Returns to education in China

Huang, Bin; Tani, Massimiliano; Wei, Yi; Zhu, Yu

Published in:
China Economic Review

DOI:
10.1016/j.chieco.2022.101804

Publication date:
2022

China Economic Review 74 (2022) 101804

Contents lists available at ScienceDirect

China Economic Review

journal homepage: www.elsevier.com/locate/chieco

Returns to education in China: Evidence from the great higher education expansion

Bin Huang [a], Massimiliano Tani [b,f], Yi Wei [c], Yu Zhu [d,e,f,*]

[a] *Institute of Education, Nanjing University, China*
[b] *School of Business, University of New South Wales Canberra, Australia*
[c] *China Institute for Educational Finance Research, Peking University, China*
[d] *School of Public Administration, Nanjing University of Finance and Economics, China*
[e] *University of Dundee School of Business, UK*
[f] *IZA - Institute of Labor Economics, Bonn, Germany*

ARTICLE INFO

JEL code:
I26
I23

Keywords:
Returns to education
2SLS
Higher education expansion
China

ABSTRACT

China experienced a near 5-fold increase in annual Higher Education (HE) enrolment in the decade starting in 1999. Using the China Household Finance Survey, we show that the Great HE Expansion has exacerbated a large pre-existing urban-rural gap in educational attainment underpinned by the *hukou* (household registration) system. We instrument the years of schooling with the interaction between urban *hukou* status during childhood and the timing of the expansion – in essence a difference-in-differences estimator using rural students to control for common time trends. We find that the Great HE raised earnings by 17% for men and 12% for women respectively, allowing for county fixed-effects. These Two Stage Least Squares (2SLS) estimates, which are robust to additional controls for *hukou* status at birth fully interacted with birth *hukou* province, can be interpreted as the Local Average Treatment Effect (LATE) of education on earnings for urban students who enrolled in HE only because of the Great HE Expansion. For the selected subsample of respondents with parental education information, we find that the 2SLS returns for students from more disadvantaged backgrounds are at least as high as their more advantaged counterparts, for both genders.

1. Introduction

Many high-income countries have experienced significant expansions of their higher education (HE) sector in recent decades, resulting in higher volumes and proportions of university graduates in the population. Within the traditional economic model of demand and supply, an increase in supply is predicted to cause a decline in the returns to university relative to high school education ('college premium'). However, a large empirical literature focusing on the US (Card & Lemieux, 2001; Goldin & Katz, 2008) and the UK (Blundell, Dearden, Goodman, & Reed, 2000; Blundell, Green, & Jin, 2016; Walker & Zhu, 2008) suggests otherwise: the returns to university degrees are largely constant over time despite the substantial increase in college enrolment – a contradiction resolved when the role of *skill-biased technological change (SBTC)* is accounted for, as SBTC increases the *relative* demand for highly educated workers (Card & DiNardo, 2002).

* Corresponding author at: 3 Perth Road, Dundee DD1 4HN, UK.
E-mail addresses: m.tani@adfa.edu.au (M. Tani), weiyipku@pku.edu.cn (Y. Wei), yuzhu@dundee.ac.uk (Y. Zhu).

https://doi.org/10.1016/j.chieco.2022.101804
Received 5 August 2021; Received in revised form 3 April 2022; Accepted 8 May 2022
Available online 11 May 2022

In contrast, fewer studies have investigated the labour market effects of a higher supply of university graduates in middle- and low-income economies, where substantive investments in tertiary education have occurred (Altbach & Balán, 2007; Bloom & Rosovsky, 2007; World Bank, 2017), despite the policy relevance of the topic. Expanding the tertiary sector absorbs considerable public resources for investments in infrastructure and workforce that yield positive returns (e.g. academic productivity) only over a long-term horizon. As these resources could be otherwise deployed to other parts of the economy (e.g. primary and secondary education, health, housing…), it is only natural to ask whether investing in HE is a wise choice vis-à-vis alternative options that are available to government. Measuring the returns to university degrees contributes to answer this fundamental public policy question.

Among middle-income countries China is the one that perhaps has carried out the largest investment in HE, and expanded the most, in recent times. As an example, the annual number of entrants into regular undergraduate higher education institutions (HEIs) grew from 1.08 million in 1998 to 6.08 million in 2008, a near five-fold increase over just a decade. We label this expansion as the Great HE Expansion, or just the Great Expansion for brevity. As a result of the Great Expansion, China's gross tertiary enrolment ratio, defined as the ratio of total enrolment, regardless of age, to the age relevant population corresponding to the HE stage, leaped from 5.86% in 1998 to 20.68% a decade later (World Bank, 2022). The most rapid growth occurred at the beginning of the period, with the number of HE entrants growing by 47% in 1999 from the year before (Che & Zhang, 2018; Wu & Zhao, 2010).

A number of studies, often relying on small surveys and qualitative analysis, have focussed on the impact of the Great Expansion on access to education and educational attainment. Both topics are particularly important given China's unique institutional settings with respect to people's movement: relocations into urban centres, where universities are typically located, are highly controlled through the nation's *hukou* (household registration) system, and rural residents, who the *hukou* penalises if they move out of their rural area of residence, are clearly disadvantaged (see e.g. Lu, Deng, & Guo, 2016; Ou & Hou, 2019; Wu, Yan, & Zhang, 2020). The general consensus is that while the Great Expansion has substantially increased educational attainment throughout the country, it has largely failed to reduce the gap in acquiring university education for the most socio-economically disadvantaged students, especially those with a rural *hukou*.

Notwithstanding these general considerations, the causal effect of the HE expansion on the returns to education and other labour market outcomes is still not very well understood. Our paper aims at filling this gap by presenting novel estimates of the causal effect of HE on the return to university degrees.

To do so, we use the 2017 China Household Finance Survey (CHFS), conducted by the Survey and Research Centre for China Household Finance at the Southwestern University of Finance and Economics (SWUFE), and develop the empirical analysis in sequential steps. First, we show that the large and unanticipated HE expansion has exacerbated the pre-existing urban-rural divide in educational attainment, as measured by average schooling by *hukou* status at age 12 - a critical age for subsequent educational choices. This finding is important as it motivates our identification strategy for developing a suitable instrument of schooling – a variable likely endogenous to wages, as is influenced by unobserved factors such as ability and motivation that also affect labour market performance. To estimate the causal returns to education we build on evidence that the variation across different birth cohorts in the exposure to the 1999 HE expansion, interacted with urban *hukou* status at age 12, can instrument the amount (years) of schooling completed. In other words, we proxy the effect of the HE on how much schooling a student acquired with a measure of that student's ability to take advantage of the HE expansion, which we derive from the student's cohort and *hukou* status. This approach is analogous to a difference-in-differences (DD) estimator that uses rural students to control for common time trends such as changes in teaching quality due to the sudden HE expansion or common macro-economic shocks such as China's accession to the WTO (World Trade Organization) at the end of 2001. This approach to identify the causal effect of the HE is novel, and, we believe, the main contribution of our paper. Using a large and nationally representative dataset is a further contribution of the paper, as, to the best of our knowledge, empirical evidence on the distributional impact of the HE expansion has so far tended to rely on small-sample and often qualitative data.

We find that the benchmark Ordinary Least Squares (OLS) estimates of returns to education, allowing for county fixed-effects (district within cities have the same status as counties in the administrative hierarchy of China), are about 5% for men and 6% for women, respectively. However, once we address the likely endogeneity of schooling choices with an IV approach, the Two Stage Least Squares (2SLS) estimates jump to 17% and 12% for men and women respectively. The 2SLS results can be interpreted as a Local Average Treatment Effect (LATE), i.e. the average treatment effect of HE attendance on earnings for urban students, defined by their *hukou* status at age 12, who enrolled in HE only as a result of the HE expansion. The 2SLS estimates satisfy all the diagnostic tests for relevance and exogeneity of instruments, and are also robust to further controls, such as the birth *hukou* status fully interacted with birth *hukou* province, and alternative specification of the age profile. While the 2SLS estimates for women are broadly comparable across geographical regions and type of areas of residence, there are significantly higher returns for men who live in the more developed Eastern Region or major cities. In addition, for a selective sample of respondents where parental education information is also collected by the CHFS, the 2SLS returns for disadvantaged students (due to their fathers' relatively low education) are at least as high as, if not higher than, their more privileged counterparts. This results arises for both genders, and is an important and novel finding (notwithstanding the selected sample from which it emerges), as it shows that acquiring HE removed the earning gap between more and less socio-economically disadvantaged students. This result sharply contrasts with that obtained by OLS, which suggests that the returns to both disadvantaged men and women are up to 40% lower than their more privileged counterparts.

Our findings shed new light on the debate regarding whether the expansion was economically justified on the grounds of returns to education and its distributional impacts. The 2SLS estimates suggest that there are very high returns to education for urban students in China, though rural students have missed out on this historical opportunity due to insufficient investment and poor quality of compulsory and upper secondary education in rural areas in the run-up to the HE. In addition, socio-economically disadvantaged students (as measured by the father's education level) who were induced to enroll in HE because of the expansion enjoy returns that are at least comparable to their more privileged counterparts, regardless of gender. However, the significantly higher returns for men living in the more developed Eastern Region or major cities indicate substantial returns to geographic mobility for graduates in China, which is consistent with the notion that the *hukou* system is a major obstacle to economic development and social mobility.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 presents the data. Section 4 introduces the 1999 Higher Education expansion and studies its distributional impacts, in the context of the unique *hukou* system. Section 5 is devoted to the discussion of the identification strategy for estimating the returns to education. The empirical results are presented and discussed in Section 6. Finally, Section 7 concludes.

2. Literature review

Due to China's rapidly increasing economic prominence since the reform started in the late 1970s, there has been growing interest in its education system and its influence on labour market outcomes. Awaworyi and Mishra (2018) conduct a meta-analysis of studies on the returns to education in China that were published in international academic journals, concluding that the returns to education in the post-reform era are around 18% per annum, with higher returns to college degrees than lower levels of education. However, most of the studies surveyed do not go beyond multivariate analyses with no control for self-selection bias, or IV estimations based on limited family characteristics, such as parental or spousal education, that tend to generate weak instruments in the second stage. Therefore one should be cautious in interpreting these estimates as causal due to the likely presence of unaddressed ability biases or remaining endogeneity problems arising from the effect of parental/spousal education on own earnings.

Compared to causal studies in other countries,[1] only few studies using Chinese data have attempted to uncover the causal returns to education using quasi-experimental methods.[2]

Some studies have exploited the introduction of the introduction of a 9-year compulsory schooling in 1986. By pooling the China Health and Nutrition Survey (CHNS) from 1997, 2000, 2004 and 2006, Fang, Rizzo, Rozelle, and Zeckhauser (2016) estimate the returns to education by instrumenting the years of schooling with the variation in the implementation of the compulsory education across provinces. Their IV estimate is as high as 20%, which is significantly above the OLS estimate. Exploiting the same policy reform on data from the China Urban Household Survey, Liu, Zhou, and Hu (2016), in the first study on returns to education in China using the Regression Discontinuity Design (RDD) approach, find a coefficient of 12.8%, which exceeds the corresponding 9.6% OLS estimate.[3]

Other studies focused on the Great HE Expansion. Using the 2002 Labour Market Survey by the Ministry of Labour and Social Security and the 2005 1% Population Survey, Wu and Zhao (2010) estimate the impact of the 1999 HE expansion on labour market participation and unemployment using DD and Difference-in-Difference-in-Differences (DDD) empirical strategies, finding an adverse employment effect of the expansion for new graduates. While new graduates do not fare worse relative to high school graduates in terms of wages, their hourly wage growth rates are more than 10 percentage points lower compared to older graduates. The estimated return to each year of university education is still high, at around 14%.[4]

Using the potential HE expansion based on provincial enrolment share in the period before the expansion as instruments for the supply of young skilled worker, Li, Ma, Meng, Qiao, and Shi (2017) find that the Great Expansion increased the college premium of workers aged 25 or above at the expense of younger graduates. Their IV estimates range from 0.164 (log points) for 20–24 year-olds to 0.474 for 40–60 year-olds. By exploring a fuzzy discontinuity in years of schooling and probability of obtaining HE qualification in the months of births, Dai, Cai, and Zhu (2021) suggest that each additional year of university education induced by the the HE expansion increases monthly wage income by 21%, compared to an OLS estimate of only 8%.

