four survey waves. We therefore report inference for all provincial-level regression results that follow using a nested bootstrap procedure. Provinces are resampled with replacement to account for the small number of clusters (Cameron and Miller, 2015); within each resampled province-year, households are then redrawn from the underlying CHFS microdata, and the educational Gini and returns to education are recalculated in each bootstrap sample. This procedure incorporates sampling variation in the generated regressors into the distribution of the regression estimates (Pagan, 1984; Murphy and Topel, 1985).[27] We construct basic (reverse-percentile) bootstrap confidence intervals from 1,000 replicates (Davison and Hinkley, 1997, p. 194), consistent with our treatment of the provincial Gini estimates in Section 4.

Table 4 reports our baseline estimates of eq. (7). We center our key variables—the educational Gini I^E_{jt} and returns to schooling (returns to high school) $\hat{\delta}_{1jt}$—by subtracting their respective means to facilitate interpretation of the main effects in specifications with interaction terms (columns 4, 8, 11, and 14). The results reveal a strong positive relationship between educational and income inequality across Chinese provinces. Our preferred specification with provincial-level covariates (column 5) indicates that a 10 percentage point (pp.) increase in the educational Gini is associated with a 4.5 pp. increase in the income Gini index, significant at the 5% level.[28]

Notably, the magnitude and significance of this relationship varies substantially depending on which returns to education we control for. When we include provincial returns to schooling (columns 7–8), the estimated association between educational inequality and income inequality attenuates modestly to 3.7 pp. but remains statistically significant. In contrast, controlling for returns to high school education (columns 13–14) substantially attenuates the coefficient to a range of 1.3 to 1.7 pp. and renders it statistically insignificant. This pattern suggests that returns to high school are associated with much of the observed relationship between educational inequality and income inequality, and represent the key correlate of provincial income inequality variation. Our regression results also document a significant positive association between provincial income inequality and returns to education. A 10 pp. increase in overall returns to schooling is associated with a 7.2 to 8 pp. increase in the income Gini (columns 6–8). Returns to high school similarly exhibit a positive and statistically significant relationship with income inequality, though with a smaller estimated coefficient of 1.3 to 1.5 pp. (columns 12–14). Despite this smaller coefficient magnitude, the importance of high school

[25] For more details refer to Coady and Dizioli (2018).

[26] We use an indicator called "Social Safety Net and Employment Effort" which reflects the government's social security and employment spending. Specifically, it includes social security and employment management services, civil administration affairs, fiscal subsidies to social insurance funds, employment subsidies, retirement resettlement, social welfare, disability services, urban minimum living security, etc.

[27] We thank an anonymous referee for raising this point.

[28] The mean provincial income Gini ranges from 0.530 to 0.568 across years, with a standard deviation of approximately 0.05 (Table D1, Appendix D).

returns as a key correlate is evidenced by their ability to correspond to the relationship between educational inequality and income inequality, as demonstrated by the attenuation results described above.

A negative sign on the interaction term in columns (4), (8), (11) and (14) suggests that educational inequality has a weaker association with income inequality in places with higher returns to education, or that returns to education have a weaker association with income inequality in areas with higher educational inequality. Of the four, only the returns-to-high-school interaction without additional covariates (column 11) is statistically significant; the remaining three are negative but not statistically different from zero. Since we centered I_{jt}^{E} and $\hat{\delta}_{1jt}$, the main effects in columns (8) and (14) indicate that a 10 pp. increase in educational Gini is associated with 3.7 pp. increase in the income Gini for a province with average returns to schooling and 1.7 pp. increase in income Gini for a province with average returns to high school. At the same time, a 10 pp. increase in returns to schooling (returns to high school) is associated with a 7.2 pp. (1.4 pp.) increase in the income Gini for a province with an average educational Gini. Contrary to the positive interaction between educational inequality and returns to education predicted by eq. (9), our initial estimates yielded a negative coefficient. To investigate this puzzle, we tested two non-mutually exclusive mechanisms. First, we hypothesized that in provinces with very high educational inequality, income inequality might already be near this ceiling, so higher returns to education can't increase it much more. We tested this by adding a quadratic term for educational Gini to our baseline specification. The results, presented in Columns (1), (2), (8), and (9) of Table I1 (Appendix I), provide no evidence for such nonlinearity. Second, we tested whether the negative interaction may stem from omitted variable bias, specifically the omission of government redistributive capacity. We tested this by introducing our key provincial covariates T_{jtm} individually. The results, reported in Columns (3)–(7) and (10)–(14) of Table I1, show that social public spending attenuates the negative interaction to insignificance in both specifications, while the government sector share does so only for returns to schooling. This suggests that the compensatory force of government redistribution attenuates the inequality-amplifying relationship of market returns to education, and helps explain the negative sign.

Our main results in Table 4 also show a significant negative association between social public spending ($Social\ spending$) and government share of the provincial economy ($SOE\ assets\ share$) on income inequality (as we suggested above). Specifically, a 1,000 CN¥ increase in local governments' expenditures on social insurance per working-age adult is associated with a 1.1 to 1.7 pp. decrease in the income Gini coefficient. A 10 pp. increase in the share of the total assets of state-holding industrial enterprises in provincial gross regional product (GRP) is associated with a 7 to 11 pp. decrease in the income Gini. Economic openness proxied by the share of imports and exports in GRP ($Foreign\ trade$) was found to be positively and insignificantly related to income inequality in most specifications. The relationship between the $Ln(GRP\ per\ capita)$ and income inequality is negative and also insignificant across most specifications.

Table 4
OLS estimates of Gini index correlates using educational Gini index calculated from CHFS.

	Dependent variable: Gini index													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Educational Gini	0.458***		0.339***	0.427***	0.454**		0.364**	0.368**		0.134	0.308*		0.133	0.166
	(0.192)		(0.192)	(0.176)	(0.153)		(0.125)	(0.131)		(0.218)	(0.190)		(0.138)	(0.142)
Returns to schooling		0.916***	0.798***	0.820***		0.805***	0.719***	0.720***						
		(0.160)	(0.157)	(0.157)		(0.144)	(0.131)	(0.132)						
Educational Gini × Returns to Schooling				−6.169				−0.232						
				(2.906)				(2.766)						
Returns to high school									0.142***	0.125***	0.147***	0.149***	0.134***	0.137***
									(0.024)	(0.030)	(0.029)	(0.022)	(0.024)	(0.025)
Educational Gini × Returns to High School											−0.920**			−0.139
											(0.450)			(0.397)
Social spending CSY					−0.012**	−0.011*	−0.015***	−0.015***				−0.016***	−0.017***	−0.016***
					(0.006)	(0.006)	(0.005)	(0.006)				(0.005)	(0.005)	(0.006)
SOE assets share CSY					−0.107***	−0.098***	−0.070*	−0.070*				−0.096***	−0.089**	−0.085**
					(0.038)	(0.035)	(0.037)	(0.037)				(0.036)	(0.036)	(0.037)
Foreign trade share CSY					0.058*	−0.004	0.023	0.023				0.002	0.013	0.015
					(0.032)	(0.027)	(0.027)	(0.028)				(0.023)	(0.025)	(0.025)
Ln(GRP per capita) CSY					−0.050**	−0.019	−0.016	−0.016				−0.003	−0.005	−0.005
					(0.021)	(0.021)	(0.021)	(0.021)				(0.020)	(0.020)	(0.020)
2014.year	−0.001	0.014	0.014	0.016	0.012	0.018	0.020*	0.020*	0.006	0.006	0.007	0.012	0.013	0.012
	(0.012)	(0.011)	(0.011)	(0.010)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
2016.year	−0.025**	−0.029**	−0.026**	−0.025**	0.006	−0.014	−0.005	−0.005	−0.019*	−0.019*	−0.019*	−0.003	0.00002	−0.001
	(0.012)	(0.011)	(0.010)	(0.010)	(0.014)	(0.015)	(0.013)	(0.013)	(0.011)	(0.011)	(0.010)	(0.013)	(0.013)	(0.013)
2018.year	−0.030**	−0.025**	−0.027**	−0.029***	0.013	−0.004	0.003	0.003	−0.025**	−0.026**	−0.029***	−0.002	0.001	−0.001
	(0.012)	(0.012)	(0.011)	(0.011)	(0.019)	(0.019)	(0.018)	(0.018)	(0.011)	(0.011)	(0.011)	(0.017)	(0.017)	(0.018)
Constant	0.558***	0.554***	0.554***	0.556***	1.106***	0.799***	0.755***	0.755***	0.553***	0.554***	0.560***	0.628***	0.653***	0.646***
	(0.008)	(0.007)	(0.006)	(0.006)	(0.220)	(0.215)	(0.214)	(0.215)	(0.006)	(0.006)	(0.006)	(0.204)	(0.209)	(0.208)
Observations	116	116	116	116	116	116	116	116	116	116	116	116	116	116
R^2	0.221	0.333	0.414	0.444	0.414	0.485	0.553	0.553	0.371	0.380	0.453	0.561	0.568	0.569

Nested (cluster + generated-regressor) bootstrap SEs
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the regression results that account for provincial income inequality based on provincial human capital factors, as outlined in eq. (7).Standard errors in parentheses are from a nested bootstrap with 1,000 replicates (Cameron and Miller, 2015).

To examine the robustness of our results, we reestimate the baseline models from columns (7) and (13) of Table 4, excluding provinces identified as influential observations or those with small sample sizes. The results, presented in Table K1 in Appendix K, demonstrate that our core findings are robust to these sample alterations. Columns (2) and (5) exclude provinces with high Cook's distance—five provinces (Beijing, Guangdong, Hainan, Sichuan, Qinghai) and six (Beijing, Tianjin, Guangdong, Sichuan, Shanxi, Qinghai), respectively. Columns (3) and (6) exclude the five provinces with the smallest sample sizes (Hainan, Guizhou, Qinghai, Inner Mongolia, and Ningxia).[29] Across all specifications, returns to schooling and returns to high school retain their signs and statistical significance. The educational Gini coefficient retains its positive sign throughout and remains significant across all returns-to-schooling specifications (columns 1–3), but is significant in only one of the five returns-to-high-school specifications. The exclusion of influential provinces in column (5) results in an increase in the magnitude of the educational Gini coefficient, which becomes statistically significant at the 10 percent level; this suggests that our baseline estimate may, if anything, be conservative. Conversely, excluding provinces with small sample sizes in column (6) reduces the point estimate on the educational Gini, rendering it statistically insignificant. However, the estimate remains positive and the reduction is not substantial—it falls within the original confidence interval. Analysis in columns (7) and (8) demonstrates that this attenuation is primarily driven by the exclusion of Qinghai province, indicating that the educational inequality association is particularly pronounced there. Throughout all these tests, the coefficient on the educational Gini remains positive and never substantially loses magnitude, consistently supporting the hypothesis of a positive association. Finally, the coefficient on $Social\ spending$ remains negative and statistically significant across all specifications. $SOE\ assets\ share$ remains negative and significant throughout the returns-to-high-school specifications (columns 4–8), though it loses significance in two of three returns-to-schooling specifications (columns 2–3), while retaining its sign.

In an additional robustness check, we also estimate eq. (7) using educational Gini indices calculated from the China Population and Employment Statistics Yearbook (CPESY). The resulting estimates are reported in Table K2, Appendix K. Table K2 suggests that the positive association between educational inequality and income inequality remains significant and of a similar magnitude, as in our baseline results: a 10 pp. increase in the educational Gini is associated with 4.5 pp. increase in the income Gini index (column 5). Controlling for returns to schooling attenuates the coefficient on educational Gini to 2.6–3.0 pp.; it remains statistically significant at the 10% level once returns to schooling alone is added (column 7), but loses significance once provincial covariates are also included (column 8). The inclusion of returns to high school, however, renders the educational Gini coefficient statistically insignificant and reduces its magnitude to 1.1-1.2 pp. (columns 13-14). The main effects of returns to education on income inequality do not change in magnitude (7.5-8 pp. for returns to schooling and 1.4-1.5 pp. for returns to high school) and remain significant in all specifications. In addition, in our preferred specification with provincial-level covariates, we find a positive but statistically insignificant effect of the interaction between the educational Gini index estimated from CPESY and returns to education on income inequality. Table K2 confirms a statistically significant negative effect of government presence in the provincial economy ($SOE\ assets\ share$) across all specifications, and of government $Social\ spending$ in most specifications (columns 6–8, 12–14).

We also trace the possible influence of migration on our results in an additional robustness check. Employing the individual-level CHFS, we assign individuals to provinces based on their *hukou* instead of residence. We then recalculate the educational Gini, assuming no migration.[30] We then re-estimate our baseline income inequality regression from eq. (7) using the adjusted educational Gini measures. Figure K1 in Appendix K shows the comparison between predicted income Gini from our baseline regression (vertical axes) with predicted values employing educational Gini based on *hukou* (horizontal axes). The predicted income Gini values are correlated at 99%, which is indicative that migration has not altered our results.

The potential impact of educational policies

Our key result on the positive association between provincial educational inequality and income inequality suggests that targeted educational policies could help ameliorate income inequality. What are the main characteristics of educational inequality in China? In Table J1 (Appendix J), we report the results of educational inequality decompositions by province and birth cohorts. To perform this decomposition, we analyze individual-level data from the CHFS. In each available year of data for each province, we define 50 birth cohorts for every birth year of working-age adults (from 16 to 60 years old). We then apply our formula from eq. (3), performing the decomposition by

[29] Provincial-level sample sizes are summarized in Figure C1 in Appendix C.

[30] The percentage of those whose *hukou* location differs from their residence is around 3-6% of the CHFS samples.

population subgroups. First, we decompose the educational Theil T index by province, and second, we decompose it by birth cohort within each province. From columns (2), (6), (10), and (14) in Table J1, we can see that the within-province component accounts for the dominant share (98%) of variation in years of schooling in China. In columns (3)-(4), (7)-(8), (11)-(12), and (15)-(16) we report how the within-province Theil T is attributable to within- and between-birth cohort sub-components. Again, the within-birth cohort component describes the most significant part (80%) of the variation in years of schooling in each province.

In order to relate current provincial inequality levels to historical data on provincial educational expenditures, we also perform the following additional analysis. First, for every birth cohort in every province/year, we estimate average government educational expenditures, employing the data from the China Compendium of Statistics 1949-2008 (CCS), and the China Statistical Yearbook (CSY). For every birth cohort in every province we calculate average government expenditures[31] on education when the birth cohort was age 6 to 12. Second, we estimate the educational Gini index for each province-birth cohort. Third, we create a data set with an observation for each year, province, and birth cohort (4 years × 29 provinces × 50 birth cohorts = 5,800 observations) and run OLS to examine the relationship between educational expenditures and educational inequality.

Table 5
OLS estimates of the association between government educational expenditures and educational inequality

	Dependent variable:			
	Educational Gini by province/birth cohort			
	(1)	(2)	(3)	(4)
Average educational expenditures		−0.890***		−0.396***
		(0.025)		(0.027)
Constant	0.248***	0.260***	0.304***	0.306***
	(0.008)	(0.008)	(0.016)	(0.018)
Province dummies	Yes	Yes	No	No
Birth cohort dummies	No	No	Yes	Yes
Observations	5,800	5,584	5,800	5,584
R^2	0.336	0.539	0.441	0.442

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

Notes: This table presents the regression results that account for provincial educational inequality based on government educational expenditures. We use heteroskedasticity-robust standard errors. Data sources: CHFS, The China Compendium of Statistics 1949-2008, China Statistical Yearbook.

Table 5 reports the results. From column (2), we can see that an increase in average annual educational expenditures per capita in 1000 CN¥ is associated with an 8.9 pp. reduction in the educational Gini index, controlling for provincial fixed effects (the coefficient is significant at the 1% level). Adding annual educational expenditures per capita in column (2) to the provincial dummies significantly increases the explanatory power of the regression (R^2 increases from 0.336 to 0.539). Overall, the results from column (2) suggest that differences in educational spending within provinces across cohorts accounts for a larger part of the variation in educational inequality. In columns (3) and (4), we report the results of a similar analysis that employs birth-cohort dummies instead of provincial dummies. We can see that the relationship between educational spending and educational inequality, controlling for birth-cohort dummies, is also significant, but half the magnitude: an increase in average annual educational expenditures per capita in 1000 CN¥ is associated with a 4 pp. reduction in the educational Gini. And the R-squared has not increased in column (4) compared to column (3). Therefore, the variation in government educational spending across provinces within birth cohorts plays a smaller role in accounting for educational inequality compared to the variation within provinces across cohorts.

These results are consistent with the decomposition results, reported in Table J1 (Appendix J): within-province variation share is 98%, and within-birth cohort share in each province is around $0.8 \times 98\% = 78.4\%$. To summarize,

[31] Government expenditures are measured in 1000 CN¥ of annual spending per capita, deflated using the retail price index.

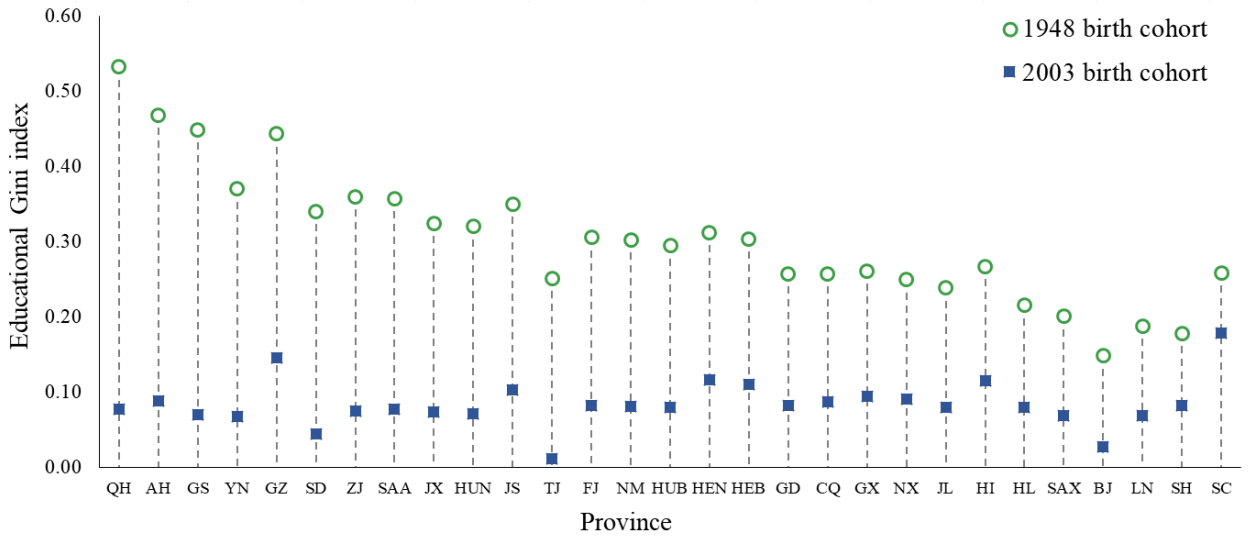


Figure 9: Provincial-level educational Gini index for the 1948 and 2003 birth cohorts. *Notes:* This figure presents the provincial educational Gini index for the earliest (1948) and the latest (2003) birth cohorts under consideration. The provinces on the x-axis are arranged in descending order of the difference in the educational Gini across birth cohorts. Data sources: The China Compendium of Statistics 1949-2008, China Statistical Yearbook.

the variation in government educational expenditures within provinces over time (across birth cohorts) accounts for the largest part of the variation in educational inequality in China. In particular, lower government spending on education for older cohorts compared to the higher spending for younger cohorts is the variable most strongly associated with educational inequality. Since we observe that increased educational government spending over time is associated with lower educational inequality, it is plausible that educational inequality may continue to decline in the future as older cohorts are replaced by younger cohorts in the labor market. This point is illustrated by Figure 9, where it can be seen that a substantial drop in educational inequality happened between the earliest (1948) and latest (2003) birth cohorts under consideration. The Educational Gini index for the 1948 birth cohort ranged between 0.149 in Beijing and 0.533 in Qinghai. Educational inequality for the 2003 birth cohorts ranged between 0.011 in Tianjin and 0.178 in Sichuan. The biggest drop in the absolute value of the Gini index happened in Qinghai and the lowest in Sichuan province. In relative terms, educational inequality dropped from 31% in Sichuan to 95% in Tianjin.

The role of secondary and tertiary education

To identify potential bottlenecks in existing educational policies, we further analyze how returns to different levels of education are associated with income inequality.[32] For this analysis, we introduce a returns to college variable, constructed analogously to the returns to high school from eq. (6) but using a dummy variable for a college degree or higher.

Table 6 presents results on the association between different educational levels and income inequality in China. Columns (1) to (3) indicate that returns to high school are a stronger correlate of income inequality than both aggregate returns to a year of education and returns to college, evident from the higher model fit (reported R^2) for the high-school specification. The inclusion of both premiums in column (4) shows that the economic magnitude of the high school premium is nearly twice that of the college premium (0.101 vs. 0.055), reinforcing the finding from the previous columns. When all three measures are included (column 5), the aggregate returns to schooling become statistically insignificant, indicating that its association in column (1) reflects primarily the underlying high school and college premiums rather than from years of education alone. Notably, in this saturated specification, returns to high school retain statistical significance at the 10% level while returns to college do not, and the point estimate of the high school premium remains larger (0.077 vs. 0.051). This finding of greater explanatory power of the high school premium for aggregate inequality is particularly salient given China's institutional context. While tertiary education falls under

[32] Table C4 in Appendix C summarizes per capita household income by educational attainment across provinces, illustrating the spatial variation in returns to different levels of education in China.

Table 6
OLS estimates of associations between different levels of education and the Gini index

	Dependent variable: Gini index				
	(1)	(2)	(3)	(4)	(5)
Educational Gini	0.364**	0.133	0.435**	0.202	0.236*
	(0.125)	(0.138)	(0.155)	(0.139)	(0.139)
Returns to schooling	0.719***				0.184
	(0.131)				(0.155)
Returns to high school		0.134***		0.101***	0.077*
		(0.024)		(0.023)	(0.028)
Returns to college			0.097***	0.055**	0.051
			(0.014)	(0.012)	(0.013)
Social spending CSY	−0.015***	−0.017***	−0.008	−0.013**	−0.013**
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
SOE assets share CSY	−0.070*	−0.089**	−0.090**	−0.084**	−0.078**
	(0.037)	(0.036)	(0.041)	(0.038)	(0.039)
Foreign trade share CSY	0.023	0.013	0.039	0.014	0.013
	(0.027)	(0.025)	(0.025)	(0.023)	(0.024)
Ln(GRP per capita) CSY	−0.016	−0.005	−0.029	−0.005	−0.005
	(0.021)	(0.020)	(0.019)	(0.019)	(0.019)
2014.year	0.020*	0.013	0.022**	0.018*	0.020**
	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)
2016.year	−0.005	0.00002	−0.008	−0.007	−0.008
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
2018.year	0.003	0.001	0.002	−0.002	−0.002
	(0.018)	(0.017)	(0.018)	(0.017)	(0.017)
Constant	0.755***	0.653***	0.881***	0.637***	0.636***
	(0.214)	(0.209)	(0.203)	(0.197)	(0.201)
Observations	116	116	116	116	116
R^2	0.553	0.568	0.532	0.597	0.600

Nested (cluster + generated-regressor) bootstrap SEs
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the regression results that account for provincial income inequality based on provincial human capital factors, as outlined in eq. (7). Standard errors in parentheses are from a nested bootstrap with 1,000 replicates (Cameron and Miller, 2015).

centralized, national-level jurisdiction, the provision and quality of K-12 education are primarily the responsibility of local governments. Therefore, while expanding college access remains important, our results indicate that policies targeting secondary education (K–12) are likely to be more relevant for reducing aggregate income inequality in China. Nevertheless, as we show in the following section, educational endowment alone is not the sole correlate of returns to education. Significant subnational variation in these returns which is associated with local labor market dynamics—including both supply- and demand-side factors—as well as the influence of internal migration.

7.2. Explaining variation in provincial returns to education

Next, we study the key factors associated with interprovincial variation in returns to schooling and returns to high school. We estimate the following model involving relative labor supply and labor demand factors:[33]

$$R_{jt} = \beta_0 + \beta_1 L_{jt} + \sum_{p=2}^{4} \beta_p D_{pjt} + \nu_t + \epsilon_{jt}, \tag{10}$$

where R_{jt} represents returns to schooling or returns to high school in province j in year t, a $\hat{\delta}_{1jt}$ estimate from eq. (6). L_{jt} is the relative supply of skilled labor in province j at year t. Following Yang (2005), for the baseline specification this variable is calculated as the share of workers with a middle-school and above education in the working-age population – $Skilled\ labor\ share$. To explain returns to high school we use the share of workers with high-school and above education – $Skilled\ labor\ share\ HS$. A larger supply of skilled labor is likely to shrink the earnings of skilled labor, and we expect to obtain a negative $\hat{\beta}_1$ estimate. D_{jt} are provincial relative labor demand factors. Including $Foreign\ trade\ share$ again to capture the openness of the local economy. As we already mentioned, in China greater trade openness is likely to increase the demand for unskilled labor. Therefore, we expect places with higher foreign trade shares in GRP to have higher returns on education. We also include $SOE\ assets\ share$ again to control for the size of the state sector in the provincial economy. A larger state sector implies a less market-driven local labor market, and therefore we expect it will be negatively associated with returns to schooling (returns to high school). Finally, we control for $Secondary\ employment$ - the share of employment in the secondary industry.[34]

Table 7 reports the OLS estimates of eq. (10). Columns (1) and (4) show the results from our baseline specification, where the $Skilled\ labor\ share$ is estimated from the CHFS. To check robustness, in columns (2) and (4) we use skilled labor share estimated from the China Population and Employment Statistics Yearbook (CPESY), and in columns (3) and (6) we used skilled labor share computed from the China Statistical Yearbook (CSY). $Skilled\ labor\ share\ CPESY$, therefore, measures the share of individuals with nine or more years of schooling among the employed population, and $Skilled\ labor\ share\ CSY$ - among those age 6 and above in the employed population. The relationship holds in sign and magnitude across all three sources of skilled labor share, and is statistically significant when measured from the independent CPESY or CSY data in both specifications. We find that provinces with a larger skilled labor share and a higher share of employment in the secondary industry have lower returns to schooling and lower returns to high school in China. A 10 pp. increase in the provincial share of skilled labor (above middle-school) is associated with a 1.2–1.8 pp. decrease in the returns to a year of schooling. The coefficient is statistically significant at the 1% level when skilled labor share is measured from the CPESY or CSY, but not when measured from the CHFS itself (column 1), where both the regressor and the dependent variable are generated from the same underlying microdata. Our nested bootstrap SEs account for this shared sampling error.[35] At the same time, a 10 pp. increase in the skilled labor share (above high-school) is associated with a 9.4–15.7 pp. decrease in the returns to high school. The coefficient is significant at the 1% level when measured from CPESY or CSY, and at the 10% level using the CHFS baseline (column 4).[36] In all specifications, the coefficients on secondary employment share are significant at the 1-5% level but have a lower magnitude than the skilled labor coefficients: a 10% increase in secondary industry employment share is associated with a 0.6-0.7 pp. decrease in returns to a year of schooling and to a 4.3-5.9 decrease in the returns to high school. We find $SOE\ assets\ share$ and $Foreign\ trade\ share$ to be insignificant but of the expected signs in most of the specifications.

Our independent variables are associated with differences in the level of provincial economic development. Considering possible factors correlated both with the level of economic development and with the returns to schooling, one source of potential omitted variable bias is educational quality.[37] Assuming that educational quality is higher in richer Chinese provinces and also increases the returns to schooling then coefficients on variables associated with the level of economic development could be upward biased.

To assess the robustness of our results, we reestimate the baseline models from columns (1) and (4) of Table 7, excluding provinces identified as influential observations or those with small sample sizes. The results, presented in Table K3 in Appendix K, indicate that our core findings are largely robust. Columns (2) and (7) exclude provinces with

[33] This approach is close to those discussed in Yang (2005), and Zhang et al. (2005).

[34] Secondary industry refers to mining, manufacturing, construction, production, and supply of electricity, water, and gas.

[35] Mean provincial returns to schooling are 0.138-0.157 in different years, and the standard deviation is around 0.03 (Table D3, Appendix D).

[36] Mean provincial returns to high school are 0.760-0.828, and the standard deviation is 0.2 (Table D3, Appendix D).

[37] Previous studies (e.g., Wei and Hao, 2011) emphasize the significant role of educational quality for human capital formation in China.

Table 7
OLS estimates of associations with returns to schooling and returns to high school.

	Dependent variable:					
	Returns to schooling			Returns to high school		
	(1)	(2)	(3)	(4)	(5)	(6)
Skilled labor share	−0.119					
	(0.065)					
Skilled labor share CPESY		−0.125***				
		(0.034)				
Skilled labor share CSY			−0.180***			
			(0.040)			
Skilled labor share HS				−1.226*		
				(0.606)		
Skilled labor share HS CPESY					−0.936***	
					(0.289)	
Skilled labor share HS CSY						−1.568***
						(0.394)
Secondary employment CSY	−0.064**	−0.058**	−0.073***	−0.426***	−0.453***	−0.588***
	(0.026)	(0.026)	(0.026)	(0.129)	(0.127)	(0.142)
SOE assets share CSY	−0.027	−0.033	−0.022	0.176	0.094	0.210
	(0.025)	(0.024)	(0.025)	(0.174)	(0.133)	(0.156)
Foreign trade share CSY	0.015	0.016	0.028**	0.116	0.136	0.213*
	(0.013)	(0.012)	(0.013)	(0.121)	(0.131)	(0.119)
2014.year	−0.015**	−0.017**	−0.016**	0.011	−0.034	−0.030
	(0.007)	(0.007)	(0.006)	(0.051)	(0.054)	(0.049)
2016.year	0.004	0.001	0.003	0.004	−0.012	−0.005
	(0.007)	(0.007)	(0.007)	(0.054)	(0.058)	(0.049)
2018.year	−0.008	−0.004	−0.002	−0.028	0.034	0.055
	(0.009)	(0.009)	(0.008)	(0.054)	(0.061)	(0.054)
Constant	0.275***	0.285***	0.308***	1.387***	1.264***	1.404***
	(0.034)	(0.026)	(0.027)	(0.146)	(0.107)	(0.123)
Observations	116	116	116	116	116	116
R^2	0.165	0.196	0.250	0.199	0.203	0.248

Nested (cluster + generated-regressor) bootstrap SEs
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the regression results on provincial returns to schooling and returns to high school based on provincial labor demand and labor supply factors, as formulated in eq. (10). Standard errors in parentheses are from a nested bootstrap with 1,000 replicates (Cameron and Miller, 2015).

high Cook's distance: four (Guangdong, Hunan, Hainan, Jilin) and three (Guangdong, Jilin, Qinghai), respectively. Columns (3) and (8) exclude the five provinces with the smallest sample sizes (Hainan, Guizhou, Qinghai, Inner Mongolia, and Ningxia).[38] Across most specifications, the coefficients on the key explanatory variables—including the share of workers with a middle-school education or above, the share with a high-school education or above in the working-age population, and the share of employment in the secondary industry—retain their signs and exhibit no substantial change in magnitude. The returns to high school coefficients retain their sign throughout, though significance is more sensitive to sample exclusions than the point estimates suggest: it holds at the 10% level or better in four of six specifications, but is lost when excluding either the smallest-sample provinces (column 8) or Qinghai alone (column 9). Similarly, the statistical significance of the key coefficients in the returns to schooling regression is sensitive to the inclusion of Qinghai province, as we show in columns (4)-(5). Qinghai has high returns to schooling and a small

[38] Provincial-level sample sizes are summarized in Figure C1 in Appendix C.

sample size, giving it high leverage.[39] A related pattern is observed in the returns to high school regression: the loss of significance in columns (8)-(9), despite the point estimates remaining within their original confidence intervals, is also primarily driven by Qinghai's exclusion. In sum, while the signs and magnitudes of the key coefficients are robust to the exclusion of influential or small-sample provinces, the statistical significance in the returns to schooling specification is not fully robust, due primarily to the influence of Qinghai province. However, since our main conclusions are based on returns to high school, and its coefficient's sign and magnitude are stable throughout — with significance held in four of six specifications, and lost only alongside Qinghai's exclusion — the core findings of the paper are not substantially affected.

