Appendix

A1. CHIP Data Weighting Procedure

To improve the representativeness of the CHIP, we weight it by based on the fraction of urban, rural, and migrant[1] residents in each region from the Chinese yearbook 2013. Our weighting process is summarized as follow: first, we categorize the overall population into 12 strata: metropolitan-urban, metropolitan-rural, metropolitan-migrant, eastern-urban, eastern-rural, eastern-migrant, central-urban, central-rural, central-migrant, western-urban, western-rural, western-migrant[2]. Second, we weight each stratum based on the corresponding population in each stratum. Since there are only 14 provinces in the CHIP, we also weight each province so that the 14 provinces can be jointly representative for their corresponding regions. Thus, the weighting process is divided into two parts: in the first part, it corrects the sample in each stratum of each province; in the second part, it further corrects the sample to be representative at the region level. Thus, the weight for a person *i* living in a province *j* (a province in the 14 provinces in the CHIP) in the stratum *k* can be written as:

$$w_i^{j,k} = \frac{N_j^k}{n_j^k} \times \frac{N^k}{\sum_j N_j^k}$$

[1] The urban residents refer to people who live in their registered address in urban areas; rural residents refer to people live in their registered address in rural areas; migrant group here refers to people who do not live at their official registered address.

[2] The eastern provinces include Hebei, Liaoning, Jilin, Heilongjiang, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong; the central provinces include Shanxi, Anhui, Jiangxi, Hubei, Henan, Hunan; western provinces include Inner Mongolia, Guangxi, Sichuan, Yunnan, Guizhou, Xizang, Shaanxi, Gansu, Ningxia, Qinghai, Xinjiang; and metropolitan areas include Beijing, Shanghai and Chongqing.

Where N_j^k is the population in province *j* of stratum k in the *Yearbook*, n_j^k is the CHIP sample size in province *j* of stratum *k*, N^k is the *Yearbook* population of stratum *k*, and $\sum_j N_j^k$ is the sum of *Yearbook* population in provinces belonging to a stratum *k* in the CHIP. For example, Liaoning, Jiangsu, Shandong and Guangdong are the four eastern provinces in the CHIP whereas people from other eastern provinces such as Zhejiang and Fujian were not included in the CHIP. Thus, the population of the eastern-urban stratum N^k includes all the urban population of the eastern provinces of China, while $\sum_j N_j^k$ includes only the urban population in Liaoning, Jiangsu, Shandong and Guangdong since they are the only eastern provinces in the CHIP.

A2. Summary Statistics for CEE Participation and CEE Scores in CHIP

The CHIP survey asked respondents and their household members whether they had ever taken CEE. If they had, they were also asked what year they took the CEE and what their CEE scores were. However, a large fraction of sample did not answer this question (59% of the sample). Moreover, we are not sure whether the CEE participation rate calculated in our sample can truly represent the actual participation rate. Therefore, we compare some descriptive statistics on CEE participation from the sample with statistics calculated from administrative data and the Chinese census and see if the CHIP is representative in terms of CEE participation.

The first step is to check the reason for the large missing proportion of respondents for the question on whether or not they took the CEE. First, we list the frequency and percentage of the sample who answer “yes”, “no", or those who did not answer this question by their birth cohort. We can see in Table A18 that the missing group is the largest group in each cohort. The missing portion is about 60% of the whole sample. We then restrict the sample to those who had at least 12 years of schooling (at least finishing high school). We find that the missing portion dramatically decreases to below 10%, indicating that most of the respondents who left the question blank did not have high school diplomas. Since a person needs to finish high school courses in order to take the CEE, those who did not answer this question were not able to take CEE. In panel (3), we compare several demographic statistics for people with at least a high school degree. We can see that the average age, average schooling, father's schooling and mother's schooling for the missing group is similar to those who answered “no". They are older, have less schooling, have less father's schooling and less mother's schooling than the “yes" group. This further confirms that the missing group is very likely to be people who did not take the CEE. Therefore, in our later analysis, we recode this missing group as “no".

Next, we compare the CEE participation rate of the 19-year-old population between the CHIP and administrative statistics. Here, we assume most people took the CEE at age 19 (the exact age does not matter if we use the same standard to compare the CHIP and the census). Therefore, the CEE participation rate for “the CHIP sample is calculated by using the number of people who report they took the CEE in the CHIP sample in a certain year divided by the number of 19-year-olds in the CHIP in that year. Similarly, we calculated the CEE participation rate from the census using the number of people who took the CEE in a certain year based on administrative data divided by the estimated number of 19-year-olds in the population in the same year. Here, we use the information from the Chinese census 2000 and an administrative online file from Baidu Wenku, and do the comparison in Figure A6. [3]

[3] “Table 3-1 Number of Population by Age Group and Gender" “Table 6-1 Number of Deaths by Province and Age Group and Gender (1999.11.1-2000.10.31)" (The Fifth Chinese Census, 2010).“Number of People Took CEE and College Acceptance Rate in 1949-2012", Baidu Wenku (in Chinese). UCL: https://wenku.baidu.com/view/e4a5434b2b160b4e777fcf04.html, last visit: May 8, 2018.

