

Butaeva, Kristina; Park, Albert

Working Paper

Provincial-level income inequality in the People's Republic of China: The role of human capital

ADB Economics Working Paper Series, No. 813

Provided in Cooperation with:

Asian Development Bank (ADB), Manila

Suggested Citation: Butaeva, Kristina; Park, Albert (2025) : Provincial-level income inequality in the People's Republic of China: The role of human capital, ADB Economics Working Paper Series, No. 813, Asian Development Bank (ADB), Manila, <https://doi.org/10.22617/WPS250426-2>

This Version is available at:

<https://hdl.handle.net/10419/336525>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/3.0/igo/>

PROVINCIAL-LEVEL INCOME INEQUALITY IN THE PEOPLE'S REPUBLIC OF CHINA

THE ROLE OF HUMAN CAPITAL

Kristina Butaeva and Albert F. Park

NO. 813

October 2025

ADB ECONOMICS
WORKING PAPER SERIES

ADB Economics Working Paper Series

Provincial-Level Income Inequality in the People's Republic of China: The Role of Human Capital

Kristina Butaeva and Albert F. Park

No. 813 | October 2025

The *ADB Economics Working Paper Series* presents research in progress to elicit comments and encourage debate on development issues in Asia and the Pacific. The views expressed are those of the authors and do not necessarily reflect the views and policies of ADB or its Board of Governors or the governments they represent.

Kristina Butaeva (kbutaeva@uchicago.edu) is a postdoctoral scholar at the Stone Center for Research on Wealth Inequality and Mobility at the University of Chicago. Albert F. Park (afpark@adb.org) is the chief economist and director general of the Economic Research and Development Impact Department, Asian Development Bank.



Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO)

© 2025 Asian Development Bank
6 ADB Avenue, Mandaluyong City, 1550 Metro Manila, Philippines
Tel +63 2 8632 4444; Fax +63 2 8636 2444
www.adb.org

Some rights reserved. Published in 2025.

ISSN 2313-6537 (print), 2313-6545 (PDF)
Publication Stock No. WPS250426-2
DOI: <http://dx.doi.org/10.22617/WPS250426-2>

The views expressed in this publication are those of the authors and do not necessarily reflect the views and policies of the Asian Development Bank (ADB) or its Board of Governors or the governments they represent.

ADB does not guarantee the accuracy of the data included in this publication and accepts no responsibility for any consequence of their use. The mention of specific companies or products of manufacturers does not imply that they are endorsed or recommended by ADB in preference to others of a similar nature that are not mentioned.

By making any designation of or reference to a particular territory or geographic area in this document, ADB does not intend to make any judgments as to the legal or other status of any territory or area.

This publication is available under the Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO) <https://creativecommons.org/licenses/by/3.0/igo/>. By using the content of this publication, you agree to be bound by the terms of this license. For attribution, translations, adaptations, and permissions, please read the provisions and terms of use at <https://www.adb.org/terms-use#openaccess>.

This CC license does not apply to non-ADB copyright materials in this publication. If the material is attributed to another source, please contact the copyright owner or publisher of that source for permission to reproduce it. ADB cannot be held liable for any claims that arise as a result of your use of the material.

Please contact pubsmarketing@adb.org if you have questions or comments with respect to content, or if you wish to obtain copyright permission for your intended use that does not fall within these terms, or for permission to use the ADB logo.

Corrigenda to ADB publications may be found at <http://www.adb.org/publications/corrigenda>.

Notes:

In this publication, “CNY” refers to yuan.

ADB recognizes “China” as the People’s Republic of China and “Hong Kong” as Hong Kong, China.

ABSTRACT

In this paper, we conduct the first systematic empirical analysis of income inequality in the People's Republic of China (PRC) at the provincial level. Using data from the China Household Finance Survey (CHFS) and a semiparametric distribution model, we estimate Gini indices for a set of provincial-level administrative units in 2012, 2014, 2016, and 2018. We find that differences in the “prices” and “quantities” of human capital are important factors in explaining differences in inequality between these provincial areas. Our findings suggest that poor areas are highly disadvantaged compared with rich ones, as they face higher income and educational inequality, as well as higher returns to education, while at the same time exhibiting lower average educational attainment. We conclude that filling existing interprovincial human capital gaps and accelerating labor market integration through appropriate government policies could reduce spatial disparities in inequality levels across regions and overall income inequality in the PRC.

Keywords: income inequality, provinces, People's Republic of China, human capital

JEL codes: D31, I24, O15

We thank Grigory Kosenok, Shlomo Weber, Steven N. Durlauf, Li Gan, Jin Wang, Li Han, and Sujata Visaria for their insightful suggestions and indispensable feedback. Additionally, we appreciate the helpful comments from participants of the Applied Micro Seminar at Hong Kong University of Science and Technology (HKUST). Butaeva gratefully acknowledges the financial support of the James M. and Cathleen D. Stone Foundation and the Greater Eurasia Research Center (GEAR) at HKUST. She also recognizes Grace Kolavo and Joseph P. Quinn for their instrumental support in the preparation of the manuscript. Corresponding author: Kristina Butaeva (kbutaeva@uchicago.edu).

1 INTRODUCTION

Over the past 4 decades, the Chinese economy has experienced impressive economic growth. At the same time, the level of spatial and individual income inequality in the People's Republic of China (PRC) has risen dramatically, and it seems that the fruits of economic success are still unevenly distributed.¹ Individual and household income inequality in the PRC, 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.³

Why is an analysis of provincial inequality important? First, intraprovincial income inequality contributes substantially to national-level inequality. The decomposition analysis from our study suggests intraprovincial income variation can explain around 90% of the nationwide variation in annual household income per capita in 2012–2018. Therefore, analysis of the possible causes of and the main contributing factors in intraprovincial income inequality, coupled with an estimation of each province's contribution to nationwide inequality, has the potential to generate appropriate 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 gross regional product (GRP) per capita or mean disposable income. Third, provincial income inequality has the capacity to impede provincial and national economic growth and political stability. The negative impact of high income inequality on growth is well documented in the literature. For example, high income inequality could create inequality of opportunities, which limits the human capital development of disadvantaged social groups (Galor and Moav, 2006; Perotti, 1996). Another plausible mechanism is that high income inequality reduces social cohesion and may increase the risk of political conflicts.⁴ Political instability creates uncertainty, which could adversely impact investments and economic growth (Alesina and Perotti, 1996;

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 Studies on the PRC's inequality trends are summarized in Kanbur et al. (2021). A comprehensive overview of the trends in the Gini index in the PRC is presented in Xie and Zhou (2014). A review of existing studies on inequality in the PRC employing top-income corrections is presented in Li, Li, and Wan (2020), whereas spatial inequality studies are described in Wang, Wan, and Yang (2014).

³ In this paper, the term “province” refers to a provincial-level administrative unit.

⁴ Previous studies find that a notion of fairness is positively related to social tolerance of 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.

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 (Chong and Gradstein, 2017; Savoia, Easaw, and McKay, 2010). Finally, studying the causes of 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 the PRC. Our purpose is fourfold. First, we aim to provide Gini index estimates for 29 Chinese provinces (excluding Tibet Autonomous Region and Xinjiang Uygur Autonomous Region, hereinafter referred to as Tibet and Xinjiang, respectively), 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 2018). Second, we seek to identify the main contributing factors that explain 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 will 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 the PRC over the past several decades.⁵ 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 in other regions (e.g., Xiang, Stillwell, Burns, and Heppenstall, 2020).

Many scholars also recognize the positive effects of Chinese institutional market reforms that encourage labor market liberalization. For example, Cai, Park, and Zhao (2008) emphasizes that internal migration in the PRC has increased substantially over time, helping 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, as pointed out by Fleisher and Yang (2008), despite the improvements, some serious challenges to the full integration of the labor market in the PRC remain. In particular, the *hukou*⁶ system still limits urbanization and industrialization in the PRC (Ngai, Pissarides, and Wang, 2019). Therefore, our third goal is to contribute to the discussion of the unequal rewards

⁵ See, e.g., Chan and Ngok (2011), Lei and Shen (2015), Mendoza (2018).

⁶ *Hukou* is a household registration system in the PRC introduced in the 1950s that imposes formal restrictions on changes of permanent residence.

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 schooling across provinces. However, we continue to observe differences in returns to schooling across provinces, which suggests that, despite the significant progress seen in labor market integration in the PRC, there still exist barriers to labor migration. Finally, we evaluate the potential effects of narrowing educational disparities and promoting labor market integration. To that end, we conduct a simulation analysis that examines how provincial and national income inequality would change in response to (i) equalization of educational structure across provinces; (ii) equalization of returns to schooling; and (iii) equalization of both educational structure and returns to schooling.

To summarize, in this study, we find there is substantial disparity in levels of provincial income inequality in the PRC. 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 schooling both seem to be the key contributing factors explaining the variation in income inequality across Chinese provinces. In turn, the factors spurring inequities in the returns to schooling include variation in the relative supply of more skilled workers, unemployment rates, and provincial industrial structure. We also find that current levels of interprovincial migration do not alter regional variation in human capital structure enough to substantially reduce the existing interprovincial differences in income inequality. Our simulation analysis reveals that, keeping all other factors constant, equalization of educational structure across provinces could lead to a 14.4% decrease in nationwide income inequality. Our findings prompt an examination of regional educational policies together with policies on labor market integration as fundamental instruments for mitigating across-provincial income disparities in the PRC.

The paper is organized as follows. In Section 2, we review the main results of previous studies on income and human capital disparities in the PRC. In Section 3, we describe the data. In Section 4, we estimate provincial Gini indices in the PRC, 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 the PRC, applying decomposition methods: (i) 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; (ii) we perform

regression-based decomposition by applying the classical Mincer model. In Section 6, we estimate provincial educational Gini indices and provincial returns to schooling. In Section 7, we study the key factors of interprovincial variation in income inequality. Namely, we regress the provincial Gini indices on provincial educational Gini indices, provincial returns to schooling, 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 schooling on different provincial labor supply and labor demand factors to identify the key factors of the existing interprovincial disparities in returns to schooling. Finally, in Section 8, we conduct a simulation analysis, which aims to study how income inequality in the PRC would change in response to (i) equalization of educational structure; (ii) labor market liberalization (which implies equalization of returns to schooling); and (iii) both equalization of educational structure and labor market liberalization. In Section 9, we conclude.

2 RELATED LITERATURE

Provincial-Level Income Inequality in the People's Republic of China

The uneven impact of economic development since the beginning of market reforms in the PRC 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).⁷ 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 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) gross domestic product (GDP) per worker fell by one-third, which they claim was mostly explained by the enhancement of internal migration.