We also repeat the robustness check that examines the potential role of the migration. We re-estimate skilled shares based on *hukou* province instead of the province of residence, and then we plugged them into eq. (10). Figures K2 and K3 in Appendix K plot predicted returns to schooling (returns to high school) from eq. (10) based on the province of residence versus based on the province of *hukou*. Predicted returns to schooling are 98% correlated. Predicted returns to high school are correlated at 97%. Therefore, we conclude that our results with and without migration, again, are very similar.

8. Simulation study

Finally, we conduct Monte Carlo simulations aimed at studying how income inequality in Chinese provinces would change under scenarios involving: (a) equalization of educational structure; (b) equalization of returns to education; and (c) both equalization of educational structure and returns to education. Our analysis uses a partial-equilibrium counterfactual simulation, in the tradition of counterfactual decompositions built on estimated earnings equations (Almeida dos Reis and Paes de Barros, 1991; Juhn, Murphy and Pierce, 1993; Sicular et al., 2007) and their extension to richer multi-equation microsimulation systems (Bourguignon, Fournier and Gurgand, 2001; Bourguignon, Ferreira and Leite, 2008); see Fortin, Lemieux and Firpo (2011) for a survey of this literature. The approach recomputes the distributional statistic of interest under a counterfactual value of a given factor — here, the provincial educational structure or the provincial returns to education — while holding the remaining estimated relationships fixed. This allows us to quantify how the distributional outcome would differ under alternative, counterfactual values of the factors considered. As a partial-equilibrium exercise, the simulation holds behavioral responses and other relationships outside the modeled framework fixed, and therefore does not capture migration responses or other general-equilibrium effects. The resulting simulated changes should consequently be interpreted as accounting relationships implied by our modeling assumptions, rather than as the causal effects of specific policies or interventions. For this analysis, we again utilize the Theil T inequality index.

We can reformulate eq. (3) for a decomposition by province:

$$T = \sum_{j=1}^{29} \frac{N_j \bar{y}_j}{N \bar{y}} T_j + \sum_{j=1}^{29} \frac{N_j \bar{y}_j}{N \bar{y}} \log\left(\frac{\bar{y}_j}{\bar{y}}\right), \tag{11}$$

where T is the national Theil T index, T_j is a Theil T index of each province j (in our case $j = 1, ..., 29$), N and $\bar{y}$ are the total number of households and the nationwide mean annual household income per capita, and N_j and $\bar{y}_j$ are the number of households and the mean household income per capita in each province j.

Each provincial Theil T index is estimated using the standard formula:

$$T_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{y_{ij}}{\bar{y}_j} \log\left(\frac{y_{ij}}{\bar{y}_j}\right), \tag{12}$$

where y_{ij} is log income of household i from province j.

[39] As shown in Table K3, Qinghai is not flagged as an outlier by Cook's distance in the returns to schooling regression—the specification where its exclusion affects significance levels. Its influence stems from high leverage (small sample size and distinctive characteristics) rather than poor fit: it follows the overall pattern but extends the range of variation in the independent variables, making its inclusion important for statistical power.

Incorporating (12) in (11) we obtain:

$$T = \sum_{j=1}^{29} \frac{\bar{y}_j}{N\bar{y}} \left(\sum_{i=1}^{N_j} \frac{y_{ij}}{\bar{y}_j} \log\left(\frac{y_{ij}}{\bar{y}_j}\right) \right) + \sum_{j=1}^{29} \frac{N_j \bar{y}_j}{N\bar{y}} \log\left(\frac{\bar{y}_j}{\bar{y}}\right) = \\ = \frac{1}{N\bar{y}} \left(\sum_{j=1}^{29} \sum_{i=1}^{N_j} y_{ij} \log\left(\frac{y_{ij}}{\bar{y}_j}\right) + \sum_{j=1}^{29} N_j \bar{y}_j \log\left(\frac{\bar{y}_j}{\bar{y}}\right) \right) \tag{13}$$

At the same time, each y_{ij} can be derived from the eq. (6) Mincer-type regression. This setup enables us to perform the following policy simulations employing the most recent (2019) wave of the CHFS.

First, to study the associations under equalization of educational structure on income inequality we randomly reassign years of schooling to individuals so that the educational structure in each province is the same as the national educational structure reported in the China Population and Employment Statistics Yearbook (CPESY). Then we aggregate the individual-level simulated data to households. This adjusts the household schooling variable $\tilde{S}_{ijt}$ as well as high-school or above dummy variable $\tilde{H}_{ijt}$. Using the updated $\tilde{S}_{ijt}$ (or $\tilde{H}_{ijt}$) and the prior estimated regression coefficients from eq. (6), we predict an adjusted income variable $\tilde{y}_{ij}$. Incorporating the new $\tilde{y}_{ij}$ into (13) we can re-estimate provincial income inequality using the Theil $\tilde{T}$ index.

Second, since equalization of education results in different skilled labor shares in each province, after obtaining new $\tilde{S}_{ijt}$ we can re-estimate skill shares $\tilde{L}_{jt}$. Again, keeping regression coefficients and the error term estimated from the initial data constant, and using adjusted $\tilde{L}_{jt}$, we can predict new returns to schooling or returns to high school $\tilde{R}_{jt}$ from (10). Then we can incorporate both adjusted schooling $\tilde{S}_{ijt}$ (or $\tilde{H}_{ijt}$) and adjusted returns to schooling $\tilde{R}_{jt}$ into (6). This again gives us new $\tilde{y}_{ij}$, from which we can re-estimate $\tilde{T}$ from eq. (13).

Third, we can consider equalization of returns to education only. This scenario implies an exogenous equalization of returns to schooling or returns to high school to the national-level returns, without any adjustments to the provincial educational structure. National-level returns to education are obtained by running the (6) OLS at the household-level data from the CHFS 2019.

Finally, we consider both educational equalization and equalization of returns to education. For that, we repeat the same educational equalization from (1) and perform an exogenous equalizing change in the returns to schooling (or returns to high school) to the national-level returns. All of the subsequent procedures remain the same as in (2). Through this simulation, we answer the question: what would happen to the Theil T index if both educational structure and returns to schooling (or returns to high school) became similar across provinces? Given our prior finding that returns to high school play a dominant role in explaining provincial income inequality, this section presents policy simulations targeting this specific premium. Results for the returns to schooling are conceptually similar but, as expected, exhibit a smaller magnitude (these are reported in Appendix L, Tables L3 and L2).

The results of the policy simulations are reported in Table 8. Columns (1) and (2) report our baseline results for the income inequality decomposition in 2018 (from Table 1). Columns (3) and (4) show the results from our first simulation scenario, where we set the distribution of education in each province equal to the nationwide educational distribution reported in CPESY. In comparison to the initial Theil T decomposition results, equalization of the educational structure is associated with a 10.8% decrease in the nationwide Theil T index from 0.462 to 0.412. The within-province component of nationwide household income per capita variation increases from 86% to 89%. The poor to rich province income inequality ratio declines narrowly from 1.34 (0.457 in Gansu divided by 0.342 in Beijing) to 1.22 (0.433 in Gansu divided by 0.354 in Beijing).

The results from our second simulation scenario are presented in columns (5) and (6). Equalization of educational structure among Chinese provinces, together with adjusting returns to high school, is associated with an almost similar decrease in nationwide income inequality (−15.2%) and an increase in the within-province variation component (from 86% to 89%). However, as opposed to the first simulation, the poor-to-rich province inequality ratio slightly increases from 1.34 to 1.40.

The results of our third simulation study are reported in columns (7) and (8). Exogenous equalization of the provincial returns to high school to the national-level returns is associated with a 1.7% increase in inequality, and to increase of the importance of the between-province component of national-level inequality, from 14% to 19%. This happens because increases in income inequality in provinces with low returns to high school (34% increase of Theil T in Shanghai, 2% increase in Hunan, etc.) slightly outweighs the inequality decline in provinces with high returns to high

Table 8
Theil T from policy simulations: equalizing provincial education structures and returns to high school to national levels.

	Baseline Estimates 2018		(a) Uniform educational structure		(b) Uniform educational structure and adjusted RS		(c) Uniform RS		(d) Uniform educational structure and uniform RS	
	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Beijing	0.342	4%	0.354	3%	0.332	3%	0.384	4%	0.360	3%
Shanghai	0.197	2%	0.259	2%	0.190	2%	0.264	3%	0.248	2%
Jiangsu	0.353	7%	0.339	6%	0.299	6%	0.343	6%	0.331	6%
Tianjin	0.256	1%	0.285	1%	0.266	1%	0.303	1%	0.286	1%
Zhejiang	0.376	6%	0.337	6%	0.304	6%	0.374	6%	0.339	6%
Fujian	0.393	3%	0.361	3%	0.350	3%	0.369	2%	0.361	3%
Guangdong	0.510	9%	0.417	9%	0.392	9%	0.496	10%	0.418	9%
Shandong	0.458	6%	0.408	6%	0.371	6%	0.405	5%	0.404	6%
Inner Mongolia	0.385	2%	0.351	2%	0.356	2%	0.343	2%	0.350	2%
Hubei	0.328	3%	0.307	3%	0.272	3%	0.326	3%	0.313	3%
Chongqing	0.454	2%	0.404	2%	0.366	2%	0.425	2%	0.406	2%
Shaanxi	0.460	2%	0.381	2%	0.415	2%	0.365	1%	0.375	2%
Liaoning	0.290	3%	0.318	3%	0.313	3%	0.291	2%	0.317	3%
Hunan	0.329	3%	0.318	4%	0.265	3%	0.337	4%	0.315	4%
Anhui	0.378	3%	0.314	4%	0.296	4%	0.348	3%	0.317	3%
Hainan	0.493	0%	0.445	0%	0.478	1%	0.435	0%	0.440	0%
Ningxia	0.350	0%	0.338	0%	0.312	0%	0.346	0%	0.335	0%
Henan	0.533	6%	0.451	6%	0.468	6%	0.467	5%	0.444	5%
Sichuan	0.506	6%	0.441	7%	0.446	7%	0.439	5%	0.432	6%
Jilin	0.341	1%	0.335	1%	0.297	1%	0.370	1%	0.335	2%
Jiangxi	0.330	2%	0.305	2%	0.276	2%	0.323	2%	0.307	2%
Hebei	0.394	3%	0.351	3%	0.318	3%	0.368	3%	0.348	3%
Shanxi	0.430	2%	0.387	2%	0.352	2%	0.392	2%	0.383	2%
Qinghai	0.477	0%	0.435	0%	0.465	1%	0.419	0%	0.430	0%
Heilongjiang	0.361	2%	0.358	3%	0.335	3%	0.363	2%	0.355	3%
Yunnan	0.491	3%	0.395	2%	0.459	3%	0.382	2%	0.370	2%
Guangxi	0.470	2%	0.437	2%	0.433	2%	0.447	2%	0.437	2%
Guizhou	0.497	2%	0.457	2%	0.435	2%	0.460	2%	0.455	2%
Gansu	0.457	1%	0.433	1%	0.464	1%	0.407	1%	0.424	1%
Within	0.397	86%	0.367	89%	0.347	89%	0.381	81%	0.362	86%
Between	0.065	14%	0.045	11%	0.044	11%	0.089	19%	0.058	14%
Total	0.462	100%	0.412	100%	0.392	100%	0.469	100%	0.421	100%

Notes: This table reports the results of the policy simulation study based on the data from the CHFS 2019. Monte Carlo Simulations were repeated 1000 times.

school (−12% in Qinghai, −22% in Yunnan, etc.). The weights of changes in provincial Theil T in national changes in inequality is based on the relative importance of each province in explaining nationwide inequality, reported in column (2). As already mentioned, that returns to high school in China are negatively associated with GRP per capita. As a result, equalization of returns to high school is associated with inequality increasing in richer provinces and decreasing in poorer provinces. This reduces by 20.6% the poor-to-rich province income inequality ratio from 1.34 to 1.06.

The last two columns in Table 8 report the results of our fourth simulation. Equalization of the educational structure together with equalization of the returns to high school is associated with a 8.9% decline in national income inequality. The between-province component of nationwide inequality remains constant. The poor-to-rich province Theil T index ratio, however, drops slightly compared to the baseline results (1.18 versus 1.34).

Instead of equalizing provincial educational structure and returns to high school at the national level, we could consider a policy that targets returns to high school and educational structure to make them similar to the current levels in one of the leading and highly developed provinces. For example, Shanghai is one of the top provinces by the level of GRP per capita with high levels of average schooling among its population (Figure 5), low educational inequality (Figure 4) and low returns to high school (Figure 7). All of these are desirable properties for reducing inequality. The results of this additional simulation analysis are reported in Table L1, Appendix L. If all provinces had an educational

structure similar to that of Shanghai, it would reduce income inequality by 16% (Scenario "a"). If the equalization of educational structure is accompanied by adjusting returns to high school (Scenario "b"), inequality would drop by 19.8%. If returns to high school in all provinces were as low as in Shanghai (Scenario "c"), income inequality would decline by 12.4%. Finally, if both returns to high school and educational structure in all provinces were the same as Shanghai (Scenario "d"), inequality would decline by 20.1%.

9. Conclusion

In this paper, we provide robust Gini index estimates for 29 mainland Chinese provinces (excluding Xinjiang and Tibet), using four waves of the China Household Finance Survey (CHFS 2013, 2015, 2017, and 2019). Building upon the first provincial-level inequality study by Hussain et al. (1994), we find that intraprovincial inequality accounts for the majority of national inequality, capturing around 90% of the variation in annual household income per capita.

Our key empirical results suggest that human capital disparities are strongly associated with differences in the levels of income inequality across Chinese provinces. Inequality in educational attainment and differences in the returns to education are positively and statistically significantly related to provincial income inequality. Using data for the period 2012-2018, we find that a 10 percentage point (pp.) increase in provincial-level educational Gini is associated with increase of 1.3-4.5 pp. in the income Gini index. We also find a 10 pp. increase in returns to schooling is associated with a 7.2–8.0 pp. increase in the income Gini index, compared to a 1.3–1.5 pp. increase for returns to high school. These results are consistent with findings from a recent study by Nie and Xing (2019), where returns to education are identified as strongly associated with income inequality in China (but only for urban households). Our results provide empirical support for Heckman (2005)'s argument that disparities in both human capital endowments and the returns to education closely linked to regional economic development gaps and inequality in China. In addition, we find that returns to high school are more strongly associated with income inequality than returns to college. We also find that the state's presence in the local economy and government social spending are negatively associated with provincial income inequality.

We argue that existing labor market frictions are associated with the interprovincial differences in returns to schooling and returns to high school and above in China. In a fully integrated labor market, returns to education across provinces are expected to be similar. However, in China, there still exists significant labor market segmentation, although it is declining. We found a negative relationship between the share of skilled labor and provincial returns to schooling and returns to high school, significant when skilled labor share is measured from CPESY or CSY. A 10 pp. increase in the skilled labor share is associated with a decrease of 1.2-1.8 pp. in returns to schooling, and a decrease of 9.4-15.7 pp. in the returns to high school and above. We also find that the secondary industry employment share is negatively associated with returns to schooling and returns to high school and above. These results align with prior findings by Meng and Wu (1998), that an important correlate of regional income disparities is the degree of regional marketization in China.

Lastly, we find that poor provinces are severely disadvantaged compared to rich provinces in China. Poor provinces experience higher income and educational inequality, and their citizens have comparatively lower levels of education. Our counterfactual simulation implies that equalizing educational structure across provinces to the national level lowers simulated income inequality by 10.8%; equalizing returns to high school in addition to educational structure lowers it by 8.9%; and equalizing both to the levels observed in Shanghai, one of the most developed provinces, lowers it by 20.1%. These are accounting exercises, detailed in Section 8, that hold the estimated Mincer relationships fixed under a partial-equilibrium assumption. These findings are particularly salient given that K-12 education falls primarily under local government jurisdiction in China, while tertiary education is centrally managed. This institutional context suggests that policies targeting high school attendance and quality at the provincial level are likely to be more relevant to aggregate income inequality. These findings are consistent with the results from previous studies showing that strengthening human capital is an important policy priority for regional development in China (Fraumeni et al., 2019; Fleisher et al., 2010). Targeted educational policies can also be justified by the desire to promote greater equality of opportunity by reducing inequities in access to public education and health services. The persistence of interprovincial differences in returns to education despite rising migration also suggests that facilitating efficient re-allocations of labor could help close such gaps which are associated with greater income equality. The magnitudes of these simulation results are based on statistical associations and so should not be interpreted as the causal impact of implementing any specific policy reform.

CRediT authorship contribution statement

Kristina Butaeva: Writing – review & editing, Writing– original draft, Conceptualization, Methodology, Formal analysis, Data curation, Visualization. **Albert F. Park:** Writing – review & editing, Writing– original draft, Conceptualization, Methodology, Formal analysis.

References

Alesina, A., Perotti, R., 1996. Income Distribution, Political Instability, and Investment. European Economic Review 40, 1203–1228.

Asadullah, M.N., Xiao, S., 2020. The Changing Pattern of Wage Returns to Education in Post-reform China. Structural Change and Economic Dynamics 53, 137–148.

Bourguignon, F., 1979. Decomposable Income Inequality Measures. Econometrica: Journal of the Econometric Society 47, 901–920.

Bourguignon, F., Ferreira, Francisco H. G., Leite, P.G., 2008. Beyond Oaxaca–Blinder: Accounting for Differences in Household Income Distributions. Journal of Economic Inequality 6, 117–148.

Bourguignon, F., Fournier, M., Gurgand, M., 2001. Fast Development with a Stable Income Distribution: Taiwan, 1979–94. Review of Income and Wealth 47, 139–163.

Butaeva, K., 2025. Income Inequality in China, 2012–2018: A New Measurement Approach. Working Paper 25-08. The University of Chicago Stone Center. URL: https://ssrn.com/abstract=5185428, doi:10.2139/ssrn.5185428.

Cai, F., Park, A., Zhao, Y., 2008. The Chinese Labor Market in the Reform Era, in: Brandt, L., Rawskin, T.G. (Eds.), China's Great Economic Transformation. Cambridge University Press Cambridge, Cambridge, pp. 167–214.

Cameron, A.C., Miller, D.L., 2015. A Practitioner's Guide to Cluster-Robust Inference. Journal of Human Resources 50, 317–372.

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. University Press, Cambridge.

Chan, W.K., Ngok, K., 2011. Accumulating Human Capital While Increasing Educational Inequality: a Study on Higher Education Policy in China. Asia Pacific Journal of Education 31, 293–310.

Charpentier, A., Flachaire, E., 2022. Pareto Models for Top Incomes. The Journal of Economic Inequality 20, 1–25.

Chong, A., Gradstein, M., 2017. Political and Economic Inequities and the Shaping of Institutions and Redistribution. Southern Economic Journal 83, 952–971.

Churchill, S.A., Mishra, V., 2018. Returns to Education in China: a Meta-analysis. Applied Economics 50, 5903–5919.

Coady, D., Dizioli, A., 2018. Income Inequality and Education Revisited: Persistence, Endogeneity and Heterogeneity. Applied Economics 50, 2747–2761.

Coles, S.G., Dixon, M.J., 1999. Likelihood-based Inference for Extreme Value Models. Extremes 2, 5–23.

Cowell, F.A., 1980. On the Structure of Additive Inequality Measures. The Review of Economic Studies 47, 521–531.

Cowell, F.A., 2000. Measurement of Inequality. Handbook of Income Distribution 1, 87–166.

Cowell, F.A., Flachaire, E., 2007. Income Distribution and Inequality Measurement: The Problem of Extreme Values. Journal of Econometrics 141, 1044–1072.

Cowell, F.A., Victoria-Feser, M.P., 2007. Robust Stochastic Dominance: A Semi-parametric Approach. The Journal of Economic Inequality 5, 21–37.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application. University Press, Cambridge.

Esteban, J., Ray, D., 1999. Conflict and Distribution. Journal of Economic Theory 87, 379–415.

Fields, G.S., 2003. Accounting for Income Inequality and Its Change: A New Method, with Application to the Distribution of Earnings in the United States, in: Polachek, S.W. (Ed.), Worker Well-being and Public Policy. Emerald Group Publishing Limited, Bingley, pp. 1–38.

Fleisher, B., Li, H., Zhao, M.Q., 2010. Human Capital, Economic Growth, and Regional Inequality in China. Journal of Development Economics 92, 215–231.

Fleisher, B., Yang, D., 2008. China's Evolving Labor Market, in: Fleisher, B., Hope, N., Pena, A., Yang, D. (Eds.), Policy Reform and Chinese Markets: Progress and Challenges. Edward Elgar Publishing, pp. 125–158.

Florida, R., Mellander, C., Rentfrow, P.J., 2013. The Happiness of Cities. Regional Studies 47, 613–627.

Fortin, N., Lemieux, T., Firpo, S., 2011. Decomposition Methods in Economics, in: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. Elsevier. volume 4A, pp. 1–102.

Fraumeni, B.M., He, J., Li, H., Liu, Q., 2019. Regional Distribution and Dynamics of Human Capital in China 1985–2014. Journal of Comparative Economics 47, 853–866.

Fu, Q., Ren, Q., 2010. Educational Inequality Under China's Rural–urban divide: The Hukou System and Return to Education. Environment and Planning A 42, 592–610.

Galor, O., Moav, O., 2006. Das Human-kapital: A Theory of the Demise of the Class Structure. The Review of Economic Studies 73, 85–117.

Gan, L., Yin, Z., Jia, N., Xu, S., Ma, S., Zheng, L., et al., 2014. Data You Need to Know About China. Springer, New York.

Gao, Q., Yang, S., Zhai, F., 2019. Social Policy and Income Inequality During the Hu–wen Era: a Progressive Legacy? The China Quarterly 237, 82–107.

Goh, C.c., Xubei, L., Nong, Z., 2009. Income Growth, Inequality and Poverty Reduction: A Case Study of Eight Provinces in China. China Economic Review 20, 485–496.

Golley, J., Kong, S.T., 2018. Inequality of Opportunity in China's Educational Outcomes. China Economic Review 51, 116128.

Gradín, C., Wu, B., 2020. Income and Consumption Inequality in China: A Comparative Approach With India. China Economic Review 62, 101463.

Gustafsson, B., Wan, H., 2020. Wage Growth and Inequality in Urban China: 1988–2013. China Economic Review 62, 101462.

Hao, T., Sun, R., Tombe, T., Zhu, X., 2020. The Effect of Migration Policy on Growth, Structural Change, and Regional Inequality in China. Journal of Monetary Economics 113, 112–134.

Hauser, S.M., Xie, Y., 2005. Temporal and regional variation in earnings inequality: Urban china in transition between 1988 and 1995. Social Science Research 34, 44–79.

Heckman, J.J., 2005. China's Human Capital Investment. China Economic Review 16, 50–70.

Heckman, J.J., Lochner, L.J., Todd, P.E., 2006. Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, in: Hanushek, E.A., Woessmann, L., Machin, S.J. (Eds.), Handbook of the Economics of Education. Elsevier. volume 1, pp. 307–458.

Heckman, J.J., Yi, J., 2012. Human Capital, Economic Growth, and Inequality in China. URL: http://www.nber.org/papers/w18100, doi:10.3386/w18100. working Paper Series, No. 18100.

Huang, C., Lin, J.G., Ren, Y.Y., 2012. Statistical Inferences for Generalized Pareto Distribution Based on Interior Penalty Function Algorithm and Bootstrap Methods and Applications in Analyzing Stock Data. Computational Economics 39, 173–193.

Hussain, A., Lanjouw, P., Stern, N., 1994. Income Inequalities in China: Evidence from Household Survey Data. World Development 22, 1947–1957.

Jain-Chandra, M.S., Khor, N., Mano, R., Schauer, J., Wingender, M.P., Zhuang, J., 2018. Inequality in China–Trends, Drivers and Policy Remedies. Working Paper No. 2018/127.

Jenkins, S.P., 2017. Pareto Models, Top Incomes and Recent Trends in UK Income Inequality. Economica 84, 261–289.

Juhn, C., Murphy, K.M., Pierce, B., 1993. Wage Inequality and the Rise in Returns to Skill. Journal of Political Economy 101, 410–442.

Kanbur, R., Wang, Y., Zhang, X., 2021. The Great Chinese Inequality Turnaround. Journal of Comparative Economics 49, 467–482.

Khan, A.R., Riskin, C., 2005. China's Household Income and its Distribution, 1995 and 2002. The China Quarterly , 356–384.

Lei, X., Shen, Y., 2015. Inequality in Educational Attainment and Expectation: Evidence from the China Family Panel Studies. China Economic Journal 8, 252–263.

Li, H., 2003. Economic transition and returns to education in china. Economics of Education Review 22, 317–328.

Li, Q., Li, S., Wan, H., 2020. Top Incomes in China: Data Collection and the Impact on Income Inequality. China Economic Review 62, 101495.

Li, S., Whalley, J., Xing, C., 2014. China's Higher Education Expansion and Unemployment of College Graduates. China Economic Review 30, 567–582.

Lü, X., 2012. Equality of Educational Opportunity and Attitudes Toward Income Inequality: Evidence from China. Quarterly Journal of Political Science 8, 271–303.

Martins, E.S., Stedinger, J.R., 2000. Generalized Maximum-likelihood Generalized Extreme-value Quantile Estimators for Hydrologic Data. Water Resources Research 36, 737–744.

Martins, E.S., Stedinger, J.R., 2001. Generalized Maximum Likelihood Pareto-Poisson Estimators for Partial Duration Series. Water Resources Research 37, 2551–2557.

Mendoza, O.M.V., 2018. Heterogeneous Determinants of Educational Achievement and Inequality Across Urban China. China Economic Review 51, 129–148.

Meng, X., Wu, H.X., 1998. Household Income Determination and Regional Income Differential in Rural China. Asian Economic Journal 12, 35–63.

Mincer, J., 1974. Schooling, Experience, and Earnings. National Bureau of Economic Research, New York.

Murphy, K.M., Topel, R.H., 1985. Estimation and Inference in Two-Step Econometric Models. Journal of Business & Economic Statistics 3, 370–379.

Naughton, B.J., 2006. The Chinese Economy: Transitions and Growth. MIT press, Cambridge, MA.

Ngai, L.R., Pissarides, C.A., Wang, J., 2019. China's Mobility Barriers and Employment Allocations. Journal of the European Economic Association 17, 1617–1653.

Nie, H., Xing, C., 2019. Education Expansion, Assortative Marriage, and Income Inequality in China. China Economic Review 55, 37–51.

Pagan, A., 1984. Econometric Issues in the Analysis of Regressions with Generated Regressors. International Economic Review 25, 221–247.

Park, A., Song, X., Zhang, J., Zhao, Y., 2003. The Growth of Wage Inequality in Urban China, 1988 to 1999. Unpublished Paper, Chinese University of Hong Kong.

Pauli, F., Coles, S., 2001. Penalized Likelihood Inference in Extreme Value Analyses. Journal of Applied Statistics 28, 547–560.

Peng, H., Qi, L., Wan, G., Li, B., Hu, B., 2020. Child Population, Economic Development and Regional Inequality of Education Resources in China. Children and Youth Services Review 110, 104819.

Perotti, R., 1996. Growth, Income Distribution, and Democracy: What the Data Say. Journal of Economic growth 1, 149–187.

Ravallion, M., Chen, S., 2009. China's (Uneven) Progress Against Poverty, in: Kanbur, R., Zhang, X. (Eds.), Governing Rapid Growth in China. Routledge, London, pp. 65–111.

Almeida dos Reis, J.G., Paes de Barros, R., 1991. Wage Inequality and the Distribution of Education: A Study of the Evolution of Regional Differences in Inequality in Metropolitan Brazil, journal = Journal of Development Economics 36, 117–143.

Savoia, A., Easaw, J., McKay, A., 2010. Inequality, Democracy, and Institutions: A Critical Review of Recent Research. World Development 38, 142–154.

Shorrocks, A.F., 1984. Inequality Decomposition by Population Subgroups. Econometrica 52, 1369–1385.

Sicular, T., Yue, X., Gustafsson, B., Li, S., 2007. The Urban–Rural Income Gap and Inequality in China. Review of Income and Wealth 53, 93–126.

Sun, W., Lu, J., Bai, C., 2007. Analysis of Income Flow in Rural China (in Chinese). Economic Research 08, 43–57.

Tombe, T., Zhu, X., 2019. Trade, Migration, and Productivity: A Quantitative Analysis of China. American Economic Review 109, 1843–1872.

Van Kerm, P., 2007. Extreme Incomes and the Estimation of Poverty and Inequality Indicators from EU-SILC. IRISS Working Papers; No. 2007-01.

Victoria-Feser, M.P., 2000. Robust Methods for the Analysis of Income Distribution, Inequality and Poverty. International Statistical Review 68, 277–293.

Victoria-Feser, M.P., Ronchetti, E., 1994. Robust Methods for Personal-income Distribution Models. Canadian Journal of Statistics 22, 247–258.

Wan, G., Zhou, Z., 2005. Income Inequality in Rural China: Regression-based Decomposition Using Household Data. Review of Development Economics 9, 107–120.

Wang, C., Wan, G., Yang, D., 2014. Income Inequality in the People's Republic of China: Trends, Determinants, and Proposed remedies. Journal of Economic Surveys 28, 686–708.