In Figure A6 we observe that the percentage of people who took the CEE in the CHIP has a similar pattern to that calculated from the census: the rate is small (<6%) before the restart of the CEE in 1977; and it surges in 1977-1981, during which people delayed by the CR were allowed to take the CEE at any age. Beginning in 1981, only those younger than 25 were allowed to take CEE, so the rate drops dramatically in that year. After 1981, the rate increases over time. Although the trend calculated from the CHIP under-represents the CEE rate in 1977-1979, 2003-2008 and over-represents the rate in 1993-2000, the general pattern is very similar to that of the census. The over-representation of the rate in the CHIP in these periods could be related to the over-representation of more educated respondents in the CHIP relative to the population, but we do not have a good explanation for the under-representation.

We also calculate the percentage of the CEE takers by their birth year using the CHIP (the horizontal axis represents the year when they were 19 years old). This series (the dotted line in Figure A6) shows that the percentage of people taking the CEE increased for those who were born after 1954 (aged 19 in 1973). We observed that the number of people who were born in 1945-1957 (aged 19 in 1964-1976) and took the CEE is greater than number of people who took the test during 1964-1976, suggesting that most people born in 1945-1957 took the exam after the CR (since the exam was suspended in 1966-1976), most likely in 1977-1981, when no age restriction applied. Also, the number of people who took the exam in 1977-1981 is much greater than the number of people who were 19 years old in that period and took the test in any period, indicating most CEE takers in this period were from previous cohorts (older than 19 years old). After 1985, the two trends become very similar, suggesting that the majority of CEE takers are those that are 19-year-olds after 1985.

Since the statistics of the CEE participation in the CHIP is similar to the population, we can thus use the sample to do the analysis on the mechanism linking parental schooling with children's college degree attainment.

Table A1 Estimated Intergenerational Educational Coefficients by Birth Cohort, Replicating Chen et al. (2015): Years of Schooling

Years of Schooling	**(1)**	**(2)**	**(3)**
1925*FS	0.660***	0.795***	0.795***
	(0.089)	(0.115)	(0.120)
1930*FS	0.422***	0.526***	0.523***
	(0.080)	(0.071)	(0.075)
1935*FS	0.537***	0.510***	0.494***
	(0.056)	(0.058)	(0.057)
1940*FS	0.449***	0.431***	0.422***
	(0.037)	(0.038)	(0.037)
1945*FS	0.389***	0.382***	0.373***
	(0.028)	(0.028)	(0.029)
1950*FS	0.334***	0.335***	0.322***
	(0.020)	(0.020)	(0.020)
1955*FS	0.303***	0.296***	0.275***
	(0.019)	(0.019)	(0.019)
1960*FS	0.282***	0.285***	0.270***
	(0.014)	(0.014)	(0.013)
1965*FS	0.309***	0.307***	0.285***
	(0.013)	(0.014)	(0.013)
1970*FS	0.341***	0.340***	0.316***
	(0.016)	(0.016)	(0.016)
1975*FS	0.404***	0.393***	0.369***
	(0.021)	(0.022)	(0.020)
1980*FS	0.361***	0.349***	0.319***
	(0.029)	(0.030)	(0.030)
1985*FS	0.474***	0.490***	0.466***
	(0.045)	(0.050)	(0.050)
Cohort FE	Yes	Yes	Yes
Baseline Control	No	Yes	Yes
Extended Controls	No	No	Yes
N	32,976	32,976	32,976
$\mathbf{R^2}$	0.2867	0.3192	0.3387

Note: The table reports the estimated intergenerational coefficients based on Chen et al. (2015) estimating the equation: $EDU_{i,t} = \sum_{t=1}^{13} \beta_t FS_i + \gamma X_i + \delta_t + \varepsilon_{i,t}$. $EDU_{i,t}$ is years of schooling of a person i born in cohort t, which is one of the five-year cohorts from 1925 to 1989 (1925-1929, …, 1985-1989, 13 cohorts in total). In the table, FS_i measures the father's years of schooling. X_i includes cohort-specific effects of gender, cohort-specific effects of father's age, father's age squared, age, and age squared. In the extensive model's specification, it also includes the cohort-specific effects of living in a coastal province. t is a full set of 5-year birth cohort dummies from 1925 to 1989. Standard errors are clustered at child's residential province level.