Data limitations mean that there are few studies on provincial-level income inequality in the PRC. (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) estimated Gini indices for most Chinese provinces in 1986 separately for rural and urban populations. They found that intraprovincial income

⁷ Provincial GRP per capita for 2012, 2014, 2016, and 2018 is presented in Table E1, Appendix E. The Appendix is available at <http://dx.doi.org/10.22617/WPS250426-2>.

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 found that, unlike in most other developing countries, income inequality in the PRC's urban areas appeared to be lower than in rural areas. Analyzing provincial income inequality in 1995 and 2002, Khan and Riskin (2005) found 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) studied the potential factors explaining provincial income inequality dynamics between 1985 and 1995. They found urban inequality increased 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. However, this fragmented evidence from 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 the PRC. Our study aims to fill this gap by providing robust Gini index estimates for 29 provincial-level administrative units (excluding Tibet and Xinjiang) for the years 2012, 2014, 2016, and 2018, and, furthermore, by identifying the key factors that shape inter and intraprovincial income inequality in the PRC.

Educational Inequality in the People's Republic of China

In this study, we find educational inequality and returns to schooling to be the main explanatory factors for the current variation in provincial income inequality levels in the PRC. 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) find a large, positive, and statistically significant, relationship between inequality of schooling and income inequality, especially in emerging and developing countries.⁸ Educational disparities in the PRC 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

⁸ The paper also provides a theoretical framework and a comprehensive literature review on the topic.

million in 2016. However, recent studies continue to provide new evidence on educational inequalities.⁹ 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 the PRC 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 owes primarily 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 regional economic development levels. Yu et al. (2015) show that the relationship between inequality in educational attainment and growth in the PRC 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 greater demand for education because they have more children to educate. In summary, there exists an extensive literature linking subnational educational inequality and economic disparities, as well as studies linking educational and income inequality in the PRC (e.g., Chan and Ngok, 2011; Gustafsson and Wan, 2020; Sicular, Ximing, Gustafsson, and Shi, 2007). 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 the PRC, poor provinces exhibit higher levels of both income and educational inequality as compared to rich provinces.

Returns to Schooling in the People's Republic China

Finally, our study provides an analysis of provincial differences in returns to schooling in the PRC. We are not the first to find that returns to schooling per se, as well as in combination with educational inequality, are significant explanatory factors for provincial income inequality variation in the PRC. However, we also seek to explain provincial

⁹ For example, Heckman and Yi (2012) emphasize substantial inequality in access to higher education in the PRC.

differences with respect to returns to schooling itself. To evaluate provincial returns to schooling, we start with the classical Mincer human capital regression model, which has been used in many studies exploring differences in returns to schooling in the PRC.¹⁰ 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 the PRC. They also document the dynamics of greater declines in returns to schooling in urban locations, in 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 the PRC's educational expansion policy predominately in non-coastal regions. The authors suggest that encouraging interregional mobility in the PRC could help better match local labor supply and demand. Overall, most existing studies, including ours, attribute differences in returns to schooling across the PRC 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 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 the PRC. 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 provincial-level administrative units (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

¹⁰Zhang, Zhao, Park, and Song (2005) provide a comprehensive review of existing estimates of returns to schooling in the PRC. Churchill and Mishra (2018) also present a broad literature review on the topic.

B1, Appendix B.¹¹ All four aforementioned survey waves are nationally and provincially representative.

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 past 12 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 nonfinancial 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 includes 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 are 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 estimates are produced employing 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 B2 and B3 in the Appendix B we report summary statistics for all household- and provincial-level variables used in the current study.

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

¹¹For 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).

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

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

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):

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

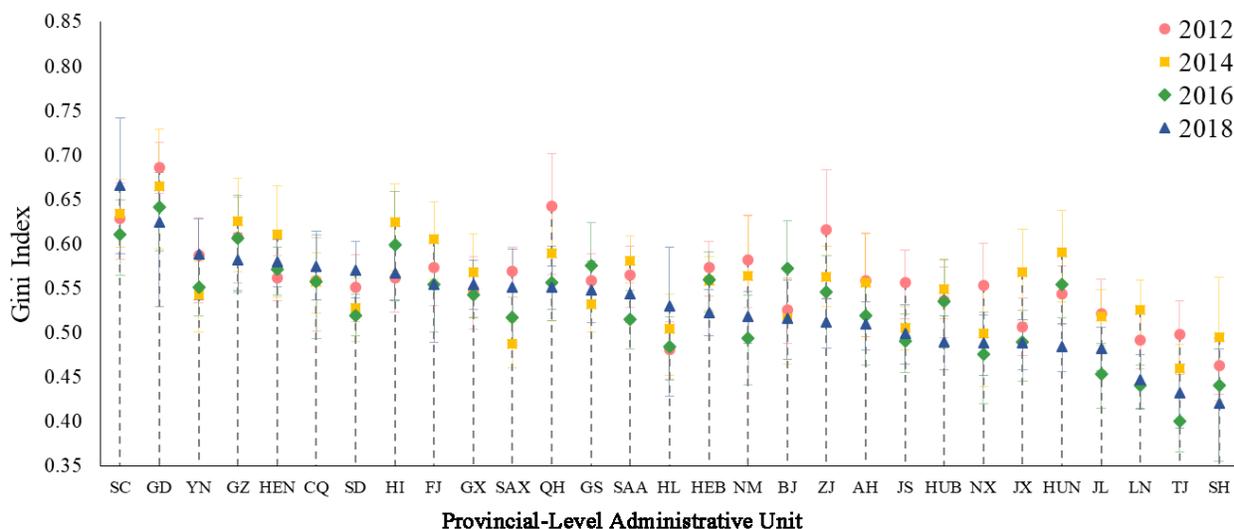
where τ is a threshold income, and p_τ is a corresponding threshold income percentile. The bottom $p_\tau\%$ of incomes are represented by the nonparametric kernel density, where N is the number of observations, h is a smoothing parameter (bandwidth), and $K(\cdot)$ is a kernel function. The upper $(1 - p_\tau)\%$ 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 1.

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 the 90th 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; Huang, Lin, and Ren, 2012; Martins and Stedinger, 2000,0; Pauli and Coles, 2001). 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, OBRE¹⁴), a MPLE exhibits high efficiency and robustness in the presence of sparse high-income observations as well as potential top-income data

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

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).

Figure 1: Provincial-Level Gini Index Estimates for the People’s Republic of China

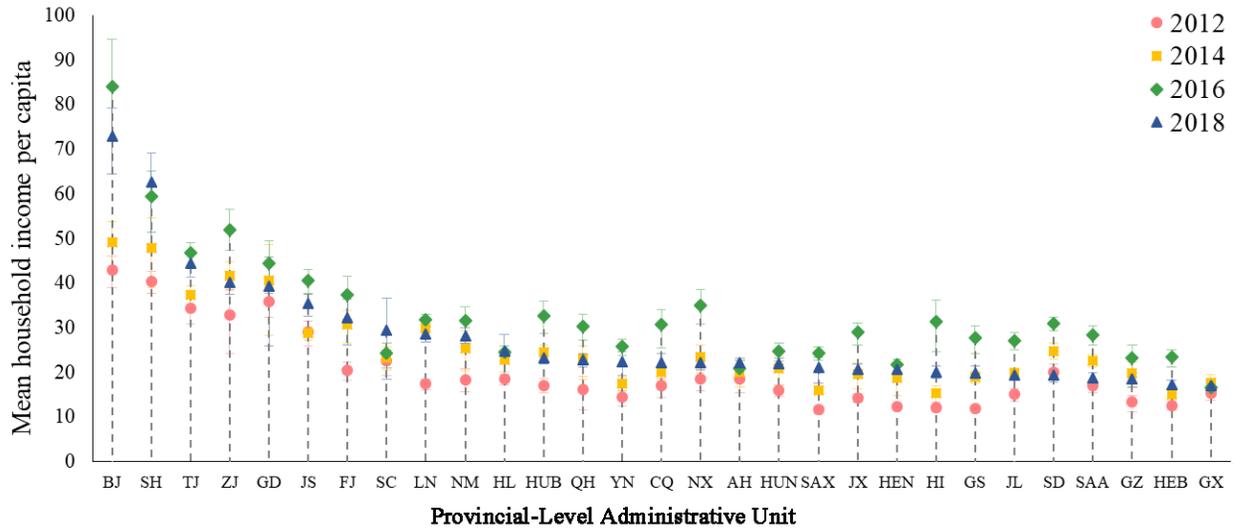


Notes: This figure shows our baseline provincial income Gini index estimates. The provinces on the x-axis are arranged in descending order of the Gini index in 2018. Error bars represent 95% basic bootstrap confidence intervals (Davison and Hinkley, 1997, p. 194) computed from 1,000 bootstrap samples. Source: Authors’ calculations using data from China Household Finance Survey.

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 C1 to C4, Appendix C. It can be seen from Figure 1 that there is a substantial variation in levels of income inequality among Chinese provinces. In 2018, the provincial Gini in the PRC 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 substantial declines

¹⁵The corresponding point estimates are reported in Table C1, Appendix C. Provincial abbreviations presented in Table C1 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).

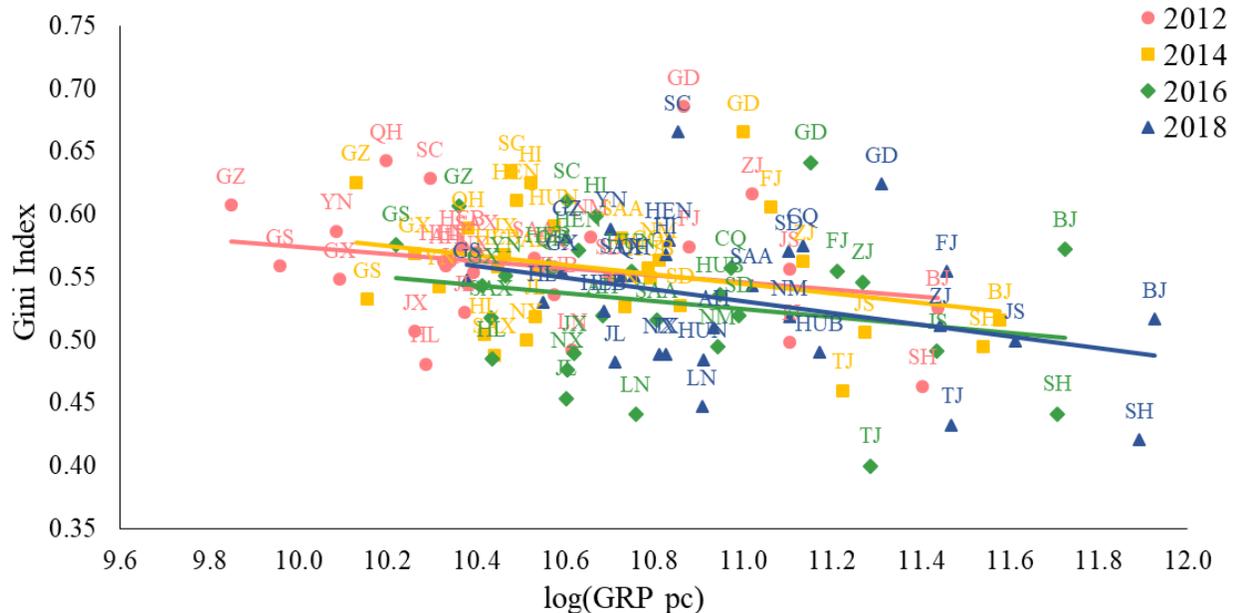
Figure 2: Provincial-Level Mean Household Income per Capita (CNY thousand)



Notes: This figure shows our provincial mean household disposable income per capita estimates. The provinces on the x-axis are arranged in descending order of the indicator in 2018. Error bars represent 95% basic bootstrap confidence intervals (Davison and Hinkley, 1997, p. 194) computed from 1,000 bootstrap samples.