3. Data

The CHFS employs a stratified three-stage probability proportional to size (PPS) random sample design, with the primary sampling units (PSU) covering all counties (including county level cities and districts within cities at the prefectural-level or above in the administrative hierarchy) in mainland China except Tibet, Xinjiang and Inner Mongolia. The second and third stage of sampling involves selecting residential committees/villages and then households respectively from the sampling units chosen at the previous stages. Every stage of samplings is carried out with PPS method and weighted by its population size.

[1] For a survey, see e.g. Harmon, Oosterbeek, & Walker, 2003; Psacharopoulos & Patrinos, 2018.

[2] Other studies using quasi-experimental methods have focused on the effect of Great HE expansion on the employment of college graduates. They find that while the expansion increased the unemployment of new college graduates in the short-run, this effect mostly vanishes after 5 year (see e.g., Li, Whalley, & Xing, 2014; Xing, Yang, & Li, 2018).

[3] Recent studies have also presented descriptive evidence of the impact of the compulsory school reform on assortative marriage and income inequality (Nie & Xing, 2019) and intergenerational education mobility (Cui, Liu, & Zhao, 2019; Guo, Song, & Chen, 2019). Liang and Dong (2019) show that the increased education induced by the compulsory education law leads to decreased religious belief.

[4] The descriptive analysis by Knight, Deng, and Li (2017) using the 2002 and 2007 China Household Income Project (CHIP) and 2002–2008 Chinese Urban Household Survey, suggest that any adverse effect of the 1999 HE expansion on wages, employment and access to managerial and professional jobs for new graduates is relatively short-lived.

Our working sample includes both male and female employees who are aged 23 or above in 2017, and who were born in 1970 or later.[5] We exclude people who work in the Agriculture, Forestry, Fishing or Animal husbandry industries or international organizations. We also exclude people who reported occupation as Agriculture, Forestry, Fishing or Animal husbandry workers or Military Personnel.

After excluding missing values on key variables, such as monthly earnings, years of education and current *hukou* status, we end up with a sample of 20, 966 individuals, of which 11,633 (55.5%) are men.

Table 1 reports the weighted unconditional sample means by gender. On average, women's monthly earnings lag their male counterparts by 0.261 log points, despite women having the same age and 0.35 years more schooling. This earnings differential is equivalent to a 23.0% gender gap in monthly earnings. About 55% of men and 48% of women currently hold a rural (formally agricultural) *hukou*. More than 60% of men and women are estimated to have a rural *hukou* status at age 12, which is a critical age for educational attainment. Less than 10% of men or women currently live in Tier 1 cities (the 4 metropolises of Beijing, Shanghai, Guangzhou and Shenzhen) while another 30% or so of each gender live in Tier 2 cities (provincial capitals). Together these two types of cities account for most of the major cities in China. While just over one-third of men and nearly 40% of women live in smaller cities or towns, men are clearly over-represented in rural areas, relative to women.

4. The 1999 higher education expansion and its distributional impact

The education system in China suffered from severe disruption during the Cultural Revolution (1966–1976), when ideology was stressed over professional or technical competence. It was not until 1978 that the university entrance exam was reintroduced, as China entered the era of reform.

The Law on Nine-Year Compulsory Education was introduced in 1986, making 6 years of primary education and 3 years of Junior High Schools mandatory for all kids. The exit examinations upon completion of compulsory schooling, known as *zhongkao*, are used to stream students who wish to continue with education into either the academically oriented senior high schools or vocational high schools, both lasting 3 years (OECD, 2016).

After obtaining senior high school qualifications (or their vocational equivalents) which would normally take 12 years of schooling in total, one can apply to colleges and universities through a centralised admissions system which proceeds sequentially in tiers on the basis of one's performance in standardized National College Entrance Examinations (*gaokao*), with little regard for gender, *hukou* status and family background.[6] Colleges and universities in China can be classified into 2 types in descending order of prestige and entry requirements, i.e. universities, and vocational training colleges. The former take at least 4 years to complete and leads to a Bachelor's degree, whereas the latter usually takes 2–3 years leading to a college diploma.

Higher Education in China was free of tuition fees but exceedingly elitist till the early 1990s, when modest tuition fees were introduced. The growth of the HE sector was tightly controlled by the Ministry of Education (MoE), which sets provincial, university and subject quotas annually (OECD, 2016). Between 1995 and 1998, college enrolment only increased by an average of 4.7% per annum (Che & Zhang, 2018).

In response to the economic slowdown and rising youth unemployment in the aftermath of the 1997 Asian Financial Crisis, the MoE suddenly announced a 47% increase in university places. There was virtually no public consultation, and HE institutions around the country were only given a few months to prepare for the surge in intake (Li et al., 2017; Wan, 2006; Wu & Zhao, 2010). This was followed by increases of 38% and 22% in 2000 and 2001 respectively, and subsequent more modest but still substantial double-digit average annual growth for the next decade (Che & Zhang, 2018). This phenomenal growth was only eased off after the Global Financial Crisis in 2008, as it became more difficult for graduates to find suitable employment. Between 1998 and 2008, annual HE enrolment grew from 1.08 million to 6.08 million.

Fig. 1 shows the conditional annual enrolment rate in China, over the period 1986–2016. It is worth noting that the probability of enrolling in universities and colleges for senior high school graduates jumped from just over 40% in 1998 to 60% in 1999, and then to almost 80% in 2001.

Another important institutional influence on educational choices is the *hukou* (household registration) system, which classifies people as rural and urban status at birth, usually according to the mother's *hukou* status.[7] It is common knowledge in China that education resources at primary and secondary level are highly unequal in favour of urban residents. For instance, despite the

[5] We exclude people who were born before 1970, who would be aged 47 or above in 2017, for several reasons. First, due to the very generous state retirement age in China, especially for blue-collar women, and the prevalence of early retirement, this group is likely to suffer from selective attrition. Second, the education of pre-1970 cohorts were severely disrupted by the Cultural Revolution (1966–1976). Third, these cohorts would have completed lower secondary education before the 9-year compulsory schooling was introduced in 1986. Moreover, Figures A1 and A2 in the Appendix show differential time trends around the 1970 birth cut-off for years of schooling and HE attainment.

[6] For details of the Chinese College Admissions system, see Zhu (2014).

[7] See Chan (2009) for a review of the history of the *hukou* system and Meng (2012) for a discussion of the key role of *hukou* in China's labour market reforms in recent decades.

Table 1
Summary Statistics by gender, weighted.

	Males	Females	Total
Log total monthly net earnings	8.186	7.925	8.072
Years of schooling	11.39	11.74	11.54
Post-1980 birth	0.618	0.606	0.613
Age	34.8	34.8	34.8
Currently rural *hukou*	0.551	0.484	0.522
Rural *hukou* at age 12	0.647	0.606	0.628
Tier 1 cities	0.082	0.098	0.089
Tier 2 cities	0.277	0.311	0.292
Smaller cities and towns	0.359	0.392	0.374
Rural areas	0.282	0.199	0.246
Observations	11,633	9333	20,966

Note: using adjusted sampling weights based on per capita GDP ranking of the county/district of residence. Table A1 in the Appendix shows the corresponding unweighted means.

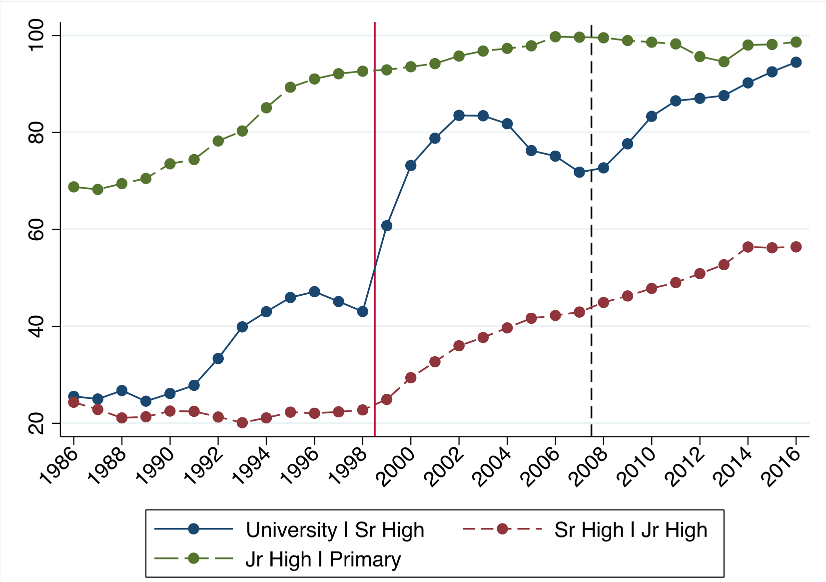

Fig. 1. Conditional enrolment (progression) rate over time, by level of qualifications. Data resources: NBS, 2017a.

significant improvements in recent years, the public expenditure per senior high school student in rural areas is only 13,360 RMB yuan in 2017, 25.2% below the national average of 16,727 RMB (NBS, 2017b).[8]

As a result, urban *hukou* holders, especially those living in major cities, enjoy much better access to HE. Intuitively, the *hukou* status in childhood determines whether the respondent attended an urban or rural secondary school, which are systematically different in quality.

For the purpose of identifying the causal effect of education on earnings, the *hukou* status that predates the schooling decision is more critical than the current status, which is subject to endogeneity bias: it is possible to change one's *hukou* status through migration, marriage, acquirement of properties in urban centres, and, most importantly, through obtaining HE qualifications (Xing, 2013). In view of these considerations, we first show the relationship between schooling and earnings by *hukou* status at age 12 for all sample members, who are all subject to the 9-year compulsory education regime by construction.

Table 2 shows the sample weighted average logarithm of the monthly income and key indicators of educational attainment by *hukou* status at age 12 by gender. More than 60% of both men and women have a rural origin. The urban-rural earnings gaps are approximately 0.17 and 0.25 log points for men and women respectively. Regardless of gender, respondents with urban origins have more than 3 years of extra schooling and are 40 percentage points more (or 3 times as) likely to have a college or above qualification (i.

[8] In 2017, 93.8% of middle school and 81.1% of senior high school teachers in rural areas (township and villages) hold associate and college degrees, respectively. The corresponding shares are 98.4% and 91.4% respectively among urban teachers. Moreover, the average values of equipment per student in rural schools are only 71.4% and 75.2% of those of their urban counterparts for primary and middle schools, respectively (MoE, 2018). These patterns indicate that the disadvantages faced by rural students are cumulative over all stages of basic education.

e. at least 14 years of schooling), compared to their rural counterparts. The 10% or so of respondents who have missing or unknown *hukou* status at age 12 seem to have characteristics more in common with people with urban origins. In the analyses that follow they will be grouped into the originally urban group to simplify the model specification and interpretation, though we will check the robustness of the results to the exclusion of this group, or imputing the missing unknown or missing childhood *hukou* status using the Multiple Imputation by Chained Equations (MICE) procedure (Royston & White, 2011).