Wang, Z., 2006. Study on the Estimation of Gini Coefficient in the Six Provinces of Central China (in Chinese). Zhongguo She Hui Ke Xue 4, 77–87.

Wei, Z., Hao, R., 2011. The Role of Human Capital in China's Total Factor Productivity Growth: A Cross-Province Analysis. The Developing Economies 49, 1–35.

Weimin, T., 2012. The Gini Coefficient of Provincial Residents' Income and Its Change Trend Analysis (in Chinese). Economic Science 02, 48–59.

Xiang, L., Stillwell, J., Burns, L., Heppenstall, A., 2020. Measuring and Assessing Regional Education Inequalities in China Under Changing Policy Regimes. Applied Spatial Analysis and Policy 13, 91–112.

Xie, Y., Hannum, E., 1996. Regional Variation in Earnings Inequality in Reform-Era Urban China. American Journal of Sociology 101, 950–992.

Xie, Y., Lu, P., 2015. The Sampling Design of the China Family Panel Studies (CFPS). Chinese Journal of Sociology 1, 471–484.

Xie, Y., Zhou, X., 2014. Income Inequality in Today's China. Proceedings of the National Academy of Sciences 111, 6928–6933.

Xu, L.C., Zou, H.f., 2000. Explaining the Changes of Income Distribution in China. China Economic Review 11, 149–170.

Xu, Q., Zhang, C., Zhou, X., Xie, Y., 2012. Comparison of Income Data from CGSS, CHIP, CHFS, and CFPS. Technical Report CFPS-16. China Family Panel Studies, Peking University. URL: https://www.isss.pku.edu.cn/cfps/docs/20180927133317005706.pdf. in Chinese.

Yang, D.T., 2005. Determinants of Schooling Returns During Transition: Evidence from Chinese Cities. Journal of Comparative Economics 33, 244–264.

Yang, J., Huang, X., Liu, X., 2014. An Analysis of Education Inequality in China. International Journal of Educational Development 37, 2–10.

Yu, N., Yu, B., de Jong, M., Storm, S., 2015. Does Inequality in Educational Attainment Matter for China's Economic Growth? International Journal of Educational Development 41, 164–173.

Zhang, J., 2021. A Survey on Income Inequality in China. Journal of Economic Literature 59, 1191–1239.

Zhang, J., Zhao, Y., Park, A., Song, X., 2005. Economic Returns to Schooling in Urban China, 1988 to 2001. Journal of Comparative Economics 33, 730–752.

Appendix

A. Empirical evidence on provincial income inequality

Table A1
Summary of provincial Gini index estimates in China.

Article	Data source	Year	Rural/ Urban	Gini summary statistics				
				Min	Max	Mean	SD	#provinces
Hussain et al. (1994)	Based on the Chinese Academy of Social Sciences (CASS) survey of households 1986.	1986	Urban	0.180	0.243	0.212	0.017	23
		1986	Rural	0.195	0.281	0.241	0.026	10
Xu and Zou (2000)	Based on the published results of average incomes of different percentiles for urban residents in each province (World Bank, 1996).	Mean 1985-1995	Urban	0.173	0.279	0.208	0.023	29
Khan and Riskin (2005)	Based on the Chinese Academy of Social Sciences (CASS) survey of households.	1995	Urban	0.272	0.432	0.331	0.045	19
		2002	Urban	0.235	0.409	0.309	0.041	21
		1995	Rural	0.223	0.482	0.280	0.074	11
		2002	Rural	0.250	0.328	0.282	0.026	11
Sun, Lu and Bai (2007)	Based on an annual fixed-point survey of farmers conducted jointly by the Central Policy Research Office and the Ministry of Agriculture.	1986	Rural	0.223	0.413	0.322	0.074	6
		1987	Rural	0.185	0.386	0.288	0.070	6
		1988	Rural	0.213	0.363	0.296	0.057	6
		1989	Rural	0.224	0.370	0.295	0.059	6
		1990	Rural	0.234	0.388	0.294	0.057	6
		1991	Rural	0.234	0.420	0.321	0.064	6
		1995	Rural	0.345	0.369	0.360	0.011	6
		1996	Rural	0.363	0.420	0.385	0.021	6
		1997	Rural	0.367	0.394	0.380	0.011	6
		1998	Rural	0.356	0.393	0.378	0.013	6
		1999	Rural	0.366	0.402	0.379	0.014	6
		2000	Rural	0.356	0.421	0.390	0.021	6
		2001	Rural	0.395	0.416	0.408	0.008	6
Wang (2006)	Based on grouped income distribution data from Statistical Yearbooks of each province.	1995	Rural and urban	0.271	0.379	0.331	0.037	6
		2000	Rural and urban	0.338	0.378	0.354	0.014	6
		2003	Rural and urban	0.385	0.428	0.407	0.015	6
Weimin (2012)	Based on grouped income distribution data from Statistical Yearbooks of each province.	1995	Rural and urban	0.241	0.438	0.348	0.056	27
		1996	Rural and urban	0.231	0.437	0.328	0.054	27
		1997	Rural and urban	0.228	0.418	0.325	0.050	27
		1998	Rural and urban	0.229	0.410	0.328	0.046	27
		1999	Rural and urban	0.248	0.421	0.343	0.043	27
		2000	Rural and urban	0.256	0.456	0.368	0.053	27
		2001	Rural and urban	0.270	0.466	0.380	0.054	27
		2002	Rural and urban	0.265	0.478	0.396	0.053	26
		2003	Rural and urban	0.267	0.483	0.406	0.052	26
		2004	Rural and urban	0.289	0.483	0.399	0.050	26
		2005	Rural and urban	0.280	0.478	0.405	0.046	26
		2006	Rural and urban	0.276	0.490	0.407	0.049	25
		2007	Rural and urban	0.280	0.491	0.406	0.052	24
		2008	Rural and urban	0.296	0.486	0.408	0.049	24
		2009	Rural and urban	0.290	0.484	0.408	0.049	24
		2010	Rural and urban	0.274	0.476	0.394	0.049	24
Nie and Xing (2019)	1% National Population Sample Survey 2005	2005	Rural and urban	0.335	0.544	0.421	0.045	31

B. Comparative Overview of Chinese Household Surveys

Table B1
Overview of CGSS, CHIP, CFPS, and CHFS household surveys

Characteristic	China General Social Survey (CGSS)	China Household Income Project (CHIP)	China Family Panel Studies (CFPS)	China Household Finance Survey (CHFS)
Data Provider	China Survey and Data Center at Renmin University of China.	Institute of Economics at the Chinese Academy of Social Sciences, with assistance from the National Bureau of Statistics.	The Institute of Social Science Survey at Peking University.	China Household Finance Survey and Research Center at Southwestern University of Finance and Economics.
Waves Available	2003, 2005, 2006, 2008, 2010, 2011, 2012, 2013, 2015, 2017, 2018, 2021	1988, 1995, 2002, 2007, 2013, and 2018	2010, 2012, 2014, 2016, and 2018	2011, 2013, 2015, 2017, and 2019
Provinces Covered	31	9-15	25 (excludes Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia, Hainan)	25 in 2011, 29 after 2011 (excludes Hong Kong, Macao, Taiwan, Tibet, and Xinjiang)
Representativeness	National-level	National-level	National-level; five large provinces oversampled to achieve provincial representativeness: Liaoning, Henan, Shanghai, Guangdong, Gansu.	National-level in 2011; national and provincial-level after 2011.
Sample Size (households)	$\approx 10,000 - 12,000$	2002-2013: $\approx 18,000 - 28,000$.	$\approx 16,000$	$\approx 8,400$ in 2011; $\approx 30,000 - 40,000$ in later waves.
Sampling Design and Stratification	Multi-stage stratified sampling; urban-rural stratification at sampling stage; confirms village/residential committee status during survey.	Pre-stratified urban and rural sampling frames; separate sampling of urban and rural households.	Urban-rural integrated sampling at district/county level; six sub-sampling frames with oversampling of five large provinces; uses village/residential committee status determined at survey time for urban-rural classification.	Urban-rural integrated sampling at district/county level; second-stage stratification at village/community level based on urbanization; proportional sampling in village and residential committee frames.
Income Definition	*Labor income:* wages, bonuses, subsidies, income from working away. *Business income:* agricultural production, non-agricultural operations (subsidiary classified as non-agricultural). *Asset income:* deposit interest, stocks, funds, bonds, land/production rental, other rent, income from selling property. *Transfer and other income:* government subsidies, pensions, minimum living allowances, gifts from relatives/friends, other reported income.	*Labor income:* wages, bonuses, subsidies, income from working away. *Business income:* primary/secondary/tertiary industry income. *Asset income:* property income (lump sum). *Transfer and other income:* government subsidies, pensions, minimum living allowances.	*Labor income:* wages, bonuses, subsidies, income from working away from home, individual dividends. *Business income:* agricultural production (agriculture, forestry, animal husbandry, subsidiary, fishery), and non-agricultural operations. *Asset income:* land/production rental, other rent, income from selling property. *Transfer and other income:* government subsidies, pensions, minimum living allowances, gifts from relatives/friends, other reported income. *Adjusted income:* includes value of self-consumed agricultural products for rural households.	*Labor income:* total after-tax monetary wages and salaries, bonuses, and subsidies from employer (both monetary and in-kind). *Business income:* net income from agricultural production and business operations. *Asset income:* net after-tax income from non-financial assets (rental income from lands, properties, and vehicles) and income from financial assets (interest on deposits, stocks, bonds, dividends from commercial life insurance, etc.). *Transfer and other income:* retirement benefits and pensions; unemployment insurance; other government compensation and subsidies; private transfers; net after-tax income from all other sources including lotteries, sale of properties, vehicles, intellectual property rights, etc.
Top Income Coverage	Xu, Zhang, Zhou and Xie (2012), based on CGSS-2010 (2009 income): top 5% of households hold 24.1% (rural) and 30.1% (urban) of total income.	Xu et al. (2012), based on CHIP-2007 (2007 income): top 5% of households hold 18.7% (rural) and 17.5% (urban) of total income—lowest inequality among surveys.	Xu et al. (2012), based on CFPS-2010 (2009 income): top 5% of households hold 27.1–27.8% of total income (adjusted).	Xu et al. (2012), based on CHFS-2011 (2010 income): top 5% of households hold 43.1% (rural) and 33.2% (urban) of total income, due to better coverage of high-income households.

Sources: CGSS: http://www.cnsda.org/; CHIP: https://www.icpsr.umich.edu/web/ICPSR/series/243/publicat; CFPS: https://www.isss.pku.edu.cn/cfps/; CHFS: https://chfs.swufe.edu.cn/; Xu et al. (2012); Xie and Lu (2015); Gao, Yang and Zhai (2019).

C. Data and summary statistics: detailed

Table C1: CHFS sample size.

	CHFS 2013	CHFS 2015	CHFS 2017	CHFS 2019
Number of provinces	29	29	29	29
Number of counties	267	351	355	343
Number of communities (villages)	1,048	1,396	1,439	1,360
Number of households	28,143	37,289	40,011	34,643

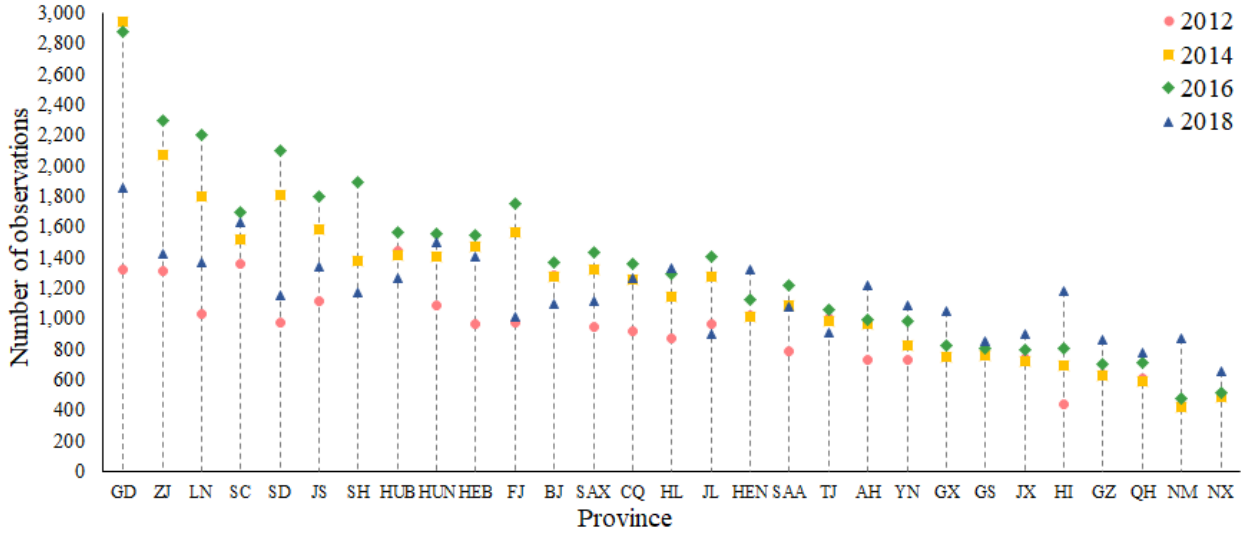


Figure C1: This figure presents the number of observations in each province based on our constructed samples from the CHFS. Provinces on the x-axis are ordered in descending order according to their average number of observations in 2012–2018.

Table C2
Summary statistics, household-level variables from the CHFS.

	2012				2014				2016				2018			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Ln(household income per capita)	9.23	1.28	4.71	12.13	9.33	1.34	5.20	12.05	9.71	1.22	5.36	12.96	9.62	1.21	5.50	12.45
Schooling per worker	10.25	3.46	0	22.00	10.25	3.49	0	22.00	10.41	3.50	0	22.00	10.30	3.51	0	22.00
Experience per worker	22.97	9.49	0	54.00	22.99	9.06	0	54.00	24.40	10.22	0	54.00	25.32	10.42	0	54.00
Experience per worker^2 : 100	6.18	5.11	0	29.16	6.11	4.83	0	29.16	7.00	5.57	0	29.16	7.50	5.78	0	29.16
Household size	3.48	1.63	1	19	3.58	1.70	1	20	3.18	1.55	1	15	3.09	1.54	1	15
Land dummy	0.39	0.49	0	1	0.43	0.50	0	1	0.33	0.47	0	1	0.44	0.50	0	1
Ln(land per capita)	0.01	0.09	0	2.71	0.38	0.59	0	3.80	0.32	0.58	0	3.75	0.06	0.17	0	2.40
Ln(business assets per capita)	0.35	1.06	0	7.43	0.35	1.08	0	7.58	0.34	1.07	0	7.13	0.30	0.95	0	6.22
Party member	0.18	0.38	0	1	0.18	0.39	0	1	0.13	0.33	0	1	0.19	0.39	0	1
Female	0.24	0.43	0	1	0.25	0.43	0	1	0.21	0.40	0	1	0.25	0.43	0	1
Rural	0.32	0.47	0	1	0.30	0.46	0	1	0.32	0.46	0	1	0.35	0.48	0	1

Table C3
Summary statistics, provincial-level variables.

	2012				2014				2016				2018			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Data: individual-level data from the CHFS, working-age population [16-60].																
Educational Gini	0.230	0.048	0.162	0.412	0.225	0.050	0.154	0.428	0.221	0.042	0.157	0.353	0.234	0.046	0.155	0.380
Average years of schooling	9.511	0.960	7.363	12.062	9.934	0.954	7.463	12.421	10.010	0.869	8.482	12.339	9.488	1.005	7.539	12.482
Skilled labor share	0.369	0.091	0.278	0.640	0.425	0.088	0.322	0.690	0.423	0.086	0.278	0.673	0.368	0.095	0.271	0.665
Data: household-level CHFS data.																
Income Gini index	0.559	0.048	0.463	0.686	0.556	0.049	0.460	0.665	0.530	0.056	0.400	0.641	0.531	0.055	0.421	0.665
Returns to schooling	0.157	0.031	0.081	0.225	0.138	0.027	0.094	0.190	0.157	0.025	0.116	0.212	0.153	0.037	0.085	0.237
Returns to high school	0.828	0.251	0.484	1.678	0.763	0.188	0.452	1.337	0.760	0.187	0.339	1.136	0.805	0.209	0.494	1.286
Returns to college	0.828	0.237	0.449	1.531	0.668	0.203	0.342	1.168	0.850	0.160	0.563	1.127	0.751	0.218	0.315	1.246
Data: China Statistical Yearbook (CSY).																
Skilled labor share CSY	0.279	0.089	0.168	0.596	0.294	0.086	0.167	0.600	0.312	0.098	0.179	0.646	0.331	0.097	0.213	0.686
Social spending CSY	1.464	0.749	0.716	4.344	1.784	0.677	0.877	3.445	2.423	1.033	1.211	5.181	2.992	1.077	1.546	5.396
SOE assets share CSY	0.257	0.127	0.041	0.490	0.252	0.131	0.046	0.542	0.248	0.136	0.049	0.638	0.216	0.113	0.051	0.519
Ln(GRP per capita) CSY	10.531	0.400	9.849	11.438	10.696	0.383	10.131	11.578	10.822	0.392	10.218	11.723	11.009	0.392	10.379	11.925
Foreign trade share CSY	0.332	0.371	0.048	1.354	0.312	0.317	0.057	1.137	0.239	0.244	0.032	0.965	0.250	0.249	0.018	0.949
Secondary employment CSY	0.448	0.104	0.206	0.672	0.458	0.113	0.197	0.682	0.431	0.113	0.178	0.659	0.398	0.118	0.157	0.647
Data: China Population and Employment Statistics Yearbook (CPESY).																
Educational Gini CPESY	0.176	0.025	0.140	0.238	0.180	0.026	0.142	0.243	0.191	0.029	0.138	0.258	0.187	0.027	0.136	0.249
Skilled labor share CPESY	0.323	0.121	0.166	0.761	0.348	0.126	0.185	0.758	0.376	0.118	0.192	0.754	0.392	0.124	0.203	0.771

Table C4
Per capita household income by educational attainment across provinces: estimates from the CHFS.

	2012				2014				2016				2018			
	Primary and below	Middle school	High school	College and above	Primary and below	Middle school	High school	College and above	Primary and below	Middle school	High school	College and above	Primary and below	Middle school	High school	College and above
Beijing	19,735	20,611	26,227	54,798	20,020	21,189	27,416	55,968	23,696	42,224	45,007	87,042	18,457	30,429	39,272	87,794
Shanghai	26,001	23,708	29,890	51,565	21,881	30,496	34,838	52,113	48,585	36,095	48,507	77,511	44,490	42,625	52,660	79,904
Jiangsu	14,378	18,276	23,498	40,604	14,140	20,098	23,990	38,035	19,861	26,284	33,292	52,023	18,034	26,343	34,936	51,342
Tianjin	10,079	19,234	23,014	41,285	17,526	17,018	27,177	40,760	19,768	29,855	33,816	51,338	22,154	21,558	31,084	56,116
Zhejiang	14,956	20,503	27,358	47,848	15,208	25,647	30,370	53,297	26,561	33,007	40,563	73,056	22,635	32,357	43,039	76,894
Fujian	9,742	13,476	17,116	29,740	11,144	16,152	23,151	42,599	18,724	20,699	29,957	51,081	15,920	20,168	24,301	47,376
Guangdong	7,316	12,243	12,810	54,707	10,901	15,497	22,849	50,125	17,053	21,467	32,789	66,936	15,392	22,799	31,352	67,853
Shandong	8,368	13,930	20,001	31,534	8,005	14,036	20,495	34,130	10,784	19,453	26,246	44,396	11,880	15,496	23,699	36,208
Inner Mongolia	8,534	11,001	20,940	32,326	7,090	12,980	20,207	32,103	15,019	19,946	23,385	39,343	13,589	21,192	23,800	44,047
Hubei	8,023	11,547	15,609	24,528	10,105	14,127	20,326	32,799	11,701	20,224	28,109	47,986	15,872	19,684	25,953	41,149
Chongqing	8,897	11,574	16,825	30,458	7,080	13,226	17,799	29,232	14,697	15,651	23,743	43,339	12,306	17,312	25,279	39,242
Shaanxi	6,802	10,506	13,516	27,325	10,022	11,036	15,729	31,150	10,271	14,366	20,942	38,468	6,820	13,336	18,553	31,209
Liaoning	10,127	11,772	15,324	23,157	11,454	17,727	21,901	35,620	15,432	20,569	26,282	40,586	11,837	19,497	23,952	39,526
Hunan	5,482	10,175	15,592	24,323	8,976	12,371	17,099	32,066	12,560	16,140	24,073	39,518	14,139	16,649	24,429	33,953
Anhui	8,268	10,878	18,924	29,047	10,590	11,664	18,897	26,009	12,913	16,863	22,813	37,074	13,533	18,906	24,949	38,268
Hainan	5,709	7,220	11,164	22,348	5,518	8,901	12,658	21,316	10,579	12,747	23,785	37,414	6,623	12,229	17,983	30,462
Ningxia	9,862	11,733	15,968	29,142	10,533	15,222	19,374	29,500	11,398	22,083	22,417	34,761	11,906	15,729	17,290	33,970
Henan	6,158	8,145	12,867	21,345	6,910	10,502	14,995	25,633	9,121	13,154	17,516	29,974	8,780	14,336	19,388	35,153
Sichuan	8,283	12,478	18,513	38,290	8,437	14,190	17,323	41,318	9,774	16,283	23,561	44,938	9,899	20,304	25,147	51,147
Jilin	8,849	11,112	14,265	22,043	9,960	12,244	16,795	24,314	13,260	15,870	21,698	34,478	11,498	15,824	18,272	28,172
Jiangxi	7,750	10,833	14,279	21,657	9,820	12,827	16,545	27,288	11,570	15,525	23,792	41,019	13,274	17,360	21,394	36,112
Hebei	5,556	8,835	12,141	21,385	7,261	9,303	14,367	24,877	11,065	14,919	19,489	37,765	10,493	14,704	18,473	35,138
Shanxi	5,549	7,123	10,709	16,942	6,449	8,992	13,178	18,152	6,052	13,146	17,485	28,375	9,020	14,521	17,623	30,094
Qinghai	4,706	9,491	17,616	28,820	5,271	14,068	24,702	36,738	9,691	18,765	26,043	47,234	11,530	18,784	24,705	45,155
Heilongjiang	8,499	12,804	15,889	21,165	10,858	15,092	19,997	25,374	12,816	17,559	23,740	33,478	12,236	18,632	23,013	34,410
Yunnan	5,874	8,287	12,695	25,977	5,873	10,041	15,642	28,668	8,564	11,825	18,845	39,848	9,823	12,593	18,602	44,708
Guangxi	5,392	8,073	13,123	25,160	6,087	9,553	14,429	27,068	6,456	12,966	17,982	32,951	8,096	12,051	19,144	33,934
Guizhou	5,151	8,020	13,544	24,464	7,451	11,638	19,641	30,477	8,703	10,853	20,531	38,518	10,832	13,323	17,213	33,722
Gansu	6,399	7,016	12,718	22,700	6,555	10,468	15,637	27,950	8,350	14,514	23,485	38,700	8,201	11,542	17,826	31,267
Mean	8,981	12,090	16,970	30,506	10,039	14,355	19,915	33,610	14,311	19,416	26,203	45,143	13,768	18,975	24,942	43,942
SD	4,638	4,392	4,999	10,607	4,206	4,955	5,288	9,758	8,206	7,542	7,795	14,208	7,147	6,841	8,461	15,743
CV	0.516	0.363	0.295	0.348	0.419	0.345	0.266	0.290	0.573	0.388	0.297	0.315	0.519	0.361	0.339	0.358

D. Provincial-level variables estimated from the CHFS

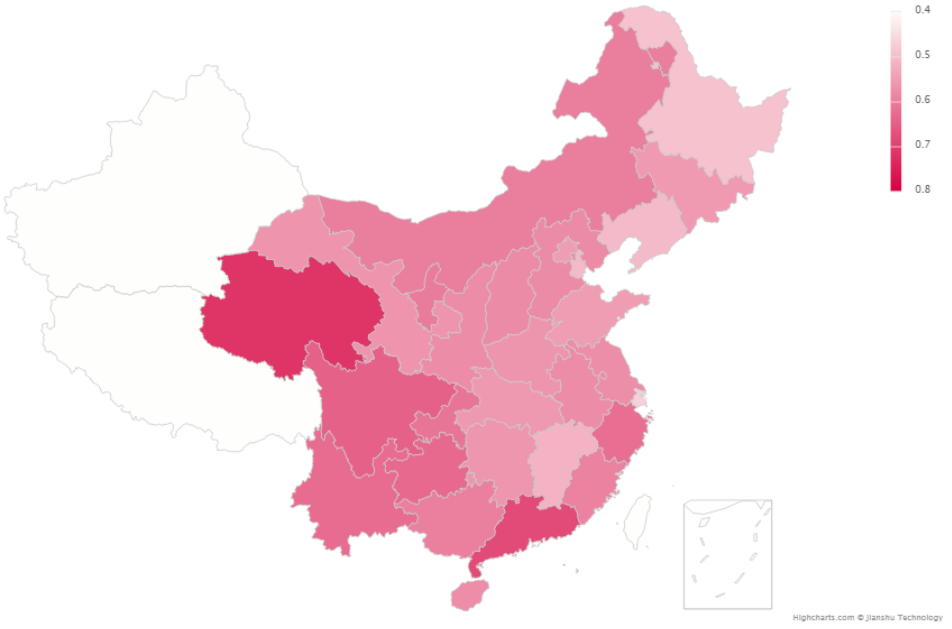


Figure D1: Provincial Gini index estimates on China's map in 2012.

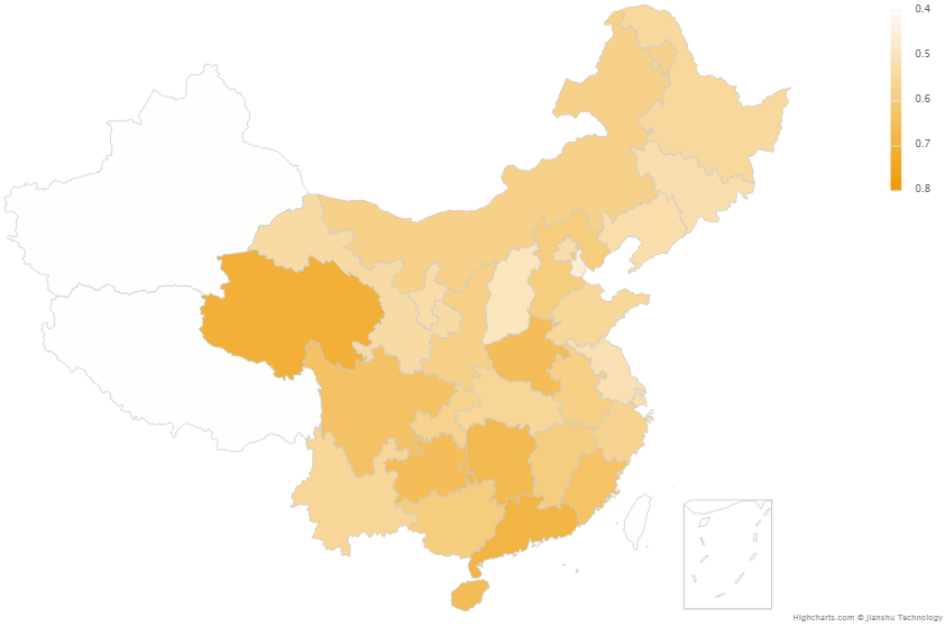


Figure D2: Provincial Gini index estimates on China's map in 2014.

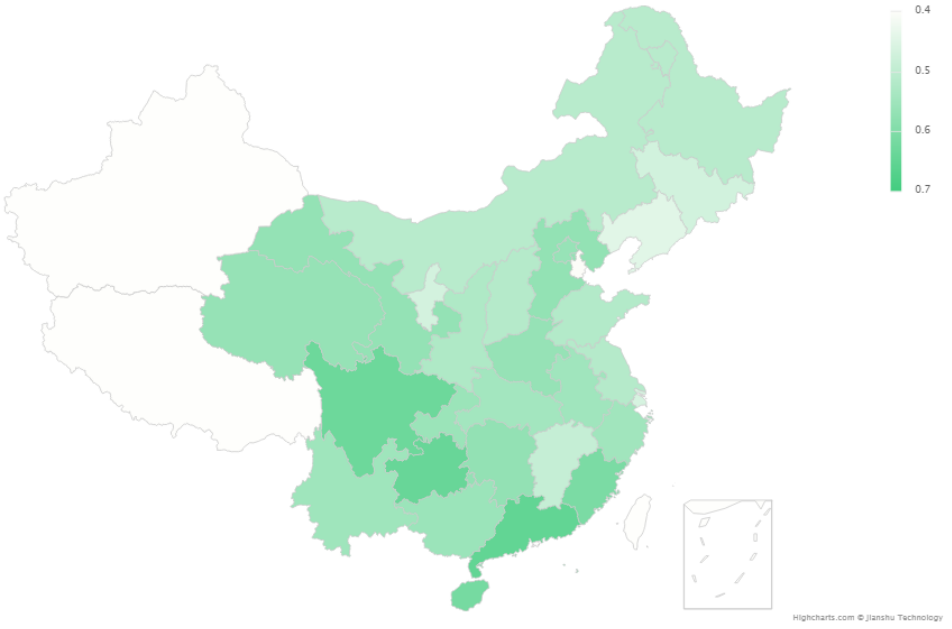


Figure D3: Provincial Gini index estimates on China's map in 2016.

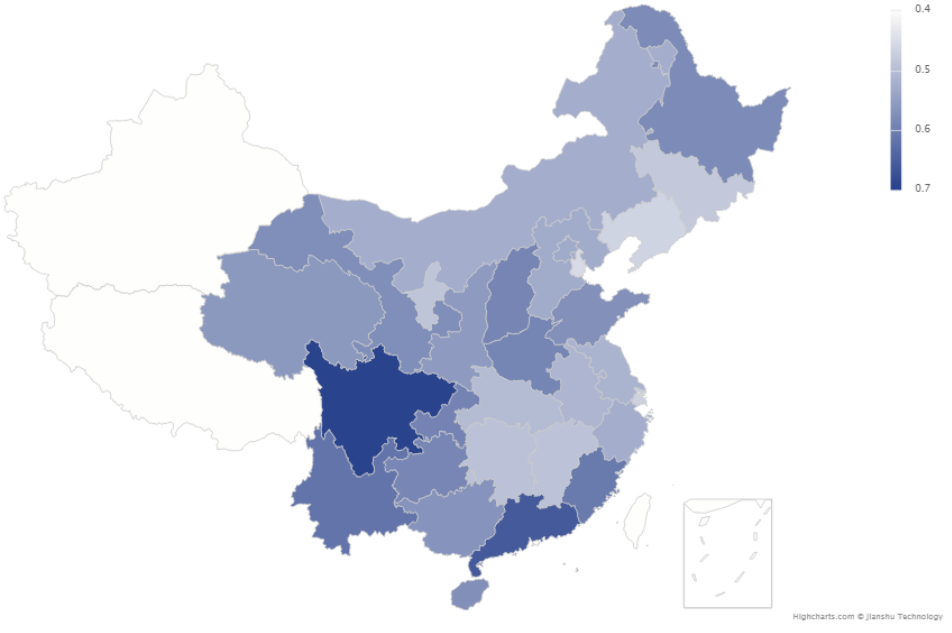


Figure D4: Provincial Gini index estimates on China's map in 2018.