Source: Authors' calculations using data from China Household Finance Survey.

Figure 3: Comparison of Provincial-Level Gini Index Estimates and Log Gross Regional Product per Capita



Notes: This figure shows our baseline provincial Gini index estimates plotted versus the log gross regional product per capita from the China Statistical Yearbook. Correlations are: -24% in 2012, -29% in 2014, -22% in 2016, and -34% in 2018.

Source: Authors' calculations using data from China Household Finance Survey.

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 ranged from around CNY17,000 in Guangxi Zhuang Autonomous Region (hereinafter referred to as Guangxi) and Hebei to CNY63,000 in Shanghai and CNY73,000 in Beijing (Figure 2). The ratio of rich to poor GRP per capita (Table E1, Appendix E) among Chinese provinces in 2018 (4.7 times) was 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 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 of the obtained income inequality estimates. It can be seen from Figure D1, Appendix D, that the resulting provincial Ginis 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 the National Bureau of Statistics of China (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 based on household surveys conducted by the NBS. The relevant correlations are 91% in 2012, 92% in 2014 and 2016, and 95% in 2018.

Figures I4–I32 in Appendix I present robustness checks with respect to the choice of threshold, bandwidth, and the likelihood penalty function. Together with the baseline threshold of the 90th percentile, we consider lower (85th percentile) and higher (95th 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 the PRC, we perform a subgroup decomposition of the Theil T inequality index by province. The Theil T index formula is:

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

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.

We choose the Theil T index because, in contrast to many other inequality measures, it can be effectively decomposed into within- and between-population subgroup components

without an overlap component (which is difficult to interpret). We decompose the Theil T index into between- and within-province shares applying the following formula:¹⁶

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

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 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 the PRC by province. We can see that within-provincial inequality explains 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 explained by the income inequality in Guangdong province (given the large share of the national population in Guangdong). The next most important provinces in explaining 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-age¹⁷ household members. We can see that the dominant part, 76%–80%, of nationwide income inequality in the PRC owes to variation within educational groups, and the remaining 20%–24% is explained by between-group variation.

¹⁶This 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).

¹⁷Here and below, we use 16–60 years old as the definition of working-age individuals.

Table 1: Decomposition of Nationwide Income Inequality by Provincial-Level

	2012		2014		2016		2018	
	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 ^a	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 ^b	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 ^c	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%

^a Refers to Inner Mongolia Autonomous Region.

^b Refers to Ningxia Hui Autonomous Region.

^c Refers to Guangxi Zhuang Autonomous Region.

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

Source: Authors' calculations using data from China Household Finance Survey.

Table 2: Decomposition of Nationwide Income Inequality by Educational Level of Working-Age (16–60) Household Members

	2012		2014		2016		2018	
	Theil T index (1)	Share (%) (2)	Theil T index (3)	Share (%) (4)	Theil T index (5)	Share (%) (6)	Theil T index (7)	Share (%) (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).

Source: Authors' calculations using data from China Household Finance Survey.

5.2 Regression-Based Decomposition of Income Inequality

To analyze the possible explanatory factors shaping the levels of income inequality in the PRC, we turn to regression-based decomposition. We apply the following variant of the Mincer model (Mincer, 1974) to each year of data:

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

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), E_i^2 is average experience squared, and ε_i is an error term that 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 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 $\ln(\text{land per capita})$, which is the log of the amount of the household's land (in mu^{18}) per capita. To control for capital endowments supporting other potential business income sources, we add $\ln(\text{business assets per capita})$, the log of total household asset value for all business projects (in CNY100,000) divided by household size. We also control for household-level covariates X_{ik} , which may also affect household income levels in the PRC (namely, sex 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):

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

where s_z is a resulting % of income inequality that can be attributed to the z th regressor in eq. (4), $\hat{\alpha}_z$ is the estimated regression coefficient on each z^{th} independent variable, $\text{cov}(V_z, \ln Y)$ 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 Ordinary Least Squares (OLS) estimates for eq. (4) controlling for provincial dummies are reported in odd columns from (1) to (7) and those controlling for city dummies 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 1-year increase in 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 explains the largest share of the variation in household income per capita (9%–17%). The positive sign of \hat{s}_z estimates for schooling suggests greater variation in years of schooling per worker among households increases 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

¹⁸One mu is equivalent to 0.06667 hectare.

are not statistically significant. The inflection point for experience per worker was around 18 years in 2014 and between 30 and 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 explains 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 household income per capita. Variation in household land explains 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 explains 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 sex of household head explain 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 explains 3%–8% of income inequality in the PRC. Provincial and city dummies explain 4%–6% and 8%–14% of inequality in the PRC, respectively.

To sum up, Table 3 shows that the most important factor explaining the variation in household income is the educational attainment of working-age household members, which explains 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 explaining household income in the PRC 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 explaining variation in household income in the PRC. It could be that the relative importance of human capital versus physical capital has increased over time in the PRC.

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

Table 3: Mincer Regression—Explaining Household Income per Capita Variation

	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*** (0.004)	13.6%	0.096*** (0.004)	10.1%	0.138*** (0.004)	17.2%	0.122*** (0.004)	14.7%	0.112*** (0.004)	12.1%	0.088*** (0.004)	9.3%	0.128*** (0.004)	15.9%	0.110*** (0.004)	13.3%
Experience per worker	0.002 (0.004)	-0.6%	0.012*** (0.004)	-2.0%	0.028*** (0.004)	-3.9%	0.022*** (0.004)	-3.8%	0.003 (0.004)	-0.7%	0.012*** (0.004)	-2.0%	0.030*** (0.004)	-4.1%	0.022*** (0.004)	-3.8%
Experience per worker ² : 100	0.006 (0.008)	-0.1%	-0.034*** (0.009)	2.4%	-0.041*** (0.007)	3.1%	-0.037*** (0.008)	3.5%	0.001 (0.008)	0.1%	-0.035*** (0.009)	2.4%	-0.043*** (0.007)	3.3%	-0.037*** (0.008)	3.5%
Ln(household size)	-0.384*** (0.025)	2.8%	-0.499*** (0.030)	4.1%	-0.271*** (0.027)	1.4%	-0.323*** (0.027)	1.4%	-0.359*** (0.024)	2.6%	-0.445*** (0.030)	3.6%	-0.228*** (0.027)	1.2%	-0.294*** (0.027)	1.3%
Land dummy	-0.084*** (0.022)	0.7%	-0.280*** (0.036)	3.6%	-0.192*** (0.032)	2.0%	-0.045* (0.025)	0.4%	-0.063*** (0.023)	0.6%	-0.270*** (0.036)	3.5%	-0.135*** (0.032)	1.4%	-0.029 (0.026)	0.2%
Ln(land per capita)	0.168** (0.079)	-0.0%	0.200*** (0.029)	-1.9%	0.125*** (0.026)	-1.2%	0.264*** (0.067)	-0.1%	0.109 (0.082)	-0.0%	0.193*** (0.030)	-1.9%	0.091*** (0.027)	-0.9%	0.203*** (0.066)	-0.1%
Ln(business assets per capita)	0.158*** (0.007)	2.3%	0.164*** (0.008)	2.5%	0.100*** (0.008)	1.2%	0.051*** (0.010)	0.4%	0.150*** (0.007)	2.2%	0.160*** (0.008)	2.4%	0.095*** (0.008)	1.1%	0.050*** (0.009)	0.4%
Party member	0.232*** (0.025)	1.0%	0.190*** (0.025)	0.8%	0.162*** (0.024)	0.6%	0.179*** (0.024)	0.8%	0.246*** (0.024)	1.1%	0.210*** (0.025)	0.9%	0.168*** (0.024)	0.6%	0.190*** (0.024)	0.8%
Female	-0.045** (0.021)	-0.1%	-0.073*** (0.022)	-0.2%	-0.126*** (0.022)	-0.2%	0.011 (0.023)	0.0%	-0.051** (0.021)	-0.1%	-0.077*** (0.022)	-0.2%	-0.134*** (0.022)	-0.2%	0.002 (0.022)	0.0%
Rural	-0.269*** (0.023)	3.5%	-0.527*** (0.031)	8.2%	-0.304*** (0.024)	3.9%	-0.336*** (0.024)	4.8%	-0.197*** (0.024)	2.5%	-0.485*** (0.033)	7.7%	-0.239*** (0.025)	3.1%	-0.230*** (0.025)	3.3%
Constant	8.468*** (0.090)		9.102*** (0.094)		8.407*** (0.090)		8.740*** (0.093)		8.150*** (0.139)		8.985*** (0.136)		8.355*** (0.140)		8.920*** (0.137)	
Provincial-level dummies	Yes	5.6%	Yes	3.7%	Yes	6.0%	Yes	6.1%	No		No		No		No	
City dummies	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	