Table A2 Estimated Intergenerational Educational Coefficients by Birth Cohort, Replicating Chen et al. (2015): Rank of Schooling

Years of Schooling	**(1)**	**(2)**	**(3)**
1925*FSR	0.570***	0.680***	0.678***
	(0.113)	(0.110)	(0.113)
1930*FSR	0.410***	0.497***	0.498***
	(0.075)	(0.070)	(0.075)
1935*FSR	0.456***	0.425***	0.494***
	(0.056)	(0.058)	(0.075)
1940*FSR	0.434***	0.414***	0.402***
	(0.038)	(0.039)	(0.038)
1945*FSR	0.414***	0.408***	0.396***
	(0.030)	(0.030)	(0.030)
1950*FSR	0.378***	0.377***	0.361***
	(0.022)	(0.022)	(0.022)
1955*FSR	0.350***	0.337***	0.310***
	(0.022)	(0.022)	(0.022)
1960*FSR	0.365***	0.375***	0.351***
	(0.018)	(0.018)	(0.018)
1965*FSR	0.395***	0.394***	0.362***
	(0.017)	(0.018)	(0.018)
1970*FSR	0.408***	0.406***	0.375***
	(0.019)	(0.020)	(0.019)
1975*FSR	0.452***	0.436***	0.409***
	(0.024)	(0.025)	(0.024)
1980*FSR	0.382***	0.370***	0.337***
	(0.029)	(0.030)	(0.030)
1985*FSR	0.501***	0.509***	0.484***
	(0.050)	(0.053)	(0.050)
Cohort FE	Yes	Yes	Yes
Baseline Control	No	Yes	Yes
Extended Controls	No	No	Yes
N	32,976	32,976	32,976
$\mathbf{R^2}$	0.1523	0.1870	0.2108

Note: The table reports the estimated intergenerational coefficients based on Chen et al. (2015) estimating the equation:$EDU_{i,t} = \sum_{t=1}^{13} \beta_t \, FSR_i + \gamma X_i + \delta_t + \varepsilon_{i,t}$. $EDU_{i,t}$ is rank of schooling of a person *i* born in cohort *t*, which is one of the five-year cohorts from 1925 to 1989 (1925-1929, …, 1985-1989, 13 cohorts in total). The rank is the relative position of a father's schooling among all fathers in the same birth cohort *t*, and it ranges between 0 to 1 (1 indicates the highest possible father's schooling in that cohort). In the table, FSR_i the rank of father's schooling. X_i includes cohort-specific effects of gender, cohort-specific effects of father's age, father's age squared, age, and age squared. In the extensive model's specification, it also includes the cohort-specific effects of living in a coastal province. *t* is a full set of 5-year birth cohort dummies from 1925 to 1989. Standard errors are clustered at child's residential province level.

Table A3 Estimated Intergenerational Educational Correlations by Birth Cohort, Replicating Chen et al. (2015): Years of Schooling

Years of Schooling	**(1)**	**(2)**	**(3)**
1925*FS	0.519***	0.626***	0.626***
	(0.055)	(0.073)	(0.075)
1930*FS	0.288***	0.360***	0.361***
	(0.037)	(0.033)	(0.035)
1935*FS	0.412***	0.391***	0.378***
	(0.033)	(0.034)	(0.034)
1940*FS	0.369***	0.355***	0.348***
	(0.025)	(0.026)	(0.025)
1945*FS	0.361***	0.354***	0.346***
	(0.024)	(0.025)	(0.025)
1950*FS	0.330***	0.331***	0.318***
	(0.019)	(0.019)	(0.019)
1955*FS	0.314***	0.306***	0.340***
	(0.020)	(0.021)	(0.020)
1960*FS	0.355***	0.360***	0.340***
	(0.022)	(0.022)	(0.021)
1965*FS	0.384***	0.383***	0.354***
	(0.021)	(0.021)	(0.021)
1970*FS	0.407***	0.405***	0.377***
	(0.023)	(0.023)	(0.023)
1975*FS	0.442***	0.430***	0.404***
	(0.025)	(0.026)	(0.024)
1980*FS	0.384***	0.371***	0.339***
	(0.032)	(0.034)	(0.034)
1985*FS	0.492***	0.508***	0.483***
	(0.048)	(0.053)	(0.054)
Cohort FE	Yes	Yes	Yes
Baseline Control	No	Yes	Yes
Extended Controls	No	No	Yes
N	32,976	32,976	32,976
R^2	0.2867	0.3192	0.3387

Note: The table reports the estimated intergenerational correlations ($\rho = \beta \frac{\sigma_1}{\sigma_0}$, where β is the estimated intergenerational coefficient, and σ_0, $\sigma_1$are standard deviations of father's schooling, child's schooling, in each of birth cohort *t*, respectively), based on Chen et al. (2015). $EDU_{i,t}$ is years of schooling of a person *i* born in cohort *t*, which is one of the five-year cohorts from 1925 to 1989 (1925-1929, …, 1985-1989, 13 cohorts in total). In the table, FSR_i measures the rank of father's schooling.X_i includes cohort-specific effects of gender, cohort-specific effects of father's age, father's age squared, age, and age squared. In the extensive model's specification, it also includes the cohort-specific effects of living in a coastal province. *t* is a full set of 5-year birth cohort dummies from 1925 to 1989. Standard errors are clustered at child's residential province level.