Fig. 2 shows the share of respondents who hold rural *hukou* status at the time of the survey (which we will refer to as 'current' throughout the rest of the paper) and when they were aged 12, by gender and current region of residence. Overall, just over 20% of rural *hukou* holders switched to urban status between age 12 and the time of the 2017 survey, when they are at least aged 23. For both genders, the gap between the share of current and age 12 rural *hukou* varies significantly by the current type of region of residence: it rises from about 4% in rural areas to 12–15% in towns and cities of various sizes. This reflects both the large rural to urban migration and an urbanization process which turn non-migrant peasants into urban residents.

Fig. 3 and Fig. 4 show the average years of schooling and college attainment (i.e. 14+ years of schooling), respectively, by year of birth and *hukou* status at age 12. The vertical lines highlight the approximate cut-off points around the birth cohort 1980, which corresponds to the Great Expansion of 1999–2008.[9] There are two important patterns to note from these figures. First, there is an enormous gap in educational attainment between rural and urban *hukou* holders, as measured by the status before age 12. This is roughly the age when primary education is completed. In the decade before the HE expansion, there was a consistent 3-year gap in years of schooling and an even growing gap in college attainment between urban and rural *hukou* holders. Second, the HE expansion has clearly had a differential impact on rural and urban people, as indicated by the shifts around the cut-offs, and the changes in the slopes before/after the reform. For both outcomes, it appears that the Great HE expansion induced only visible discontinuities for respondents with an urban *hukou*, by about 0.54 more years of schooling and 6.5 percentage points higher HE attainment. One plausible explanation is that the secondary schools in the cities were of much better quality compared to rural schools, and this led graduates from urban areas to benefit disproportionately from the expansion. According to Zhao, Ye, Li, and Xue (2017), "In 1990, 2000 and 2010, the transition rates from junior to senior high school in cities were 40.41%, 66.71% and 88.11% respectively, while in the same years, the transition rates in towns and rural areas were 18.96%, 22.06% and 38.36%, respectively". It is worthnoting that the urban-rural gap in junior-to-high school transition rates reached the peak of 44.7 percentage points (or a ratio of 3 to 1) in 2000, which was the year immediately after the Great HE Expansion.[10]

On the other hand, while there appears to be no immediate impact of the HE expansion on either outcome for rural students, rural students manage to catch up over time, as evidenced by the steeper time trends after the expansion for both outcomes compared to their urban counterparts. These empirical facts suggest us to allow for the full interaction of the *hukou* origin with the time trend (i.e. the intercept dummy, as well as their interaction term, which captures the change in the trend), in the empirical modelling.

Table 3 presents the OLS and Linear Probability Model estimates of the effect of the HE expansion on the years of schooling and on attainment of HE, allowing for differences in *hukou* status at age 12. This is effectively a parametrisation of Figs. 3 and 4, but broken down by gender and controlling for heterogeneity in individual characteristics and county and district fixed-effects. Given China's vast diversity in geography and stages of economic development, we control for unobservable variation across geographical areas by adding 353 county (or city district) fixed-effects.[11]

To parameterize the potential change in both the intercept and the slope induced by the HE expansion we include three main likely effect variables in the regressions:

a) *Birth year trend (T)*: a linear time trend before the expansion as the baseline;
b) *Post-1980 birth (D)*: a dummy for being born in 1980 or later to capture the instantaneous effect of the HE expansion (on the intercept);
c) *Post-1980 birth year trend (TD)*: an interaction term between the linear *birth year trend* and the *post-1980 birth* dummy to capture the change in the time trend from the pre-expansion baseline.

In addition, a dummy for urban *hukou* at age 12, U, is included to capture any persistent urban-rural gap before the expansion.

To enable the impact of the HE expansion to differ between rural and urban *hukou* holders, we interact an urban *hukou* at age 12 dummy with each of the 3 main HE expansion variables above.

Formally, the OLS wage equation is specified as:

[9] Dai et al. (2021) use RDD for identification, with September 1979 as the optimal birth cut-off. Unfortunately, we only have calendar year of birth but not month of birth in CHFS. A dummy for born 1980 or later effectively treats January 1980 as the cut-off.

[10] Using the 2001 and 2005 China Urban Labour Survey, Wang (2009) shows that while the returns to education for urban students rises sharply at college level or above compared to earlier stages, the turning point occurs at the upper secondary school level for rural students. This is consistent with the notion that prior to the Great HE Expansion, the vocational secondary school route which charges no tuition fees, was perceived by many academically able rural students as a cost-effective route to urban *hukou* status and guaranteed public-sector jobs, due to credit constraints and more limited resources in the academic-oriented senior high education in rural areas. This implies that rural students as a whole would benefit less from an unanticipated HE expansion in short-run, compared to their privileged urban counterparts.

[11] A county is the third highest level of administrative division of China, under the provincial and prefectural levels. China administers 2851 county-level divisions, including city districts, at the end of 2017 (NBS, 2018).

Table 2
Weighted Earnings and Educational Attainment at age 12 *Hukou* Status and gender.

	Men			Women		
	Rural	Urban	Missing/Unknown	Rural	Urban	Missing/Unknown
Log monthly income	8.127	8.294	8.298	7.832	8.084	8.017
Years of schooling	10.331	13.506	12.789	10.537	13.855	12.746
9+ years of schooling	0.848	0.982	0.969	0.833	0.988	0.931
14+ years of schooling	0.221	0.626	0.516	0.279	0.678	0.535
Observations	7196	3303	1134	5322	3017	994
Weighted Share (%)	64.7	26.2	9.1	60.4	29.6	10.0

Note: Means using adjusted sampling weights based on per capita GDP ranking of the county/district of residence.

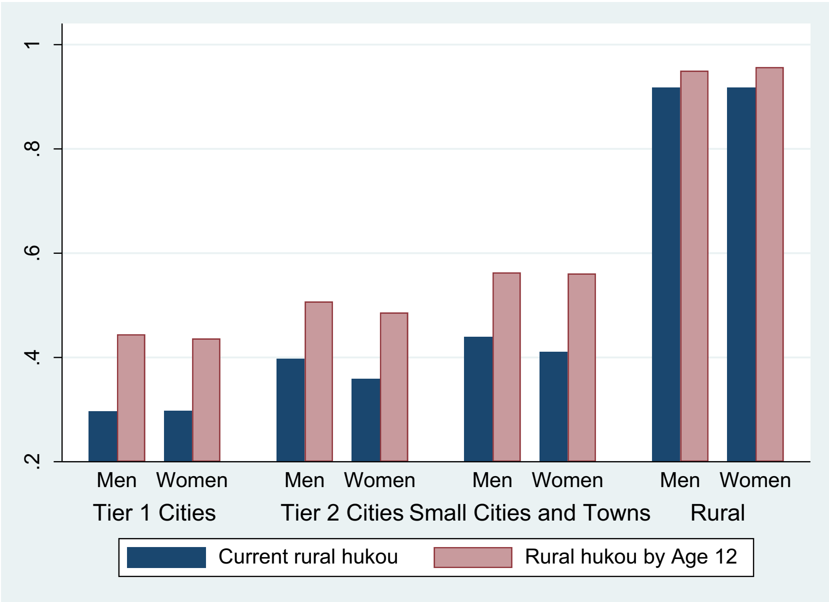

Fig. 2. Current and historical rural *hukou* status by region of residence and gender.

$$lnW_{ij} = \alpha + \beta S_{ij} + \gamma X_{ij} + \delta_0 U_{ij} + \delta_1 T_{ij} + \delta_2 D_{ij} + \delta_3 \left(TD_{ij}\right) + \theta C_j + \varepsilon_{ij} \quad (1)$$

where $\ln W_{ij}$ denotes log monthly earnings for individual *i* in county *j* (the dependent variable), S_{ij} is years of schooling (the endogenous variable), X_{ij} is a vector of control variables, C_j is a dummy for county *j* which captures the county fixed-effect, and ε_{ij} is the error term. U, T, D and TD denote the urban *hukou, Birth year trend, Post-1980 birth and Post-1980 birth year trend* as defined above respectively.

For both educational outcomes and genders, the first two rows of Table 3 indicate that there is no statistically significant change in either intercept (*post-1980 birth* dummy) or slope (*post-1980 birth year trend*) relating to the HE expansion for respondents with rural *hukou* status at age 12 (the reference category). This is in sharp contrast to men and women with urban *hukou* at age 12, who enjoy a 2.6- and 3.0-years advantage in years of schooling, respectively, before the expansion and more than 20 percentage point advantage in terms of college attainment. The urban-rural gaps in years of schoolings widened by 1.2 and 2.3 years after the HE expansion for men and women, respectively. In terms of HE attainment, the Great Expansion increased the urban-rural gaps by 21 and 45 percentage points for men and women in the short-run. However, the widened urban-rural education attainment gap gradually narrowed over time, as evidenced by the negative coefficient of the interaction term between the *post-1980 birth* (i.e. post HE expansion) *trend* and urban *hukou* indicator.[12] It is important to note that for both outcomes and genders resuming their relative pre-expansion positions will take rural students at least a decade. In the case of rural women, the recovery would take 15 years, i.e. 2.302/0.154.

Overall, the parametric estimates are totally consistent with the earlier graphic presentations, and clearly show that the Great HE Expansion had a large and unintended adverse effect on the pre-existing educational attainment gap in favour of urban *hukou* holders,

[12] Dai, Xu, and Zhu (2022) show that the continued college expansion has contributed to the rapid narrowing of the huge urban-rural gap in the proportion of college-educated teachers in both primary and secondary education through increased supply of college graduates. Moreover, with rural school consolidations and implementation of more inclusive education policies, many rural students now attend schools in urban areas where their parents work (Tani, Xu, & Zhu, 2021; Xie & Zhang, 2022).

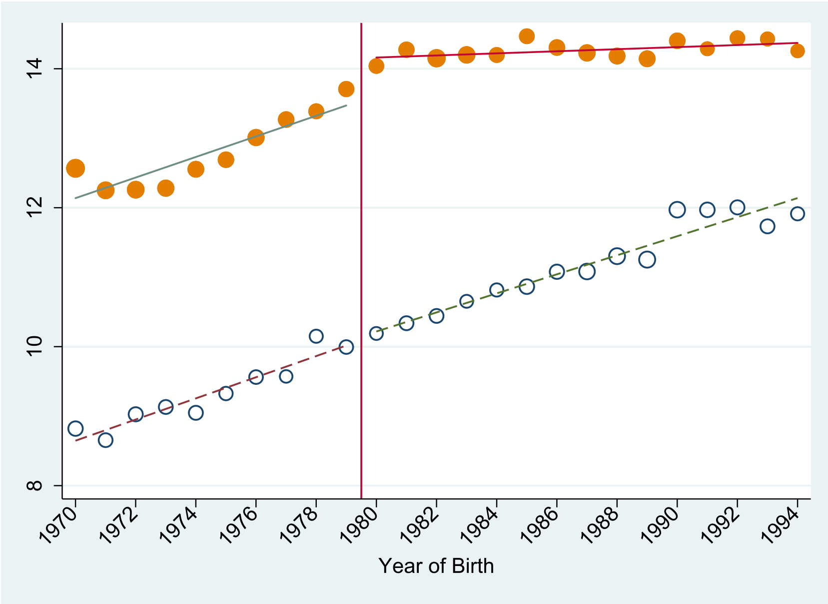

Fig. 3. Years of Schooling by birth year and *hukou* status at age 12. Note: Solid dots/lines denote urban *hukou* at age 12. Hollowed dots and dashed lines denote rural *hukou* at age 12. The marker size is proportional to the number of observations.