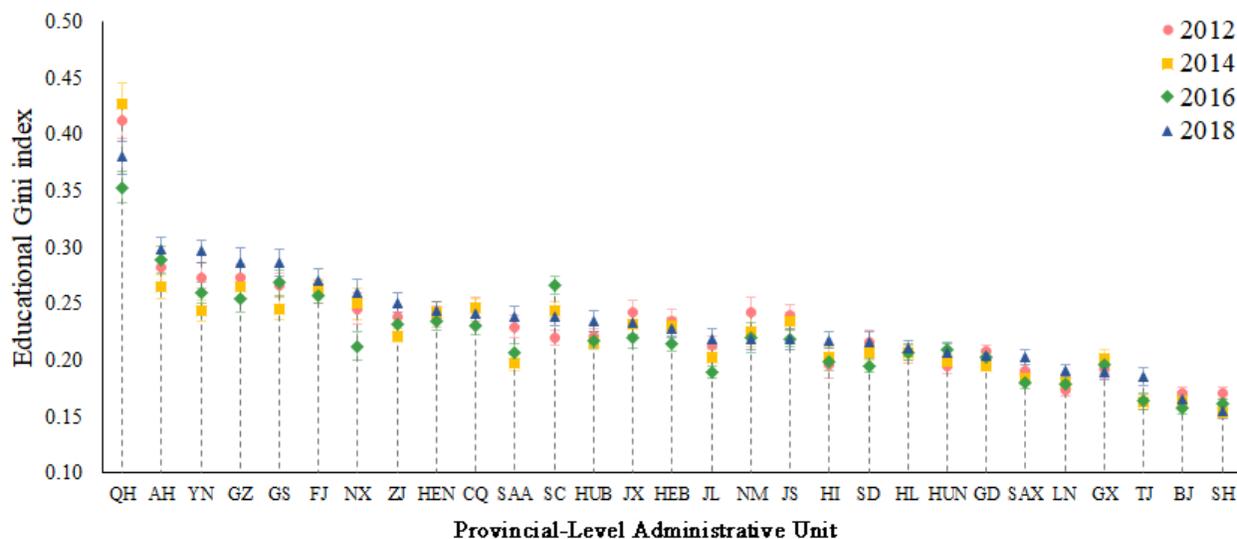
Notes: This table reports the results of the Mincer-type regression formulated in eq. (4), controlling for provincial dummies in columns (1) to (7) and for city dummies in columns (9) to (15). We employ heteroskedasticity-robust standard errors. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: Authors' calculations using data from China Household Finance Survey.

6 EDUCATIONAL INEQUALITY AND RETURNS TO SCHOOLING

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: 0 years for no schooling; 6 years for primary school; 9 years for junior school; 12 years for regular or technical high school; 15 years for vocational college; 16 years for regular college and bachelor's; 19 years for a master's degree; and 22 years for a doctoral degree.

Figure 4: Provincial-Level Educational Gini Index in the People's Republic of China



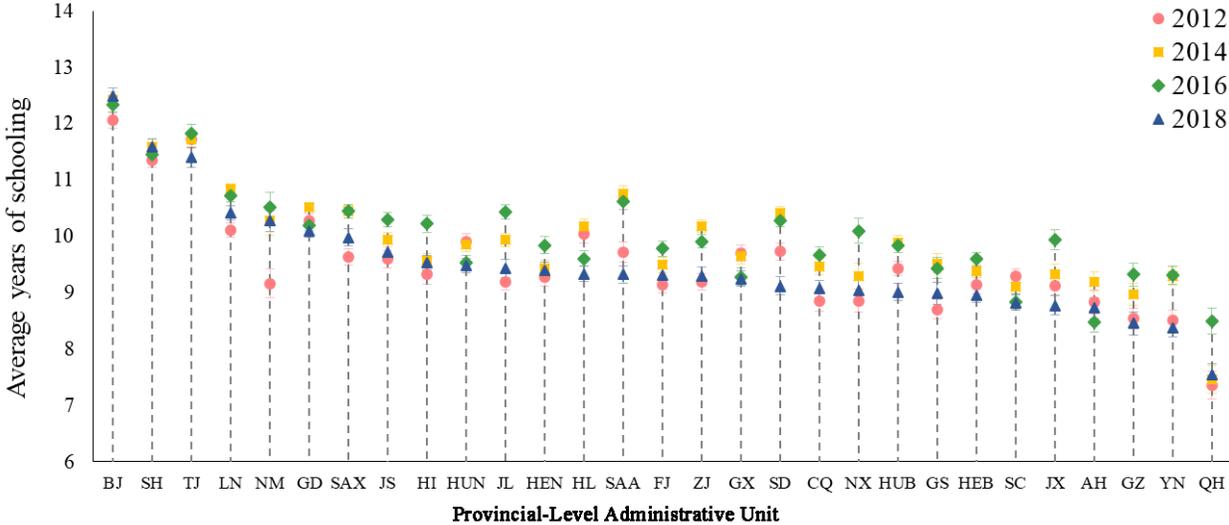
Notes: This figure shows the provincial educational Gini index estimates. The provinces on the x-axis are arranged in descending order of the indicator in 2018. Error bars represent 95% basic bootstrap confidence intervals (Davison and Hinkley, 1997, p. 194) computed from 1,000 bootstrap samples.

Source: Authors' calculations using data from China Household Finance Survey.

Figure 4 reports our estimates of the provincial-level educational Gini indices using the CHFS data.²⁰ The educational Gini index ranges 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. 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 level of economic development. Richer

²⁰The corresponding point estimates of all educational variables presented in the current section are reported in Table C2, Appendix C.

Figure 5: Provincial-Level Average Years of Schooling in the People’s Republic of China

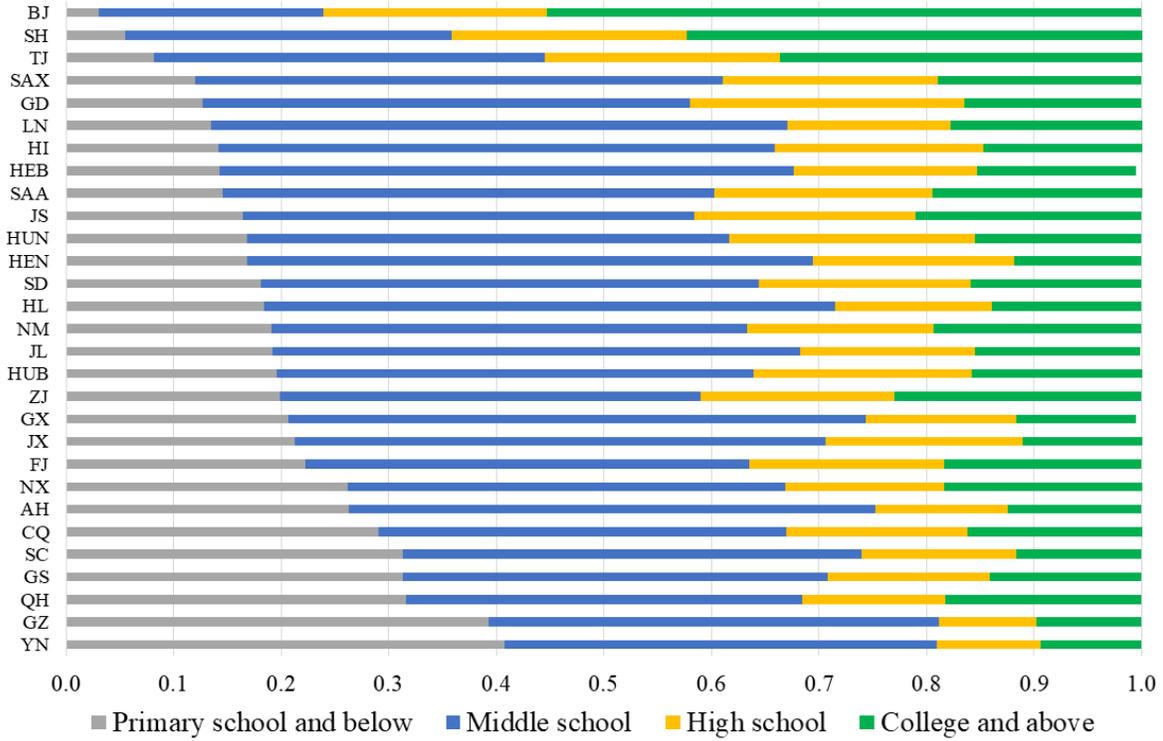


Notes: This figure shows the provincial average years of schooling estimates. The provinces on the x-axis are arranged in descending order of the indicator in 2018. Error bars represent 95% basic bootstrap confidence intervals (Davison and Hinkley, 1997, p. 194) computed from 1,000 bootstrap samples. Source: Authors’ calculations using data from China Household Finance Survey.

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 Gini 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 data from the China Population and Employment Statistics Yearbook (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 is high-skilled compared to in other provinces. It can also be noticed that, except in Beijing, Shanghai, and Tianjin, high

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

Figure 6: Average Provincial-Level Educational Structure, 2012–2018



Notes: This figure shows the average educational structure of the employed population in Chinese provinces. The estimates are ordered in the “Primary school and below” category.
 Source: Authors’ calculations using data from China Population and Employment Statistics Yearbook.

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 the PRC are driven mostly by the variation in the share of less skilled labor.

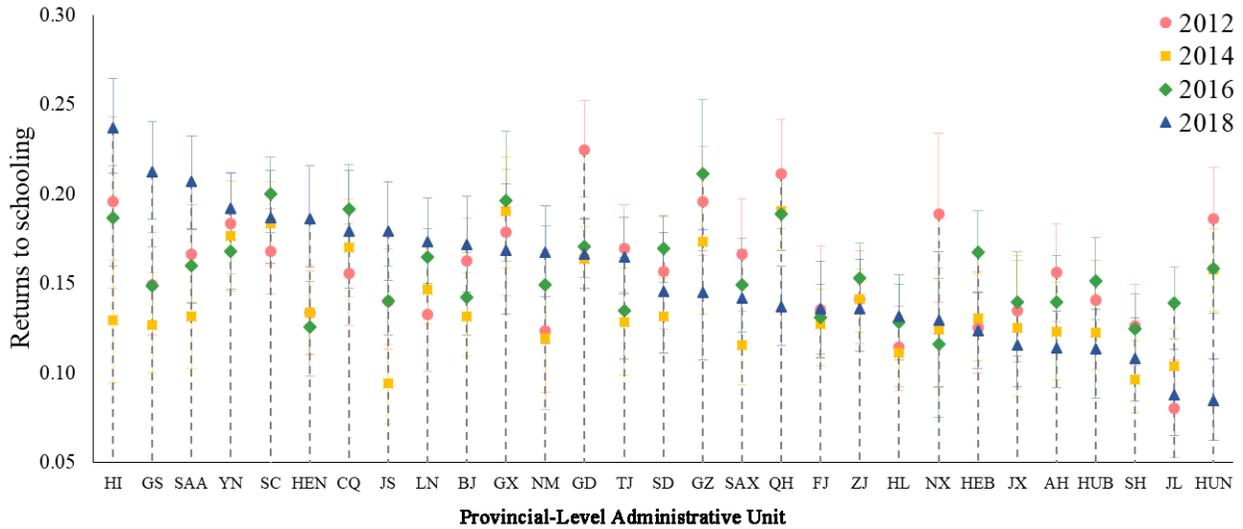
Next, we evaluate differences in returns to human capital across provinces. To do so, we estimate provincial returns to schooling and returns to high school and above for each province and year. For returns to schooling, we apply the following provincial-level version of eq. (4)²² to each year of the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

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

²²To evaluate provincial returns to schooling and returns to high school, we focus on the importance of human capital in explaining the variation in 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).