Table D1
Provincial-level estimates of income variables derived from the CHFS.

		Household income per capita (CN¥)				Gini index			
		2012	2014	2016	2018	2012	2014	2016	2018
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Beijing	BJ	42,959	49,130	84,114	72,860	0.526	0.516	0.572	0.517
		[39,032; 46,053]	[42,884; 53,725]	[72,211; 94,609]	[64,598; 79,328]	[0.488; 0.562]	[0.464; 0.561]	[0.523; 0.626]	[0.470; 0.559]
Shanghai	SH	40,410	47,870	59,438	62,767	0.463	0.494	0.441	0.421
		[37,717; 42,720]	[39,118; 54,665]	[48,389; 65,229]	[51,387; 69,277]	[0.430; 0.495]	[0.424; 0.563]	[0.355; 0.493]	[0.333; 0.481]
Jiangsu	JS	29,102	28,843	40,653	35,438	0.557	0.506	0.491	0.499
		[25,976; 31,514]	[27,179; 30,336]	[37,673; 43,172]	[32,584; 37,585]	[0.515; 0.593]	[0.480; 0.530]	[0.455; 0.521]	[0.465; 0.531]
Tianjin	TJ	34,528	37,357	46,928	44,556	0.498	0.460	0.400	0.432
		[30,981; 37,181]	[35,020; 39,447]	[43,893; 49,175]	[41,341; 47,264]	[0.453; 0.536]	[0.430; 0.487]	[0.366; 0.430]	[0.393; 0.464]
Zhejiang	ZJ	32,818	41,748	52,015	40,221	0.616	0.562	0.546	0.512
		[24,270; 38,559]	[38,438; 44,878]	[47,287; 56,475]	[37,450; 42,519]	[0.539; 0.683]	[0.530; 0.597]	[0.506; 0.586]	[0.483; 0.538]
Fujian	FJ	20,554	30,761	37,492	32,348	0.574	0.606	0.555	0.555
		[18,339; 22,348]	[26,688; 33,964]	[31,883; 41,554]	[26,226; 36,436]	[0.531; 0.611]	[0.560; 0.647]	[0.501; 0.600]	[0.489; 0.609]
Guangdong	GD	35,990	40,641	44,475	39,356	0.686	0.665	0.641	0.624
		[32,351; 39,073]	[28,224; 48,692]	[37,726; 49,409]	[25,907; 45,869]	[0.657; 0.714]	[0.594; 0.729]	[0.592; 0.681]	[0.530; 0.686]
Shandong	SD	20,141	24,880	30,948	19,406	0.551	0.528	0.520	0.570
		[18,195; 21,812]	[22,883; 26,637]	[29,332; 32,442]	[17,846; 20,810]	[0.512; 0.588]	[0.490; 0.562]	[0.496; 0.543]	[0.539; 0.603]
Inner Mongolia	NM	18,421	25,361	31,672	28,304	0.582	0.564	0.494	0.518
		[15,681; 20,721]	[20,438; 29,349]	[27,429; 34,833]	[26,504; 29,974]	[0.528; 0.633]	[0.493; 0.631]	[0.441; 0.543]	[0.488; 0.547]
Hubei	HUB	17,174	24,548	32,787	23,276	0.536	0.549	0.536	0.490
		[15,446; 18,594]	[22,525; 26,139]	[28,668; 36,083]	[21,809; 24,634]	[0.494; 0.575]	[0.515; 0.581]	[0.483; 0.583]	[0.458; 0.519]
Chongqing	CQ	17,006	20,163	30,678	22,322	0.559	0.557	0.557	0.575
		[14,489; 18,773]	[18,336; 21,719]	[25,486; 34,162]	[20,089; 24,234]	[0.502; 0.607]	[0.522; 0.590]	[0.493; 0.611]	[0.536; 0.614]
Shaanxi	SAA	17,000	22,721	28,469	18,810	0.565	0.581	0.516	0.544
		[15,539; 18,272]	[21,229; 24,128]	[26,256; 30,358]	[17,492; 20,049]	[0.528; 0.597]	[0.556; 0.609]	[0.482; 0.548]	[0.512; 0.576]
Liaoning	LN	17,410	29,930	31,773	28,735	0.492	0.526	0.441	0.447
		[16,124; 18,514]	[27,623; 32,046]	[30,183; 33,071]	[26,919; 30,253]	[0.459; 0.525]	[0.492; 0.559]	[0.414; 0.464]	[0.415; 0.475]
Hunan	HUN	15,918	20,966	24,788	22,005	0.544	0.591	0.555	0.485
		[14,554; 17,103]	[17,249; 23,543]	[22,637; 26,635]	[20,699; 23,140]	[0.509; 0.575]	[0.534; 0.638]	[0.516; 0.589]	[0.456; 0.510]
Anhui	AH	18,579	20,028	20,932	22,156	0.559	0.557	0.519	0.510
		[15,390; 20,745]	[16,714; 22,533]	[17,970; 22,733]	[20,643; 23,272]	[0.496; 0.612]	[0.496; 0.611]	[0.464; 0.561]	[0.481; 0.535]
Hainan	HI	12,256	15,362	31,346	20,129	0.562	0.625	0.599	0.568
		[11,165; 13,244]	[12,732; 17,074]	[24,597; 36,158]	[18,597; 21,562]	[0.523; 0.600]	[0.570; 0.668]	[0.536; 0.659]	[0.537; 0.597]
Ningxia	NX	18,614	23,544	35,161	22,292	0.554	0.500	0.476	0.488
		[15,907; 20,601]	[20,030; 26,212]	[30,907; 38,575]	[20,682; 23,670]	[0.498; 0.600]	[0.439; 0.553]	[0.420; 0.523]	[0.452; 0.520]
Henan	HEN	12,453	18,791	21,690	20,701	0.562	0.611	0.572	0.580
		[11,627; 13,179]	[14,762; 21,549]	[20,274; 22,970]	[19,265; 21,929]	[0.536; 0.587]	[0.540; 0.665]	[0.542; 0.597]	[0.552; 0.605]
Sichuan	SC	22,712	23,309	24,296	29,414	0.629	0.634	0.611	0.665
		[19,235; 25,187]	[20,638; 25,753]	[20,956; 26,670]	[18,513; 36,716]	[0.583; 0.669]	[0.596; 0.673]	[0.565; 0.649]	[0.588; 0.742]
Jilin	JL	15,131	19,782	27,244	19,467	0.522	0.519	0.454	0.482
		[13,605; 16,387]	[18,333; 21,067]	[25,041; 28,955]	[18,451; 20,402]	[0.478; 0.560]	[0.486; 0.549]	[0.415; 0.488]	[0.458; 0.506]
Jiangxi	JX	14,294	19,708	29,005	20,720	0.507	0.568	0.489	0.488
		[13,300; 15,185]	[16,433; 22,190]	[26,287; 31,130]	[19,480; 21,796]	[0.474; 0.539]	[0.511; 0.616]	[0.445; 0.525]	[0.458; 0.515]
Hebei	HEB	12,590	15,127	23,467	17,304	0.574	0.558	0.559	0.523
		[11,538; 13,494]	[13,949; 16,040]	[21,336; 25,211]	[16,274; 18,233]	[0.542; 0.602]	[0.526; 0.586]	[0.523; 0.591]	[0.496; 0.548]
Shanxi	SAX	11,725	16,047	24,366	21,234	0.570	0.488	0.517	0.551
		[10,784; 12,448]	[15,024; 16,910]	[22,791; 25,853]	[17,523; 23,410]	[0.540; 0.596]	[0.461; 0.513]	[0.489; 0.545]	[0.485; 0.594]
Qinghai	QH	16,309	23,125	30,429	22,813	0.643	0.589	0.556	0.551
		[11,516; 19,134]	[18,370; 25,940]	[27,281; 33,129]	[21,234; 24,175]	[0.566; 0.701]	[0.513; 0.642]	[0.514; 0.598]	[0.526; 0.576]
Heilongjiang	HL	18,486	22,964	24,493	24,888	0.481	0.504	0.485	0.530
		[17,190; 19,626]	[20,264; 24,971]	[22,438; 26,050]	[18,304; 28,465]	[0.448; 0.512]	[0.451; 0.543]	[0.447; 0.518]	[0.429; 0.596]
Yunnan	YN	14,547	17,538	25,755	22,401	0.587	0.543	0.551	0.588
		[12,425; 16,008]	[15,967; 18,918]	[23,740; 27,505]	[19,324; 24,581]	[0.534; 0.630]	[0.501; 0.581]	[0.518; 0.583]	[0.537; 0.629]
Guangxi	GX	15,381	17,649	16,569	17,185	0.548	0.569	0.543	0.554
		[13,739; 16,624]	[15,336; 19,494]	[15,481; 17,621]	[16,047; 18,317]	[0.504; 0.586]	[0.518; 0.611]	[0.517; 0.571]	[0.526; 0.582]
Guizhou	GZ	13,392	19,892	23,273	18,580	0.607	0.626	0.606	0.582
		[11,150; 14,932]	[16,468; 22,402]	[18,895; 26,217]	[16,856; 19,913]	[0.556; 0.653]	[0.569; 0.674]	[0.545; 0.654]	[0.548; 0.613]
Gansu	GS	11,941	19,062	27,720	19,928	0.559	0.532	0.576	0.548
		[11,113; 12,719]	[17,605; 20,346]	[24,222; 30,538]	[18,262; 21,403]	[0.528; 0.589]	[0.501; 0.562]	[0.528; 0.624]	[0.512; 0.580]
Mean		20,270	25,408	33,172	27,918	0.559	0.556	0.530	0.531
SD		8,846	9,389	13,723	13,237	0.048	0.049	0.056	0.055
CV		0.436	0.370	0.414	0.474	0.086	0.089	0.106	0.103

Table D2
Provincial-level estimates of educational variables derived from the CHFS.

	Average years of schooling				Educational Gini index			
	2012	2014	2016	2018	2012	2014	2016	2018
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Beijing	12.1	12.4	12.3	12.5	0.171	0.165	0.157	0.166
	[11.9; 12.2]	[12.3; 12.6]	[12.2; 12.5]	[12.3; 12.6]	[0.166; 0.177]	[0.159; 0.170]	[0.152; 0.162]	[0.160; 0.172]
Shanghai	11.4	11.6	11.4	11.6	0.171	0.154	0.162	0.155
	[11.2; 11.5]	[11.5; 11.7]	[11.3; 11.6]	[11.4; 11.7]	[0.165; 0.176]	[0.150; 0.159]	[0.157; 0.166]	[0.149; 0.161]
Jiangsu	9.6	9.9	10.3	9.7	0.240	0.235	0.219	0.218
	[9.4; 9.7]	[9.8; 10.1]	[10.2; 10.4]	[9.6; 9.9]	[0.230; 0.249]	[0.227; 0.242]	[0.212; 0.226]	[0.210; 0.228]
Tianjin	11.7	11.7	11.8	11.4	0.162	0.163	0.164	0.185
	[11.6; 11.9]	[11.6; 11.9]	[11.7; 12.0]	[11.2; 11.6]	[0.156; 0.169]	[0.156; 0.169]	[0.157; 0.171]	[0.178; 0.193]
Zhejiang	9.2	10.2	9.9	9.3	0.239	0.222	0.232	0.251
	[9.1; 9.3]	[10.1; 10.3]	[9.8; 10.0]	[9.1; 9.5]	[0.231; 0.247]	[0.216; 0.228]	[0.225; 0.238]	[0.242; 0.260]
Fujian	9.1	9.5	9.8	9.3	0.260	0.264	0.258	0.271
	[9.0; 9.3]	[9.4; 9.6]	[9.6; 9.9]	[9.1; 9.5]	[0.251; 0.270]	[0.256; 0.271]	[0.250; 0.265]	[0.261; 0.282]
Guangdong	10.3	10.5	10.2	10.1	0.208	0.195	0.203	0.204
	[10.2; 10.4]	[10.4; 10.6]	[10.1; 10.3]	[10.0; 10.2]	[0.203; 0.214]	[0.191; 0.198]	[0.199; 0.207]	[0.198; 0.209]
Shandong	9.7	10.4	10.3	9.1	0.217	0.207	0.195	0.216
	[9.6; 9.9]	[10.3; 10.5]	[10.2; 10.4]	[9.0; 9.3]	[0.208; 0.226]	[0.201; 0.213]	[0.190; 0.201]	[0.206; 0.226]
Inner Mongolia	9.2	10.3	10.5	10.3	0.242	0.226	0.221	0.219
	[8.9; 9.4]	[10.0; 10.5]	[10.2; 10.8]	[10.1; 10.5]	[0.228; 0.256]	[0.212; 0.239]	[0.207; 0.234]	[0.209; 0.228]
Hubei	9.4	9.9	9.8	9.0	0.221	0.216	0.217	0.235
	[9.3; 9.6]	[9.8; 10.0]	[9.7; 10.0]	[8.9; 9.2]	[0.215; 0.228]	[0.209; 0.221]	[0.210; 0.225]	[0.226; 0.243]
Chongqing	8.9	9.5	9.7	9.1	0.245	0.246	0.231	0.241
	[8.7; 9.0]	[9.3; 9.6]	[9.5; 9.8]	[8.9; 9.2]	[0.234; 0.255]	[0.238; 0.254]	[0.223; 0.238]	[0.232; 0.250]
Shaanxi	9.7	10.7	10.6	9.3	0.229	0.198	0.207	0.238
	[9.5; 9.9]	[10.6; 10.9]	[10.5; 10.8]	[9.2; 9.5]	[0.220; 0.240]	[0.191; 0.205]	[0.200; 0.215]	[0.229; 0.248]
Liaoning	10.1	10.8	10.7	10.4	0.174	0.181	0.178	0.191
	[10.0; 10.2]	[10.7; 10.9]	[10.6; 10.8]	[10.3; 10.5]	[0.168; 0.180]	[0.177; 0.185]	[0.174; 0.182]	[0.186; 0.196]
Hunan	9.9	9.8	9.5	9.5	0.195	0.199	0.209	0.207
	[9.8; 10.0]	[9.7; 10.0]	[9.4; 9.7]	[9.4; 9.6]	[0.188; 0.202]	[0.193; 0.205]	[0.202; 0.215]	[0.200; 0.214]
Anhui	8.8	9.2	8.5	8.7	0.283	0.266	0.290	0.298
	[8.6; 9.1]	[9.0; 9.4]	[8.3; 8.7]	[8.5; 8.9]	[0.269; 0.296]	[0.255; 0.276]	[0.278; 0.301]	[0.288; 0.309]
Hainan	9.3	9.6	10.2	9.5	0.196	0.203	0.199	0.218
	[9.2; 9.5]	[9.4; 9.7]	[10.1; 10.4]	[9.4; 9.7]	[0.184; 0.206]	[0.194; 0.212]	[0.191; 0.206]	[0.210; 0.226]
Ningxia	8.9	9.3	10.1	9.0	0.246	0.250	0.212	0.259
	[8.7; 9.1]	[9.1; 9.5]	[9.9; 10.3]	[8.8; 9.2]	[0.231; 0.260]	[0.236; 0.264]	[0.200; 0.225]	[0.247; 0.272]
Henan	9.3	9.4	9.8	9.4	0.239	0.242	0.235	0.243
	[9.1; 9.4]	[9.3; 9.6]	[9.7; 10.0]	[9.2; 9.5]	[0.230; 0.248]	[0.233; 0.252]	[0.226; 0.244]	[0.234; 0.252]
Sichuan	9.3	9.1	8.8	8.8	0.220	0.244	0.266	0.238
	[9.2; 9.4]	[9.0; 9.3]	[8.7; 9.0]	[8.7; 9.0]	[0.213; 0.226]	[0.237; 0.252]	[0.259; 0.275]	[0.230; 0.246]
Jilin	9.2	9.9	10.4	9.4	0.213	0.203	0.190	0.219
	[9.0; 9.3]	[9.8; 10.1]	[10.3; 10.6]	[9.3; 9.6]	[0.205; 0.221]	[0.198; 0.209]	[0.184; 0.195]	[0.210; 0.227]
Jiangxi	9.1	9.3	9.9	8.8	0.243	0.233	0.220	0.233
	[9.0; 9.3]	[9.2; 9.5]	[9.8; 10.1]	[8.6; 8.9]	[0.233; 0.253]	[0.223; 0.242]	[0.211; 0.231]	[0.223; 0.243]
Hebei	9.1	9.4	9.6	9.0	0.234	0.231	0.215	0.228
	[9.0; 9.3]	[9.3; 9.5]	[9.5; 9.7]	[8.8; 9.1]	[0.224; 0.245]	[0.224; 0.238]	[0.208; 0.222]	[0.219; 0.238]
Shanxi	9.6	10.4	10.4	10.0	0.190	0.184	0.180	0.203
	[9.5; 9.8]	[10.3; 10.6]	[10.3; 10.6]	[9.8; 10.1]	[0.183; 0.198]	[0.179; 0.189]	[0.175; 0.185]	[0.196; 0.210]
Qinghai	7.4	7.5	8.5	7.5	0.412	0.428	0.353	0.380
	[7.1; 7.6]	[7.2; 7.7]	[8.3; 8.7]	[7.3; 7.7]	[0.396; 0.429]	[0.410; 0.446]	[0.339; 0.367]	[0.365; 0.394]
Heilongjiang	10.0	10.2	9.6	9.3	0.206	0.207	0.207	0.211
	[9.9; 10.2]	[10.0; 10.3]	[9.5; 9.7]	[9.2; 9.5]	[0.198; 0.214]	[0.200; 0.213]	[0.200; 0.215]	[0.203; 0.218]
Yunnan	8.5	9.3	9.3	8.4	0.274	0.244	0.259	0.297
	[8.3; 8.7]	[9.1; 9.5]	[9.1; 9.5]	[8.2; 8.5]	[0.263; 0.286]	[0.235; 0.253]	[0.250; 0.269]	[0.287; 0.307]
Guangxi	9.7	9.6	9.3	9.2	0.192	0.202	0.196	0.190
	[9.6; 9.8]	[9.5; 9.8]	[9.1; 9.4]	[9.1; 9.4]	[0.185; 0.199]	[0.194; 0.209]	[0.188; 0.204]	[0.183; 0.197]
Guizhou	8.5	9.0	9.3	8.5	0.273	0.265	0.255	0.287
	[8.3; 8.7]	[8.8; 9.2]	[9.1; 9.5]	[8.3; 8.7]	[0.261; 0.285]	[0.254; 0.275]	[0.243; 0.266]	[0.275; 0.300]
Gansu	8.7	9.5	9.4	9.0	0.266	0.246	0.269	0.286
	[8.5; 8.9]	[9.3; 9.7]	[9.3; 9.6]	[8.8; 9.2]	[0.256; 0.277]	[0.236; 0.256]	[0.257; 0.280]	[0.274; 0.299]
Mean	9.5	9.9	10.0	9.5	0.230	0.225	0.221	0.234
SD	1.0	1.0	0.9	1.0	0.048	0.050	0.042	0.046
CV	0.101	0.096	0.087	0.106	0.209	0.221	0.190	0.196

Table D3
Provincial-level estimates of returns to schooling and returns to high school derived from the CHFS.

	Returns to schooling				Returns to high school			
	2012	2014	2016	2018	2012	2014	2016	2018
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Beijing	0.163	0.132	0.143	0.172	0.827	0.654	0.635	0.789
	[0.141; 0.186]	[0.110; 0.156]	[0.121; 0.167]	[0.145; 0.199]	[0.704; 0.963]	[0.523; 0.799]	[0.509; 0.757]	[0.620; 0.962]
Shanghai	0.126	0.097	0.125	0.108	0.559	0.452	0.505	0.498
	[0.104; 0.149]	[0.078; 0.118]	[0.106; 0.144]	[0.085; 0.131]	[0.446; 0.672]	[0.349; 0.561]	[0.413; 0.595]	[0.389; 0.611]
Jiangsu	0.140	0.094	0.140	0.180	0.621	0.475	0.717	0.652
	[0.113; 0.170]	[0.072; 0.116]	[0.121; 0.160]	[0.152; 0.207]	[0.468; 0.789]	[0.345; 0.620]	[0.611; 0.811]	[0.512; 0.795]
Tianjin	0.170	0.129	0.135	0.165	0.633	0.556	0.689	0.808
	[0.146; 0.194]	[0.099; 0.159]	[0.108; 0.165]	[0.144; 0.187]	[0.471; 0.779]	[0.408; 0.703]	[0.555; 0.820]	[0.675; 0.956]
Zhejiang	0.141	0.142	0.153	0.136	0.646	0.702	0.710	0.609
	[0.117; 0.168]	[0.123; 0.158]	[0.135; 0.173]	[0.112; 0.163]	[0.476; 0.827]	[0.595; 0.805]	[0.613; 0.810]	[0.460; 0.765]
Fujian	0.136	0.128	0.131	0.136	0.623	0.590	0.769	0.839
	[0.104; 0.171]	[0.109; 0.147]	[0.110; 0.150]	[0.109; 0.163]	[0.426; 0.829]	[0.444; 0.726]	[0.644; 0.884]	[0.645; 1.029]
Guangdong	0.225	0.164	0.171	0.166	1.285	0.831	0.878	0.767
	[0.199; 0.253]	[0.147; 0.182]	[0.154; 0.187]	[0.148; 0.186]	[1.113; 1.456]	[0.729; 0.940]	[0.776; 0.986]	[0.657; 0.887]
Shandong	0.157	0.132	0.170	0.145	0.647	0.788	0.708	0.615
	[0.124; 0.188]	[0.111; 0.153]	[0.151; 0.188]	[0.111; 0.179]	[0.460; 0.829]	[0.662; 0.919]	[0.586; 0.811]	[0.416; 0.835]
Inner Mongolia	0.124	0.119	0.150	0.168	0.783	0.621	0.753	0.942
	[0.080; 0.169]	[0.089; 0.151]	[0.117; 0.182]	[0.143; 0.193]	[0.491; 1.079]	[0.351; 0.878]	[0.513; 0.982]	[0.779; 1.107]
Hubei	0.141	0.122	0.152	0.114	0.615	0.724	0.706	0.603
	[0.120; 0.163]	[0.102; 0.144]	[0.130; 0.176]	[0.086; 0.136]	[0.480; 0.755]	[0.596; 0.860]	[0.580; 0.839]	[0.456; 0.748]
Chongqing	0.156	0.170	0.192	0.180	0.893	0.847	0.879	0.619
	[0.127; 0.184]	[0.143; 0.197]	[0.168; 0.217]	[0.147; 0.214]	[0.690; 1.064]	[0.672; 1.018]	[0.740; 1.023]	[0.407; 0.816]
Shaanxi	0.167	0.132	0.160	0.207	0.994	0.780	0.733	1.090
	[0.140; 0.194]	[0.103; 0.160]	[0.139; 0.180]	[0.181; 0.232]	[0.808; 1.181]	[0.595; 0.964]	[0.592; 0.882]	[0.913; 1.261]
Liaoning	0.133	0.147	0.165	0.174	0.625	0.751	0.751	0.907
	[0.101; 0.164]	[0.127; 0.165]	[0.150; 0.181]	[0.149; 0.198]	[0.467; 0.771]	[0.621; 0.872]	[0.666; 0.841]	[0.774; 1.045]
Hunan	0.186	0.158	0.158	0.085	0.774	0.749	0.637	0.494
	[0.156; 0.215]	[0.135; 0.181]	[0.134; 0.186]	[0.062; 0.108]	[0.620; 0.934]	[0.610; 0.912]	[0.497; 0.786]	[0.369; 0.617]
Anhui	0.156	0.123	0.140	0.114	0.871	0.702	0.616	0.742
	[0.129; 0.183]	[0.096; 0.151]	[0.112; 0.166]	[0.092; 0.135]	[0.697; 1.050]	[0.524; 0.880]	[0.440; 0.780]	[0.581; 0.918]
Hainan	0.196	0.130	0.187	0.237	0.986	0.764	0.837	1.092
	[0.147; 0.243]	[0.095; 0.163]	[0.160; 0.216]	[0.212; 0.265]	[0.693; 1.265]	[0.533; 0.987]	[0.678; 1.002]	[0.905; 1.268]
Ningxia	0.189	0.124	0.116	0.129	0.845	0.706	0.339	0.770
	[0.139; 0.234]	[0.092; 0.159]	[0.075; 0.153]	[0.093; 0.168]	[0.497; 1.139]	[0.519; 0.904]	[0.109; 0.543]	[0.564; 0.978]
Henan	0.134	0.134	0.126	0.187	0.805	1.027	0.590	1.034
	[0.111; 0.157]	[0.110; 0.160]	[0.098; 0.151]	[0.159; 0.216]	[0.634; 0.977]	[0.843; 1.199]	[0.401; 0.753]	[0.842; 1.224]
Sichuan	0.168	0.183	0.200	0.187	0.921	1.068	1.131	0.999
	[0.144; 0.192]	[0.160; 0.207]	[0.179; 0.221]	[0.162; 0.213]	[0.752; 1.084]	[0.889; 1.238]	[0.984; 1.288]	[0.825; 1.176]
Jilin	0.081	0.104	0.139	0.088	0.484	0.614	0.675	0.544
	[0.053; 0.107]	[0.082; 0.125]	[0.119; 0.159]	[0.065; 0.113]	[0.282; 0.678]	[0.480; 0.759]	[0.556; 0.789]	[0.375; 0.721]
Jiangxi	0.135	0.126	0.140	0.115	0.746	0.741	0.701	0.685
	[0.106; 0.163]	[0.087; 0.166]	[0.110; 0.168]	[0.092; 0.140]	[0.553; 0.934]	[0.529; 0.969]	[0.533; 0.851]	[0.510; 0.872]
Hebei	0.126	0.131	0.168	0.124	0.705	0.845	0.715	0.621
	[0.100; 0.156]	[0.107; 0.156]	[0.145; 0.191]	[0.103; 0.145]	[0.507; 0.916]	[0.674; 1.011]	[0.560; 0.865]	[0.467; 0.791]
Shanxi	0.167	0.115	0.150	0.142	0.853	0.598	0.652	0.719
	[0.135; 0.197]	[0.093; 0.139]	[0.123; 0.175]	[0.117; 0.169]	[0.677; 1.044]	[0.463; 0.725]	[0.525; 0.778]	[0.570; 0.870]
Qinghai	0.211	0.190	0.189	0.137	1.678	1.337	1.079	1.143
	[0.181; 0.242]	[0.169; 0.212]	[0.168; 0.211]	[0.115; 0.160]	[1.396; 1.949]	[1.086; 1.583]	[0.876; 1.270]	[0.936; 1.368]
Heilongjiang	0.115	0.111	0.129	0.132	0.550	0.654	0.613	0.722
	[0.092; 0.137]	[0.090; 0.132]	[0.107; 0.150]	[0.111; 0.155]	[0.394; 0.695]	[0.524; 0.772]	[0.480; 0.740]	[0.585; 0.866]
Yunnan	0.184	0.177	0.168	0.192	1.060	0.909	1.116	1.286
	[0.155; 0.212]	[0.146; 0.207]	[0.147; 0.189]	[0.175; 0.212]	[0.808; 1.287]	[0.706; 1.115]	[0.955; 1.281]	[1.110; 1.449]
Guangxi	0.179	0.190	0.196	0.169	0.815	0.934	0.994	0.874
	[0.143; 0.214]	[0.159; 0.221]	[0.163; 0.235]	[0.133; 0.206]	[0.598; 1.040]	[0.720; 1.125]	[0.756; 1.225]	[0.650; 1.090]
Guizhou	0.196	0.173	0.212	0.145	1.116	1.013	1.136	0.774
	[0.166; 0.227]	[0.133; 0.212]	[0.169; 0.253]	[0.107; 0.180]	[0.854; 1.374]	[0.679; 1.269]	[0.903; 1.401]	[0.538; 1.019]
Gansu	0.149	0.127	0.149	0.213	1.056	0.698	0.775	1.107
	[0.121; 0.179]	[0.101; 0.156]	[0.128; 0.170]	[0.186; 0.241]	[0.829; 1.260]	[0.505; 0.879]	[0.604; 0.937]	[0.891; 1.312]
Mean	0.157	0.138	0.157	0.153	0.828	0.763	0.760	0.805
SD	0.031	0.027	0.025	0.037	0.251	0.188	0.187	0.209
CV	0.200	0.195	0.159	0.243	0.303	0.246	0.246	0.260

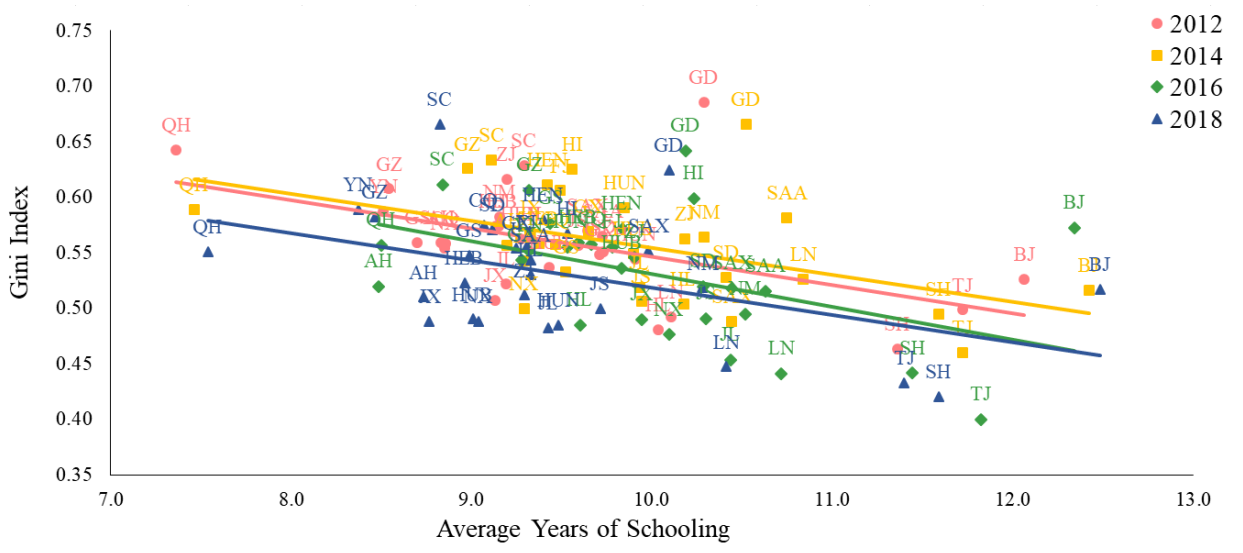


Figure D5: Gini index and average years of schooling. *Notes:* This figure presents provincial Gini index and average years of schooling estimated from the CHFS. Correlations are: -51% in 2012, -47% in 2014, -46% in 2016, and -45% in 2018.