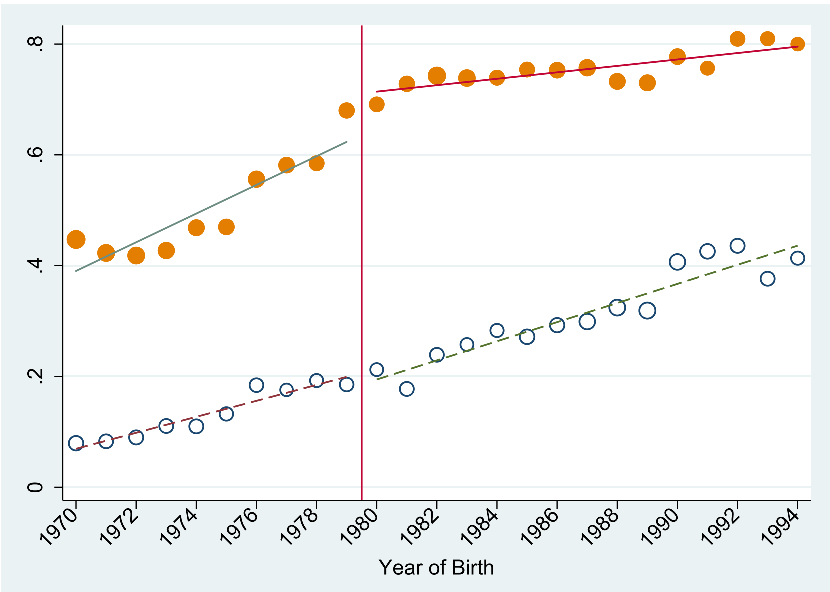

Fig. 4. Higher Education qualifications by birth year and *hukou* at age 12. Note: Solid dots/lines denote urban *hukou* at age 12. Hollowed dots and dashed lines denote rural *hukou* at age 12. The marker size is proportional to the number of observations.

Table 3
OLS/LPM estimates of educational attainment by gender, with county/district FE.

	Years of schooling		HE attainment	
	Men	Women	Men	Women
	(1)	(2)	(3)	(4)
Post-1980 birth	0.764	−0.038	0.041	−0.032
	(0.595)	(0.676)	(0.090)	(0.095)
Post-1980 birth year trend	−0.053	0.001	−0.003	0.001
	(0.053)	(0.061)	(0.008)	(0.009)
Urban *hukou* at age 12	2.568***	3.041***	0.232***	0.207***
	(0.183)	(0.208)	(0.027)	(0.028)
Post-1980 birth X Urban *hukou*	1.231***	2.302***	0.207***	0.453***
	(0.346)	(0.383)	(0.054)	(0.056)
Birth year trend X Urban *hukou*	0.012	−0.016	0.009**	0.016***
	(0.030)	(0.033)	(0.004)	(0.005)
Post-1980 birth year trend X Urban *hukou*	−0.106***	−0.154***	−0.016***	−0.034***
	(0.034)	(0.038)	(0.005)	(0.005)
Age	−0.264	−0.335*	−0.025	−0.060**
	(0.165)	(0.187)	(0.026)	(0.028)
Age squared	0.002	0.002	0.000	0.001*
	(0.002)	(0.002)	(0.000)	(0.000)
Constant	16.334***	18.831***	0.768	1.544**
	(3.594)	(4.075)	(0.559)	(0.602)
County/District FE	✓	✓	✓	✓
Observations	11,633	9333	11,633	9333
R^2	0.370	0.411	0.295	0.341

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Columns 1 and 2 are estimated by OLS. Columns 3 and 4 are estimated by Linear Probability Model (LPM). The baseline (pre-1980) birth year trend is omitted due to perfect multi-collinearity with age in cross-sectional data.

at least in the short-run.

5. Identification strategy

The main challenge in estimating the returns to education is the endogeneity of educational attainment, which is subject to self-selection, ability bias and measurement errors. All three factors are relevant in our case. Unlike compulsory education, HE is certainly a matter of individual choice. In CHFS, we also do not have a direct measure of the respondents' ability such as scores in standardized exams or test, or good measures of the quality of the education qualifications held.[13] Moreover, the variable measuring the years of schooling is imputed from people's self-reported highest qualification obtained or attempted rather than institutional sources.

To address endogeneity and measurement error in self-reported years of schooling converted from attempted/completed level of qualifications, we exploit the Great HE Expansion as a source of exogenous variation in educational attainment, and combine it with people's original *hukou* status (given the strong evidence of differential impact of the HE expansion on educational attainment) to generate an instrumental variable and carry out a regression using Two-Stage-Least-Squares (2SLS). In line with Figs. 3 and 4, which suggest that the impact of the HE expansion differs by childhood *hukou* status, we interact the dummy variable capturing urban *hukou* at age 12 with each of the 3 main HE expansion variables described above.

As a result, the identification of the causal effect of education on earnings only relies on the interactions of the urban *hukou* at age 12 indicator with these variables. In other words, we assume that the interaction effects of the *hukou* system and the HE expansion has no direct effect on earnings over and above their impact on educational attainment. Compared to an instrumental variable strategy, which relies on the main effects of HE expansion variables and the urban *hukou* dummy, our identification strategy is less restrictive. Our model is over-identified. Hence, we are able to formally test the exogeneity of the instruments using the over-identification tests. If any of the main HE expansion effects above is statistically significant in the second stage, that will lend support to our identification strategy, which builds on the interaction terms only.

Formally, the 2SLS is a two-equation system whereby the first stage involves estimating an equation by OLS in which the years of

[13] Using China Family Panel Studies and Inverse Probability Weighted Regression Adjustment (IPWRA), Kang, Peng, and Zhu (2019) show that returns to the more selective key universities are higher than ordinary universities or vocational colleges regardless of subjects studied. However, IPWRA estimates can only be interpreted as causal if there is no selection on unobservables.

schooling are regressed on exogenous controls, as well as the instrumental variables $U_{ij}T_{ij}$, $U_{ij}D_{ij}$, and $U_{ij}TD_{ij}$ (the excluded variables):

$$S_{ij} = \alpha + \gamma X_{ij} + \delta_0 U_{ij} + \delta_1 T_{ij} + \delta_2 D_{ij} + \delta_3 \left(TD_{ij}\right) + \left[\delta_{10}\left(U_{ij}T_{ij}\right) + \delta_{20}\left(U_{ij}D_{ij}\right) + \delta_{30}\left(U_{ij}TD_{ij}\right)\right] + \theta C_j + \varepsilon_{ij} \quad (2)$$

The second stage is similar to eq. (1), but it replaces the observed S_{ij} with the fitted value estimated from the first stage:

$$lnW_{ij} = \alpha + \beta\widehat{S_{ij}} + \gamma X_{ij} + \delta_0 U_{ij} + \delta_1 T_{ij} + \delta_2 D_{ij} + \delta_3 \left(TD_{ij}\right) + \theta C_j + \varepsilon_{ij} \quad (3)$$

Technically, the IV estimator we construct is analogous to a DD estimator of the returns to education when the earnings trend is common for both rural and urban *hukou* holders, in the absence of the Great HE Expansion. The fact that rural *hukou* holders share a common trend in years of schooling with their urban counterparts before the expansion ensures that they can serve as an ideal control group for the urban *hukou* holders.

6. Empirical results

We first present the benchmark OLS estimates of returns to years of schooling by gender, with county/district fixed-effects (FEs). Then we present the main 2SLS estimates, followed by analysis of heterogeneous treatment effects, whereby the sample is separately analysed by type of area or geographical region of residence. Finally, for the selective subsample of members living with parents recorded in the 2017 CHFS, we check whether the returns to education vary by parental education. However, we stress that any lessons drawn using this non-random subsample must be treated as strictly tentative, as the external validity cannot be guaranteed.

6.1. OLS baseline

Table A2 shows the OLS estimates of returns to schooling, which serve as the baseline against which the 2SLS results are compared.[14] Columns 1 and 2 present our preferred gender-specific specification of returns to schooling, which control for 353 county/district FEs.[15] Columns 3, 4 and 5 present pooled-gender specifications, without and with the interaction between a female dummy and years of schooling and age and age squared. According to the formal statistical tests, the pooled-gender specifications are overwhelmingly rejected, due to significant differences in both the returns and the age-earning profiles across gender.

Apart from urban *hukou* at age 12, our baseline model also controls for being born in 1980 or later, and an interaction term with the time trend (denoted as "post-1980 birth year trend"). These are intended to capture the main effect of the HE expansion. The pre-expansion birth year trend variable is omitted due to perfect multi-collinearity with age in our single cross-sectional data. It is interesting to note that urban *hukou* at age 12 has a negative and significant effect on earnings for men, conditional on years of education and other regressors.

6.2. 2SLS main results

Table 4 presents the 2SLS estimates by gender. The first-stage results in columns (1) and (3) are identical to the first two columns of Table 3. The second stages of the 2SLS in columns (2) and (4) have the same specification as the baseline OLS in Table A2, apart from replacing the observed years of schooling with the corresponding predicted values from the first stage. However, the coefficient for urban *hukou* at age 12 has a substantial negative influence on earnings for both genders. This is consistent with an education system that is inherently biased against rural *hukou* holders: conditional on educational attainment rural students might have higher cognitive and/or non-cognitive abilities than their urban counterparts.

Allowing for endogeneity and measurement error in the years of schooling increases the returns from 4.9% in OLS to 16.5% for men, and from 6.2% in OLS to 12.4% for women. Consistent with Figs. 3 and 4, the large positive coefficients of the interaction between post-1980 birth and urban *hukou* origin suggests that the HE expansion itself had significantly widened the urban-rural gap in years of schooling by approximately 1.2 years for men and 2.3 years for women. On the other hand, the interaction between childhood urban *hukou* and its pre-expansion time trend is statistically insignificant, reflecting the parallel time trends between urban and rural *hukou* holders before the HE expansion. Finally, the negative coefficient of the post-1980 birth year trend interacted with urban *hukou* origin indicates that the urban-rural gap in years of schooling is narrowed at an annual rate of 0.10–0.15 year after the initial shock, again fully consistent with Fig. 3. The main effects of the HE expansion variables as well as the urban *hukou* at age 12 dummy are all included as controls in the second stage. Because of the perfect multi-collinearity between birth year trend and age in a single cross-

[14] While all our regression analysis is unweighted, we report robust standard errors throughout. Moreover, allowing for county-FE should capture much of the heterogeneity arising from the multi-stage cluster sampling design.

[15] A more parsimonious specification (not reported to save space) replacing 353 county/district FEs with 3 region-type dummies (with Tier 1 city as the reference) would increase the returns by 0.9 and 0.5 percentage points for men and women respectively. Moreover, the R^2 measures would also decrease by around 10 percentage points for both genders, suggesting that local labour market conditions account for a significant proportion of the variation in individual earnings.

Table 4
2SLS estimates of returns to schooling by gender, with county/district FE.