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.²³ In case of $Z_{ijt} = H_{ijt}$, the estimate of $\hat{\delta}_{1jt}$ is called “returns to high school.”

Figure 7: Provincial-Level Returns to Schooling in the People’s Republic of China

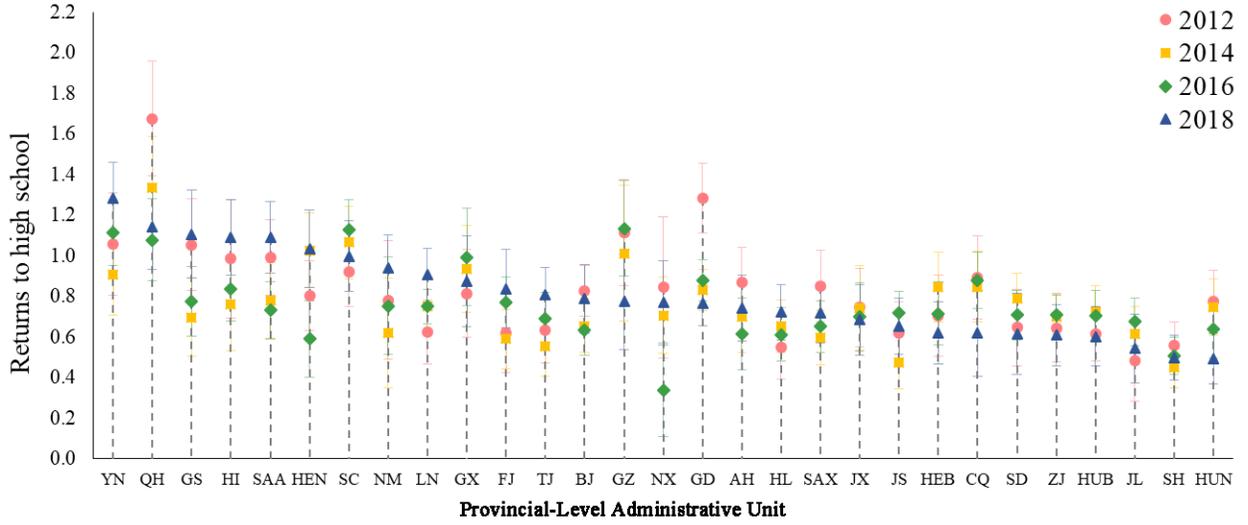


Notes: This figure shows the provincial returns to schooling estimates. The provinces on the x-axis are arranged in descending order of the indicator in 2018. Error bars represent 95% basic bootstrap confidence intervals (Davison and Hinkley, 1997, p. 194) computed from 1,000 bootstrap samples. Source: Authors’ calculations using data from China Household Finance Survey.

Figures 7 and 8 report estimated provincial returns to schooling and returns to high school (respectively). On average during 2012–2018, 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). The returns to high school vary from 50.4% in Shanghai to 130.9% in Qinghai. 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

²³Our definition of returns to schooling is close to the “ex-post returns to schooling” from Heckman, Lochner, and Todd (2006).

Figure 8: Provincial-Level Returns to High School in the People’s Republic of China



Notes: This figure shows the provincial returns to high school estimates. The provinces on the x-axis are arranged in descending order of the indicator in 2018. Error bars represent 95% basic bootstrap confidence intervals (Davison and Hinkley, 1997, p. 194) computed from 1,000 bootstrap samples.

Source: Authors’ calculations using data from China Household Finance Survey.

consistent with richer provinces having a greater relative supply of skilled labor than poor provinces.

In Figures C5-C8 (Appendix C), 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-LEVEL VARIATION IN INCOME INEQUALITY

7.1 Provincial-Level Income Inequality Determinants

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:

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

where I_{jt} is the income inequality Gini index in province j at year t , I_{jt}^E is the educational Gini index, $\hat{\delta}_{1jt}$ is estimated returns to schooling, 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):

$$\ln Y_i = \delta_0 + \delta_1 S_i + u_i, \quad (8)$$

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:

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

where the variance in household income per capita $\text{Var}(\ln Y)$ measures income inequality, and $\text{Var}(S)$ captures educational inequality. If we believe 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 schooling, and especially their interaction, should all increase provincial income inequality among households within the same province. Therefore, if we estimate eq. (7) with provincial fixed effects, we expect to obtain positive coefficients for the educational Gini, returns to schooling, and their interaction term.²⁴

To capture other provincial-specific factors that could be important for explaining variation in provincial income inequality, we further control eq. (7) for four T_{jtm} provincial-level covariates collected from the CSY. First, we include the *Social spending* variable, which measures local governments' expenditures on social insurance per working-age adult (in thousand CNY).²⁵ 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

²⁴For more details refer to Coady and Dizioli (2018).

²⁵We 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.

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 the PRC greater trade openness is likely associated with increased demand and prices for unskilled labor, negatively impacting income inequality. Finally, we include the $\ln(\text{GRP per capita})$ 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 $\ln(\text{GRP per capita})$ coefficient estimate.

Table 4 reports our baseline estimates of eq. (7). We normalize our key educational Gini I_{jt}^E and returns to schooling $\hat{\delta}_{1jt}$ variables by subtracting the mean in each case to make the main effects in the specification with interaction terms in columns (5), (8), (12), and (16) more interpretable. The OLS results (columns 1–8) suggest a strong positive relationship between educational and income inequality in Chinese provinces. A 10% increase in the educational Gini is associated with a 3.5–4.6 percentage point increase in the income Gini index. The coefficient is significant at the 1% level in all OLS specifications. The mean provincial income Gini is 0.530–0.568 in different years, and the standard deviation is around 0.05 (Table C1, Appendix C). Similarly, a significant and positive relationship is found between provincial income Gini indices and returns to schooling. A 10% increase in returns to schooling is associated with a 7.0–9.2 percentage point increase in the income Gini. A negative sign on the interaction term in columns (4) and (8) suggests educational inequality has a lower effect on income inequality in places with higher returns to schooling, or returns to schooling have a lower effect on income inequality in areas with higher educational inequality. Since we centered I_{jt}^E and $\hat{\delta}_{1jt}$, the main effects in columns (4) and (8) indicate that a 10% increase in educational Gini is associated with 3.7–4.3 percentage point increase in the income Gini for a province with average returns to schooling. At the same time, a 10% increase in returns to schooling is associated with a 7.2–8.2 percentage point increase in the income Gini for a province with an average educational Gini.

Table 4: Regression Estimates of Gini Index Determinants Using Educational Gini Index Calculated from the China Household Finance Survey

	Dependent variable: Gini index															
	(OLS) (1)	(OLS) (2)	(OLS) (3)	(OLS) (4)	(OLS) (5)	(OLS) (6)	(OLS) (7)	(OLS) (8)	(FE) (9)	(FE) (10)	(FE) (11)	(FE) (12)	(FE) (13)	(FE) (14)	(FE) (15)	(FE) (16)
Educational Gini	0.458*** (0.097)		0.349*** (0.095)	0.427*** (0.095)	0.454*** (0.089)		0.364*** (0.089)	0.368*** (0.092)	0.392** (0.187)		0.326 (0.207)	0.268 (0.221)	0.398** (0.195)		0.353* (0.214)	0.292 (0.224)
Returns to schooling		0.916*** (0.125)	0.696*** (0.132)	0.820*** (0.126)		0.805*** (0.123)	0.719*** (0.115)	0.720*** (0.115)		0.277** (0.111)	0.255** (0.114)	0.237* (0.128)		0.291** (0.113)	0.275** (0.117)	0.259** (0.128)
Educational Gini × Returns to schooling				-6.169*** (2.244)				-0.232 (2.353)				1.771 (1.644)				2.153 (1.627)
Social spending CSY					-0.012** (0.006)	-0.011* (0.006)	-0.015*** (0.005)	-0.015*** (0.006)					0.003 (0.011)	0.002 (0.011)	0.004 (0.012)	0.006 (0.013)
SOE assets share CSY					-0.107*** (0.038)	-0.098*** (0.035)	-0.070* (0.037)	-0.070* (0.037)					-0.071 (0.087)	-0.044 (0.089)	-0.016 (0.092)	0.005 (0.088)
Foreign trade share CSY					0.058* (0.032)	-0.004 (0.027)	0.023 (0.027)	0.023 (0.028)					-0.032 (0.052)	-0.036 (0.048)	-0.035 (0.049)	-0.028 (0.050)
Ln(GRP per capita) CSY					-0.050** (0.021)	-0.019 (0.021)	-0.016 (0.021)	-0.016 (0.021)					0.050 (0.058)	0.061 (0.057)	0.069 (0.057)	0.080 (0.059)
2014.year	-0.001 (0.012)	0.014 (0.011)		0.016 (0.010)	0.012 (0.011)	0.018 (0.011)	0.020* (0.010)	0.020* (0.010)	-0.001 (0.007)	0.002 (0.007)	0.003 (0.007)	0.002 (0.007)	-0.011 (0.013)	-0.009 (0.012)	-0.010 (0.013)	-0.013 (0.014)
2016.year	-0.025** (0.012)	-0.029** (0.011)		-0.025** (0.010)	0.006 (0.014)	-0.014 (0.015)	-0.005 (0.013)	-0.005 (0.013)	-0.025*** (0.007)	-0.029*** (0.007)	-0.026*** (0.007)	-0.026*** (0.007)	-0.046** (0.023)	-0.052** (0.022)	-0.053** (0.023)	-0.058** (0.025)
2018.year	-0.030** (0.012)	-0.025** (0.012)		-0.029*** (0.011)	0.013 (0.019)	-0.004 (0.019)	0.003 (0.018)	0.003 (0.018)	-0.030*** (0.007)	-0.027*** (0.006)	-0.029*** (0.006)	-0.028*** (0.007)	-0.063* (0.036)	-0.064* (0.035)	-0.072* (0.037)	-0.078** (0.039)
Constant	0.558*** (0.008)	0.554*** (0.007)	0.544*** (0.004)	0.556*** (0.006)	1.106*** (0.220)	0.799*** (0.215)	0.755*** (0.214)	0.755*** (0.215)								
Observations	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116	116
R ²	0.221	0.333	0.312	0.444	0.414	0.485	0.553	0.553	0.291	0.312	0.327	0.332	0.310	0.332	0.350	0.357
Number of provincial-level administrative units									29	29	29	29	29	29	29	29

Notes: This table presents the regression results that explain provincial income inequality based on provincial human capital factors, as outlined in eq. (7). Columns (1) to (8) display results from the Ordinary Least Squares, and columns (9) to (16) the results from the fixed effects model. We employ heteroskedasticity-robust standard errors. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: Authors' calculations using data from China Household Finance Survey.