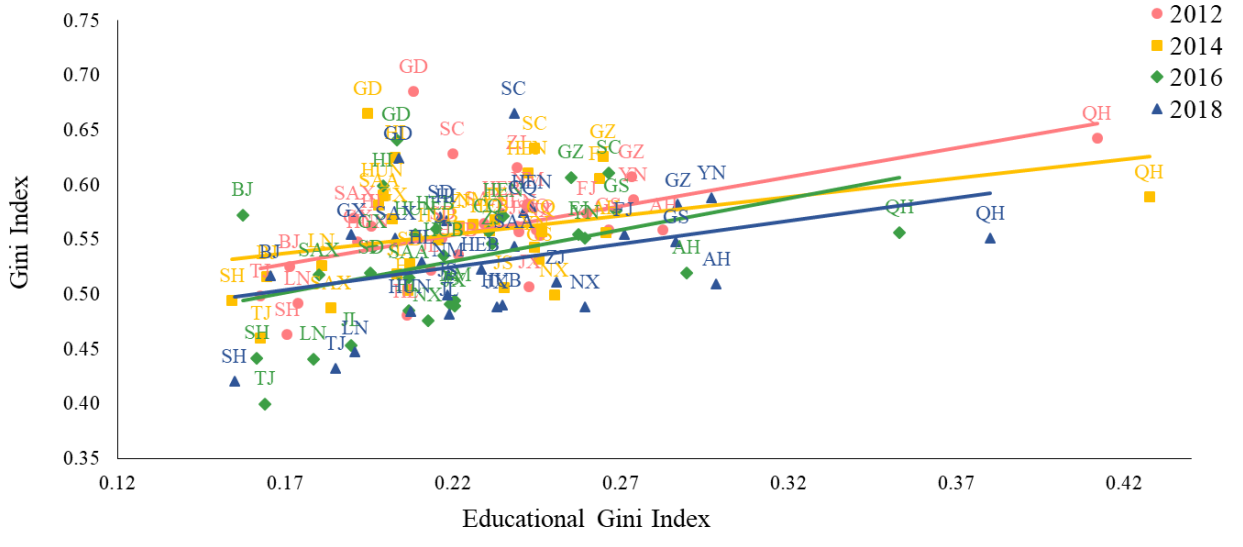


Figure D6: Gini index and educational Gini index. *Notes:* This figure presents provincial Gini index and educational Gini index estimated from the CHFS. Correlations are: 53% in 2012, 35% in 2014, 43% in 2016, and 35% in 2018.



Figure D7: Gini index and returns to schooling. *Notes:* This figure presents provincial Gini index and returns to schooling estimated from the CHFS. Correlations are: 58% in 2012, 56% in 2014, 58% in 2016, and 48% in 2018.

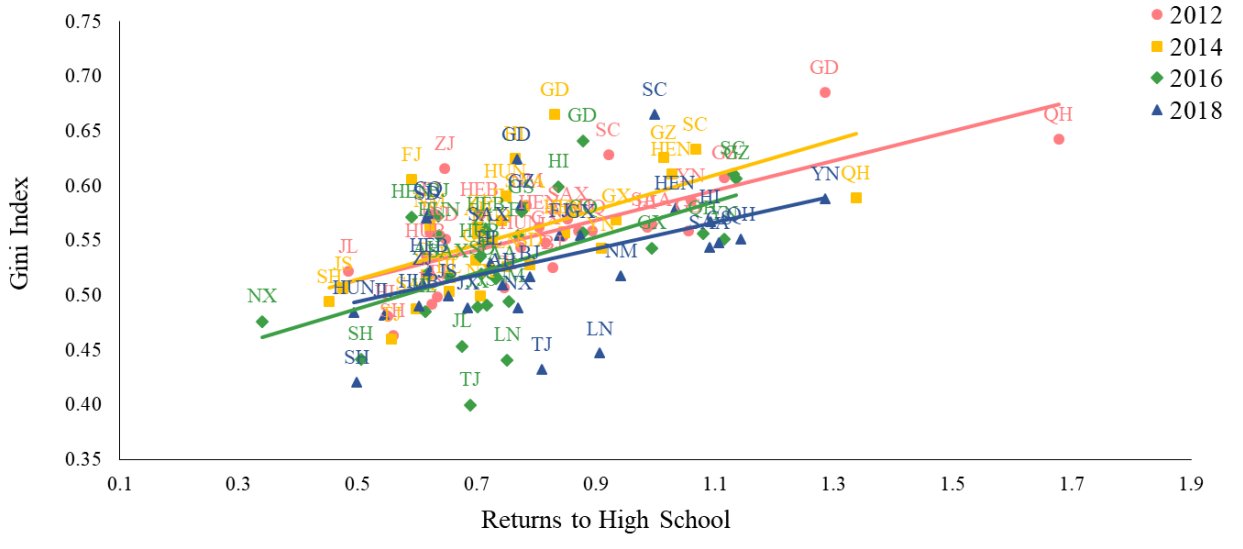
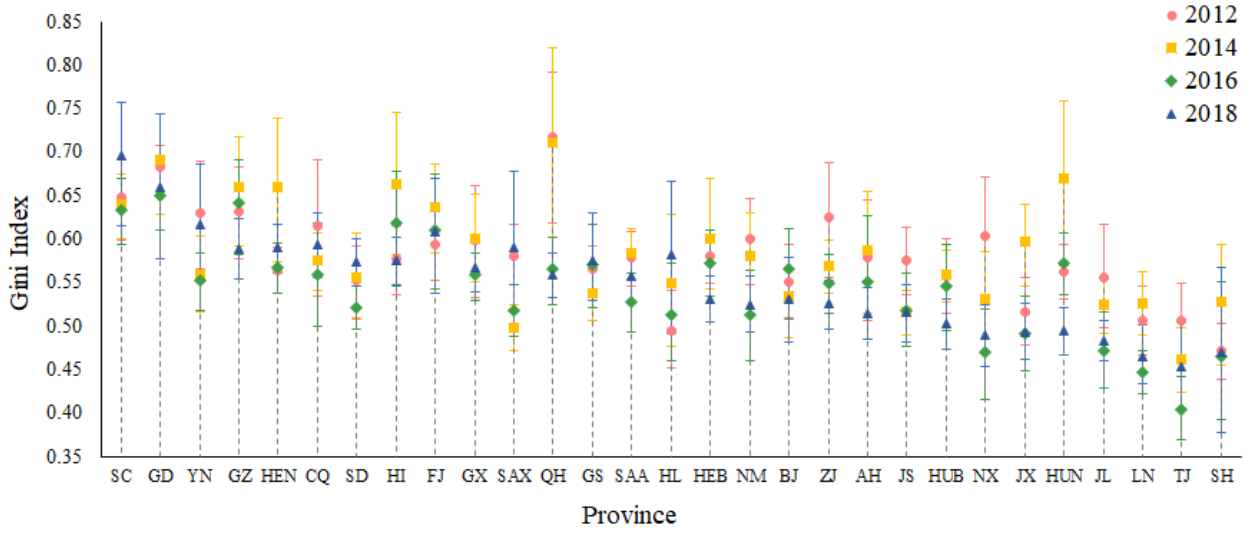
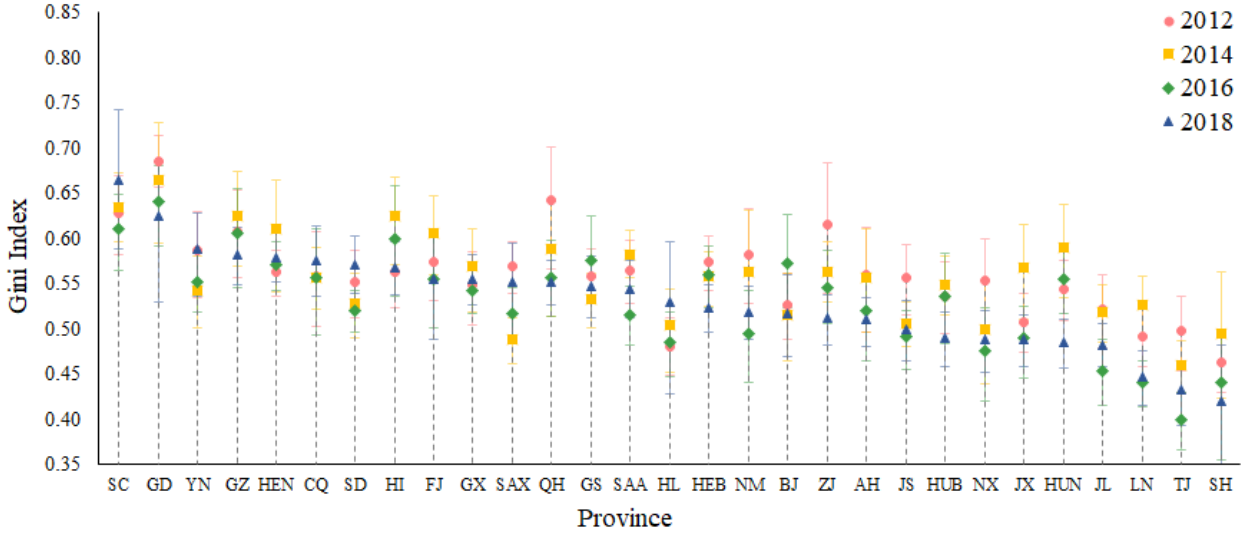


Figure D8: Gini index and returns to high school. *Notes:* This figure presents provincial Gini index and returns to high school estimated from the CHFS. Correlations are: 71% in 2012, 61% in 2014, 54% in 2016, and 46% in 2018.

E. Raw Gini estimates



(a) Raw data estimates.



(b) Baseline estimates.

Figure E1: Baseline versus raw data estimates of provincial Gini index. *Notes:* These figures illustrate the substantial reduction of variation in our baseline estimated provincial Gini indices (b) compared to the raw weighted data estimates (a). Error bars represent 95% confidence intervals computed from 1000 bootstrap samples. Basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) are used for our estimates, and percentile bootstrap CIs are applied to the raw weighted data results reported here.

F. Comparison to the NBS official statistics

Table F1
Demographic and economic statistics on Chinese provinces from the China Statistical Yearbook.

	Population (10,000 individuals)				Rural population share (%)				GRP per capita (CN¥)				Per capita disposable income (CN¥)			
	2012	2014	2016	2018	2012	2014	2016	2018	2012	2014	2016	2018	2012	2014	2016	2018
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Beijing	2,078	2,171	2,195	2,192	0.137	0.135	0.133	0.129	92,758	106,732	123,391	150,962	36,817	44,489	52,530	62,361
Shanghai	2,399	2,467	2,467	2,475	0.107	0.107	0.110	0.109	89,613	102,827	121,369	145,767	38,550	45,966	54,305	64,183
Jiangsu	8,120	8,281	8,381	8,446	0.370	0.343	0.311	0.288	66,533	78,711	92,658	110,508	22,432	27,173	32,070	38,096
Tianjin	1,378	1,429	1,443	1,383	0.184	0.174	0.167	0.161	66,517	74,960	79,647	95,689	24,030	28,832	34,074	39,506
Zhejiang	5,685	5,890	6,072	6,273	0.371	0.350	0.323	0.300	61,097	68,569	78,384	93,230	27,020	32,658	38,529	45,840
Fujian	3,841	3,945	4,016	4,104	0.407	0.380	0.356	0.330	52,959	63,709	74,024	94,719	19,141	23,331	27,608	32,644
Guangdong	11,041	11,489	11,908	12,348	0.329	0.314	0.299	0.282	52,308	59,909	69,671	81,625	21,268	25,685	30,296	35,810
Shandong	9,708	9,808	9,973	10,077	0.480	0.452	0.409	0.385	44,348	51,933	59,239	66,284	17,127	20,864	24,685	29,205
Inner Mongolia	2,464	2,449	2,436	2,422	0.416	0.390	0.366	0.345	42,441	49,585	56,560	66,491	16,800	20,559	24,127	28,376
Hubei	5,781	5,816	5,885	5,917	0.468	0.443	0.414	0.390	39,149	48,635	56,844	71,097	14,809	18,283	21,787	25,815
Chongqing	2,975	3,043	3,110	3,163	0.434	0.403	0.367	0.334	39,180	48,311	58,327	68,460	14,924	18,352	22,034	26,386
Shaanxi	3,787	3,827	3,874	3,931	0.503	0.470	0.436	0.403	37,453	45,610	49,341	61,115	12,885	15,837	18,874	22,528
Liaoning	4,375	4,358	4,327	4,291	0.344	0.330	0.311	0.297	40,778	45,915	47,069	54,657	18,761	22,820	26,040	29,701
Hunan	6,590	6,611	6,625	6,635	0.538	0.510	0.473	0.439	32,203	39,181	46,606	54,763	14,391	17,622	21,115	25,241
Anhui	5,978	5,997	6,033	6,076	0.537	0.507	0.474	0.444	30,697	37,580	43,686	56,063	13,593	16,796	19,998	23,984
Hainan	910	936	957	982	0.490	0.467	0.433	0.408	30,993	37,166	43,009	50,263	14,180	17,476	20,653	24,579
Ningxia	659	678	695	710	0.489	0.451	0.413	0.379	32,609	36,815	40,339	49,614	13,104	15,907	18,832	22,400
Henan	9,532	9,645	9,778	9,864	0.580	0.550	0.512	0.478	30,497	35,982	41,326	50,714	12,772	15,695	18,443	21,964
Sichuan	8,085	8,139	8,251	8,321	0.566	0.535	0.500	0.465	29,627	35,563	40,297	51,658	12,753	15,749	18,808	22,461
Jilin	2,698	2,642	2,567	2,484	0.455	0.432	0.413	0.391	32,005	37,539	40,259	44,925	14,395	17,520	19,967	22,798
Jiangxi	4,475	4,480	4,496	4,513	0.526	0.494	0.460	0.427	28,624	34,988	40,950	50,347	13,567	16,734	20,110	24,080
Hebei	7,262	7,323	7,375	7,426	0.534	0.506	0.461	0.427	31,844	34,507	38,688	43,808	13,647	16,647	19,725	23,446
Shanxi	3,548	3,528	3,514	3,502	0.487	0.457	0.427	0.401	32,864	34,248	33,972	45,517	13,592	16,538	19,049	21,990
Qinghai	571	576	582	587	0.522	0.491	0.464	0.428	26,839	32,218	38,968	46,854	11,470	14,374	17,302	20,757
Heilongjiang	3,724	3,608	3,463	3,327	0.431	0.408	0.389	0.365	29,352	33,464	34,025	38,199	14,302	17,404	19,838	22,726
Yunnan	4,631	4,653	4,677	4,703	0.615	0.588	0.554	0.526	23,992	30,217	35,051	44,446	11,233	13,772	16,720	20,084
Guangxi	4,694	4,770	4,857	4,947	0.565	0.535	0.508	0.482	24,181	28,603	33,340	39,837	12,644	15,557	18,305	21,485
Guizhou	3,587	3,677	3,758	3,822	0.637	0.597	0.544	0.505	18,947	25,101	31,589	40,271	9,850	12,371	15,121	18,430
Gansu	2,550	2,531	2,520	2,515	0.612	0.577	0.539	0.503	21,141	25,724	27,396	32,178	9,768	12,185	14,670	17,488
Mean	4,591	4,647	4,698	4,739	0.453	0.427	0.399	0.373	40,743	47,735	54,346	65,519	16,891	20,593	24,332	28,771
SD	2,791	2,850	2,919	2,986	0.134	0.126	0.115	0.106	18,725	21,141	24,861	30,064	7,072	8,368	9,832	11,604
CV	0.608	0.613	0.621	0.630	0.296	0.294	0.288	0.285	0.460	0.443	0.457	0.459	0.419	0.406	0.404	0.403

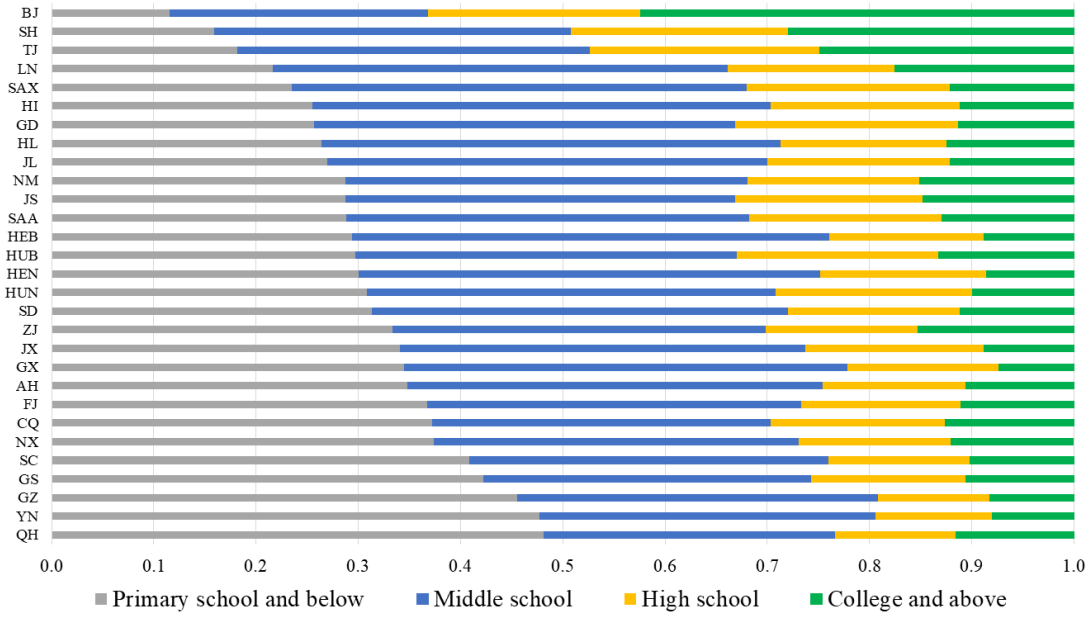


Figure E1: Average educational structure in Chinese provinces, 2012-2018. *Notes:* This figure presents the average educational structure of the six-years and above population in Chinese provinces for 2012-2018 estimated from the China Statistical Yearbook (CSY). The estimates are ordered in the "Primary school and below" category.

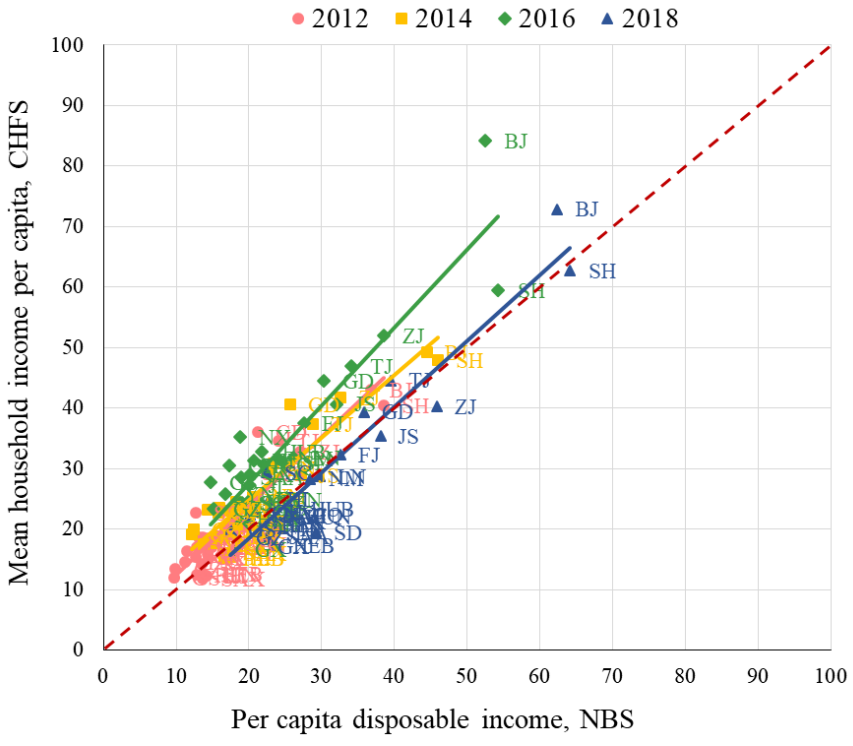


Figure E2: Mean household income per capita versus the NBS statistics (in 1000 CN¥). *Notes:* This figure presents provincial-level mean household income per capita estimated from the CHFS versus per capita disposable income from the China Statistical Yearbook (CSY). Correlations are: 89% in 2012 and 2014, 91% in 2016, and 95% in 2018.

G. Mincer regression for the returns to schooling

Table G1
Mincer regression: explaining household income per capita variation in China.

	Dependent variable: Ln(household income per capita)															
	(2012)	($\hat{s}_z$)	(2014)	($\hat{s}_z$)	(2016)	($\hat{s}_z$)	(2018)	($\hat{s}_z$)	(2012)	($\hat{s}_z$)	(2014)	($\hat{s}_z$)	(2016)	($\hat{s}_z$)	(2018)	($\hat{s}_z$)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Schooling per worker	0.157***	17.0%	0.140***	14.9%	0.160***	19.8%	0.152***	18.1%	0.135***	14.5%	0.123***	13.1%	0.143***	17.7%	0.129***	15.4%
	(0.003)		(0.003)		(0.003)		(0.004)		(0.003)		(0.004)		(0.003)		(0.004)	
Experience per worker	0.004	-0.8%	0.013***	-2.0%	0.032***	-4.2%	0.022***	-3.7%	0.005	-0.8%	0.013***	-1.9%	0.033***	-4.3%	0.022***	-3.7%
	(0.005)		(0.005)		(0.004)		(0.004)		(0.005)		(0.005)		(0.004)		(0.004)	
Experience per worker^2 : 100	0.003	0.1%	−0.036***	2.3%	−0.048***	3.4%	−0.036***	3.2%	−0.0002	0.2%	−0.035***	2.3%	−0.049***	3.5%	−0.035***	3.2%
	(0.008)		(0.010)		(0.007)		(0.008)		(0.008)		(0.009)		(0.007)		(0.008)	
Ln(household size)	−0.439***	3.2%	−0.647***	5.2%	−0.330***	1.8%	−0.381***	1.7%	−0.385***	2.8%	−0.540***	4.4%	−0.257***	1.4%	−0.316***	1.4%
	(0.024)		(0.028)		(0.025)		(0.026)		(0.024)		(0.029)		(0.025)		(0.026)	
Constant	8.140***		8.684***		8.074***		8.441***		7.818***		8.424***		8.009***		8.656***	
	(0.087)		(0.097)		(0.084)		(0.088)		(0.135)		(0.132)		(0.128)		(0.136)	
Provincial dummies	Yes	5.9%	Yes	4.4%	Yes	6.3%	Yes	6.7%	No		No		No		No	
City dummies	No	No	No	No	No	No	No	No	Yes	12.0%	Yes	10.4%	Yes	12.2%	Yes	15.1%
Residual		74.6%		75.1%		72.9%		73.9%		71.3%		71.8%		69.5%		68.5%
Observations	23,410		29,882		30,130		24,581		23,410		29,882		30,130		24,581	
R^2	0.279		0.272		0.293		0.286		0.313		0.305		0.326		0.341	

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

Notes: This table reports the results of the Mincer-type regression formulated in eq. (6), controlling for provincial dummies in columns (1) to (7) and controlling for city dummies in columns (9) to (15). We employ heteroskedasticity-robust standard errors.

H. Mincer regression for the returns to high school

Table H1
Mincer regression: explaining household income per capita variation in China.

	Dependent variable: Ln(household income per capita)															
	(2012)	($\hat{s}_z$)	(2014)	($\hat{s}_z$)	(2016)	($\hat{s}_z$)	(2018)	($\hat{s}_z$)	(2012)	($\hat{s}_z$)	(2014)	($\hat{s}_z$)	(2016)	($\hat{s}_z$)	(2018)	($\hat{s}_z$)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
High school per worker	0.807***	10.1%	0.762***	9.7%	0.761***	11.0%	0.773***	10.3%	0.666***	8.4%	0.657***	8.4%	0.654***	9.4%	0.611***	8.1%
	(0.020)		(0.023)		(0.019)		(0.021)		(0.020)		(0.023)		(0.019)		(0.022)	
Experience per worker	0.011**	-1.5%	0.028***	-3.6%	0.042***	-5.4%	0.039***	-6.2%	0.009**	-1.3%	0.025***	-3.3%	0.041***	-5.2%	0.035***	-5.5%
	(0.005)		(0.005)		(0.004)		(0.004)		(0.005)		(0.005)		(0.004)		(0.004)	
Experience per worker^2 : 100	−0.044***	2.4%	−0.089***	5.2%	−0.100***	6.7%	−0.099***	8.3%	−0.038***	2.1%	−0.080***	4.7%	−0.094***	6.3%	−0.085***	7.3%
	(0.008)		(0.010)		(0.007)		(0.008)		(0.008)		(0.009)		(0.007)		(0.008)	
Ln(household size)	−0.530***	4.0%	−0.738***	6.0%	−0.444***	2.5%	−0.496***	2.3%	−0.436***	3.2%	−0.597***	4.8%	−0.336***	1.8%	−0.396***	1.8%
	(0.024)		(0.029)		(0.025)		(0.026)		(0.024)		(0.029)		(0.025)		(0.026)	
Constant	9.626***		9.892***		9.616***		9.837***		9.007***		9.391***		9.371***		9.774***	
	(0.077)		(0.083)		(0.072)		(0.073)		(0.129)		(0.131)		(0.124)		(0.127)	
Provincial dummies	Yes	6.9%	Yes	4.9%	Yes	7.1%	Yes	7.3%	No		No		No		No	
City dummies	No	No	No	No	No	No	No	No	Yes	14.2%	Yes	11.7%	Yes	14.3%	Yes	16.9%
Residual		78.1%		77.9%		78.0%		77.9%		73.5%		73.7%		73.3%		71.4%
Observations	23,410		29,882		30,130		24,581		23,410		29,882		30,130		24,581	
R^2	0.242		0.243		0.239		0.246		0.289		0.285		0.286		0.312	

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

Notes: This table reports the results of the Mincer-type regression formulated in eq. (6), controlling for provincial dummies in columns (1) to (7) and controlling for city dummies in columns (9) to (15). We employ heteroskedasticity-robust standard errors.

I. Factors associated with provincial income inequality: additional results

Table I1
Test results for negative sign in interaction term in the Gini correlates model.

	Dependent variable: Gini index													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Educational Gini	0.368***	0.412**	0.427***	0.363***	0.375***	0.449***	0.420***	0.166	0.199	0.308***	0.157	0.202*	0.384***	0.345***
	(0.092)	(0.166)	(0.095)	(0.081)	(0.080)	(0.107)	(0.097)	(0.105)	(0.171)	(0.115)	(0.104)	(0.106)	(0.120)	(0.112)
Educational Gini^2		−0.591							−0.838					
		(3.219)							(3.397)					
Returns to schooling	0.720***	0.726***	0.820***	0.787***	0.755***	0.819***	0.816***							
	(0.115)	(0.132)	(0.126)	(0.115)	(0.122)	(0.124)	(0.130)							
Educational Gini × Returns to schooling	−0.232	0.292	−6.169***	−0.798	−2.741	−6.374***	−6.107***							
	(2.353)	(2.983)	(2.244)	(2.407)	(1.960)	(2.402)	(2.260)							
Returns to high school								0.137***	0.136***	0.147***	0.140***	0.142***	0.148***	0.153***
								(0.020)	(0.025)	(0.021)	(0.020)	(0.019)	(0.020)	(0.022)
Educational Gini × Returns to high school								−0.139	0.015	−0.920***	−0.200	−0.502***	−1.008***	−0.981***
								(0.216)	(0.492)	(0.240)	(0.245)	(0.195)	(0.267)	(0.248)
Social spending CSY	−0.015***	−0.014**		−0.020***				−0.016***	−0.016**		−0.021***			
	(0.006)	(0.006)		(0.005)				(0.006)	(0.006)		(0.005)			
SOE assets share CSY	−0.070*	−0.066*			−0.113***			−0.085**	−0.083**			−0.128***		
	(0.037)	(0.038)			(0.033)			(0.037)	(0.037)			(0.034)		
Foreign trade share CSY	0.023	0.026				0.006		0.015	0.018				0.019	
	(0.028)	(0.028)				(0.017)		(0.025)	(0.026)				(0.016)	
Ln(GRP per capita) CSY	−0.016	−0.016					−0.002	−0.005	−0.005					0.010
	(0.021)	(0.021)					(0.010)	(0.020)	(0.020)					(0.010)
2014.year	0.020*	0.020*	0.016	0.020**	0.013	0.016	0.016	0.012	0.013	0.007	0.013	0.006	0.007	0.005
	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
2016.year	−0.005	−0.005	−0.025**	−0.007	−0.027***	−0.025**	−0.025**	−0.001	−0.001	−0.019*	0.002	−0.020**	−0.017*	−0.022**
	(0.013)	(0.014)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.013)	(0.013)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
2018.year	0.003	0.001	−0.029***	0.003	−0.033***	−0.029***	−0.028**	−0.001	−0.002	−0.029***	0.006	−0.033***	−0.028***	−0.034***
	(0.018)	(0.018)	(0.011)	(0.012)	(0.011)	(0.011)	(0.012)	(0.018)	(0.018)	(0.011)	(0.012)	(0.010)	(0.011)	(0.012)
Constant	0.755***	0.751***	0.556***	0.583***	0.584***	0.554***	0.576***	0.646***	0.646***	0.560***	0.586***	0.590***	0.554***	0.453***
	(0.215)	(0.215)	(0.006)	(0.009)	(0.011)	(0.008)	(0.107)	(0.208)	(0.209)	(0.006)	(0.009)	(0.010)	(0.007)	(0.107)
Observations	116	116	116	116	116	116	116	116	116	116	116	116	116	116
R^2	0.553	0.554	0.444	0.531	0.505	0.445	0.444	0.569	0.570	0.453	0.534	0.529	0.461	0.457

Nested (cluster + generated-regressor) bootstrap SEs
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the regression results that account for provincial income inequality based on provincial human capital factors, as outlined in eq. (7). Standard errors in parentheses are from a nested bootstrap with 1,000 replicates (Cameron and Miller, 2015).