Table A4 Population Distribution by Region in the CHIP Sample and the Chinese Yearbook 2013

	CHIP (1)	YB13 (2)	WCHIP-YB13 (3)
Panel A: Total Sample			
Metropolitan	5.1	3.3	0.0
Eastern	31.2	43.0	0.0
Central	37.7	26.7	0.0
Western	25.9	27.0	0.0
Panel B: Urban Sample			
Metropolitan	9.6	5.4	0.0
Eastern	31.5	47.8	0.0
Central	34.6	23.8	0.0
Western	24.3	22.9	0.0
Panel C: Rural Sample			
Metropolitan	3.1	0.9	0.0
Eastern	31.0	37.2	0.0
Central	39.4	30.0	0.0
Western	26.5	31.9	0.0
Panel D: Urbanization Rate			
Metropolitan	58.2	88.1	0.0
Eastern	31.5	60.4	0.0
Central	28.5	48.5	0.0
Western	29.4	46.0	0.0

Note: Panel A shows the regional distribution of the CHIP sample (column (1)), the distribution in the Chinese Yearbook 2013 (column (2)), and the differences in regional distribution between the weighted CHIP sample and the yearbook (column (3)). Panel B shows regional distribution for urban residents in the CHIP sample and the Chinese Yearbook, and the difference between weighted CHIP and the yearbook; Panel C shows the regional distribution for rural residents in the CHIP sample and the Chinese Yearbook, and the difference between weighted CHIP and the yearbook; Panel D shows the urbanization rates in each region in the CHIP and the yearbook, and the difference between the weighted CHIP and the yearbook. Column (3) is 0 by construction of the weights.

Table A5 Household Distribution and Average Household Size of the 14 CHIP Provinces in CHIP Sample and the Chinese Yearbook 2013

Province	% of Households			Avg. HH. Size			Diff in HH. Size			
	CHIP (1)	YB13 (2)	WCHIP-YB13 (3)	CHIP (4)	YB13 (5)	WCHIP-YB13 (6)	1 per (7)	2 pers (8)	3 pers (9)	4 pers (10)
Panel A: Eastern Provinces										
Beijing	6.4	2.7	2.3	3.1	2.6	0.4	-11.6	3.9	11.2	-3.6
Jiangsu	8.3	9.5	0.5	4.0	2.9	1.0	-8.9	-4.4	4.0	9.3
Shandong	8.5	12.4	1.9	3.5	2.8	0.6	-8.5	-5.4	7.4	6.6
Guangdong	8.6	11.7	0.7	4.7	3.2	1.1	-14.9	-11.5	4.5	21.9
Liaoning	6.2	5.7	1.4	3.2	2.7	0.4	-6.6	5.4	0.2	1.0
Panel B: Central Provinces										
Shanxi	7.1	4.3	-1.2	3.6	3.0	0.6	-5.7	-1.5	4.1	3.0
Anhui	6.7	7.0	-2.0	4.0	3.0	0.9	-7.3	-8.7	3.8	12.1
Henan	9.0	10.2	-2.7	4.2	3.3	0.8	-6.8	-7.0	2.2	11.6
Hubei	7.2	6.9	-2.0	4.1	2.9	1.0	-6.0	-9.8	0.9	14.9
Hunan	7.1	7.3	-1.8	4.1	3.3	0.7	-10.7	-1.7	5.4	7.0
Panel C: Western Provinces										
Chongqing	5.5	3.9	0.4	3.8	2.7	1.0	-15.9	-5.7	9.3	12.3
Sichuan	7.6	10.5	1.5	3.8	2.8	1.0	-13.3	-5.2	5.8	12.6
Yunnan	6.2	5.1	0.9	4.3	3.3	0.8	-9.6	-5.3	3.3	11.7
Gansu	5.6	2.7	0.4	4.5	3.4	0.9	-7.7	-9.6	0.0	17.2

Note: Columns (1) to (3) show the proportions of household in each province in the CHIP, in the yearbook, and the difference between weighted CHIP and the yearbook; Columns (4) to (6) show the average household size in each province of the CHIP sample, in the yearbook, and the difference between weighted CHIP and the yearbook; Columns (7) to (10) show the differences of the sample distribution by household size (1 person, 2 persons, 3 persons, 4+ persons, respectively) between the CHIP and the yearbook.

Table A6 Demographic Comparison between CHIP Sample and Chinese Population by Province