	Men		Women	
	First-stage	Second-stage	First-stage	Second-stage
	S	lnW	S	lnW
	(1)	(2)	(3)	(4)
Years of schooling (S)		0.165***		0.124***
		(0.028)		(0.017)
Post-1980 birth X Urban *hukou*	1.231***		2.302***	
	(0.346)		(0.383)	
Birth year trend X Urban *hukou*	0.012		−0.016	
	(0.030)		(0.033)	
Post-1980 birth year trend X Urban *hukou*	−0.106***		−0.154***	
	(0.034)		(0.038)	
Controls				
Age	−0.264	0.147***	−0.335*	0.106***
	(0.165)	(0.036)	(0.187)	(0.034)
Age squared	0.002	−0.002***	0.002	−0.001***
	(0.002)	(0.000)	(0.002)	(0.000)
Post-1980 birth	0.764	−0.265**	−0.038	−0.181
	(0.595)	(0.129)	(0.675)	(0.116)
Post-1980 Birth year trend	−0.053	0.023**	0.001	0.019*
	(0.053)	(0.011)	(0.061)	(0.010)
Urban *hukou* at age 12	2.568***	−0.313***	3.041***	−0.157***
	(0.183)	(0.065)	(0.208)	(0.045)
Constant	16.334***	3.303***	18.831***	4.274***
	(3.594)	(0.887)	(4.075)	(0.794)
County/District FE	✓	✓	✓	✓
Diagnostic Tests:				
Joint significance of IVs: $F_{(3)}$ (*p*-value)		19.995 (0.000)		41.327 (0.000)
Endogeneity of years of schooling: $\chi^2_{(1)}$ (*p*-value)		24.124 (0.000)		15.648 (0.000)
Over-identification: $\chi^2_{(2)}$ (*p*-value)		0.698 (0.706)		0.839 (0.657)
Observations	11,633		9333	

Note: 353 county/district dummies. Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The baseline (pre-1980) birth year trend is omitted due to perfect multi-collinearity with age in cross-sectional data.

section, the birth year time trend variable is omitted from the second-stage wage regression. It turns out that all these main effects are statistically significant individually in Table 4, thus precluding their use as instruments.

For both genders, our model specifications satisfy all the textbook diagnostic tests. The instruments are jointly significant with the F-statistics with 3 degrees of freedom above the normal threshold of 10, thus indicating a very unlikely presence of weak instruments. The endogeneity tests for years of schooling are easily rejected with *p*-values at 0.000. On the other hand, we cannot reject the over-identification tests even at the 50% level of significance across all specifications. This means that if at least one of our 3 instruments is exogenous (the maintained hypothesis), then we cannot reject the hypothesis that all 3 instruments are exogenous.[16]

6.2.1. Robustness of 2SLS results

In Table A3 in the Appendix, we test the sensitivity of our headline 2SLS estimates with respect to an alternative specification of the age-earnings profiles. Instead of the conventional Mincer wage specification with a quadratic term for age (Mincer, 1958), we control for a full set of age dummies. It turns out the estimates for returns to education are virtually identical.

Although we define the post-expansion cohorts as being born in 1980 or later, we would like to use September 1979 as the ideal starting point for defining post-expansion cohorts (Dai et al., 2021; Wu & Zhao, 2010). Unfortunately, we do not have month of birth in the CHFS data. Nevertheless, in Table A4 in the Appendix, we show the results of a test of robustness of the main 2SLS results with respect to being born in 1979 (rather than in 1980). The 2SLS estimates turn out to be virtually identical, implying that our choice of a 1980 cut-off date does not influence the results presented so far.

An additional concern is the 10% or so of respondents with unknown or missing age 12 *hukou* status that have been treated as urban residents in Table 4. In Table A5 in the Appendix, we impute these unknown or missing values using the Multiple Imputation by Chained Equations (MICE) procedure in Stata (Royston & White, 2011). The resulting 2SLS estimates are, if anything, more

[16] The returns to education estimates are virtually unchanged when using the Limited-Information Maximum Likelihood (LIML) or Generalized Method of Moments (GMM) estimators. These results are not reported but are available upon request.

pronounced than the headline estimates in Table 4. Despite the increase in the bootstrapped robust standard errors, the estimated returns to education remain statistically significant at the 1% level regardless of gender.[17]

6.2.2. Local Average Treatment effects (LATE)

Our 2SLS estimates are significantly higher than their OLS counterparts. However, they are not out of line with the literature: Fang et al. (2016), Li et al. (2017), and Dai et al. (2021) all report higher estimates. We highlight however that the 2SLS estimates should be interpreted as the Local Average Treatment Effect (LATE), i.e. the returns to (one-year of) schooling for marginal students with an urban *hukou* at age 12 who invested in their education only because of the HE expansion. This is plausible if marginal students not undertaking college education in the absence of the HE expansion possess favourable characteristics, such as valuable unobservable skills like academic ability and motivation, or more social capital as proxied by parental education. An alternative explanation is that the coefficients obtained by OLS might be downward biased because of errors in self-reporting, e.g. misclassifying dropouts and non-completions.

The 2SLS effectively compares younger and older birth cohorts with an urban *hukou* at age 12, after controlling for an HE expansion effect assumed to be common for all younger birth cohorts regardless of their *hukou* status. Compared to a conventional IV, the advantage of our approach is that we can use rural *hukou* holders as a control to net out any common birth cohort-related effect such as a deterioration in HE quality directly caused by the rapid HE expansion or any common macro-economic shocks such as China's accession to the WTO at the end of 2001.

We recognise that while it is unfortunate from an equity perspective that the rural origin students largely missed out on the HE expansion, at least in the short run, from an econometric point of view they serve as an ideal control group for urban *hukou* holders in the sense of indicating the potential trend in the absence of the HE expansion.

6.2.3. Threats to identification

One main concern for the chosen identification strategy is that the instruments might also pick up other policy changes around the period, such as the introduction of the 9-year compulsory education in 1986 or the family planning policy (often referred to as the One Child Policy). By selecting a sample of people born in 1970 or later, all sample members would have been exposed to the compulsory education reform given that they would be no more than 16 years old in 1986. Therefore, we would not expect the Compulsory Education Law to have any significant effect on our outcomes.

We address the potential endogeneity with respect to family planning policies by additionally controlling for the full interaction of *hukou* type at birth and *hukou* province at birth. One key feature of China's family planning policy is that the policy regime effectively varies at the province level by *hukou* status because of large variations in population density and level of economic development across the country (Wang, Zhao, & Zhao, 2017; Weng, Gao, He, & Li, 2019; Yan, 2011).

Table 5 present both stages of the 2SLS results, in an extended model specification with the full interaction of *hukou* type at birth and *hukou* province at birth. Compared to Table 4, there is remarkable consistency in the 2SLS estimates of the returns to education. Moreover, the modified model easily passes all the diagnostic tests.

6.3. Heterogeneous treatment effects

We further explore heterogeneous treatment effects, with respect to the type of area of residence and geographical region. In Table 6, we repeat the analysis by conditioning on the current type of region, by grouping Tier 1 and Tier 2 cities together, and the rest together. Consistent with the pooled sample analysis, 2SLS estimates are significantly higher than their OLS counterparts conditional on the type of area of residence, for both genders. While the OLS estimates are significantly higher regardless of gender for residents of major cities, the gap becomes insignificant in 2SLS for women. As for men, the causal estimates of returns for residents of major cities are 9 percentage points higher than (or twice as high as) those living in smaller cities, towns or rural areas.

In Table 7, we re-run OLS and 2SLS regressions for each of the 3 main geographic regions of China, namely Eastern, Central (including Northeast) and Western. Whereas the 2SLS returns for women are similar across the regions, men in the Eastern region enjoy a return of almost 20%. This exceeds the comparable return for men in other regions by at least 7 percentage points. The Eastern region (excluding the Northeast) is China's most developed area, and also hosts all 4 Tier 1 cities (Beijing, Shanghai, Guangzhou and Shenzhen). Therefore, it is perhaps not surprising that the pattern presented in Table 7 overlaps with that in Table 6.

One obvious explanation is that the OLS estimates might be contaminated by endogenous migration decisions. In particular, people who expect higher returns to their education may choose to migrate to major cities in order to take advantage of the greater economic opportunities in terms of earnings and career progression.[18] In equilibrium, the returns to education should be equalized across different area types or geographical regions in the absence of institutional barriers. Our finding implies that the *hukou* system has acted

[17] Excluding observations with unknown or missing age 12 *hukou* status would only change the headline returns to education estimates marginally to 0.169 and 0.126 for men and women respectively and the model still passes all the diagnostic tests. In the interest of space, we don't report the full set of estimates.

[18] Our estimates might be biased upwards if workers in major cities work longer hours. Unfortunately, we are unable to estimate the returns to education on hourly wages due to lack of information on working hours.

Table 5
2SLS returns to schooling by gender, with county FEs and birth province-*hukou* FEs.

	Men		Women	
	First-stage	Second-stage	First-stage	Second-stage
	S	lnW	S	lnW
	(1)	(2)	(3)	(4)
Years of schooling (S)		0.174***		0.121***
		(0.032)		(0.016)
Post-1980 birth X Urban *hukou*	1.287***		2.414***	
	(0.357)		(0.402)	
Birth year trend X Urban *hukou*	0.010		−0.017	
	(0.029)		(0.033)	
Post-1980 birth year trend X Urban *hukou*	−0.103***		−0.161***	
	(0.034)		(0.038)	
Controls				
Age	−0.332**	0.148***	−0.249	0.095***
	(0.169)	(0.038)	(0.193)	(0.033)
Age squared	0.002	−0.002***	0.001	−0.001***
	(0.002)	(0.000)	(0.002)	(0.000)
Post-1980 birth	0.972	−0.268**	−0.252	−0.162
	(0.606)	(0.135)	(0.689)	(0.115)
Post-1980 Birth year trend	−0.079	0.024**	0.024	0.016
	(0.054)	(0.012)	(0.062)	(0.010)
Urban *hukou* at age 12	1.740***	−0.243***	1.762***	−0.096***
	(0.198)	(0.054)	(0.230)	(0.030)
Constant	18.785***	2.704***	20.139***	4.407***
	(3.714)	(1.007)	(4.241)	(0.796)
County/District FE	✓	✓	✓	✓
Birth *hukou*-province FE	✓	✓	✓	✓
Diagnostic Tests:				
Joint significance of IVs: $F_{(3)}$ (*p*-value)		15.093 (0.000)		38.465 (0.000)
Endogeneity of years of schooling: $\chi^2_{(1)}$ (*p*-value)		22.404 (0.000)		14.630 (0.000)
Over-identification: $\chi^2_{(2)}$ (*p*-value)		1.309 (0.520)		0.513 (0.774)
Observations	11,496		9154	

Note: 353 county/district dummies. Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The baseline (pre-1980) birth year trend is omitted due to perfect multi-collinearity with age in cross-sectional data.

as a severe barrier for geographic mobility, even for college graduates who are expected to be the most responsive demographic group to take advantage of opportunities in distant labour markets.[19]

6.4. Heterogenous effect with respect to parental education

For roughly one-third of the respondents in our sample who live together with their parents, there is information on parental education. For this non-random subsample we can test for heterogeneous treatment effects with respect to parental education as a proxy for socio-economic background. Following the literature, we focus on father's education, which turns out to be available for 44.5% of men and 20.8% of women in the original sample. Given this much smaller subsample, we can only distinguish between low and high father's education, depending on whether the father's qualification is higher than junior high school.

Figs. A3 and A4 in the Appendix show the change in the educational qualification distribution of respondents before and after the Great HE expansion by father's education, and *hukou* status at age 12. From the literature on intergenerational mobility of education, we expect parental education to be a key determinant of the child's education. The figures suggest a different pattern between the urban and the rural *hukou* holders: while in urban areas the relatively more disadvantaged (i.e. with fathers holding no more than junior high school qualifications) benefited more from the HE, the opposite occurs among people with rural origins. In other words, while the Great HE expansion has widened between-group educational inequality across the urban-rural divide, the within-group inequality has narrowed for urban residents but widened for rural residents.

In Table 8 we report the OLS and 2SLS estimates of the returns to schooling by gender for students with low education fathers, high education fathers, as well the large reference group with missing father's education. Co-residence with parents in contemporary China is itself a selective outcome. Indeed, the summary statistics in Table A6 in the Appendix highlight the selective nature of this sub-sample, by comparing key characteristics between the individuals with parental information versus those who have missing parental information. Men who have parental information are generally disadvantaged relative to their counterparts with missing parental

[19] Using 3 decades of census data, Wozniak (2010) shows that US college graduates are several times more responsive in residential location choices to variations in labour market conditions across states than High School graduates.