The fixed effects results (columns 9–16) indicate positive effects of educational inequality, returns to schooling, and their interaction with income inequality within each province, as expected from eq. (9). Educational inequality has a larger effect on income inequality in the fixed effects specification: a 10% increase in the educational Gini within a province is associated with a 2.7–3.9 percentage point increase in the income Gini index. Returns to schooling has a lower impact on income inequality in the fixed effects model: a 10% increase in returns to schooling in a province is associated with a 2.4–2.9 percentage point increase in income Gini. The main effects are significant at the 1%–5% level, and the coefficient on the interaction term is not statistically significant.

For other covariates, in the OLS specification we find a significant negative impact of social public spending (*Social spending*) and government share of the provincial economy (*SOE assets share*) on income inequality (as we suggested above). Specifically, a CNY1,000 increase in local governments' expenditures on social insurance per working-age adult is associated with a 1.1–1.5 percentage point decrease in the income Gini coefficient. A 10% increase in the share of the total assets of state-holding industrial enterprises in provincial GRP is associated with a 7–11 percentage point decrease in the income Gini. Economic openness proxied by the share of imports and exports in GRP (*Foreign trade*) was found to be positively related to income inequality in the OLS specification and negatively in the fixed effects one. However, the coefficients for this variable are not statistically significant in most specifications. Similarly, the impact of the $\ln(\text{GRP per capita})$ is also insignificant across most specifications.

To examine the robustness of these results, we also estimate eq. (7) using educational Gini indices calculated from the CPESY. The resulting estimates are reported in Table I1, Appendix I. Table I1 suggests the positive impact of educational inequality on income inequality remains significant and of a similar magnitude, as in our baseline results: a 10% increase in the educational Gini is associated with about 3.2–5.2 percentage point increase in the income Gini index. However, in the fixed effects specification, we find effects that are one-third of the size (compared to our baseline results) and insignificant. The main effects of returns to schooling on income inequality do not change in magnitude (2.4–2.9 percentage points) and remain significant in all specifications. In addition, we found a positive and significant effect of the interaction between the educational Gini index estimated from the CPESY and returns to schooling on income inequality in the fixed effects specification. Table I1 confirms statistically significant effect of the state's presence in the local economy and government social spending to reduce provincial income inequality in the OLS specification.

We also trace the possible impact 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.²⁶ We then re-estimate our baseline income inequality regression from eq. (7) using the adjusted educational Gini measures. Figure I1 in Appendix I 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 Impact of Educational Policies

Our key result on the negative impact of provincial educational inequality on income inequality suggests targeted educational policies are promising instruments for reducing income inequality. What are the main characteristics of educational inequality in the PRC? In Table H1 (Appendix H), 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 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 H1, we can see that the within-province component explains the dominant share (98%) of variation in years of schooling in the PRC. 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 subcomponents. 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,

²⁶The percentage of those whose *hukou* location differs from their residence is around 3%–6% of the CHFS samples.

we calculate average government expenditures²⁷ 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 dataset with an observation for each year, province, and birth cohort (4 years × 29 provincial-level administrative units × 50 birth cohorts = 5,800 observations) and run OLS to identify the impact of educational expenditures on educational inequality.

Table 5: OLS Estimates of the Impact of Government Educational Expenditures on Educational Inequality

	<i>Dependent variable:</i>			
	Educational Gini by provincial-level/birth cohort			
	(1)	(2)	(3)	(4)
Average educational expenditures		-0.890*** (0.025)		-0.396*** (0.027)
Constant	0.248*** (0.008)	0.260*** (0.008)	0.304*** (0.016)	0.306*** (0.018)
Provincial-level 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

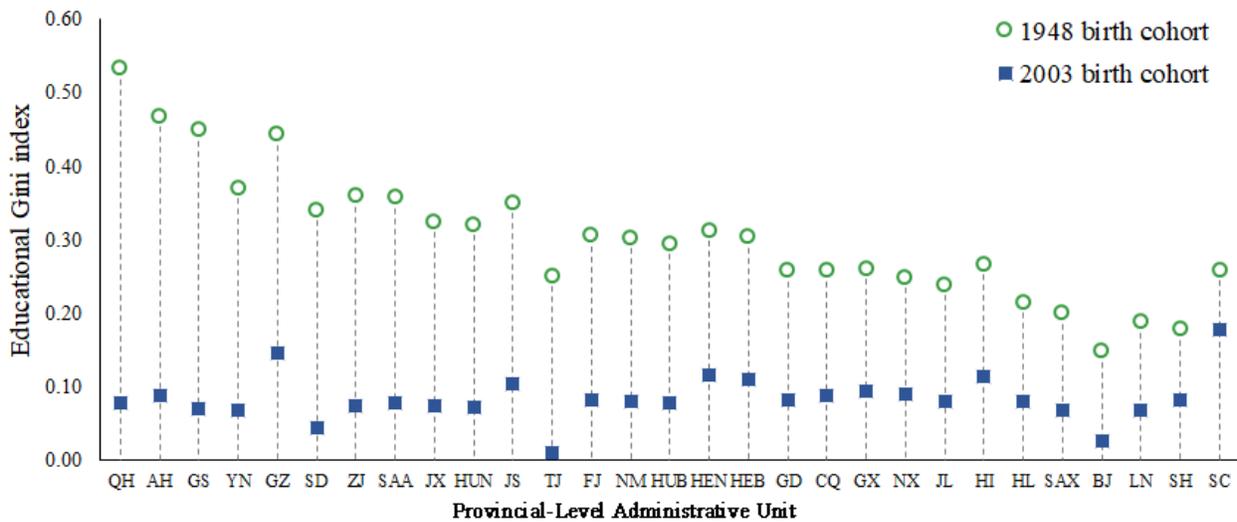
Notes: This table presents the regression results that explain provincial educational inequality based on government educational expenditures. We use heteroskedasticity-robust standard errors. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Sources: Authors' calculations using data from China Household Finance Survey, 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 thousand CNY is associated with an 8.9 percentage point 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 squared increases from 0.336 to 0.539). Overall, the results from column (2) suggest differences in educational spending within provinces across cohorts explains 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 impact of educational spending controlling for the birth-cohort dummies is also significant, but half the magnitude: an increase in average annual educational expenditures per capita in thousand CNY is associated with a 4 percentage point reduction in the educational Gini.

²⁷Government expenditures are measured in thousand CNY 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 shows 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.

Source: Authors' calculations using data from China Household Finance Survey.

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 explaining educational inequality compared to the variation within provinces across cohorts.

These results are consistent with the decomposition results, reported in Table H1 (Appendix H): within-province variation share is 98%, and within-birth cohort share in each province is around $0.8 \times 98\% = 78.4\%$. To summarize, the variation in government educational expenditures within provinces over time (across birth cohorts) explains the largest part of the variation in educational inequality in the PRC. In particular, lower government spending on education for older cohorts compared to the higher spending for younger cohorts is the key factor shaping educational inequality. Since we observe that increased educational government spending over time helped reduce educational inequality, we should expect that educational inequality will proceed 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 we can see 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.

7.2 Provincial-Level Returns to Schooling Determinants

Next, we study the key factors contributing to interprovincial variation in returns to schooling and returns to high school. We estimate the following model involving relative labor supply and labor demand factors:²⁸

$$R_{jt} = \beta_0 + \beta_1 L_{jt} + \sum_{p=2}^4 \beta_p D_{pjt} + \nu_t + \varepsilon_{jt}, \quad (10)$$

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. We include *Foreign trade share* again to capture the openness of the local economy. As we already mentioned, in the PRC 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 have a negative impact on returns to schooling (returns to high school). Finally, we control for *Secondary employment*—the share of employment in the secondary industry.²⁹

Table 6 reports the OLS estimates of eq. (10). Columns (1) and (4) show the results from our baseline specification, where *Skilled labor share* is estimated from the CHFS. To check robustness, in columns (2) and (4) we use skilled labor share estimated from the CPESY, and in columns (3) and (6) we use skilled labor share computed from the CSY. *Skilled labor share CPESY*, therefore, measures the share of individuals with 9 or more years of schooling among the employed population, and *Skilled labor share CSY-*

²⁸This approach is close to those discussed in Yang (2005), and Zhang et al. (2005).

²⁹Secondary industry refers to mining, manufacturing, construction, production, and supply of electricity, water, and gas.