Province	% of Male			% of Same Hukou			Diff in HH. Size		
	CHIP (1)	YB13 (2)	WCHIP-YB13 (3)	CHIP (4)	YB13 (5)	WCHIP-YB13 (6)	CHIP (7)	YB13 (8)	WCHIP-YB13 (9)
Panel A: Eastern Provinces									
Beijing	48.9	51.8	-5.0	79.3	45.6	3.0	80.5	81.5	-1.5
Jiangsu	51.3	50.0	-1.3	91.4	75.9	0.8	74.4	74.5	0.0
Shandong	50.7	50.9	-0.4	92.0	84.7	0.9	76.9	73.6	3.4
Guangdong	52.1	52.7	-1.6	88.6	63.1	1.4	74.1	76.0	-2.0
Liaoning	50.1	50.5	-1.3	85.1	78.1	0.9	75.3	79.4	-4,3
Panel B: Central Provinces									
Shanxi	52.0	51.4	0.7	82.3	89.1	-4.1	77.5	76.3	1.3
Anhui	52.4	51.0	-0.2	93.1	87.6	2.4	74.7	71.0	4.0
Henan	52.2	50.3	1.7	92.2	94.4	-0.6	74.2	70.3	3.9
Hubei	51.6	50.8	0.5	88.6	83.5	2.1	74.2	75.2	-0.3
Hunan	51.7	51.4	0.8	90.6	84.5	6.8	74.4	71.1	3.5
Panel C: Western Provinces									
Chongqing	50.6	51.0	-0.9	87.8	78.9	4.0	73.4	71.2	2.5
Sichuan	51.4	51.0	-0.5	92.6	87.6	-0.5	73.7	70.7	3.8
Yunnan	50.8	51.8	-1.2	93.4	86.3	1.6	72.2	72.0	1.2
Gansu	52.2	51.4	0.6	95.3	82.6	6.1	76.3	74.2	3.4

Note: Columns (1) to (3) show the fraction of male in each province of the CHIP, the fraction in the yearbook, and the difference between the weighted CHIP and the yearbook statistics; Columns (4) to (6) show the fraction of non-migrant[4] sample (more detail see footnote 2) in each province of the CHIP, the fraction in the yearbook, and the difference between weighted CHIP and the yearbook; Columns (7) to (9) show the fraction of 15-64 year-olds in each province of the CHIP, the fraction in the yearbook, and the difference between weighted CHIP and the yearbook

[4] Same Hukou is defined as that the registered resident address for a person is in the same county as the address he/she currently lives in. This is the common way to classify the migration status for a person in China

Table A7 Demographic Comparison between CHIP Sample and Chinese Population by Province

Age Group	**Educational Level**						
	None	Primary	Junior HS	Senior HS	Associate	Undergraduate	Graduate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
25-29	-0.2	-1.7	-8.7	3.7	2.8	3.6	0.4
	(-0.3)	(-3.0)	(-12.7)	(4.4)	(3.8)	(6.9)	(0.8)
30-34	-0.2	-1.9	-7.0	2.6	2.2	3.8	0.5
	(-0.3)	(-3.1)	(-12.4)	(4.4)	(4.2)	(6.3)	(0.8)
35-39	0.1	-1.5	-8.1	3.9	0.9	3.9	0.9
	(-0.3)	(-4.4)	(-12.5)	(6.3)	(3.1)	(6.7)	(1.1)
40-44	1.2	-1.0	-5.9	2.3	1.1	1.9	0.3
	(0.5)	(-4.1)	(-8.5)	(4.7)	(2.7)	(4.0)	(0.1)
45-49	1.2	0.7	-5.3	1.4	0.7	1.4	0.1
	(0.3)	(-1.8)	(-6.8)	(3.6)	(2.2)	(2.5)	(0.1)
50-54	1.5	-4.2	-1.4	2.6	0.9	0.6	0.1
	(0.6)	(-6.3)	(-2.4)	(4.5)	(2.1)	(1.5)	(0.1)
55-59	4.0	-8.9	-1.2	6.0	-0.1	0.1	0.1
	(2.8)	(-10.2)	(-1.3)	(7.3)	(0.9)	(0.5)	(0.1)
60-64	5.4	-8.4	-1.0	3.1	0.6	0.1	0.1
	(4.1)	(-10.8)	(-0.2)	(5.0)	(1.5)	(0.3)	(0.1)
65-69	5.0	-6.1	-1.4	1.6	0.3	0.6	0.0
	(3.2)	(-7.9)	(-1.2)	(3.9)	(0.8)	(1.2)	(0.0)
70-74	2.7	-8.6	0.9	3.9	0.3	0.9	0.0
	(-1.5)	(-10.3)	(2.9)	(6.5)	(0.9)	(1.6)	(0.0)
75-79	4.8	-11.3	1.7	3.5	0.3	1.0	0.0
	(3.0)	(-13.9)	(3.1)	(5.3)	(0.8)	(1.7)	(0.0)
80-84	11.7	-15.9	1.5	1.7	-0.1	1.0	0.1
	(10.0)	(-15.9)	(2.5)	(2.1)	(-0.2)	(1.4)	(0.1)
>=85	14.4	-15.8	-2.1	1.9	0.6	0.6	0.3
	(11.0)	(-15.0)	(-2.2)	(4.1)	(0.9)	(0.9)	(0.2)

Note: The table shows the differences (in percentage points) of the educational distributions between the CHIP (before and after weighting) and the Census 2010 by age group. The numbers in parentheses are the differences between the weighted CHIP and the Census.