J. Educational Theil T index decomposition

Table J1
Decomposition of educational Theil T index by province and birth cohorts (BC).

	2012				2014				2016				2018			
	Theil T index	Share (%)	Within BC (%)	Between BC (%)	Theil T index	Share (%)	Within BC (%)	Between BC (%)	Theil T index	Share (%)	Within BC (%)	Between BC (%)	Theil T index	Share (%)	Within BC (%)	Between BC (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Beijing	0.088	2.0%	1.6%	0.4%	0.084	1.1%	0.8%	0.3%	0.075	1.1%	0.8%	0.3%	0.076	1.0%	0.8%	0.3%
Shanghai	0.088	1.4%	1.0%	0.3%	0.094	1.3%	1.0%	0.3%	0.093	1.4%	1.0%	0.4%	0.092	1.1%	0.8%	0.3%
Jiangsu	0.166	3.6%	2.7%	0.9%	0.203	7.4%	5.6%	1.8%	0.165	6.5%	5.2%	1.3%	0.157	5.7%	4.6%	1.1%
Tianjin	0.088	1.5%	1.2%	0.3%	0.091	0.8%	0.6%	0.1%	0.082	0.8%	0.7%	0.1%	0.095	0.8%	0.7%	0.1%
Zhejiang	0.183	5.0%	3.5%	1.5%	0.156	3.9%	2.7%	1.1%	0.174	4.2%	3.0%	1.2%	0.208	4.9%	3.4%	1.5%
Fujian	0.212	6.5%	5.2%	1.3%	0.213	3.5%	2.8%	0.7%	0.191	3.7%	2.9%	0.8%	0.204	3.3%	2.6%	0.7%
Guangdong	0.121	3.3%	2.9%	0.5%	0.125	6.0%	5.1%	0.9%	0.134	6.6%	5.6%	1.0%	0.137	6.1%	5.1%	1.0%
Shandong	0.158	2.5%	1.9%	0.6%	0.155	6.9%	5.3%	1.6%	0.137	4.7%	3.7%	1.0%	0.194	7.5%	6.0%	1.5%
Inner Mongolia	0.187	2.2%	1.7%	0.5%	0.153	1.9%	1.5%	0.3%	0.156	1.5%	1.2%	0.2%	0.144	1.7%	1.3%	0.4%
Hubei	0.157	4.4%	3.9%	0.5%	0.168	4.6%	4.0%	0.6%	0.160	4.3%	3.5%	0.7%	0.196	4.5%	3.7%	0.8%
Chongqing	0.193	4.6%	3.5%	1.1%	0.207	2.5%	1.9%	0.6%	0.166	2.4%	1.7%	0.6%	0.218	2.6%	1.8%	0.8%
Shaanxi	0.165	3.6%	3.1%	0.5%	0.145	2.6%	2.2%	0.4%	0.131	2.5%	2.2%	0.3%	0.184	3.1%	2.6%	0.5%
Liaoning	0.110	2.2%	1.8%	0.4%	0.096	2.1%	1.7%	0.4%	0.085	2.2%	1.8%	0.4%	0.091	2.0%	1.6%	0.5%
Hunan	0.143	3.7%	3.1%	0.6%	0.149	4.4%	3.6%	0.7%	0.157	4.9%	4.1%	0.8%	0.151	4.4%	3.6%	0.8%
Anhui	0.260	3.3%	2.6%	0.7%	0.261	6.2%	5.1%	1.2%	0.301	6.8%	5.5%	1.3%	0.290	6.8%	5.0%	1.8%
Hainan	0.139	2.1%	1.9%	0.2%	0.150	0.5%	0.5%	0.0%	0.122	0.6%	0.5%	0.1%	0.141	0.5%	0.4%	0.1%
Ningxia	0.216	3.1%	2.5%	0.6%	0.215	0.6%	0.5%	0.1%	0.163	0.6%	0.5%	0.1%	0.218	0.6%	0.4%	0.1%
Henan	0.199	2.7%	2.3%	0.5%	0.227	8.5%	7.2%	1.4%	0.201	7.7%	6.4%	1.3%	0.209	7.8%	6.4%	1.3%
Sichuan	0.138	5.0%	3.8%	1.2%	0.195	6.4%	5.0%	1.4%	0.219	6.9%	5.7%	1.2%	0.194	6.2%	4.9%	1.3%
Jilin	0.130	2.4%	2.1%	0.4%	0.121	1.5%	1.3%	0.2%	0.103	1.6%	1.4%	0.2%	0.130	1.6%	1.3%	0.3%
Jiangxi	0.196	3.5%	2.8%	0.7%	0.186	3.4%	2.7%	0.7%	0.155	3.3%	2.7%	0.6%	0.175	3.2%	2.5%	0.7%
Hebei	0.169	2.9%	2.4%	0.5%	0.180	5.2%	4.4%	0.8%	0.154	5.2%	4.2%	1.0%	0.181	5.1%	4.4%	0.8%
Shanxi	0.130	2.8%	2.4%	0.4%	0.116	2.0%	1.6%	0.4%	0.103	2.1%	1.7%	0.4%	0.129	2.2%	1.8%	0.4%
Qinghai	0.460	6.1%	5.5%	0.6%	0.447	0.8%	0.7%	0.1%	0.317	0.9%	0.8%	0.1%	0.370	0.8%	0.7%	0.1%
Heilongjiang	0.139	2.2%	2.0%	0.2%	0.119	2.2%	1.9%	0.3%	0.125	2.4%	2.1%	0.3%	0.122	2.1%	1.9%	0.3%
Yunnan	0.220	3.6%	3.0%	0.6%	0.205	3.9%	3.4%	0.5%	0.197	4.2%	3.7%	0.6%	0.240	4.2%	3.6%	0.6%
Guangxi	0.108	2.8%	2.5%	0.3%	0.140	2.7%	2.3%	0.4%	0.136	2.7%	2.3%	0.5%	0.118	2.1%	1.7%	0.4%
Guizhou	0.227	3.9%	3.2%	0.7%	0.224	3.0%	2.4%	0.6%	0.207	3.2%	2.6%	0.6%	0.255	3.3%	2.5%	0.8%
Gansu	0.232	4.3%	3.8%	0.6%	0.190	2.1%	1.9%	0.2%	0.219	2.8%	2.4%	0.3%	0.257	2.6%	2.2%	0.4%
Within	0.166	97.3%	79.7%	17.6%	0.167	98.0%	79.9%	18.1%	0.159	97.9%	80.1%	17.8%	0.175	97.8%	78.2%	19.6%
Between	0.005	2.7%			0.003	2.0%			0.003	2.1%			0.004	2.2%		
Total	0.171	100%			0.170	100%			0.162	100%			0.179	100%		

Notes: This table reports the results of the Theil T index (2) decomposition by province and birth cohorts, as formulated in eq. (3).

K. Robustness checks

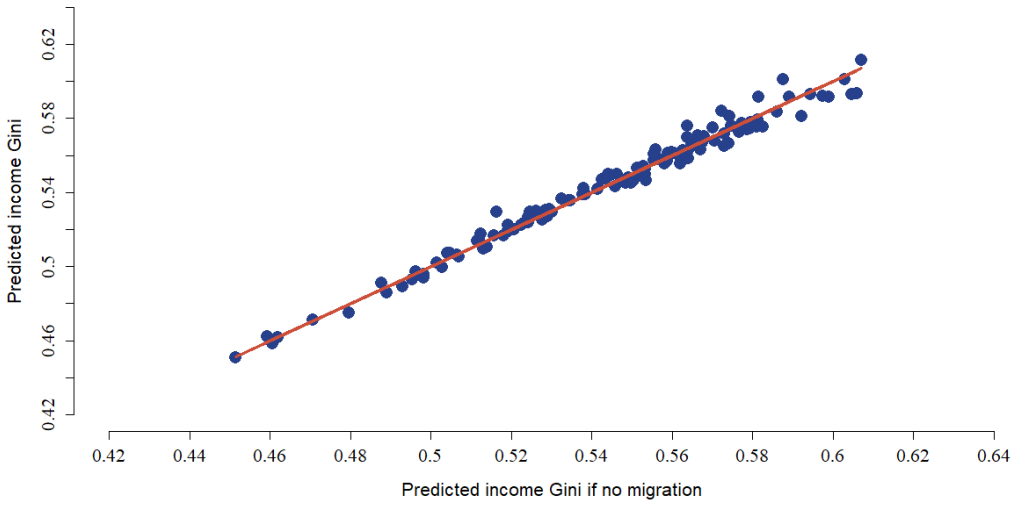


Figure K1: Predicted Gini index with/without migration: pooled OLS. *Notes:* This figure presents predicted provincial-level Gini index with and without migration estimated from the CHFS. Each dot represents an estimate for each province/year. Correlation is 99%.

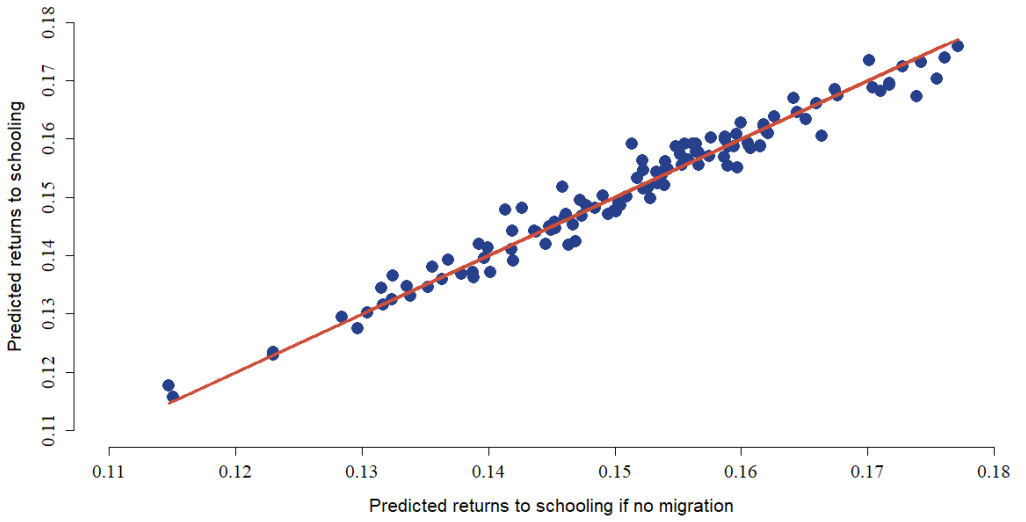


Figure K2: Predicted returns to schooling with/without migration: pooled OLS. *Notes:* This figure presents predicted provincial-level returns to schooling with and without migration estimated from the CHFS. Each dot represents an estimate for each province/year. Correlation is 98%.

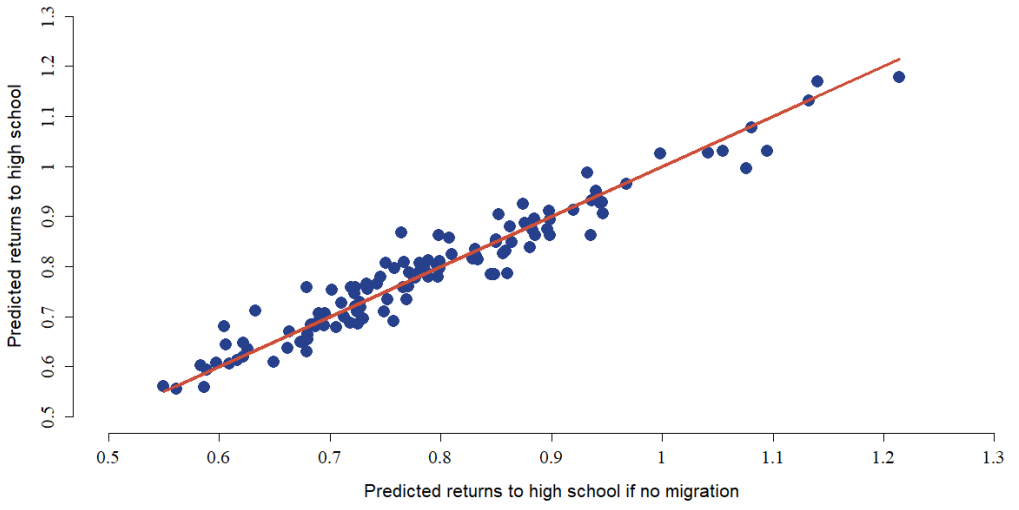


Figure K3: Predicted returns to high school with/without migration: pooled OLS. *Notes:* This figure presents predicted provincial-level returns to high school with and without migration estimated from the CHFS. Each dot represents an estimate for each province/year. Correlation is 97%.