Table 6: Ordinary Least Squares Estimates of Returns to Schooling and Returns to High School Determinants

	<i>Dependent variable:</i>					
	Returns to schooling			Returns to high school		
	(1)	(2)	(3)	(4)	(5)	(6)
Skilled labor share	-0.119*** (0.045)					
Skilled labor share CPESY		-0.125*** (0.034)				
Skilled labor share CSY			-0.180*** (0.040)			
Skilled labor share HS				-1.226*** (0.411)		
Skilled labor share HS CPESY					-0.936*** (0.289)	
Skilled labor share HS CSY						-1.568*** (0.394)
Secondary employment CSY	-0.064** (0.026)	-0.058** (0.026)	-0.073*** (0.026)	-0.426*** (0.129)	-0.453*** (0.127)	-0.588*** (0.142)
SOE assets share CSY	-0.027 (0.025)	-0.033 (0.024)	-0.022 (0.025)	0.176 (0.174)	0.094 (0.133)	0.210 (0.156)
Foreign trade share CSY	0.015 (0.013)	0.016 (0.012)	0.028** (0.013)	0.116 (0.121)	0.136 (0.131)	0.213* (0.119)
2014.year	-0.015** (0.007)	-0.017** (0.007)	-0.016** (0.006)	0.011 (0.051)	-0.034 (0.054)	-0.030 (0.049)
2016.year	0.004 (0.007)	0.001 (0.007)	0.003 (0.007)	0.004 (0.054)	-0.012 (0.058)	-0.005 (0.049)
2018.year	-0.008 (0.009)	-0.004 (0.009)	-0.002 (0.008)	-0.028 (0.054)	0.034 (0.061)	0.055 (0.054)
Constant	0.275*** (0.034)	0.285*** (0.026)	0.308*** (0.027)	1.387*** (0.146)	1.264*** (0.107)	1.404*** (0.123)
Observations	116	116	116	116	116	116
R ²	0.165	0.196	0.250	0.199	0.203	0.248

Notes: This table presents the regression results that explain provincial returns to schooling and returns to high school based on provincial labor demand and labor supply factors, as formulated in eq. (10). We use heteroskedasticity-robust standard errors. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: Authors' calculations using data from China Household Finance Survey.

among those aged 6 and above in the employed population. The signs, magnitude, and significance level of the skilled labor variables from the NBS statistical yearbooks are close to our baseline result. We find 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 the PRC. A 10% increase in the provincial share of skilled labor (above middle school) is associated with a 1.2–1.8 percentage point decrease in the returns to a year of schooling. The coefficients are statistically significant at the 1%–5% level. Mean provincial returns to schooling are 0.138–0.157 in different years,

and the standard deviation is around 0.03 (Table C2, Appendix C). At the same time, a 10% increase in the skilled labor share (above high school) is associated with a 9.4–15.7 percentage point decrease in the returns to high school. The coefficients are significant at the 1% level. Mean provincial returns to high school are 0.760–0.828, and the standard deviation is 0.2 (Table C2, Appendix C). 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 percentage point decrease in returns to a year of schooling and to a 4.3–5.9 percentage point 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 with both level of economic development and returns to schooling, one source of potential omitted variable bias is educational quality.³⁰ 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.

We also repeat the robustness check that examines the potential role of migration. We re-estimate skilled shares based on *hukou* province instead of the province of residence, and then we plug them into eq. (10). Figures I2 and I3 in Appendix I plot predicted returns to schooling (returns to high school) from eq. (10) based on province of residence versus based on 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 could be impacted by (i) equalization of educational structure; (ii) labor market integration (equalization of returns to schooling); and (iii) both equalization of educational structure and labor market integration. For this analysis, we again analyze the Theil T inequality index.

³⁰Previous studies (e.g., Wei and Hao, 2011) emphasize the significant role of educational quality for human capital formation in the PRC.

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

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

where T is the national Theil T index, T_j is a Theil T index of each province j (in our case $j = 1, \dots, 29$), 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:

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

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

Incorporating (12) in (11) we obtain:

$$\begin{aligned} T &= \sum_{j=1}^{29} \frac{\bar{y}_j}{N \bar{y}} \left(\sum_{i=1}^{N_j} \frac{y_{ij}}{\bar{y}_j} \log \left(\frac{y_{ij}}{\bar{y}_j} \right) \right) + \sum_{j=1}^{29} \frac{N_j \bar{y}_j}{N \bar{y}} \log \left(\frac{\bar{y}_j}{\bar{y}} \right) = \\ &= \frac{1}{N \bar{y}} \left(\sum_{j=1}^{29} \sum_{i=1}^{N_j} y_{ij} \log \left(\frac{y_{ij}}{\bar{y}_j} \right) + \sum_{j=1}^{29} N_j \bar{y}_j \log \left(\frac{\bar{y}_j}{\bar{y}} \right) \right). \end{aligned} \quad (13)$$

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 impact of equalization of educational structure on income inequality we randomly reassign years of schooling to individuals so the educational structure in each province is the same as the national educational structure reported in the CPESY. Then we aggregate the individual-level simulated data to households. This adjusts the household schooling variable \tilde{S}_{ijt} . Using the updated \tilde{S}_{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 leads to 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 \tilde{R}_{jt} from (10). Then we can

incorporate both adjusted schooling \tilde{S}_{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 labor market integration only. This scenario implies an exogenous equalization of returns to schooling to the national-level returns to schooling, without any adjustments to the provincial educational structure. National-level returns to schooling are obtained by running the (6) OLS at the household-level data from the CHFS 2019.

Finally, we consider both educational equalization and labor market integration. For that, we repeat the same educational equalization from (1) and perform an exogenous equalizing change in the returns to schooling to the national-level returns to schooling. 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 became similar across provinces?

The results of the policy simulations are reported in Table 7. 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 the CPESY. In comparison to the initial Theil T decomposition results, equalization of the educational structure leads to a 14.4% decrease in the nationwide Theil T index from 0.462 to 0.395. The within-province component of nationwide household income per capita variation increases from 86% to 91%. The poor to rich province income inequality ratio declines narrowly, from 1.34 (0.457 in Gansu divided by 0.342 in Beijing) to 1.20 (0.389 in Gansu divided by 0.323 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 schooling, leads to an almost similar decrease in nationwide income inequality (–18.4%) and an increase in the within-province variation component (from 86% to 91%). However, unlike in the first simulation, the poor-to-rich province inequality ratio slightly increases, from 1.34 to 1.38.

The results of our third simulation study are reported in columns (7) and (8). Exogenous equalization of the provincial returns to schooling to the national-level returns leads to a 11.7% increase in inequality, and an almost doubling of the importance of the between-province component of national-level inequality, from 14% to 27%. This happens because increases in income inequality in provinces with low returns to schooling (a 39% increase of Theil T in Shanghai, a 21% increase in Jilin, a 7% increase in Hunan, etc.) outweigh the inequality decline in provinces with high returns to schooling (–25% in Theil T in Shaanxi,

Table 7: Theil T Indices Resulting from Policy Simulations

	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 ^a	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 ^b	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 ^c	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%

^a Refers to Inner Mongolia Autonomous Region.

^b Refers to Ningxia Hui Autonomous Region.

^c Refers to Guangxi Zhuang Autonomous Region.

Notes: This table reports the results of the policy simulation study. Monte Carlo simulations were repeated 1,000 times.

Source: Authors' calculations using data from China Household Finance Survey.

–21% in Theil T in Hainan, –16% in Gansu and Yunnan, etc.). The weights of changes in provincial Theil T in national changes in inequality are based on the importance of

each province in explaining nationwide inequality, reported in column (2). As already mentioned, returns to schooling in the PRC are negatively associated with GRP per capita. As a result, equalization of returns to schooling leads to inequality increasing in richer provinces and decreasing in poorer provinces. This reduces by 20.8% the poor-to-rich province income inequality ratio from 1.34 to 1.06.

The last two columns in Table 7 report the results of our fourth simulation. Equalization of the educational structure together with equalization of the returns to schooling leads to a 2.7% decline in national income inequality. The between-province component of nationwide inequality again increases to 23%, which is similar to the result from the previous simulation. The poor-to-rich province Theil T index ratio, however, drops slightly compared to the baseline results (1.27 versus 1.34).

Instead of equalizing provincial educational structure and returns to schooling at the national level, we could consider a policy that targets returns to schooling and educational structure to make them similar to current levels in one of the leading and highly developed provinces. For example, Shanghai is one of the top provinces by 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 schooling (Figure 7). All of these are desirable properties for reducing inequality. The results of this additional simulation analysis are reported in Table J1, Appendix J. If all provinces had an educational structure similar to that of Shanghai, it would reduce income inequality by 14.7% (Scenario “a”). If the equalization of educational structure were accompanied by adjusting returns to schooling (Scenario “b”), inequality would drop by 17.4%. If returns to schooling in all provinces were as low as in Shanghai (Scenario “c”), income inequality would decline by 4.4%. Finally, if both returns to schooling and educational structure in all provinces were the same as Shanghai (Scenario “d”), inequality would decline by 11.4%.

9 CONCLUSION

In this paper, we provide robust Gini index estimates for 29 Chinese provinces (excluding Tibet and Xinjiang), 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 is the main contributing factor to national inequality, explaining around 90% of the variation in annual household income per capita.

Our key empirical results suggest human capital disparities play a significant role in explaining differences in levels of income inequality across Chinese provinces.

Inequality in educational attainment and differences in returns to schooling are positively and statistically significantly related to provincial income inequality. Using data for the period 2012–2018, we find that a 10% increase in provincial-level educational Gini and provincial returns to schooling is associated with increases of 3.5–4.6 and 7.0–9.2 percentage points in the income Gini index. This result is consistent with findings from a recent study by Nie and Xing (2019), where returns to education are identified as a key explanatory factor explaining income inequality in the PRC (but only for urban households). We also find that the state's presence in the local economy and government social spending reduce provincial income inequality.

We argue that existing labor market frictions explain much of the interprovincial differences in returns to schooling and returns to high school and above in the PRC. In a fully integrated labor market, returns to schooling across provinces are expected to be similar. However, in the PRC, there still exists significant labor market segmentation, although it is declining. We found a statistically significant negative relationship between the share of skilled labor and provincial returns to schooling (as well as returns to high school). A 10% increase in the skilled labor share is associated with a decrease of 1.2–1.8 percentage point in returns to schooling, and a decrease of 9.4–15.7 percentage points in returns to high school and above. We also find that secondary industry employment share has a significant negative impact on returns to schooling and returns to high school and above. These results align with prior findings by Meng and Wu (1998), that an important determinant of regional income disparities is the degree of regional marketization in the PRC.

Lastly, we find that poor provinces are severely disadvantaged compared to rich provinces in the PRC. Poor provinces experience higher income and educational inequality, and their citizens have comparatively lower levels of education. The results from our simulation analysis suggest that equalization of educational structure across provinces to the national level could lead to a 14.4% decrease in income inequality. Equalization of returns to schooling in provinces to the national level, accompanied by equalization of the educational structure, could lead to a 2.7% decrease in income inequality. Equalization of returns to schooling to the low levels of highly developed Shanghai, together with equalization of the educational structure, could lead to an 11.4% decrease in national income inequality. These findings support results from previous studies showing that strengthening human capital is an important policy priority for regional development in the PRC (Fraumeni et al., 2019). The benefits of targeting educational policies can also be justified by the conviction that promoting education equalizes opportunities for citizens. Government policies that foster educational

development in disadvantaged provinces and reduce barriers to labor mobility can support labor market integration, reduce income disparities, and enhance sustainable and balanced regional economic development in the PRC.