Table A8 Robustness Check: Estimated Coefficients of Equation (2), Dropping Elder Sample

	High School Degree		College Degree	
	(1)	(2)	(3)	(4)
Ftype1*YA	-0.019**	-0.018**	-0.018*	-0.015
	(0.007)	(0.008)	(0.009)	(0.011)
Ftype2*YA	-0.001	-0.002	0.003	-0.004
	(0.004)	(0.008)	(0.005)	(0.007)
Ftype3*YA	-0.003	-0.007*	-0.001	-0.004
	(0.002)	(0.003)	(0.003)	(0.003)
YA	-0.016**	-0.014***	-0.00005	-0.002
	(0.005)	(0.003)	(0.003)	(0.002)
Ftype1	0.392***	0.390***	0.312***	0.310***
	(0.014)	(0.014)	(0.024)	(0.024)
Ftype2	0.247***	0.248***	0.137***	0.141***
	(0.019)	(0.019)	(0.018)	(0.018)
Ftype3	0.113***	0.115***	0.041***	0.042***
	(0.015)	(0.015)	(0.008)	(0.007)
Measure	Meng	Zhou	Meng	Zhou
R^2	0.3009	0.3013	0.2485	0.2482
N	30,877	30,877	30,877	30,877

Note: This table lists the regression results for obtaining high school and college degree based on equation (1), restricting the sample to people younger than 67 years old. Controls include child's birth cohort dummies, gender, age and age squared, father's age and father's age squared; the standard errors are clustered at child's residential province level.

Table A9 Estimated Marginal Effects of Equation (1), Using Logit and Probit Models

	High School (Logit)		**College Degree (Logit)**		**High School (Probit)**		**College Degree (Probit)**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ftype1*YA	-0.026***	-0.021***	-0.003	-0.001	-0.027***	-0.023***	-0.007*	-0.005
	(0.005)	(0.008)	(0.004)	(0.005)	(0.006)	(0.008)	(0.004)	(0.005)
Ftype2*YA	-0.001	0.002	0.008**	0.006	-0.001	0.001	0.005	0.004
	(0.004)	(0.006)	(0.003)	(0.004)	(0.004)	(0.006)	(0.003)	(0.003)
Ftype3*YA	0.001	0.002	0.002	-0.001	0.001	0.001	0.0004	-0.002
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)
YA	0.002	-0.007**	-0.003	-0.003	0.002	-0.006**	-0.002	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Ftype1	0.472***	0.464***	0.246***	0.246***	0.485***	0.478***	0.252***	0.251***
	(0.023)	(0.023)	(0.021)	(0.020)	(0.022)	(0.022)	(0.023)	(0.022)
Ftype2	0.280***	0.278***	0.155***	0.158***	0.285***	0.284***	0.149***	0.151***
	(0.018)	(0.018)	(0.015)	(0.015)	(0.022)	(0.019)	(0.016)	(0.015)
Ftype3	0.141***	0.139***	0.085***	0.087***	0.138***	0.137***	0.077***	0.077***
	(0.009)	(0.009)	(0.006)	(0.006)	(0.010)	(0.010)	(0.006)	(0.006)
Measure	Meng	Zhou	Meng	Zhou	Meng	Zhou	Meng	Zhou
N	32,976	32,976	32,976	32,976	32,976	32,976	32,976	32,976
R^2	0.1410	0.1409	0.1904	0.1903	0.1414	0.1413	0.1847	0.1907

Note: This table lists the regression results on obtaining high school and college degree based on equation (2), using Logit (Columns (1)-(4)) and Probit (Columns (5)-(8)) models. Controls include child's birth cohort dummies, gender, age and age squared, father's age and father's age squared (results were not shown in this table); the standard errors are clustered at child's residential province level.

Table A10 Estimated Results of Equation (1), Using Years of Schooling as Measure of Father's Education

	High School Degree		**College Degree**	
	(1)	(2)	(3)	(4)
FS*YA	-0.0013***	-0.0016***	-0.0014***	-0.0017***
	(0.0003)	(0.0005)	(0.0005)	(0.0006)
YA	0.014**	-0.012***	0.003	0.002
	(0.006)	(0.004)	(0.003)	(0.002)
FS	0.028***	0.028***	0.020***	0.020***
	(0.001)	(0.001)	(0.002)	(0.002)
Measure	Meng	Zhou	Meng	Zhou
N	32,976	32,976	32,976	32,976
R^2	0.2990	0.2993	0.2354	0.2355

Note: This table lists the regression results predicting high school and college degree attainment based on equation (1), using years of schooling as the measure of father's education. Controls include child's birth cohort, gender, age and age squared, father's age and father's age squared (results were not shown in this table); the standard errors are clustered at child's residential province level.