Table 6
OLS/2SLS estimates of returns to schooling by area type.

	OLS		2SLS	
	Men	Women	Men	Women
	(1)	(2)	(3)	(4)
Panel A: Metropolises and Provincial Capitals				
Second-stage				
Years of schooling	0.069***	0.070***	0.187***	0.133***
	(0.003)	(0.003)	(0.036)	(0.023)
First-stage (partial effects of instruments):				
Post-1980 birth X Urban *hukou*			0.549	1.290**
			(0.525)	(0.548)
Birth year trend X Urban *hukou*			−0.012	−0.033
			(0.044)	(0.050)
Post-1980 birth year trend X Urban *hukou*			−0.069	−0.104*
			(0.050)	(0.055)
Diagnostic Tests:				
Joint significance of IVs: $F_{(3)}$ (*p*-value)			12.862 (0.000)	21.287 (0.000)
Endogeneity of years of schooling: $\chi^2_{(1)}$ (*p*-value)			15.202 (0.000)	8.994 (0.003)
Over-identification: $\chi^2_{(2)}$ (*p*-value)			0.191 (0.909)	0.669 (0.716)
Observations	5060	4561	5060	4561
R^2	0.344	0.412	0.086	0.330
Panel B: Small Cities/towns and rural areas				
Second-stage				
Years of schooling	0.034***	0.056***	0.093**	0.112***
	(0.002)	(0.003)	(0.041)	(0.024)
First-stage (partial effects of instruments):				
Post-1980 birth X Urban *hukou*			1.453***	3.023***
			(0.500)	(0.568)
Birth year trend X Urban *hukou*			0.036	−0.014
			(0.041)	(0.047)
Post-1980 birth year trend X Urban *hukou*			−0.123***	−0.175***
			(0.047)	(0.054)
Diagnostic Tests:				
Joint significance of IVs: $F_{(3)}$ (*p*-value)			6.072 (0.000)	18.097 (0.000)
Endogeneity of years of schooling: $\chi^2_{(1)}$ (*p*-value)			2.249 (0.134)	5.917 (0.015)
Over-identification: $\chi^2_{(2)}$ (*p*-value)			3.109 (0.211)	3.954 (0.139)
Observations	6573	4772	6573	4772
R^2	0.177	0.293		

Note: Same controls as in Table 4, including county/district dummies. Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The baseline (pre-1980) birth year trend is omitted due to perfect multi-collinearity.

information, in terms of education, *hukou* status and earnings. However, they are nearly seven years younger. On the other hand, women who have parental information have very similar earnings and almost two years more education, compared to their counterparts with missing parental information.

Compared to the main analysis in Tables 4 and 5 using the full sample, it is also no longer feasible to control for county/district FEs due to the much smaller sample sizes. Instead, we use a parsimonious specification with dummies for the type of area of residence (tiers of cities) and geographical regions. The upper panel of Table 8 presents the results for men. The OLS coefficients suggest that disadvantaged family background gives a 2.6 percentage points (or 40%) penalty relative to having a more privileged upbringing. The opposite holds for the 2SLS estimates, even though the causal estimates for students with better-educated fathers are statistically insignificant due to the presence of weak instruments. At the same time, the returns for respondents who live independently, and hence have missing parental education information, have the highest returns.

Panel B reports the corresponding results for women. Traditionally a woman lives with her husband's family after marriage in China, and this largely explains the much smaller sample of women with non-missing information on fathers, compared to men. Nevertheless, a similar pattern to men's emerges. Women with a disadvantaged family background appear to have much lower returns to education based on OLS estimates while the corresponding 2SLS estimates are virtually identical. Not living with one's own parents (and hence missing father's education) does not produce different results in the case of 2SLS estimates for women.

To the extent that the results arising from this non-random subsample can be generalized to the wider population, the 2SLS estimates suggest that students who are from a disadvantaged family background stand to gain most from the HE expansion. This is in sharp contrast to the common belief that better-off children have higher returns to schooling. However, one must treat this evidence as tentative, as the external validity cannot be guaranteed in the non-random sample captured by the 2017 CHFS.

Table 7
OLS/2SLS estimates of returns to schooling by geographical regions.

	Eastern		Central & Northeast		Western	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Men						
Second-stage						
Years of schooling	0.060***	0.199***	0.038***	0.100**	0.045***	0.129***
	(0.003)	(0.050)	(0.004)	(0.042)	(0.003)	(0.040)
First-stage (partial effects of instruments):						
Post-1980 birth X Urban *hukou*		0.654		1.964***		1.712**
		(0.495)		(0.646)		(0.734)
Birth year trend X Urban *hukou*		−0.049		0.103*		0.020
		(0.044)		(0.052)		(0.061)
Post-1980 birth year trend X Urban *hukou*		−0.027		−0.206***		−0.149**
		(0.049)		(0.061)		(0.070)
Diagnostic Tests:						
Joint significance of IVs: $F_{(3)}$ (*p*-value)		6.983 (0.000)		6.944 (0.000)		8.241 (0.000)
Endogeneity of years of schooling: $\chi^2_{(1)}$ (*p*-value)		11.867 (0.000)		2.469 (0.116)		5.218 (0.022)
Over-identification: $\chi^2_{(2)}$ (*p*-value)		3.103 (0.212)		2.999 (0.223)		0.332 (0.847)
Observations	5380	5380	3526	3526	2727	2727
R^2	0.370		0.157		0.204	
Panel B: Women						
Second-stage						
Years of schooling	0.070***	0.134***	0.055***	0.122***	0.057***	0.095***
	(0.003)	(0.024)	(0.003)	(0.038)	(0.004)	(0.024)
First-stage (partial effects of instruments):						
Post-1980 birth X Urban *hukou*		1.526***		2.835***		3.850***
		(0.528)		(0.711)		(0.890)
Birth year trend X Urban *hukou*		−0.023		0.041		−0.053
		(0.048)		(0.061)		(0.076)
Post-1980 birth year trend X Urban *hukou*		−0.118**		−0.199***		−0.209**
		(0.053)		(0.068)		(0.086)
Diagnostic Tests:						
Joint significance of IVs: $F_{(3)}$ (*p*-value)		20.925 (0.000)		8.360 (0.000)		15.285 (0.000)
Endogeneity of years of schooling: $\chi^2_{(1)}$ (*p*-value)		8.133 (0.004)		3.662 (0.056)		3.028 (0.022)
Over-identification: $\chi^2_{(2)}$ (*p*-value)		0.307 (0.858)		3.983 (0.137)		1.281 (0.527)
Observations	4615	4615	2639	2639	2079	2079
R^2	0.438		0.267		0.303	

Note: Same controls as in Table 4, including county/district dummies. Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The baseline (pre-1980) birth year trend is omitted due to perfect multi-collinearity.

7. Concluding remarks

Using the large nationally representative 2017 China Household Finance Survey (CHFS), we document that the unanticipated massive HE expansion of 1999–2008 has had a detrimental effect on the already substantial pre-existing urban-rural gap in educational attainment. We build on this evidence to develop a novel identification strategy to estimate the returns to education by gender, by instrumenting the years of schooling with the interaction of childhood urban *hukou* status and the time of the expansion. Our 2SLS results can be interpreted as a Local Average Treatment Effect of education on earnings for urban students who enrolled in HE only because of the Great HE Expansion.

We find returns of 17% and 12% for men and women respectively in the 2SLS. While these are substantially larger than their OLS counterparts of 5% and 6%, both allowing for county fixed-effects, they are not out of line with the existing estimates on returns to HE in China using difference-in-differences or regression discontinuity design (Dai et al., 2021; Wu & Zhao, 2010). Our 2SLS estimates are also consistent with the IV estimate of college premium in excess of 0.41 log points for workers over 30 in Li et al. (2017). The results are robust to additional controls for *hukou* status at birth fully interacted with birth *hukou* province.

We also find evidence of heterogenous returns with regards to the type of residency area and geographical region of residence. After accounting for endogeneity of education, the returns for women are broadly comparable across geographical regions and residence, whereas the returns for men who are currently living in the more developed Eastern Region and/or major cities are significantly higher. One possible interpretation of this asymmetry is that the *hukou* system acts as a severe barrier for geographic mobility, even for college graduates who are expected to be the most responsive to take up opportunities in geographically distant labour markets. For respondents with parental education information, we find that returns for students with less educated fathers are at least as high as

Table 8
OLS/2SLS returns to schooling by father's education.

	Father Low Education		Father High Education		Father Education Missing	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Men						
Second-stage						
Years of schooling	0.038***	0.095***	0.064***	0.040	0.060***	0.206***
	(0.003)	(0.036)	(0.006)	(0.069)	(0.003)	(0.063)
First-stage (partial effects of instruments):						
Post-1980 birth X Urban *hukou*		2.905***		0.183		1.123***
		(0.832)		(1.338)		(0.499)
Birth year trend X Urban *hukou*		0.138		−0.011		0.019
		(0.095)		(0.149)		(0.034)
Post-1980 birth year trend X Urban *hukou*		−0.267***		−0.057		−0.099**
		(0.099)		(0.155)		(0.043)
Diagnostic Tests:						
Joint significance of IVs: $F_{(3)}$ (*p*-value)		8.680 (0.000)		2.883 (0.035)		4.843 (0.002)
Endogeneity of years of schooling: $\chi^2_{(1)}$ (*p*-value)		2.683 (0.101)		0.122 (0.763)		8.900 (0.003)
Over-identification: $\chi^2_{(2)}$ (*p*-value)		0.457 (0.796)		0.869 (0.668)		0.158 (0.924)
Observations	3808	3808	1368	1368	6457	6457
R^2	0.086		0.175		0.264	
Panel B: Women						
Second-stage						
Years of schooling	0.048***	0.138***	0.087***	0.144**	0.069***	0.128***
	(0.005)	(0.034)	(0.008)	(0.059)	(0.002)	(0.031)
First-stage (partial effects of instruments):						
Post-1980 birth X Urban *hukou*		1.557		1.211		1.806***
		(1.497)		(2.096)		(0.464)
Birth year trend X Urban *hukou*		−0.280		0.257		−0.042
		(0.180)		(0.317)		(0.035)
Post-1980 birth year trend X Urban *hukou*		0.045		−0.387		−0.095**
		(0.185)		(0.322)		(0.043)
Diagnostic Tests:						
Joint significance of IVs: $F_{(3)}$ (*p*-value)		12.769 (0.000)		6.228 (0.000)		12.726 (0.000)
Endogeneity of years of schooling: $\chi^2_{(1)}$ (*p*-value)		9.059 (0.003)		0.992 (0.319)		4.382 (0.036)
Over-identification: $\chi^2_{(2)}$ (*p*-value)		1.100 (0.578)		1.468 (0.480)		0.101 (0.951)
Observations	1218	1218	715	715	7400	7400
R^2	0.183		0.353		0.330	

Note: Low/High education indicates up to junior high school and senior high school or above respectively. Same controls as in Table 4 are used, except for city tier and region instead of county/district dummies. Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The baseline (pre-1980) birth year trend is omitted due to perfect multi-collinearity.

their more advantaged counterparts for both genders.

Overall the Great HE Expansion has resulted in higher returns for urban *hukou* holders who would not have had the opportunity to enroll in colleges and universities in the absence of this large HE expansion. The 2SLS estimates suggests that students from disadvantaged family backgrounds experience returns that are at least as high as their more privileged counterparts. The strict rationing of HE places under an elitist central planning model before the expansion had resulted in large excess demand for high skills by the rapidly growing economy, regardless of family background.