REFERENCES

- Alesina, A. and R. Perotti (1996). Income Distribution, Political Instability, and Investment. *European Economic Review* 40(6), 1203–1228.
- Asadullah, M. N. and S. Xiao (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(4), 901–920.
- Butaeva, K. (2025). Income Inequality in China, 2012-2018: A New Measurement Approach.
- Cai, F., A. Park, and Y. Zhao (2008). The Chinese Labor Market in the Reform Era. In L. Brandt and T. G. Rawski (Eds.), *China's Great Economic Transformation*, pp. 167–214. Cambridge: Cambridge University Press Cambridge.
- Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge: University Press.
- Chan, W. K. and K. Ngok (2011). Accumulating Human Capital While Increasing Educational Inequality: A Study on Higher Education Policy in China. *Asia Pacific Journal of Education* 31(3), 293–310.
- Charpentier, A. and E. Flachaire (2022). Pareto Models for Top Incomes. *The Journal of Economic Inequality* 20, 1–25.
- Chong, A. and M. Gradstein (2017). Political and Economic Inequities and the Shaping of Institutions and Redistribution. *Southern Economic Journal* 83(4), 952–971.
- Churchill, S. A. and V. Mishra (2018). Returns to Education in China: A Meta-Analysis. *Applied Economics* 50(54), 5903–5919.
- Coady, D. and A. Dizioli (2018). Income Inequality and Education Revisited: Persistence, Endogeneity and Heterogeneity. *Applied Economics* 50(25), 2747–2761.
- Coles, S. G. and M. J. Dixon (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(3), 521–531.
- Cowell, F. A. (2000). Measurement of Inequality. *Handbook of Income Distribution* 1, 87–166.
- Cowell, F. A. and E. Flachaire (2007). Income Distribution and Inequality Measurement: The Problem of Extreme Values. *Journal of Econometrics* 141(2), 1044–1072.
- Cowell, F. A. and M.-P. Victoria-Feser (2007). Robust Stochastic Dominance: A Semi-parametric Approach. *The Journal of Economic Inequality* 5(1), 21–37.

- Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and Their Application*. Cambridge: University Press.
- Esteban, J. and D. Ray (1999). Conflict and Distribution. *Journal of Economic Theory* 87(2), 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 S. W. Polachek (Ed.), *Worker Well-Being and Public Policy*, pp. 1–38. Bingley: Emerald Group Publishing Limited.
- Fleisher, B. and D. Yang (2008). China's Evolving Labor Market. In B. Fleisher, N. Hope, A. Pena, and D. Yang (Eds.), *Policy Reform and Chinese Markets: Progress and Challenges*, pp. 125–158. Edward Elgar Publishing.
- Florida, R., C. Mellander, and P. J. Rentfrow (2013). The Happiness of Cities. *Regional Studies* 47(4), 613–627.
- Fraumeni, B. M., J. He, H. Li, and Q. Liu (2019). Regional Distribution and Dynamics of Human Capital in China 1985–2014. *Journal of Comparative Economics* 47(4), 853–866.
- Fu, Q. and Q. Ren (2010). Educational Inequality Under China's Rural–Urban Divide: The Hukou System and Return to Education. *Environment and Planning A* 42(3), 592–610.
- Galor, O. and O. Moav (2006). Das Human-kapital: A Theory of the Demise of the Class Structure. *The Review of Economic Studies* 73(1), 85–117.
- Gan, L., Z. Yin, N. Jia, S. Xu, S. Ma, L. Zheng, et al. (2014). *Data You Need to Know About China*. New York: Springer.
- Goh, C.-c., L. Xubei, and Z. Nong (2009). Income Growth, Inequality and Poverty Reduction: A Case Study of Eight Provinces in China. *China Economic Review* 20(3), 485–496.
- Golley, J. and S. T. Kong (2018). Inequality of Opportunity in China's Educational Outcomes. *China Economic Review* 51, 116128.
- Gradín, C. and B. Wu (2020). Income and Consumption Inequality in China: A Comparative Approach with India. *China Economic Review* 62, 101463.
- Gustafsson, B. and H. Wan (2020). Wage Growth and Inequality in Urban China: 1988–2013. *China Economic Review* 62, 101462.
- Hao, T., R. Sun, T. Tombe, and X. Zhu (2020). The Effect of Migration Policy on Growth, Structural Change, and Regional Inequality in China. *Journal of Monetary Economics* 113, 112–134.

- Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In E. A. Hanushek, L. Woessmann, and S. J. Machin (Eds.), *Handbook of the Economics of Education*, Volume 1, pp. 307–458. Elsevier.
- Heckman, J. J. and J. Yi (2012). Human Capital, Economic Growth, and Inequality in China. Working Paper Series, No. 18100.
- Huang, C., J.-G. Lin, and Y.-Y. Ren (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., P. Lanjouw, and N. Stern (1994). Income Inequalities in China: Evidence from Household Survey Data. *World Development* 22(12), 1947–1957.
- Jain-Chandra, M. S., N. Khor, R. Mano, J. Schauer, M. P. Wingender, and J. Zhuang (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(334), 261–289.
- Kanbur, R., Y. Wang, and X. Zhang (2021). The Great Chinese Inequality Turnaround. *Journal of Comparative Economics* 49(2), 467–482.
- Khan, A. R. and C. Riskin (2005). China's Household Income and its Distribution, 1995 and 2002. *The China Quarterly* (182), 356–384.
- Lei, X. and Y. Shen (2015). Inequality in Educational Attainment and Expectation: Evidence from the China Family Panel Studies. *China Economic Journal* 8(3), 252–263.
- Li, Q., S. Li, and H. Wan (2020). Top Incomes in China: Data Collection and the Impact on Income Inequality. *China Economic Review* 62, 101495.
- Li, S., J. Whalley, and C. Xing (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(3), 271–303.
- Martins, E. S. and J. R. Stedinger (2000). Generalized Maximum-Likelihood Generalized Extreme-Value Quantile Estimators for Hydrologic Data. *Water Resources Research* 36(3), 737–744.
- Martins, E. S. and J. R. Stedinger (2001). Generalized Maximum-Likelihood Pareto-Poisson Estimators for Partial Duration Series. *Water Resources Research* 37(10), 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. and H. X. Wu (1998). Household Income Determination and Regional Income Differential in Rural China. *Asian Economic Journal* 12(1), 35–63.
- Mincer, J. (1974). Schooling, Experience, and Earnings. *Human Behavior & Social Institutions* 2.
- Naughton, B. J. (2006). *The Chinese Economy: Transitions and Growth*. Cambridge, MA: MIT press.
- Ngai, L. R., C. A. Pissarides, and J. Wang (2019). China's Mobility Barriers and Employment Allocations. *Journal of the European Economic Association* 17(5), 1617–1653.
- Nie, H. and C. Xing (2019). Education Expansion, Assortative Marriage, and Income Inequality in China. *China Economic Review* 55, 37–51.
- Park, A., X. Song, J. Zhang, and Y. Zhao (2003). The Growth of Wage Inequality in Urban China, 1988 to 1999. Unpublished Paper, Chinese University of Hong Kong.
- Pauli, F. and S. Coles (2001). Penalized Likelihood Inference in Extreme Value Analyses. *Journal of Applied Statistics* 28(5), 547–560.
- Peng, H., L. Qi, G. Wan, B. Li, and B. Hu (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(2), 149–187.
- Ravallion, M. and S. Chen (2009). China's (Uneven) Progress Against Poverty. In R. Kanbur and X. Zhang (Eds.), *Governing Rapid Growth in China*, pp. 65–111. London: Routledge.
- Savoia, A., J. Easaw, and A. McKay (2010). Inequality, Democracy, and Institutions: A Critical Review of Recent Research. *World Development* 38(2), 142–154.
- Shorrocks, A. F. (1984). Inequality Decomposition by Population Subgroups. *Econometrica* 52(6), 1369–1385.
- Sicular, T., Y. Ximing, B. Gustafsson, and L. Shi (2007). The Urban-Rural Income Gap and Inequality in China. *Review of Income and Wealth* 53(1), 93–126.
- Sun, W., J. Lu, and C. Bai (2007). Analysis of Income Flow in Rural China (in Chinese). *Economic Research* 08, 43–57.
- 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(3), 277–293.
- Victoria-Feser, M.-P. and E. Ronchetti (1994). Robust Methods for Personal-Income Distribution Models. *Canadian Journal of Statistics* 22(2), 247–258.
- Wan, G. and Z. Zhou (2005). Income Inequality in Rural China: Regression-Based Decomposition Using Household Data. *Review of Development Economics* 9(1), 107–120.
- Wang, C., G. Wan, and D. Yang (2014). Income Inequality in the People's Republic of China: Trends, Determinants, and Proposed Remedies. *Journal of Economic Surveys* 28(4), 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. and R. Hao (2011). The Role of Human Capital in China's Total Factor Productivity Growth: A Cross-Province Analysis. *The Developing Economies* 49(1), 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., J. Stillwell, L. Burns, and A. Heppenstall (2020). Measuring and Assessing Regional Education Inequalities in China Under Changing Policy Regimes. *Applied Spatial Analysis and Policy* 13(1), 91–112.
- Xie, Y. and X. Zhou (2014). Income Inequality in Today's China. *Proceedings of the National Academy of Sciences* 111(19), 6928–6933.
- Xu, L. C. and H.-f. Zou (2000). Explaining the Changes of Income Distribution in China. *China Economic Review* 11(2), 149–170.
- Yang, D. T. (2005). Determinants of Schooling Returns During Transition: Evidence from Chinese Cities. *Journal of Comparative Economics* 33(2), 244–264.
- Yang, J., X. Huang, and X. Liu (2014). An Analysis of Education Inequality in China. *International Journal of Educational Development* 37, 2–10.
- Yu, N., B. Yu, M. de Jong, and S. Storm (2015). Does Inequality in Educational Attainment Matter for China's Economic Growth? *International Journal of Educational Development* 41, 164–173.
- Zhang, J., Y. Zhao, A. Park, and X. Song (2005). Economic Returns to Schooling in Urban China, 1988 to 2001. *Journal of Comparative Economics* 33(4), 730–752.

Provincial-Level Income Inequality in the People’s Republic of China

The Role of Human Capital

This paper examines income inequality across provincial-level administrative units in the People’s Republic of China (PRC) using a semiparametric distribution model. It finds that disparities in human capital—particularly educational inequality, returns to schooling, and average attainment—contribute to higher inequality in poor areas. The results suggest that bridging interprovincial gaps in human capital and enhancing labor market integration through targeted government policies could reduce regional disparities and overall income inequality in the PRC.

About the Asian Development Bank

ADB is a leading multilateral development bank supporting inclusive, resilient, and sustainable growth across Asia and the Pacific. Working with its members and partners to solve complex challenges together, ADB harnesses innovative financial tools and strategic partnerships to transform lives, build quality infrastructure, and safeguard our planet. Founded in 1966, ADB is owned by 69 members—50 from the region.



ASIAN DEVELOPMENT BANK

6 ADB Avenue, Mandaluyong City

1550 Metro Manila, Philippines

www.adb.org