Table A11 Estimated Results of Equation (1), Using Sibling Sample

	High School Degree		College Degree	
	(1)	(2)	(3)	(4)
Ftype1*YA	-0.022***	-0.030***	-0.021***	-0.024***
	(0.005)	(0.006)	(0.004)	(0.005)
Ftype2*YA	-0.002	-0.007	-0.001	-0.005*
	(0.004)	(0.004)	(0.002)	(0.003)
Ftype3*YA	-0.001	-0.003*	-0.001	-0.002**
	(0.002)	(0.002)	(0.001)	(0.001)
YA	-0.003	-0.004**	0.001	0.001
	(0.005)	(0.002)	(0.001)	(0.001)
Ftype1	0.328***	0.329***	0.239***	0.236***
	(0.013)	(0.013)	(0.011)	(0.011)
Ftype2	0.168***	0.171***	0.075***	0.077***
	(0.010)	(0.009)	(0.006)	(0.006)
Ftype3	0.069***	0.072***	0.021***	0.021***
	(0.006)	(0.006)	(0.003)	(0.003)
Measure	Meng	Zhou	Meng	Zhou
R^2	0.1950	0.1953	0.1644	0.1643
N	54,113	54,113	54,113	54,113

Note: ***stands for significance level <0.001; **stands for significance level <0.005; *stands for significance level <0.01. This table reports the regression results for obtaining high school and college degrees based on equation (1) using sibling sample. Ftype1 denotes whether a person has a father with high school degree; Ftype2 denotes whether father has middle school diploma; Ftype3 denotes whether father has primary school diploma. YA is the measure of years of schooling affected by the CR, using either Meng's measure or Zhou's measure. Control variables include child's birth cohort dummies, gender, age and age squared, father's age and father's age squared (results were not shown in this table); The standard errors were clustered at the child's residential province level.

Table A12 Estimated Results of Equation (1), Adding Family Fixed Effects (Sibling Sample)

	High School Degree		College Degree	
	(1)	(2)	(3)	(4)
Ftype1*YA	-0.014*	-0.022**	-0.014***	-0.016***
	(0.007)	(0.010)	(0.004)	(0.005)
Ftype2*YA	-0.006	-0.007	-0.001	-0.004*
	(0.006)	(0.008)	(0.002)	(0.002)
Ftype3*YA	-0.0002	-0.001	-0.001	-0.002*
	(0.003)	(0.004)	(0.001)	(0.001)
YA	-0.0003	-0.004	0.002	0.002*
	(0.004)	(0.003)	(0.001)	(0.001)
Measure	Meng	Zhou	Meng	Zhou
R^2	0.0309	0.3012	0.0245	0.0246
N	36,361	36,361	36,361	36,361

Note: ***stands for significance level <0.001; **stands for significance level <0.005; *stands for significance level <0.01. This table lists the regression results for obtaining high school and college degree based on equation (1), adding family fixed effect dummies into the model, using sibling data. Controls include child's birth cohort dummies, gender, age and age squared, father's age and father's age squared; the standard errors are clustered at child's household level.

Table A13 Estimated Results of Equation (1), Changing School Entry Age

	High School Degree		College Degree	
	Entry Age: 6	Entry Age: 8	Entry Age: 6	Entry Age: 8
Ftype1*YA	-0.021***	-0.019**	-0.022**	-0.016*
	(0.007)	(0.007)	(0.008)	(0.009)
Ftype2*YA	0.001	-0.003	0.006	-0.001
	(0.004)	(0.006)	(0.005)	(0.005)
Ftype3*YA	-0.001	-0.006**	-0.001	-0.003
	(0.002)	(0.002)	(0.002)	(0.003)
YA	-0.003	-0.023***	-0.002	0.0001
	(0.008)	(0.005)	(0.003)	(0.002)
Ftype1	0.394***	0.396***	0.314***	0.311***
	(0.013)	(0.014)	(0.022)	(0.024)
Ftype2	0.250***	0.254***	0.136***	0.141***
	(0.017)	(0.017)	(0.016)	(0.017)
Ftype3	0.112***	0.119***	0.042***	0.045***
	(0.013)	(0.014)	(0.007)	(0.008)
Measure	Meng	Meng	Meng	Meng
R^2	0.3017	0.3024	0.2456	0.2455
N	32,976	32,976	32,976	32,976

Note: ***stands for significance level <0.001; **stands for significance level <0.005; *stands for significance level <0.01. This table reports the regression results for obtaining high school and college degrees based on equation (1), changing the entry school age in Meng's measure. Ftype1 denotes whether a person has a father with high school degree; Ftype2 denotes whether father has middle school diploma; Ftype3 denotes whether father has primary school diploma. YA is the measure of years of schooling affected by the CR, using either Meng's measure or Zhou's measure. Control variables include child's birth cohort dummies, gender, age and age squared, father's age and father's age squared (results were not shown in this table); The standard errors were clustered at the child's residential province level.

Table A14 Estimated Coefficients of the Impact of father's schooling and the Impact of the CR on High School and College Attainments Using All Sample, Urban Sample, and Sample of Urban Residents Who Have Children in the CHIP

	Full Sample (1)	**Urban Sample (2)**	**Urban with Child (3)**
Panel A: High School Degree			
FS*CR	-0.009***	-0.003	-0.007*
	(0.002)	(0.002)	(0.004)
CR	0.142***	0.126***	0.517
	(0.021)	(0.037)	(0.319)
FS	0.040***	0.036***	0.039***
	(0.002)	(0.002)	(0.004)
N	32,976	10,787	6,182
	0.1587	0.1838	0.1365
Panel B: College Degree			
FS*CR	-0.012***	-0.010***	-0.008***
	(0.002)	(0.002)	(0.003)
CR	0.013	-0.019	0.103
	(0.016)	(0.034)	(0.193)
FS	0.027***	0.033***	0.025***
	(0.002)	(0.002)	(0.003)
N	32,976	10,787	6,182
R^2	0.1456	0.1814	0.0910

Note: This table reports the regression results of the effect of father's schooling on the high school and college degree attainment for the full sample, the urban sample, and the urban sample who had children in the CHIP. FS denotes years of father's schooling. CR is a dummy for whether the person received his/her prime education in the CR. The regression also controls for the age, age squared, gender, father's age and father's age squared. Standard errors were clustered at child's residential province level.