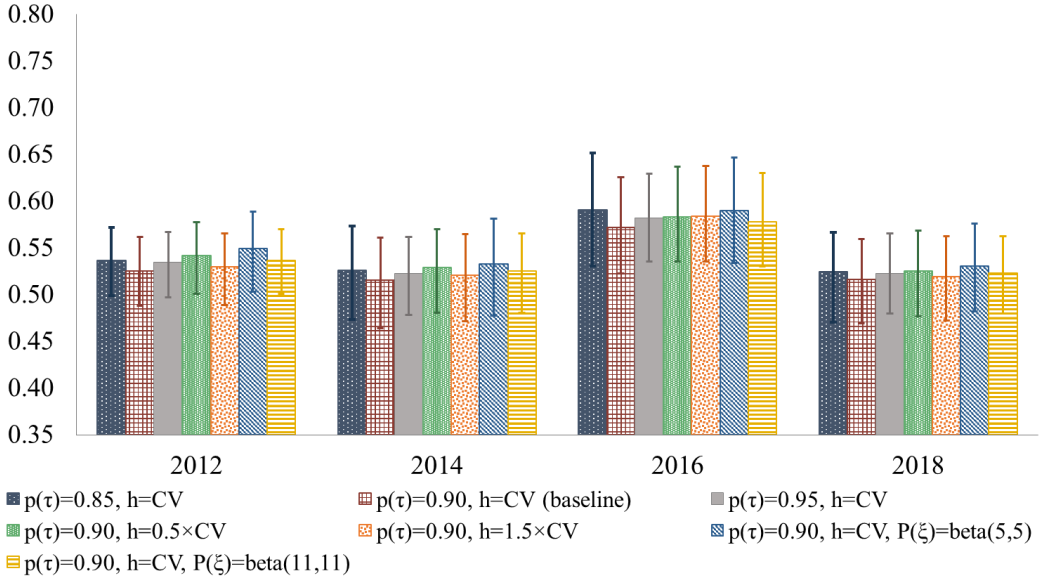


Figure K4: Beijing Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

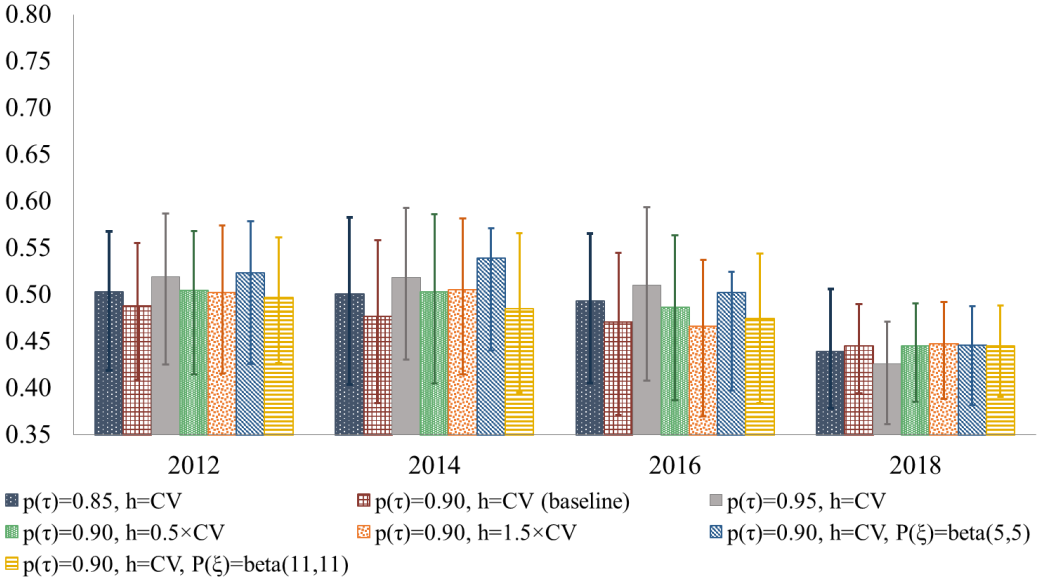


Figure K5: Shanghai Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

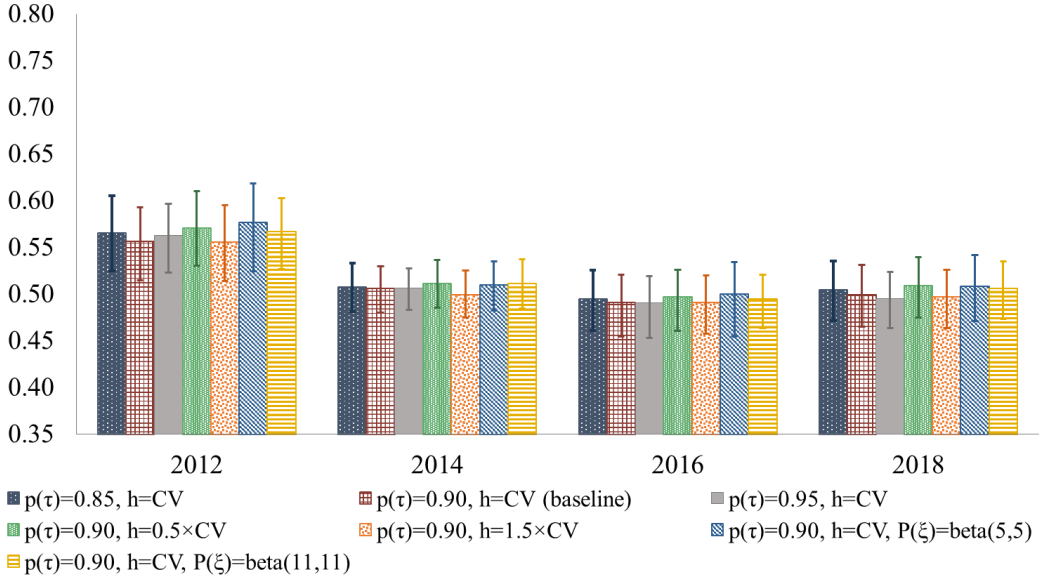


Figure K6: Jiangsu Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

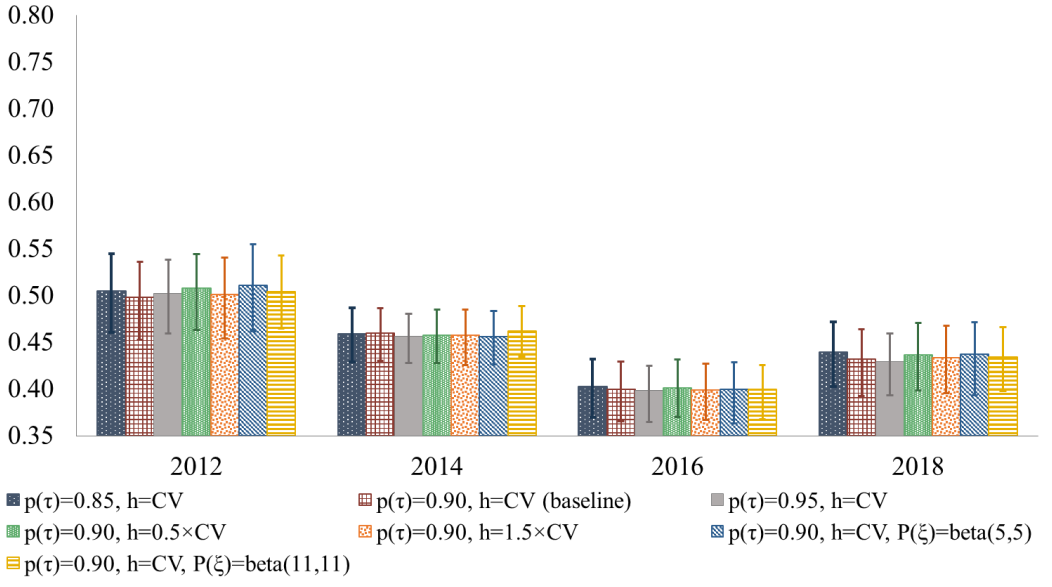


Figure K7: Tianjin Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

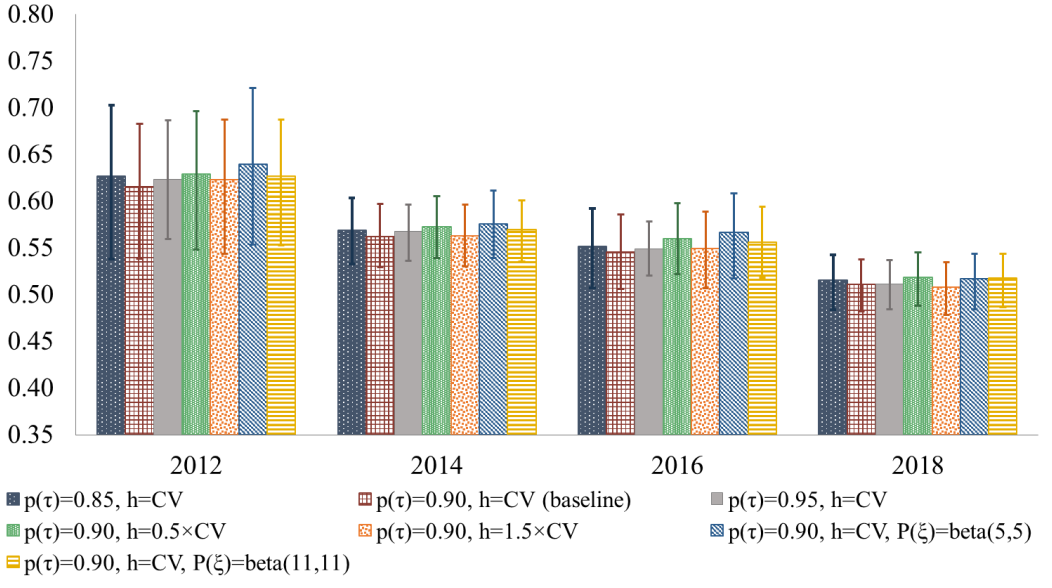


Figure K8: Zhejiang Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

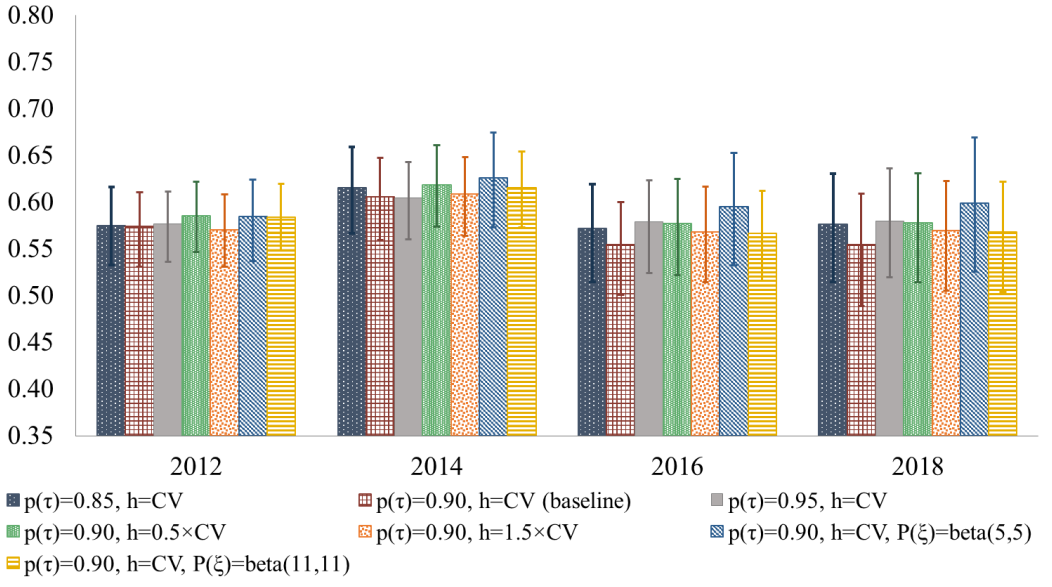


Figure K9: Fujian Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

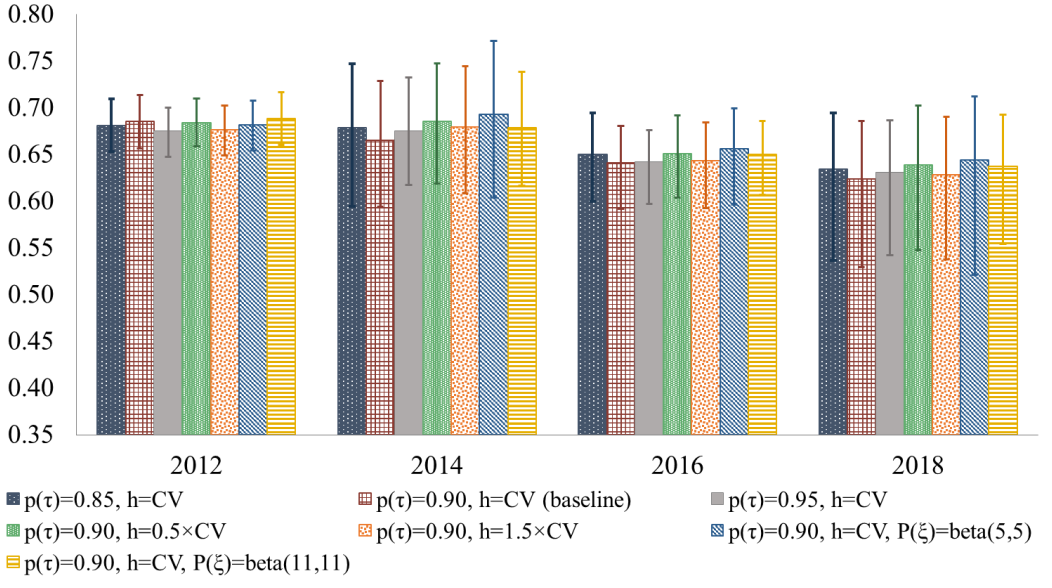


Figure K10: Guangdong Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

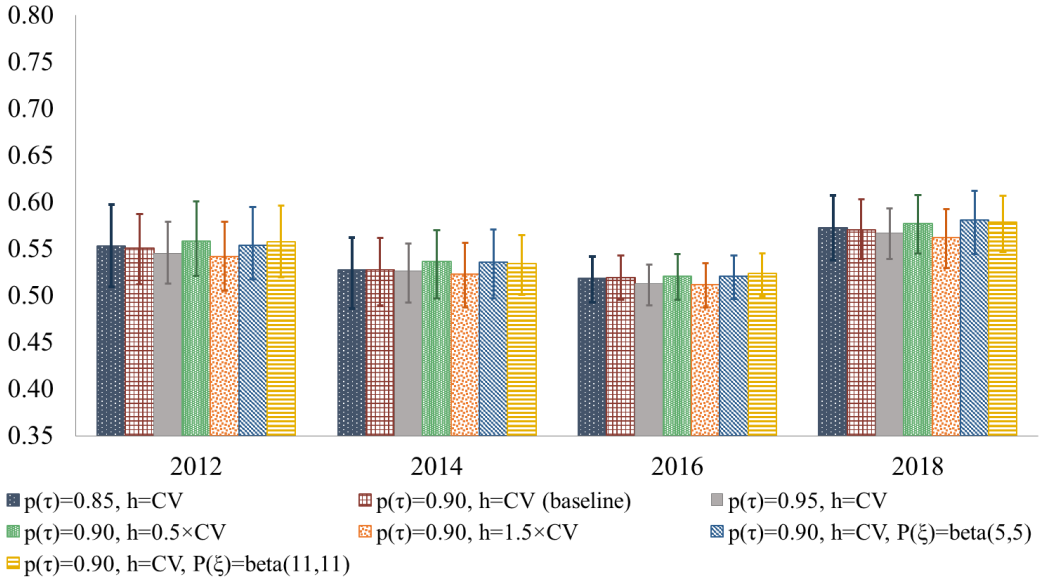


Figure K11: Shandong Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

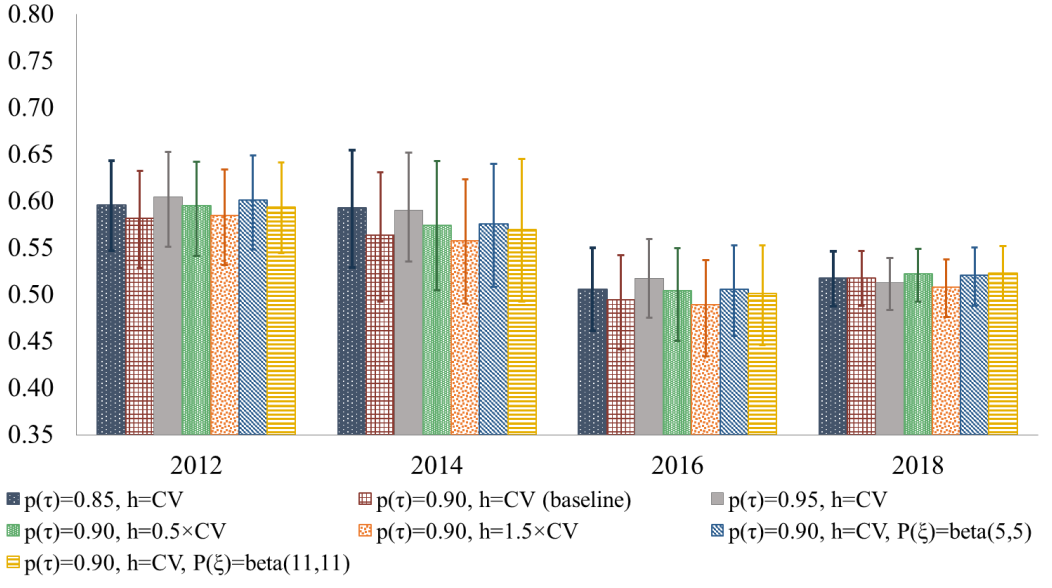


Figure K12: Inner Mongolia Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

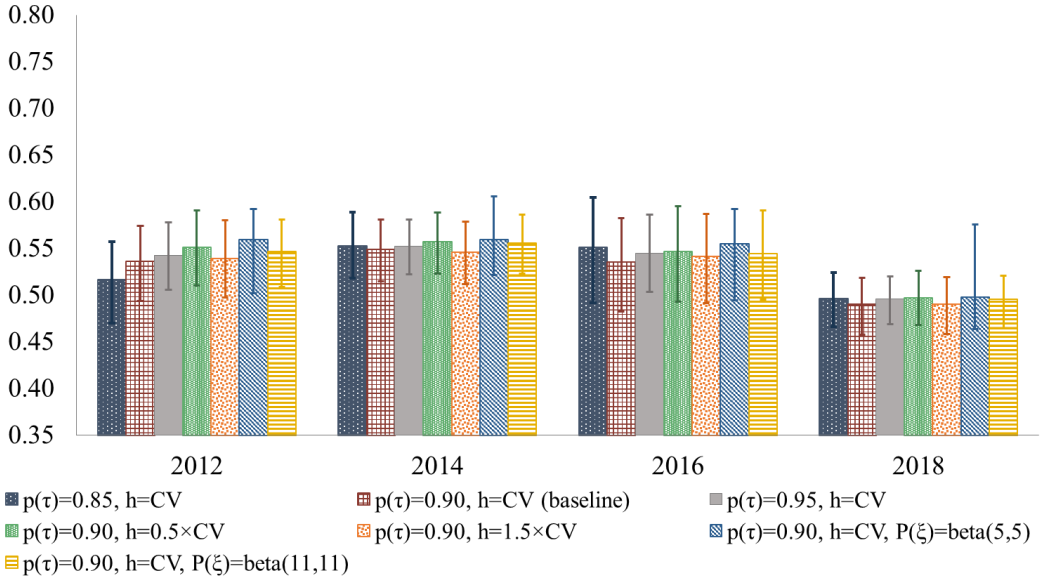


Figure K13: Hubei Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

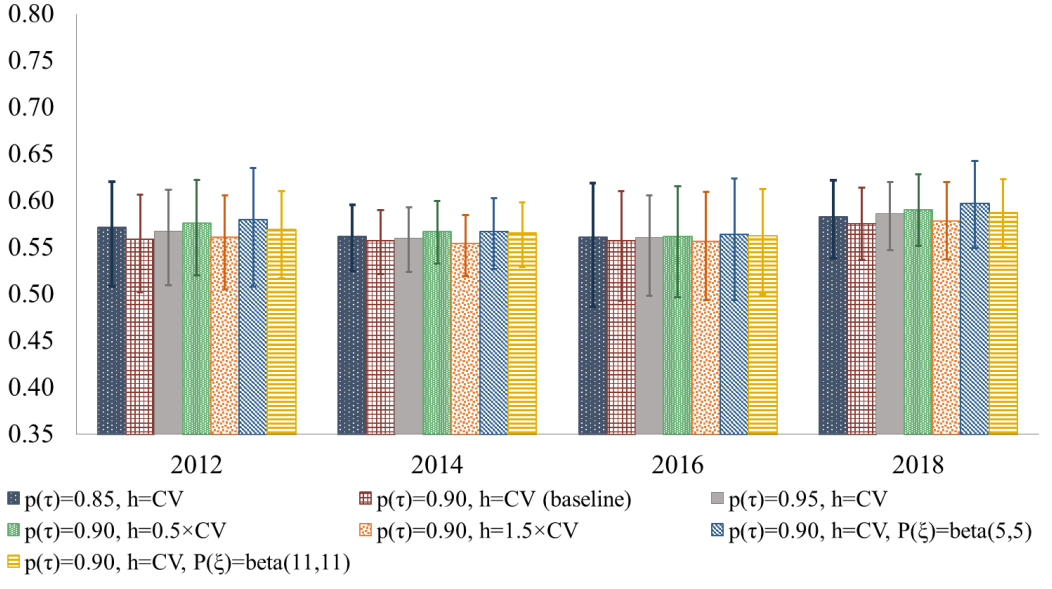


Figure K14: Chongqing Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

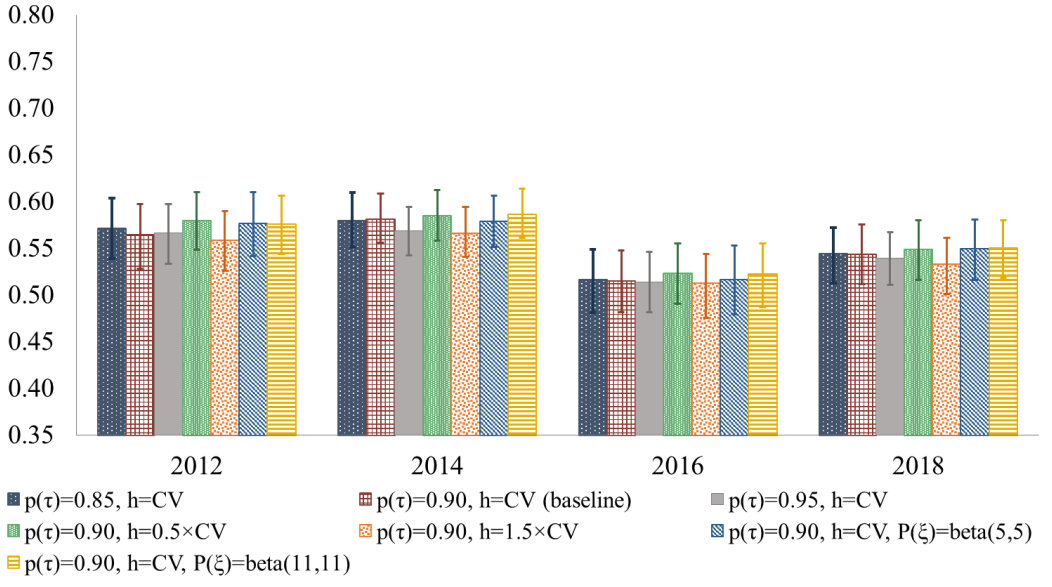


Figure K15: Shaanxi Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

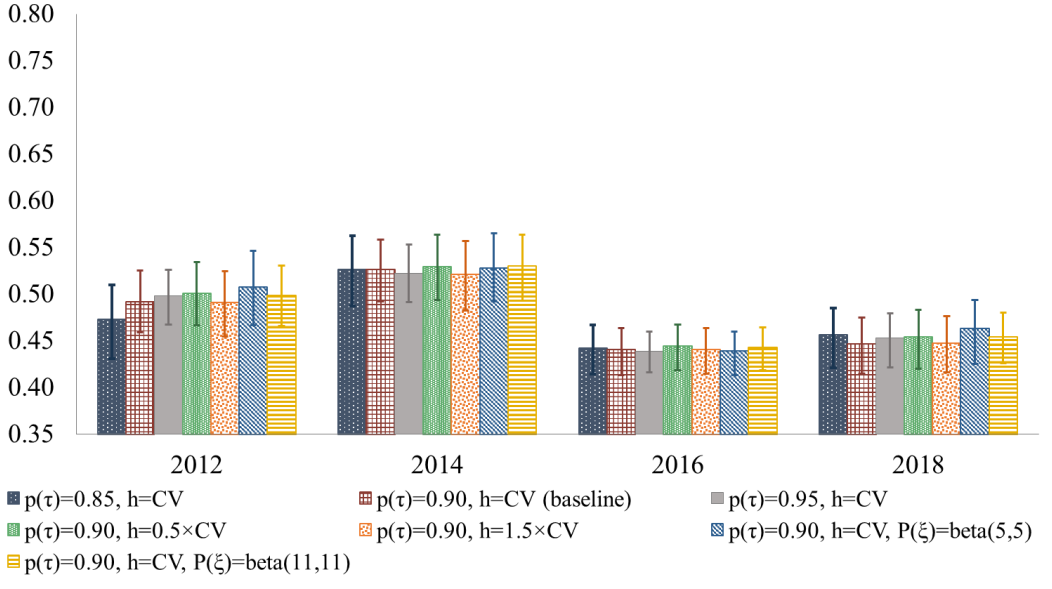


Figure K16: Liaoning Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

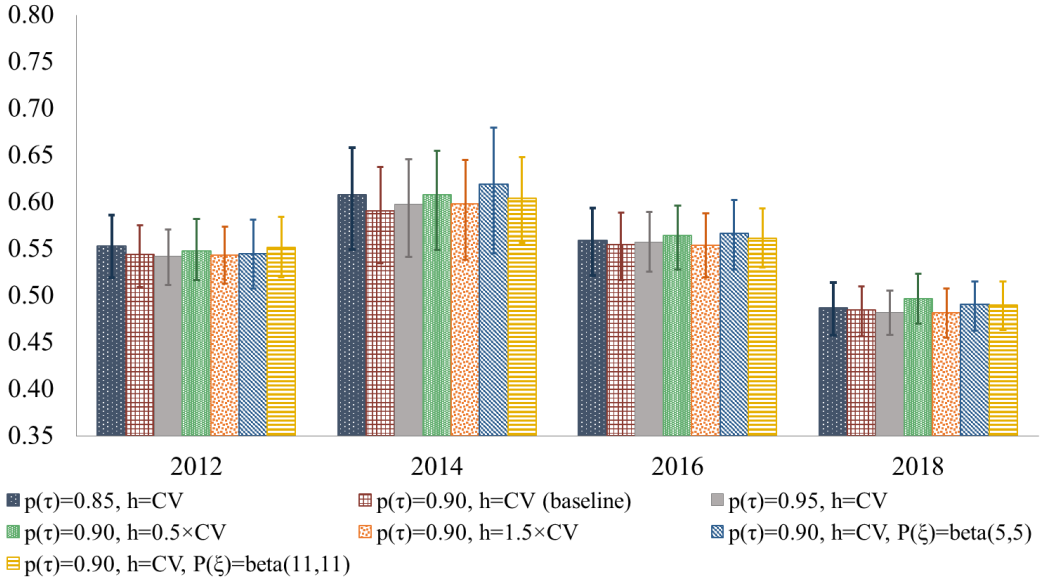


Figure K17: Hunan Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

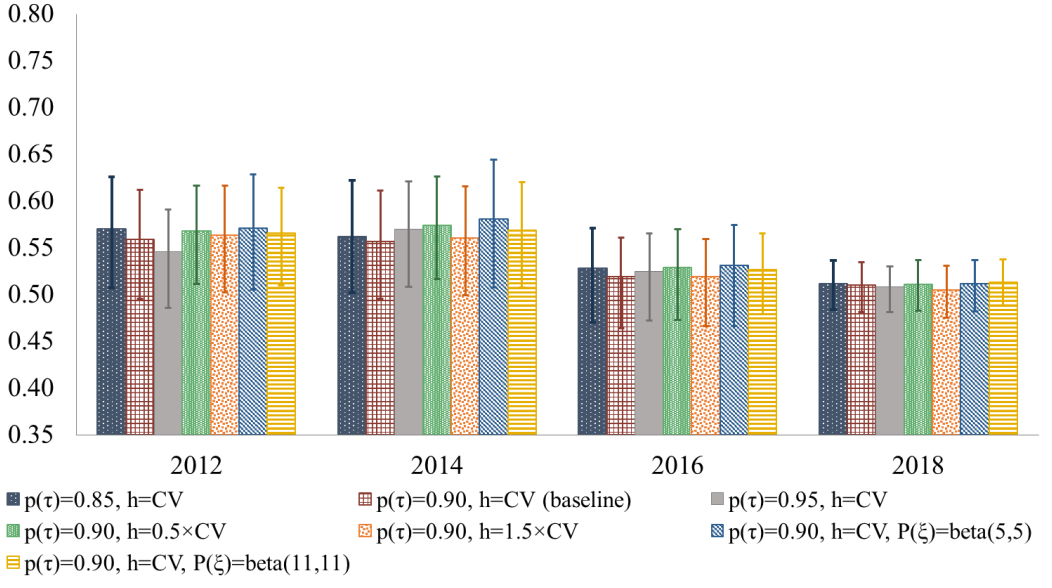


Figure K18: Anhui Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

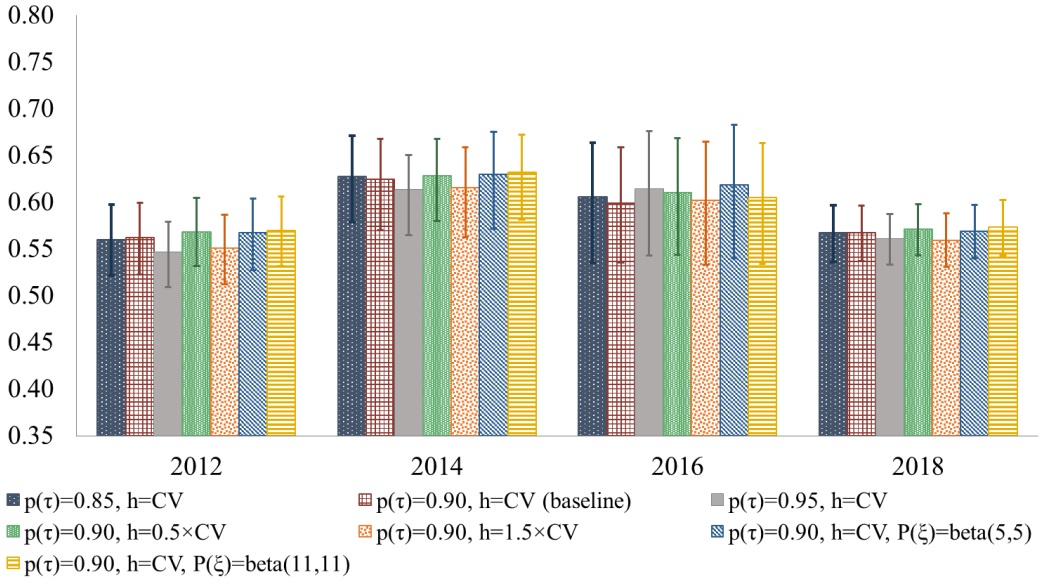


Figure K19: Hainan Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

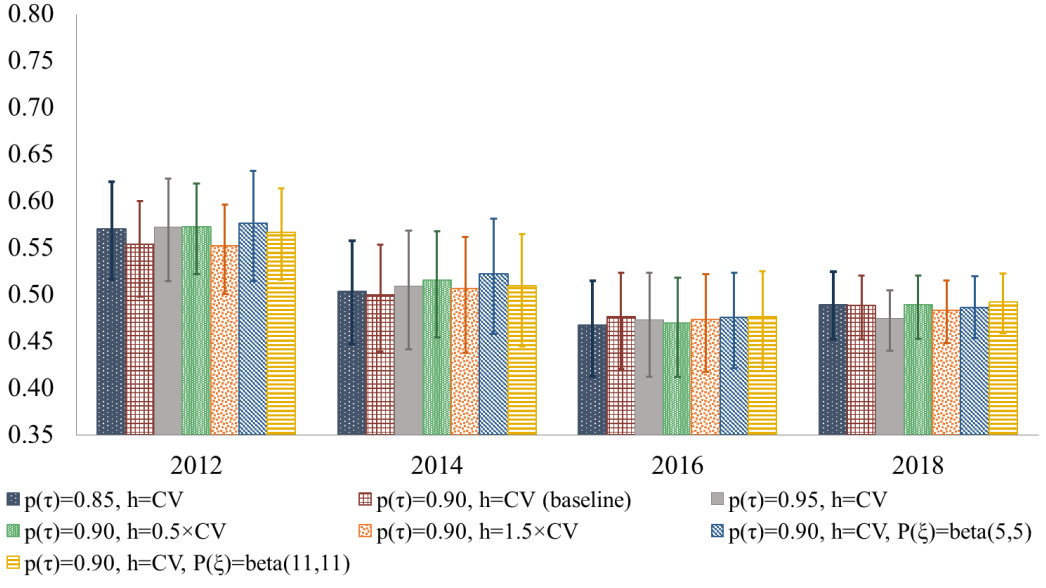


Figure K20: Ningxia Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

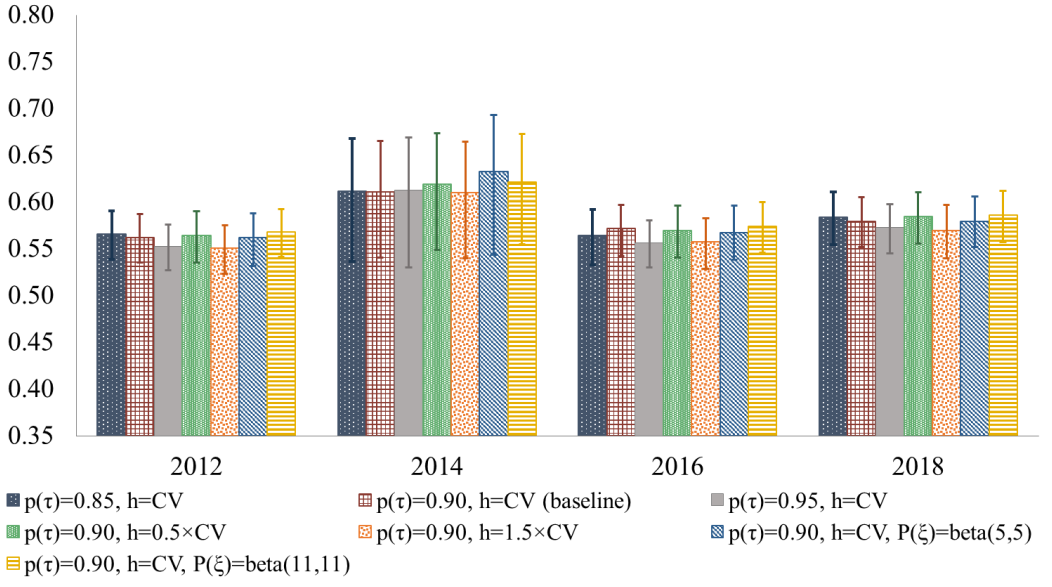


Figure K21: Henan Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

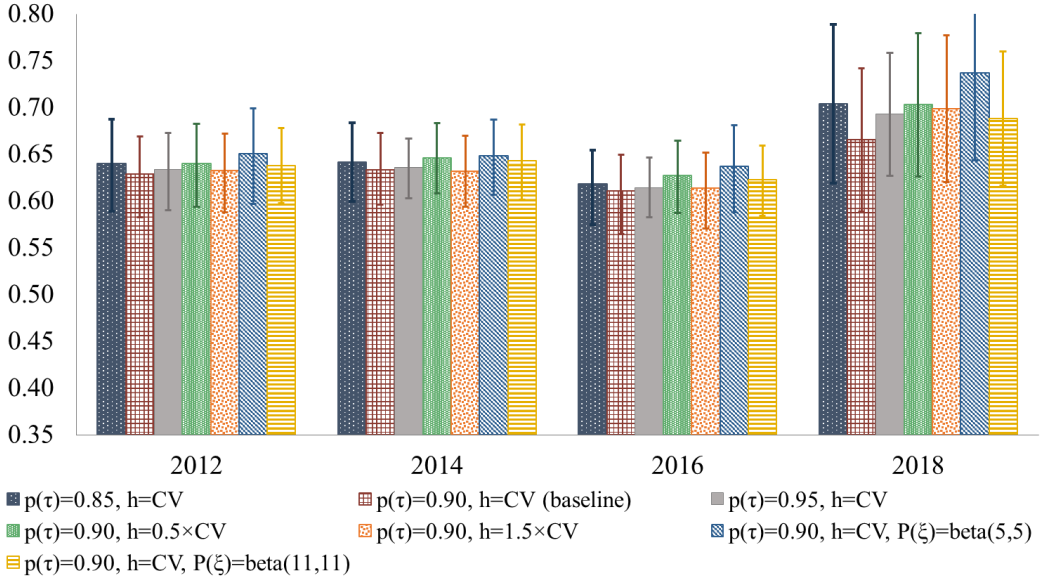


Figure K22: Sichuan Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

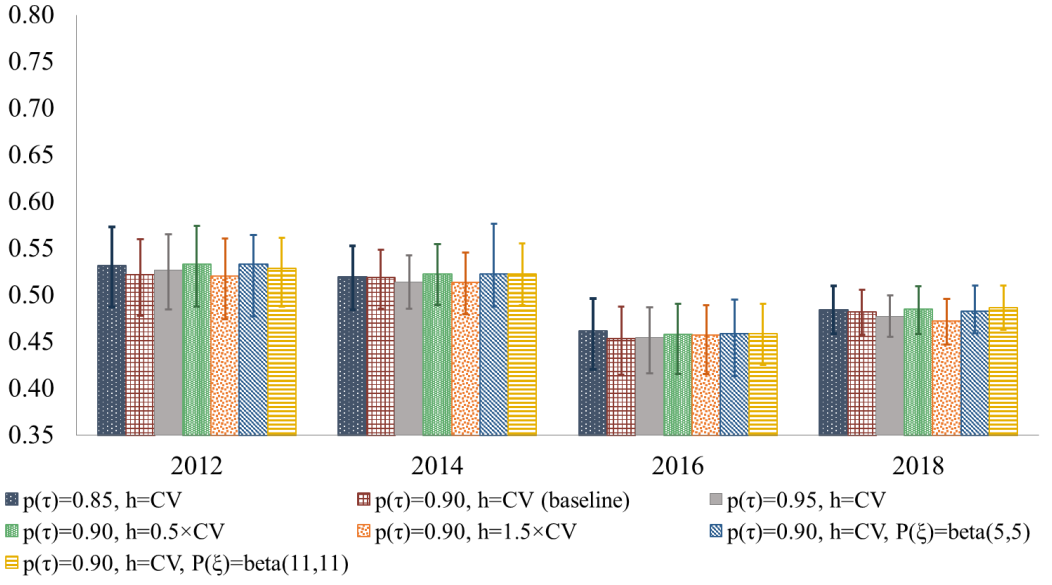


Figure K23: Jilin Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

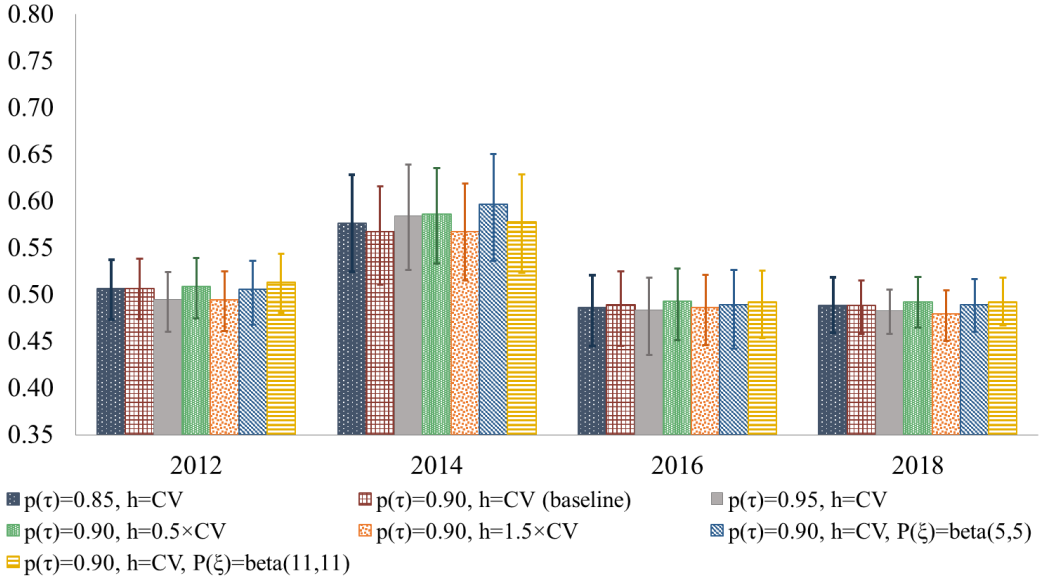


Figure K24: Jiangxi Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

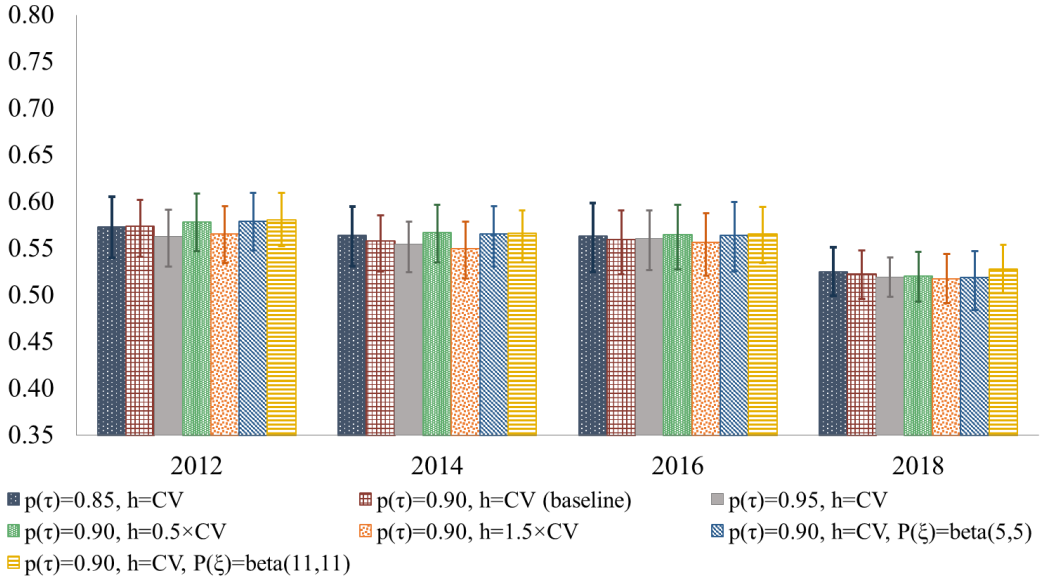


Figure K25: Hebei Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

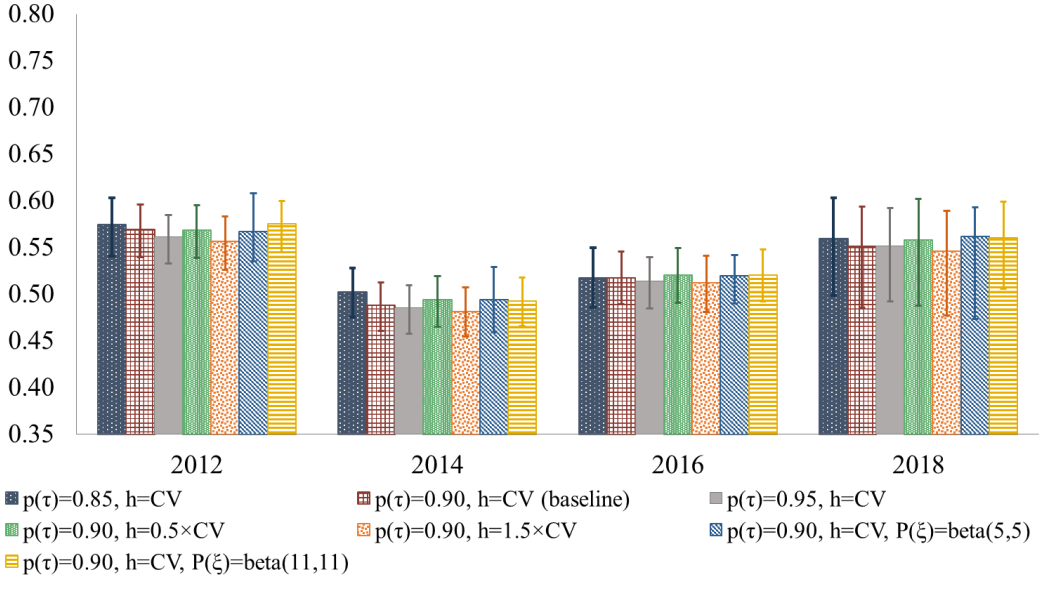


Figure K26: Shanxi Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau)$ = 0.85, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