However, we also find evidence that rural *hukou* holders generally fail to benefit from the HE expansion, at least in the short-run. Although the thorough understanding of the underlying causes of this likely unintended consequence is beyond the scope of this study, we outline some potential contributing factors:

(1) a widening urban-rural gap in accessing senior high school around the time of HE expansion rate means fewer rural students would be able to benefit from a sudden surge in HE enrollment (Wang, 2009; Zhao et al., 2017).
(2) a rising cost of attending senior high school and HE around the time of the expansion is another important factor to discourage rural students from undertaking HE, as their household income per capita is less than half of their urban counterpart (NBS, various years). For rural students who fail to enter colleges, the returns to senior high school are particularly low, as the *hukou* system limits their job opportunities in the formal sector (Wu, Wei, Zhang, & Zhou, 2019).
(3) rising wages for migrant workers in the urban areas over this period also increase the opportunity cost of staying beyond compulsory education for rural students. Indeed, our analysis indicates significant returns to geographic mobility in China, especially for men.

Due to the rapidly declining fertility rate, new entrants to the labour market will come primarily from the rural *hukou* population (Meng, 2012). Therefore, reducing the urban-rural educational attainment gap at all levels of education is key to China's future economic growth. To the extent that people with rural *hukou* or from less developed areas are held back by inferior school quality and less disadvantaged family background, targeted public investments in primary and secondary education in rural areas are particularly needed to improve social equality and increase productivity.

Declaration of Competing Interest

None.

Acknowledgement

We thank the Survey and Research Centre for China Household Finance at the Southwestern University of Finance and Economics (SWUFE) and the China Institute for Educational Finance Research (CIEFR) Peking University for data access. This work has been supported by the National Social Science Foundation of China in Education (Grant No.: BFA190054). Yu Zhu thanks the UNSW Rector-funded Visiting Fellowship for financial support. We are grateful to seminar participants at the Nanjing University of Finance and Economics of China, CIDE Mexico, University of Birmingham, Queen Mary University of London, the 2019 WPEG Meeting at Sheffield University, the EALE-SOLE-AASLE World Conference 2020 and particularly to Marco Ercolani, Pedro Martins, Alfonso Miranda and Peter Sinclair, for helpful comments. All errors are our own.

Appendix A. Appendix

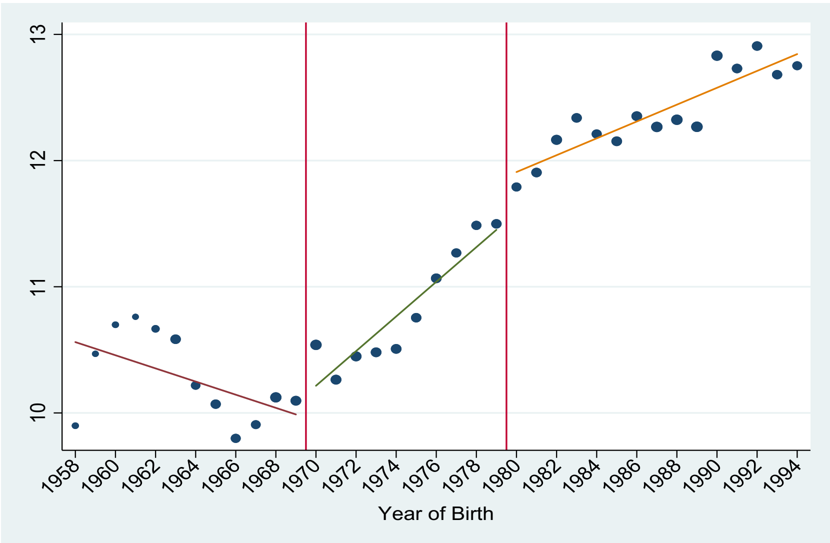

Fig. A1. Years of schoolings of sample members, by year of birth. Note: Sample of men aged 23–59 and women aged 23–54. Marker size proportional to number of observations.

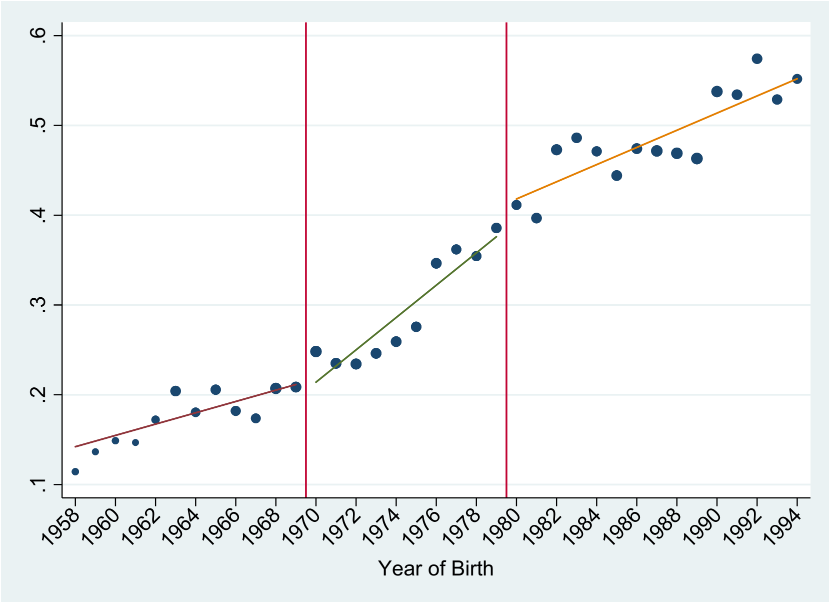

Fig. A2. Proportion of sample members with HE qualifications, by year of birth. Note: Sample of men aged 23–59 and women aged 23–54. Marker size proportional to number of observations.

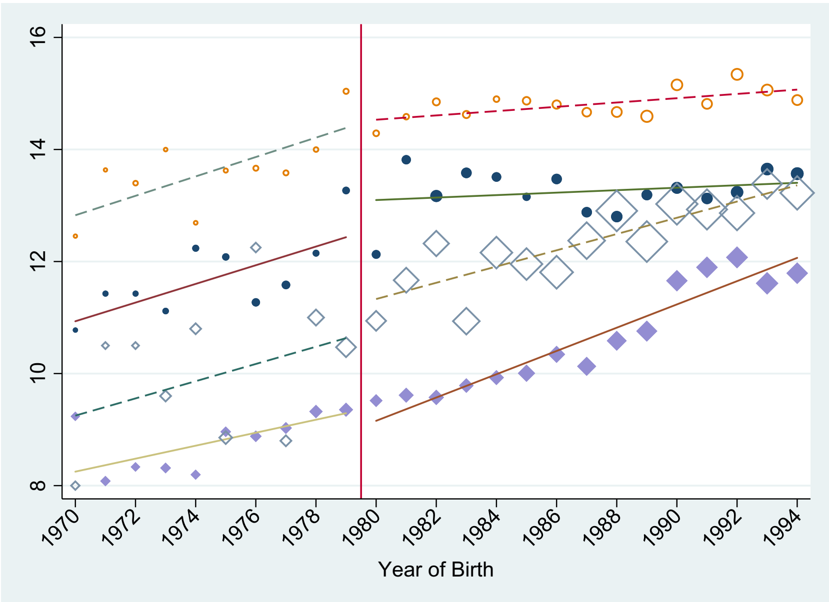

Fig. A3. Years of schooling by father's education and *hukou* status at age 12.

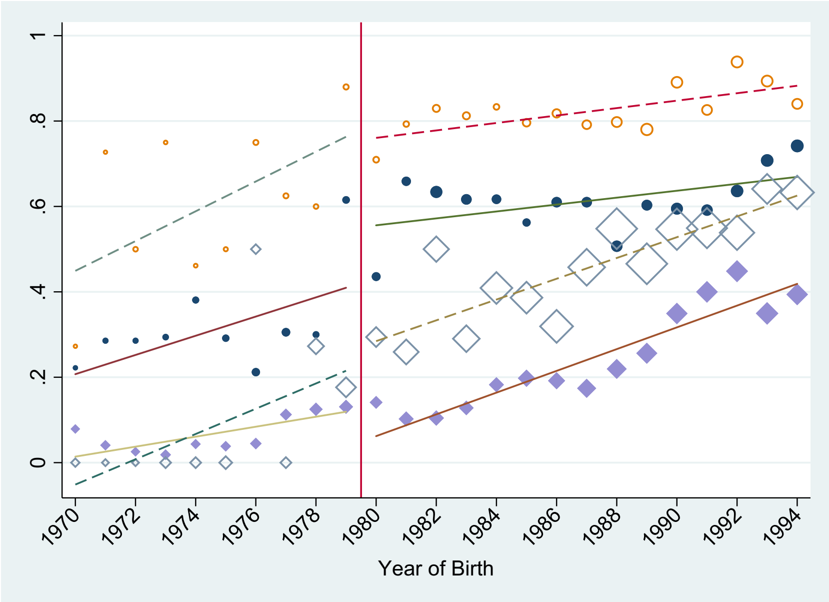

Fig. A4. HE qualifications by father's education and *hukou* status at age 12. Note: Solid/hollow dots denote urban residents whose fathers hold at most junior high and at least senior high school qualifications respectively; solid/hollow diamonds denote rural residents whose fathers hold at most junior high and at least senior high school qualifications respectively. Marker size proportional to number of observations.

Table A1
Summary Statistics by gender, Unweighted.

	Males	Females	Total
Log total monthly net earnings	8.219	7.981	8.113
Years of schooling	11.61	11.94	11.75
Post-1980 birth	0.611	0.593	0.603
Age	35.0	35.2	35.0
Currently rural *hukou*	0.522	0.455	0.492
Rural *hukou* at age 12	0.619	0.570	0.597
Tier 1 cities	0.126	0.149	0.137
Tier 2 cities	0.309	0.339	0.322
Smaller cities and towns	0.333	0.348	0.340
Rural areas	0.232	0.163	0.201
Observations	11,633	9333	20,966

Table A2
OLS estimates of returns to schooling by gender, with county/district fixed-effects.

	Gender-Specific		Pooled-gender		
	Men	Women	No interaction	Int. with schooling only	Int. with schooling & age profiles
	(1)	(2)	(3)	(4)	(5)
Years of schooling	0.049***	0.062***	0.055***	0.055***	0.047***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Age	0.114***	0.083***	0.096***	0.116***	0.116***
	(0.030)	(0.031)	(0.022)	(0.022)	(0.022)
Age squared	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female			−0.280***	0.432**	0.240
			(0.007)	(0.178)	(0.179)
Female x Years of schooling					0.016***
					(0.002)
Female x Age				−0.035***	−0.037***
				(0.010)	(0.010)
Female x Age squared				0.000***	0.000***
				(0.000)	(0.000)
Post-1980 birth	−0.119	−0.118	−0.105	−0.115	−0.111
	(0.107)	(0.108)	(0.076)	(0.076)	(0.076)

(continued on next page)

Table A2 (*continued*)

	Gender-Specific		Pooled-gender		
	Men	Women	No interaction	Int. with schooling only	Int. with schooling & age profiles
	(1)	(2)	(3)	(4)	(5)
Post-1980 birth year trend	0.012	0.014	0.012*	0.013*	0.013*
	(0.009)	(0.010)	(0.007)	(0.007)	(0.007)
Urban *hukou* at age 12	−0.044***	0.002	−0.023***	−0.022**	−0.022**
	(0.012)	(0.012)	(0.009)	(0.009)	(0.009)
County/District FE	✓	✓	✓	✓	✓
Constant	5.274***	5.485***	5.571***	5.194***	5.292***
	(0.655)	(0.672)	(0.470)	(0.477)	(0.475)
F-test of gender-pooling:					
F-statistic				$F_2 = 27.30$	$F_3 = 34.21$
P-value				0.000	0.000
Observations	11,633	9333	20,966	20,966	20,966
R^2	0.304	0.405	0.356	0.358	0.360

Note: 353 county/district dummies. Robust standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The baseline (pre-1980) birth year trend is omitted due to perfect multi-collinearity with age in cross-sectional data.

Table A3
Robustness of 2SLS estimates with respect to specification of age profile.

	Men	Women
	(1)	(2)
Second-stage		
Years of schooling	0.166***	0.124***
	(0.028)	(0.017)
Age dummies	✓	✓
First-stage (partial effects of instruments):		
Post-1980 birth X Urban *hukou*	1.250***	2.333***
	(0.346)	(0.384)
Birth year trend X Urban *hukou*	0.015	−0.014
	(0.030)	(0.033)
Post-1980 birth year trend X Urban *hukou*	−0.110***	−0.156***
	(0.034)	(0.038)
Diagnostic Tests:		
Joint significance of IVs: $F_{(3)}$ (*p*-value)	20.158 (0.000)	41.120 (0.000)
Endogeneity of years of schooling: $\chi^2_{(1)}$ (*p*-value)	24.788 (0.000)	15.730 (0.000)
Over-identification: $\chi^2_{(2)}$ (*p*-value)	0.601 (0.741)	0.851 (0.654)
Observations	11,633	9333

Note: Specification same as in Table 4, but replacing age and age squared with a full set of age dummies. The post-1980 birth dummy and all time trend variables are omitted due to perfect multi-collinearity with age in cross-sectional data.

Table A4
Robustness of 2SLS estimates of returns to schooling, with respect to the 1979 birth cutoff.

	Men	Women
	(1)	(2)
Second-stage		
Years of schooling	0.166***	0.123***
	(0.027)	(0.017)
County/District FE	✓	✓
Joint significance of IVs: $F_{(3)}$ (*p*-value)	20.165 (0.000)	39.974 (0.000)
Observations	11,633	9333

Note: Post-expansion defined as born in 1979 or later (instead of 1980 or later as in Table 4). Robust standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5
Robustness of 2SLS estimates of returns to schooling, with imputed age 12 urban *hukou* status.

	Men	Women
	(1)	(2)
Second-stage		
Years of schooling	0.183***	0.127***
	(0.040)	(0.018)
County/District FE	✓	✓
Joint significance of IVs: $F_{(3)}$ (*p*-value)	11.767 (0.000)	29.987 (0.000)
Observations	11,633	9333
Percentage of imputed urban *hukou* at age 12	9.75	10.65

Note: Unknown or missing urban *hukou* status at age 12 imputed using the Multiple Imputation by Chained Equations (MICE) procedure in Stata (Royston & White, 2011). Bootstrapped robust standard errors with 200 repetitions. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6
Summary statistics by gender and whether parental information is missing.

	Males		Females	
	Non-missing	Missing	Non-missing	Missing
Log total monthly net earnings	8.092	8.320	8.009	7.973
Years of schooling	11.401	11.769	13.406	11.553
Post-1980 birth	0.832	0.434	0.895	0.514
Age	31.1	38.0	28.9	36.8
Currently rural *hukou*	0.625	0.438	0.483	0.447
Rural *hukou* at age 12	0.688	0.563	0.550	0.575
Tier 1 cities	0.088	0.157	0.135	0.153
Tier 2 cities	0.259	0.349	0.315	0.346
Smaller cities and towns	0.300	0.359	0.316	0.356
Rural areas	0.353	0.135	0.234	0.145
Observations	5176	6457	1931	7400

Note: unweighted means.

References

Altbach, P. G., & Balán, J. (Eds.). (2007). *World class worldwide: Transforming research universities in Asia and Latin America*. JHU Press.
Awaworyi, S., & Mishra, V. (2018). Returns to education in China: A meta-analysis. *Applied Economics, 50*(54), 5903–5919.
Bloom, D. E., & Rosovsky, H. (2007). Higher education in developing countries. In *International Handbook of Higher Education* (pp. 443–459).
Blundell, R. W., Dearden, L., Goodman, A., & Reed, H. (2000). The returns to higher education in Britain. *Economic Journal, 110*, F82–F99.
Blundell, R. W., Green, D. A., & Jin, W. (2016). *The UK wage premium puzzle: How did a large increase in university graduates leave the education premium unchanged?* (IFS Working Paper W16/01).
Card, D., & DiNardo, J. E. (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics, 20*(4), 733–783.
Card, D., & Lemieux, T. (2001). Can falling supply explain the rising return to college for young men? *Quarterly Journal of Economics, 116*, 705–746.
Chan, K. W. (2009). The Chinese *Hukou* system at 50. *Eurasian Geography and Economics, 50*(2), 197–221.
Che, Y., & Zhang, L. (2018). Human capital, technology adoption and firm performance: Impacts of China's Higher Education expansion in the late 1990s. *Economic Journal, 128*, 2282–2320.
Cui, Y., Liu, H., & Zhao, L. (2019). Mother's education and child development: Evidence from the compulsory school reform in China. *Journal of Comparative Economics, 47*, 669–692.
Dai, F., Cai, F., & Zhu, Y. (2021). Returns to higher education in China - evidence from the 1999 higher education expansion using fuzzy regression discontinuity. *Applied Economics Letters, 29*(6), 489–494. https://doi.org/10.1080/13504851.2020.1871465 (online first).
Dai, F., Xu, L., & Zhu, Y. (2022). Higher education expansion and supply of teachers in China. *China Economic Review, 71*. https://doi.org/10.1016/j.chieco.2021.101732
Fang, H., Rizzo, J. A., Rozelle, S., & Zeckhauser, R. J. (2016). The returns to education in China: Evidence from the 1986 compulsory education law. In K. Eggleston (Ed.), *Policy challenges from demographic change in China and India*. Stanford University Walter H. Shorenstein Asia-Pacific Research Center series with Brookings Institution.
Goldin, C., & Katz, L. (2008). *The race between education and technology*. Harvard: Harvard University Press.
Guo, Y., Song, Y., & Chen, Q. (2019). Impacts of education policies on intergenerational education mobility in China. *China Economic Review, 55*, 124–142.
Harmon, C., Oosterbeek, H., & Walker, I. (2003). The returns to education: Microeconomics. *Journal of Economic Surveys, 17*(2), 115–155.
Kang, L., Peng, F., & Zhu, Y. (2019). Returns to university subjects and tiers in China - evidence from the China Family Panel Studies. *Studies in Higher Education*. https://doi.org/10.1080/03075079.2019.1698538
Knight, J., Deng, Q., & Li, S. (2017). China's expansion of higher education: The labour market consequences of a supply shock. *China Economic Review, 43*, 127–141.
Li, H., Ma, Y., Meng, L., Qiao, X., & Shi, X. (2017). Skill complementarities and returns to higher education: Evidence from college enrollment expansion in China. *China Economic Review, 46*, 10–26.

Li, S., Whalley, J., & Xing, C. (2014). China's higher education expansion and unemployment of college graduates. *China Economic Review, 30*, 567–582.
Liang, Y., & Dong, Z. (2019). Has education led to secularization? Based on the study on compulsory education las in China. *China Economic Review, 54*, 324–336.
Liu, S., Zhou, S., & Hu, A. (2016). Compulsory education law and economic return to education in urban China: Based on regression discontinuity design. *Economic Research, 2016*(2), 154–167 (in Chinese).
Lu, X. F., Deng, F., & Guo, J. R. (2016). The expansion and inequality in access to higher education in China. *Peking University Education Review, 14*(3), 131–145 (in Chinese).
Meng, X. (2012). Labor market outcomes and reforms in China. *Journal of Economic Perspectives, 26*(4), 75–102.
Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy, 66*(4), 281–302.
MoE. (2018). *Overview of educational achievements in China in 2017*. Beijing: Ministry of Education. http://en.moe.gov.cn/documents/reports/201907/t20190708_389354.html (Accessed 30th March 2022).
NBS. (2017a). *China statistic yearbook 2016*. Beijing: National Bureau of Statistics.
NBS. (2017b). *China educational finance statistical yearbook 2016*. Beijing: National Bureau of Statistics.
NBS. (2018). *China statistic yearbook 2017*. Beijing: National Bureau of Statistics.
Nie, H., & Xing, C. (2019). Education expansion, assortative marriage, and income inequality in China. *China Economic Review, 55*, 37–51.
OECD. (2016). *Education in China – A snap shot*. Paris: OECD.
Ou, D., & Hou, Y. (2019). Bigger pie, bigger slice? The impact of higher education expansion on educational opportunity in China. *Research in Higher Education, 60*(3), 358–391.
Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics, 26*(5), 445–458.
Royston, P., & White, I. R. (2011). Multiple imputation by chained equations (MICE): Implementation in Stata. *Journal of Statistical Software, 45*(4), 1–20.
Tani, M., Xu, L., & Zhu, Y. (2021). The impact of an un(der)funded inclusive education policy: Evidence from the 2013 China education panel survey. *Journal of Economic Behavior & Organization, 190*, 768–784. https://doi.org/10.1016/j.jbusres.2021.06.044
Walker, I., & Zhu, Y. (2008). The college wage premium and the expansion of higher education in the UK. *Scandinavian Journal of Economics, 110*(4), 695–709.
Wan, Y. (2006). Expansion of Chinese higher education since 1998: Its causes and outcomes. *Asia Pacific Education Review, 7*(1), 19–31.
Wang, F., Zhao, L., & Zhao, Z. (2017). China's family planning policies and their labor market consequences. *Journal of Population Economics, 30*(1), 31–68.
Wang, M. (2009). Educational returns and resource allocation between rural and urban areas. *The World Economy, 5*, 3–17 (in Chinese).
Weng, Q., Gao, X., He, H., & Li, S. (2019). Family size, birth order and educational attainment: Evidence from China. *China Economic Review, 57*, Article 101346.
World Bank. (2017). *World development report 2018: Learning to realize education's promise*. The World Bank.
World Bank. (2022). School enrolment, tertiary (% gross) – China (Accessed 30th March 2022) https://data.worldbank.org/indicator/SE.TER.ENRR?locations=CN.
Wozniak, A. (2010). Are college graduates more responsive to distant labor market opportunities? *Journal of Human Resources, 45*(4), 944–970.
Wu, J., Wei, X., Zhang, H., & Zhou, X. (2019). Elite schools, magnet classes, and academic performances: Regression-discontinuity evidence from China. *China Economic Review, 55*, 143–167.
Wu, L., Yan, K., & Zhang, Y. (2020). Higher education expansion and inequality in educational opportunities in China. *Higher Education, 80*(3), 549–570.
Wu, Y., & Zhao, Q. (2010). Higher education expansion and employment of university graduates. *Economic Research, 2010*(9), 93–108 (in Chinese).
Xie, G., & Zhang, L. (2022). Effect of school closure on household labor supply: Evidence from rural China. *China Economic Review, 71*. https://doi.org/10.1016/j.chieco.2021.101737
Xing, C. (2013). Education expansion, migration and rural-urban education gap: A case study on the effect of university expansion. *China Economic Quarterly, 13*(1), 207–232 (in Chinese).
Xing, C., Yang, P., & Li, Z. (2018). The medium-run effect of China's higher education expansion on the unemployment of college graduates. *China Economic Review, 51*, 181–193.
Yan, C. (2011). *Reproduction within different population policy environments in rural China 1979–2000*. Doctoral dissertation. London School of Hygiene & Tropical Medicine.
Zhao, G., Ye, J., Li, Z., & Xue, S. (2017). How and why do Chinese urban students outperform their rural counterparts? *China Economic Review, 45*, 103–123.
Zhu, M. (2014). College admissions in China: A mechanism design perspective. *China Economic Review, 30*, 618–631.