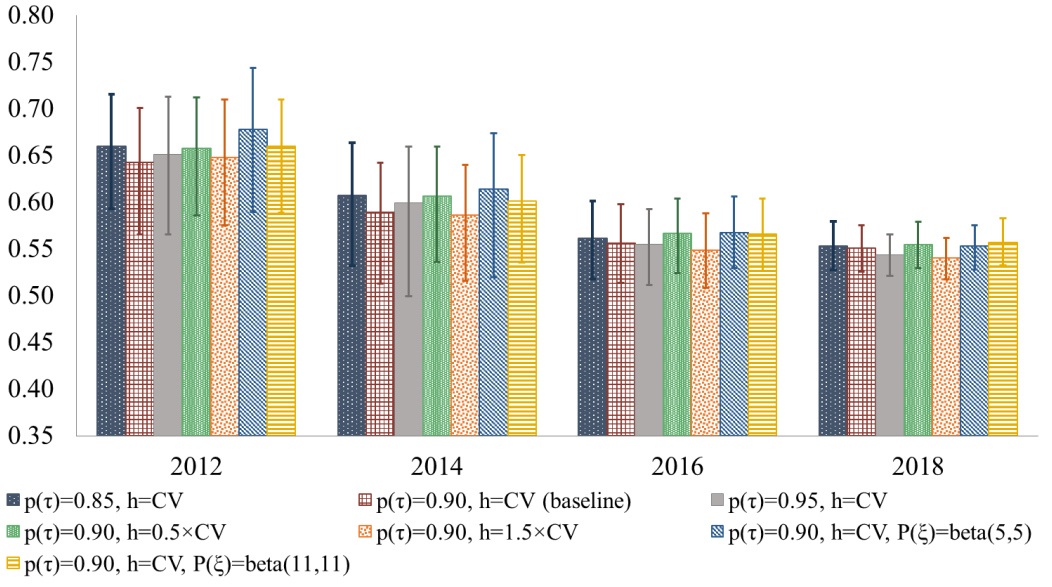


Figure K27: Qinghai Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

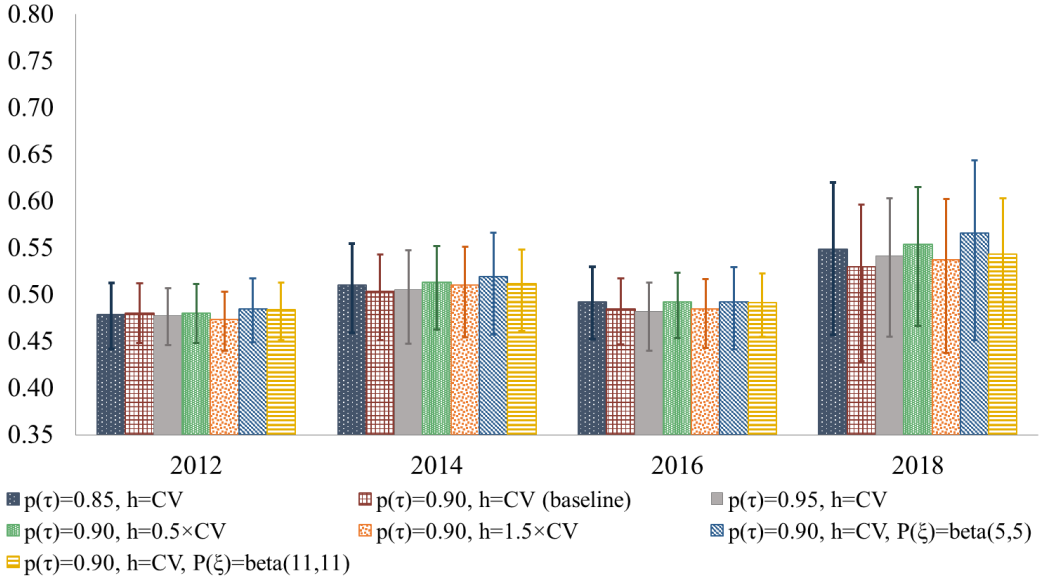


Figure K28: Heilongjiang Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

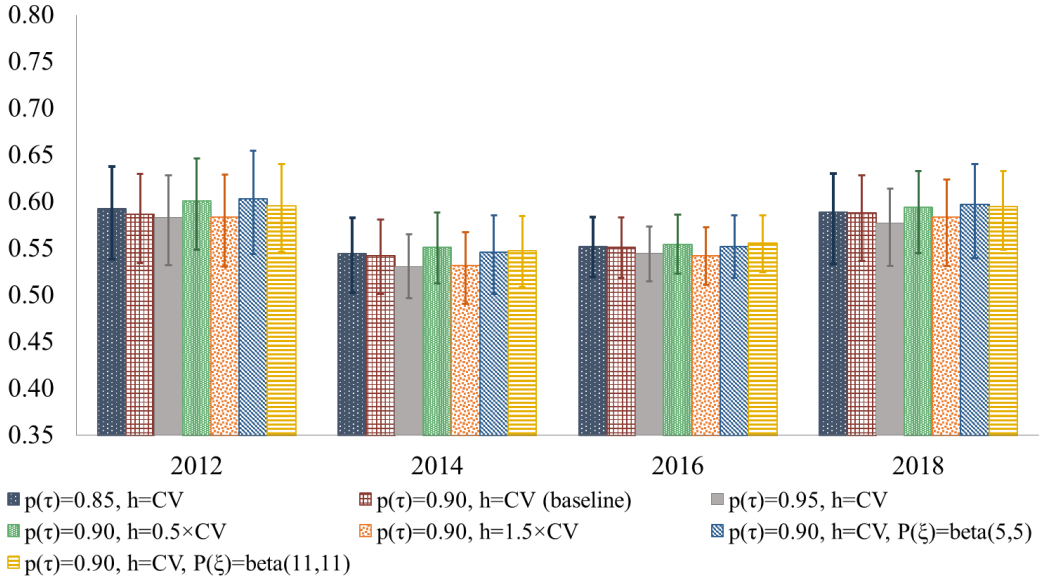


Figure K29: Yunnan Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

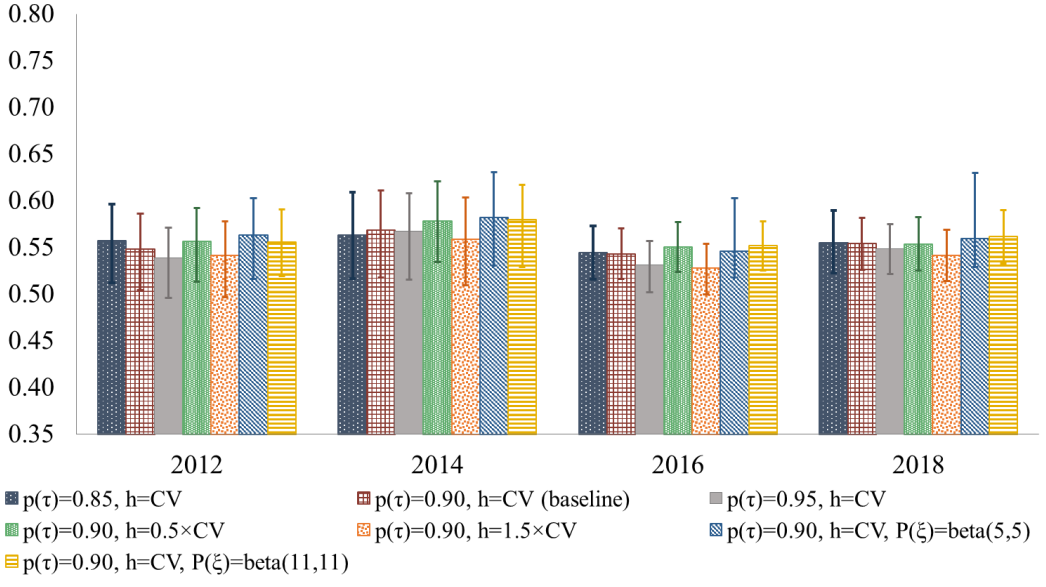


Figure K30: Guangxi Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

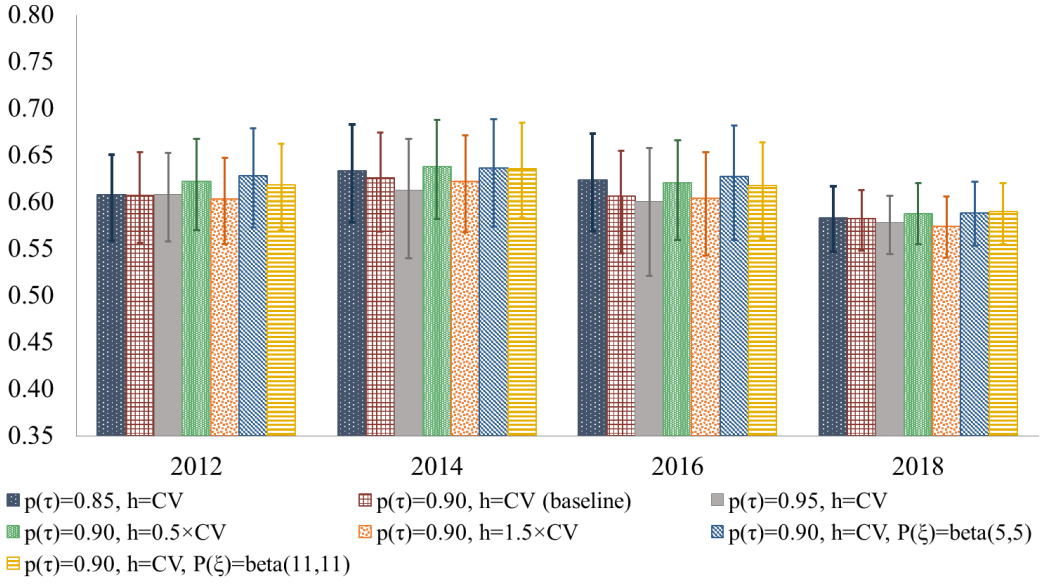


Figure K31: Guizhou Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

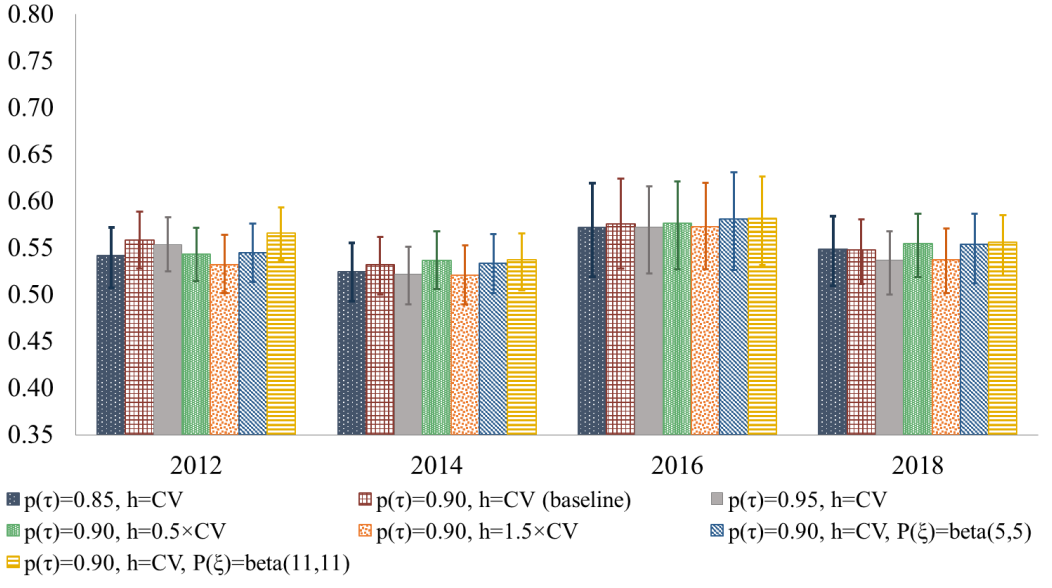


Figure K32: Gansu Gini estimates under the alternative parameter choices. *Notes:* This figure presents the results of the robustness check on different chosen threshold percentiles ($p(\tau) = 0.85$, 0.90, and 0.95), bandwidths (h = 0.5×CV, CV, and 1.5×CV, where CV is optimal bandwidth estimated using cross-validation), and alternative penalty functions ($P(\xi) = beta(5,5)$, or $beta(11,11)$). Error bars represent 95% basic bootstrap CIs (Davison and Hinkley, 1997, p. 194) computed from 1000 samples.

Table K1
Correlates of the provincial Gini index: robustness check on small-sample and influential provinces.

	Dependent variable: Gini index							
	Baseline	W/o Influential	W/o Small	Baseline	W/o Influential	W/o Small	W/o QH	W/o other Small
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Educational Gini	0.364**	0.416***	0.305*	0.133	0.258*	0.074	0.104	0.140
	(0.125)	(0.147)	(0.190)	(0.138)	(0.183)	(0.206)	(0.177)	(0.162)
Returns to schooling	0.719***	0.511***	0.779***					
	(0.131)	(0.101)	(0.168)					
Returns to high school				0.134***	0.089**	0.139***	0.134***	0.141***
				(0.024)	(0.019)	(0.031)	(0.024)	(0.029)
Social spending CSY	−0.015***	−0.019***	−0.021***	−0.017***	−0.010**	−0.022***	−0.019***	−0.020***
	(0.005)	(0.006)	(0.007)	(0.005)	(0.005)	(0.008)	(0.007)	(0.006)
SOE assets share CSY	−0.070*	−0.020	−0.058	−0.089**	−0.113***	−0.077*	−0.089**	−0.076*
	(0.037)	(0.032)	(0.038)	(0.036)	(0.043)	(0.040)	(0.037)	(0.040)
Foreign trade share CSY	0.023	−0.037*	0.033	0.013	−0.030	0.022	0.013	0.022
	(0.027)	(0.022)	(0.030)	(0.025)	(0.020)	(0.030)	(0.025)	(0.030)
Ln(GRP per capita) CSY	−0.016	0.007	−0.019	−0.005	0.0001	−0.007	−0.005	−0.006
	(0.021)	(0.016)	(0.024)	(0.020)	(0.016)	(0.024)	(0.021)	(0.024)
2014.year	0.020*	0.013	0.023**	0.013	0.010	0.013	0.014	0.011
	(0.010)	(0.009)	(0.011)	(0.009)	(0.008)	(0.010)	(0.009)	(0.009)
2016.year	−0.005	−0.016	0.003	0.00002	−0.014	0.004	0.001	0.002
	(0.013)	(0.011)	(0.015)	(0.013)	(0.012)	(0.015)	(0.014)	(0.014)
2018.year	0.003	−0.008	0.017	0.001	−0.021	0.013	0.003	0.008
	(0.018)	(0.015)	(0.021)	(0.017)	(0.014)	(0.021)	(0.019)	(0.020)
Constant	0.755***	0.517***	0.790***	0.653***	0.598***	0.667***	0.652***	0.658***
	(0.214)	(0.163)	(0.251)	(0.209)	(0.168)	(0.251)	(0.213)	(0.244)
Observations	116	96	96	116	96	96	112	100
R^2	0.553	0.669	0.558	0.568	0.624	0.558	0.557	0.571

Nested (cluster + generated-regressor) bootstrap SEs
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table reports robustness checks on the regression estimates of provincial income inequality, as specified in equation (7), using provincial human capital variables. Columns (2) and (5) exclude provinces identified as highly influential by Cook's distance: five provinces (Beijing, Guangdong, Hainan, Sichuan, Qinghai) and six (Beijing, Tianjin, Guangdong, Sichuan, Shanxi, Qinghai), respectively. Columns (3) and (6) exclude the five provinces with the smallest sample sizes: Hainan, Guizhou, Qinghai, Inner Mongolia, and Ningxia. Standard errors in parentheses are from a nested bootstrap with 1,000 replicates (Cameron and Miller, 2015).

Table K2
OLS estimates of Gini index correlates using educational Gini index calculated from CPESY.

	Dependent variable: Gini index													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Educational Gini CPESY	0.515***		0.315**	0.326**	0.454**		0.293*	0.263		0.094	0.170		0.120	0.113
	(0.186)		(0.160)	(0.166)	(0.188)		(0.165)	(0.164)		(0.173)	(0.181)		(0.152)	(0.157)
Returns to schooling		0.916***	0.860***	0.863***		0.805***	0.762***	0.753***						
		(0.160)	(0.164)	(0.168)		(0.144)	(0.140)	(0.140)						
Educational Gini CPESY × Returns to schooling				−1.539				4.453						
				(5.071)				(4.580)						
Returns to high school									0.142***	0.137***	0.159***	0.149***	0.144***	0.142***
									(0.024)	(0.027)	(0.029)	(0.022)	(0.022)	(0.024)
Educational Gini CPESY × Returns to high school											−1.539			0.131
											(0.736)			(0.674)
Social spending CSY					−0.008	−0.011*	−0.012*	−0.012**				−0.016***	−0.017***	−0.017***
					(0.007)	(0.006)	(0.006)	(0.006)				(0.005)	(0.005)	(0.005)
SOE assets share CSY					−0.140***	−0.098***	−0.095***	−0.100***				−0.096***	−0.095***	−0.097***
					(0.039)	(0.035)	(0.037)	(0.037)				(0.036)	(0.036)	(0.037)
Foreign trade share CSY					0.044	−0.004	0.008	0.008				0.002	0.007	0.007
					(0.034)	(0.027)	(0.028)	(0.029)				(0.023)	(0.025)	(0.025)
Ln(GRP per capita) CSY					−0.056**	−0.019	−0.019	−0.019				−0.003	−0.004	−0.004
					(0.022)	(0.021)	(0.021)	(0.022)				(0.020)	(0.020)	(0.020)
2014.year	−0.005	0.014	0.012	0.012	0.007	0.018	0.017	0.017	0.006	0.006	0.005	0.012	0.012	0.012
	(0.012)	(0.011)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
2016.year	−0.036***	−0.029**	−0.033***	−0.033***	−0.009	−0.014	−0.016	−0.016	−0.019*	−0.021*	−0.021*	−0.003	−0.004	−0.004
	(0.014)	(0.011)	(0.012)	(0.012)	(0.016)	(0.015)	(0.015)	(0.015)	(0.011)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
2018.year	−0.034***	−0.025**	−0.028**	−0.029**	0.004	−0.004	−0.005	−0.003	−0.025**	−0.026**	−0.027**	−0.002	−0.002	−0.002
	(0.013)	(0.012)	(0.012)	(0.012)	(0.021)	(0.019)	(0.019)	(0.019)	(0.011)	(0.011)	(0.011)	(0.017)	(0.017)	(0.017)
Constant	0.563***	0.554***	0.557***	0.557***	1.184***	0.799***	0.800***	0.797***	0.553***	0.554***	0.558***	0.628***	0.638***	0.641***
	(0.008)	(0.007)	(0.007)	(0.007)	(0.230)	(0.215)	(0.219)	(0.224)	(0.006)	(0.007)	(0.006)	(0.204)	(0.208)	(0.208)
Observations	116	116	116	116	116	116	116	116	116	116	116	116	116	116
R^2	0.130	0.333	0.357	0.357	0.346	0.485	0.501	0.506	0.371	0.373	0.407	0.561	0.564	0.564

Nested (cluster + generated-regressor) bootstrap SEs
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the regression results that account for provincial income inequality based on provincial human capital factors, as outlined in eq. (7). Standard errors in parentheses are from a nested bootstrap with 1,000 replicates (Cameron and Miller, 2015).

Table K3
Correlates of the provincial returns to education: robustness check on small-sample and influential provinces.

	Dependent variable:										
	Returns to schooling					Returns to high school					
	Baseline	W/o Influential	W/o Small	W/o QH	W/o other Small	Baseline	W/o Influential	W/o Small	W/o QH	W/o other Influential	W/o other Small
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Skilled labor share	−0.119	−0.139**	−0.084	−0.083	−0.125						
	(0.065)	(0.054)	(0.091)	(0.080)	(0.075)						
Skilled labor share HS						−1.226*	−0.732*	−0.771	−0.846	−1.190***	−1.106**
						(0.606)	(0.478)	(0.566)	(0.524)	(0.375)	(0.440)
Secondary employment CSY	−0.064**	−0.037	−0.031	−0.059**	−0.035	−0.426***	−0.415***	−0.339**	−0.373***	−0.416***	−0.467***
	(0.026)	(0.028)	(0.029)	(0.026)	(0.029)	(0.129)	(0.120)	(0.156)	(0.120)	(0.132)	(0.174)
SOE assets share CSY	−0.027	0.030	−0.017	−0.037	−0.004	0.176	0.073	−0.062	−0.057	0.182	0.310
	(0.025)	(0.025)	(0.031)	(0.028)	(0.027)	(0.174)	(0.130)	(0.146)	(0.138)	(0.175)	(0.190)
Foreign trade share CSY	0.015	0.009	0.012	0.010	0.018	0.116	−0.050	0.054	0.064	0.105	−0.0001
	(0.013)	(0.011)	(0.015)	(0.014)	(0.014)	(0.121)	(0.098)	(0.127)	(0.119)	(0.100)	(0.129)
2014.year	−0.015**	−0.012*	−0.012	−0.016**	−0.011	0.011	−0.005	0.008	−0.003	0.022	0.008
	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.051)	(0.042)	(0.046)	(0.043)	(0.050)	(0.056)
2016.year	0.004	0.007	0.006	0.003	0.006	0.004	−0.020	0.003	−0.005	0.007	−0.014
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.054)	(0.052)	(0.051)	(0.050)	(0.056)	(0.058)
2018.year	−0.008	−0.002	−0.003	−0.006	−0.006	−0.028	−0.011	−0.019	−0.020	−0.033	−0.021
	(0.009)	(0.009)	(0.010)	(0.009)	(0.010)	(0.054)	(0.047)	(0.055)	(0.049)	(0.052)	(0.062)
Constant	0.275***	0.262***	0.228***	0.249***	0.257***	1.387***	1.248***	1.224***	1.272***	1.369***	1.361***
	(0.034)	(0.030)	(0.046)	(0.041)	(0.035)	(0.146)	(0.120)	(0.164)	(0.133)	(0.135)	(0.184)
Observations	116	100	96	112	100	116	104	96	112	100	108
R^2	0.165	0.228	0.101	0.148	0.133	0.199	0.251	0.134	0.166	0.188	0.292

Nested (cluster + generated-regressor) bootstrap SEs
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table reports robustness checks on the regression estimates of provincial returns to schooling and returns to high school, as specified in equation (10), using provincial labor demand and labor supply factors. Columns (2) and (7) exclude provinces identified as highly influential by Cook's distance: four provinces (Guangdong, Hunan, Hainan, Jilin) and three (Guangdong, Jilin, Qinghai), respectively. Columns (3) and (8) exclude the five provinces with the smallest sample sizes: Hainan, Guizhou, Qinghai, Inner Mongolia, and Ningxia. Standard errors in parentheses are from a nested bootstrap with 1,000 replicates (Cameron and Miller, 2015).

L. Simulation study: additional results

Table L1
Theil T from policy simulations: equalizing provincial education structures and returns to high school to Shanghai levels.

	Baseline Estimates 2018		(a) Uniform educational structure		(b) Uniform educational structure and adjusted RS		(c) Uniform RS		(d) Uniform educational structure and uniform RS	
	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Beijing	0.342	4%	0.350	3%	0.337	3%	0.346	4%	0.320	3%
Shanghai	0.197	2%	0.228	2%	0.196	2%	0.224	2%	0.204	2%
Jiangsu	0.353	7%	0.312	6%	0.284	6%	0.294	6%	0.279	6%
Tianjin	0.256	1%	0.266	1%	0.257	1%	0.255	1%	0.235	1%
Zhejiang	0.376	6%	0.327	7%	0.293	5%	0.318	6%	0.293	6%
Fujian	0.393	3%	0.335	3%	0.328	3%	0.320	2%	0.306	3%
Guangdong	0.510	9%	0.398	9%	0.392	10%	0.425	10%	0.375	9%
Shandong	0.458	6%	0.391	6%	0.359	5%	0.376	6%	0.358	6%
Inner Mongolia	0.385	2%	0.305	2%	0.305	2%	0.289	1%	0.278	1%
Hubei	0.328	3%	0.291	3%	0.259	3%	0.274	3%	0.256	3%
Chongqing	0.454	2%	0.393	2%	0.357	2%	0.373	2%	0.357	2%
Shaanxi	0.460	2%	0.357	2%	0.383	2%	0.320	1%	0.334	2%
Liaoning	0.290	3%	0.283	3%	0.284	3%	0.249	2%	0.257	2%
Hunan	0.329	3%	0.298	4%	0.257	3%	0.289	4%	0.262	4%
Anhui	0.378	3%	0.301	4%	0.285	4%	0.288	3%	0.271	3%
Hainan	0.493	0%	0.414	0%	0.440	1%	0.382	0%	0.391	0%
Ningxia	0.350	0%	0.310	0%	0.296	0%	0.298	0%	0.281	0%
Henan	0.533	6%	0.418	6%	0.425	7%	0.406	5%	0.386	5%
Sichuan	0.506	6%	0.395	6%	0.403	7%	0.376	5%	0.377	6%
Jilin	0.341	1%	0.316	1%	0.287	1%	0.320	2%	0.287	1%
Jiangxi	0.330	2%	0.286	2%	0.263	2%	0.275	2%	0.255	2%
Hebei	0.394	3%	0.323	3%	0.300	3%	0.328	3%	0.297	3%
Shanxi	0.430	2%	0.351	2%	0.334	2%	0.351	2%	0.324	2%
Qinghai	0.477	0%	0.414	0%	0.444	1%	0.383	0%	0.396	0%
Heilongjiang	0.361	2%	0.328	3%	0.311	3%	0.316	2%	0.302	3%
Yunnan	0.491	3%	0.371	3%	0.394	4%	0.324	2%	0.322	2%
Guangxi	0.470	2%	0.410	2%	0.408	3%	0.390	2%	0.387	2%
Guizhou	0.497	2%	0.423	2%	0.410	2%	0.412	2%	0.397	2%
Gansu	0.457	1%	0.397	1%	0.422	2%	0.359	1%	0.374	1%
Within	0.397	86%	0.343	89%	0.332	90%	0.330	82%	0.312	85%
Between	0.065	14%	0.044	11%	0.038	10%	0.074	18%	0.057	15%
Total	0.462	100%	0.388	100%	0.370	100%	0.404	100%	0.369	100%

Notes: This table reports the results of the policy simulation study based on the data from the CHFS 2019. Monte Carlo Simulations were repeated 1000 times.

Table L2
Theil T from policy simulations: equalizing provincial education structures and returns to schooling to national levels.

	Baseline Estimates 2018		(a) Uniform educational structure		(b) Uniform educational structure and adjusted RS		(c) Uniform RS		(d) Uniform educational structure and uniform RS	
	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Beijing	0.342	4%	0.323	2%	0.331	3%	0.363	2%	0.323	1%
Shanghai	0.197	2%	0.247	2%	0.194	2%	0.273	4%	0.245	2%
Jiangsu	0.353	7%	0.323	6%	0.335	7%	0.298	3%	0.328	4%
Tianjin	0.256	1%	0.260	1%	0.261	1%	0.295	1%	0.261	1%
Zhejiang	0.376	6%	0.360	7%	0.313	6%	0.364	6%	0.341	7%
Fujian	0.393	3%	0.367	3%	0.328	3%	0.402	3%	0.354	3%
Guangdong	0.510	9%	0.375	8%	0.374	9%	0.476	7%	0.375	6%
Shandong	0.458	6%	0.399	6%	0.378	6%	0.403	4%	0.395	5%
Inner Mongolia	0.385	2%	0.360	2%	0.362	2%	0.354	1%	0.363	1%
Hubei	0.328	3%	0.323	4%	0.263	3%	0.335	4%	0.305	4%
Chongqing	0.454	2%	0.395	2%	0.408	2%	0.373	1%	0.400	1%
Shaanxi	0.460	2%	0.330	2%	0.364	2%	0.345	1%	0.331	1%
Liaoning	0.290	3%	0.310	3%	0.319	3%	0.288	2%	0.314	2%
Hunan	0.329	3%	0.322	4%	0.252	3%	0.354	6%	0.308	6%
Anhui	0.378	3%	0.321	4%	0.265	3%	0.402	5%	0.300	4%
Hainan	0.493	0%	0.414	0%	0.498	1%	0.389	0%	0.439	0%
Ningxia	0.350	0%	0.349	0%	0.304	0%	0.369	0%	0.333	0%
Henan	0.533	6%	0.431	6%	0.454	7%	0.466	3%	0.437	3%
Sichuan	0.506	6%	0.416	6%	0.435	7%	0.433	3%	0.423	4%
Jilin	0.341	1%	0.352	2%	0.281	1%	0.413	3%	0.338	2%
Jiangxi	0.330	2%	0.329	3%	0.270	2%	0.353	3%	0.310	3%
Hebei	0.394	3%	0.335	3%	0.293	3%	0.386	3%	0.326	3%
Shanxi	0.430	2%	0.381	2%	0.352	2%	0.397	2%	0.374	2%
Qinghai	0.477	0%	0.424	1%	0.388	1%	0.517	0%	0.409	1%
Heilongjiang	0.361	2%	0.364	3%	0.325	3%	0.378	3%	0.354	3%
Yunnan	0.491	3%	0.341	2%	0.369	3%	0.413	1%	0.349	1%
Guangxi	0.470	2%	0.443	2%	0.446	3%	0.444	1%	0.445	2%
Guizhou	0.497	2%	0.465	2%	0.436	2%	0.463	1%	0.454	2%
Gansu	0.457	1%	0.389	1%	0.455	2%	0.384	0%	0.409	1%
Within	0.397	86%	0.359	91%	0.343	91%	0.377	73%	0.344	77%
Between	0.065	14%	0.036	9%	0.034	9%	0.139	27%	0.105	23%
Total	0.462	100%	0.395	100%	0.377	100%	0.516	100%	0.449	100%

Notes: This table reports the results of the policy simulation study based on the data from the CHFS 2019. Monte Carlo Simulations were repeated 1000 times.

Table L3
Theil T from policy simulations: equalizing provincial education structures and returns to schooling to Shanghai levels.

	Baseline Estimates 2018		(a) Uniform educational structure		(b) Uniform educational structure and adjusted RS		(c) Uniform RS		(d) Uniform educational structure and uniform RS	
	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)	Theil T index	Share (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Beijing	0.342	4%	0.322	2%	0.326	3%	0.292	2%	0.278	1%
Shanghai	0.197	2%	0.240	2%	0.196	1%	0.224	3%	0.198	2%
Jiangsu	0.353	7%	0.321	6%	0.331	7%	0.257	3%	0.288	4%
Tianjin	0.256	1%	0.257	1%	0.257	1%	0.231	1%	0.217	1%
Zhejiang	0.376	6%	0.359	7%	0.314	6%	0.298	6%	0.302	6%
Fujian	0.393	3%	0.368	3%	0.327	3%	0.330	3%	0.314	3%
Guangdong	0.510	9%	0.375	8%	0.373	9%	0.380	6%	0.337	6%
Shandong	0.458	6%	0.402	6%	0.375	6%	0.355	5%	0.358	5%
Inner Mongolia	0.385	2%	0.362	2%	0.361	2%	0.298	1%	0.328	1%
Hubei	0.328	3%	0.319	4%	0.263	3%	0.284	4%	0.263	4%
Chongqing	0.454	2%	0.392	2%	0.403	2%	0.333	1%	0.361	1%
Shaanxi	0.460	2%	0.333	2%	0.365	2%	0.291	1%	0.297	1%
Liaoning	0.290	3%	0.307	3%	0.316	3%	0.237	1%	0.268	2%
Hunan	0.329	3%	0.320	4%	0.252	3%	0.304	6%	0.268	6%
Anhui	0.378	3%	0.318	4%	0.264	3%	0.319	5%	0.264	4%
Hainan	0.493	0%	0.414	0%	0.496	1%	0.356	0%	0.406	0%
Ningxia	0.350	0%	0.352	0%	0.307	0%	0.312	0%	0.297	0%
Henan	0.533	6%	0.424	6%	0.448	7%	0.393	3%	0.399	3%
Sichuan	0.506	6%	0.418	7%	0.436	8%	0.368	3%	0.385	4%
Jilin	0.341	1%	0.351	2%	0.280	1%	0.347	3%	0.294	2%
Jiangxi	0.330	2%	0.330	3%	0.270	2%	0.292	3%	0.268	3%
Hebei	0.394	3%	0.333	3%	0.293	3%	0.327	3%	0.288	3%
Shanxi	0.430	2%	0.376	2%	0.350	2%	0.342	2%	0.332	2%
Qinghai	0.477	0%	0.428	1%	0.390	1%	0.419	0%	0.376	1%
Heilongjiang	0.361	2%	0.361	3%	0.324	3%	0.323	3%	0.311	3%
Yunnan	0.491	3%	0.340	2%	0.366	3%	0.314	1%	0.315	1%
Guangxi	0.470	2%	0.444	2%	0.447	3%	0.390	1%	0.409	2%
Guizhou	0.497	2%	0.465	2%	0.436	2%	0.404	2%	0.414	2%
Gansu	0.457	1%	0.392	1%	0.449	2%	0.334	0%	0.368	1%
Within	0.397	86%	0.357	91%	0.346	91%	0.316	72%	0.305	75%
Between	0.065	14%	0.036	9%	0.035	9%	0.126	28%	0.104	25%
Total	0.462	100%	0.394	100%	0.381	100%	0.441	100%	0.409	100%

Notes: This table reports the results of the policy simulation study based on the data from the CHFS 2019. Monte Carlo Simulations were repeated 1000 times.