Table A15 Occupation as the Measure of Father's Status: Estimated Coefficients of the Effect of Affected Years in the CR on High School and College Degree Attainments

	High School Degree		College Degree	
	(1)	(2)	(3)	(4)
Ftype1*YA	-0.014***	-0.026***	-0.024***	-0.035***
	(0.007)	(0.008)	(0.005)	(0.008)
Ftype2*YA	-0.014***	-0.018***	-0.019***	-0.020***
	(0.006)	(0.007)	(0.005)	(0.006)
Ftype3*YA	-0.010***	-0.015**	-0.003	-0.005*
	(0.003)	(0.004)	(0.002)	(0.003)
YA	0.001	-0.008***	0.004***	0.003*
	(0.002)	(0.003)	(0.001)	(0.002)
Ftype1	0.479***	0.488***	0.350***	0.348***
	(0.014)	(0.013)	(0.016)	(0.016)
Ftype2	0.416***	0.426***	0.350***	0.255***
	(0.014)	(0.013)	(0.016)	(0.014)
Ftype3	0.121***	0.128***	0.045***	0.052***
	(0.012)	(0.011)	(0.009)	(0.008)
Measure	Meng	Zhou	Meng	Zhou
R^2	0.1902	0.1775	0.1673	0.1468
N	31,356	31,356	31,356	31,356

Note: This table reports the regression results for obtaining high school and college degrees based on equation (2), using father's occupation as their types. Ftype1 denotes father's occupation as principal, manager, technician or professional; Ftype2 includes clerk, commercial and service personnel; Ftype3 are farmers or manufacturing related personnel. Controls include child's birth cohort dummies, gender, age and age squared, father's age and father's age squared (results were not shown in this table); the standard errors are clustered at the child's residential province level.

Table A16 Occupation as Measure of Father's Status: Estimated Coefficients of the Impact of the CR on Multi-Generational Mobility for High School and College Degree Attainment

	High School			**College**		
	(1)	(2)	(3)	(4)	(5)	(6)
GFtype1*YA	-0.004	-	-0.004	0.009	-	0.007
	(0.011)		(0.012)	(0.018)		(0.021)
GFtype2*YA	-0.020**	-	-0.020**	-0.013**	-	-0.015**
	(0.007)		(0.007)	(0.004)		(0.006)
GFtype3*YA	0.011	-	0.005	0.007	-	0.002
	(0.044)		(0.011)	(0.012)		(0.011)
Ftype1*YA	-	0.004	0.017	-	0.006	0.016
		(0.009)	(0.009)		(0.011)	(0.011)
Ftype2*YA	-	0.022	0.0010	-	0.020*	0.021*
		(0.017)	(0.008)		(0.010)	(0.010)
Ftype3*YA	-	0.022	0.018	-	0.013	0.007
		(0.017)	(0.018)		(0.011)	(0.012)
YA	-0.003	-0.028	-0.003	-0.005	-0.003	-0.004
	(0.010)	(0.050)	(0.010)	(0.009)	(0.008)	(0.008)
GFtype1	0.154***	-	0.110**	0.186***	-	0.147*
	(0.043)		(0.047)	(0.075)		(0.083)
GFtype2	0.184***	-	0.088**	0.226***	-	0.214***
	(0.040)		(0.036)	(0.049)		(0.053)
GFtype3	-0.009	-	0.001	0.032	-	0.041
	(0.044)		(0.042)	(0.062)		(0.053)
Ftype1	-	0.152***	0.179***	-	0.190***	0.163***
		(0.043)	(0.041)		(0.038)	(0.037)
Ftype2	-	0.100**	0.088**	-	0.059	0.050
		(0.035)	(0.036)		(0.035)	(0.035)
Ftype3	-	-0.121	-0.105	-	-0.072	-0.048
		(0.069)	(0.074)		(0.050)	(0.052)
N	4,684	4,684	4,684	4,684	4,684	4,684
R^2	0.0959	0.1202	0.1346	0.1382	0.1468	0.1708

Note: This table reports the regression results for obtaining high school and college degrees based on equations (4)-(6), using father's occupation as their types. Controls include child's birth cohort dummies, gender, age and age squared, father's age and father's age squared (results were not shown in this table); the standard errors are clustered at the child's residential province level.

Table A17 Estimated Coefficients of the Effect of Mother's Schooling and Father's Schooling on High School and College Attainments: Based on Equations (8), (9) and (10)

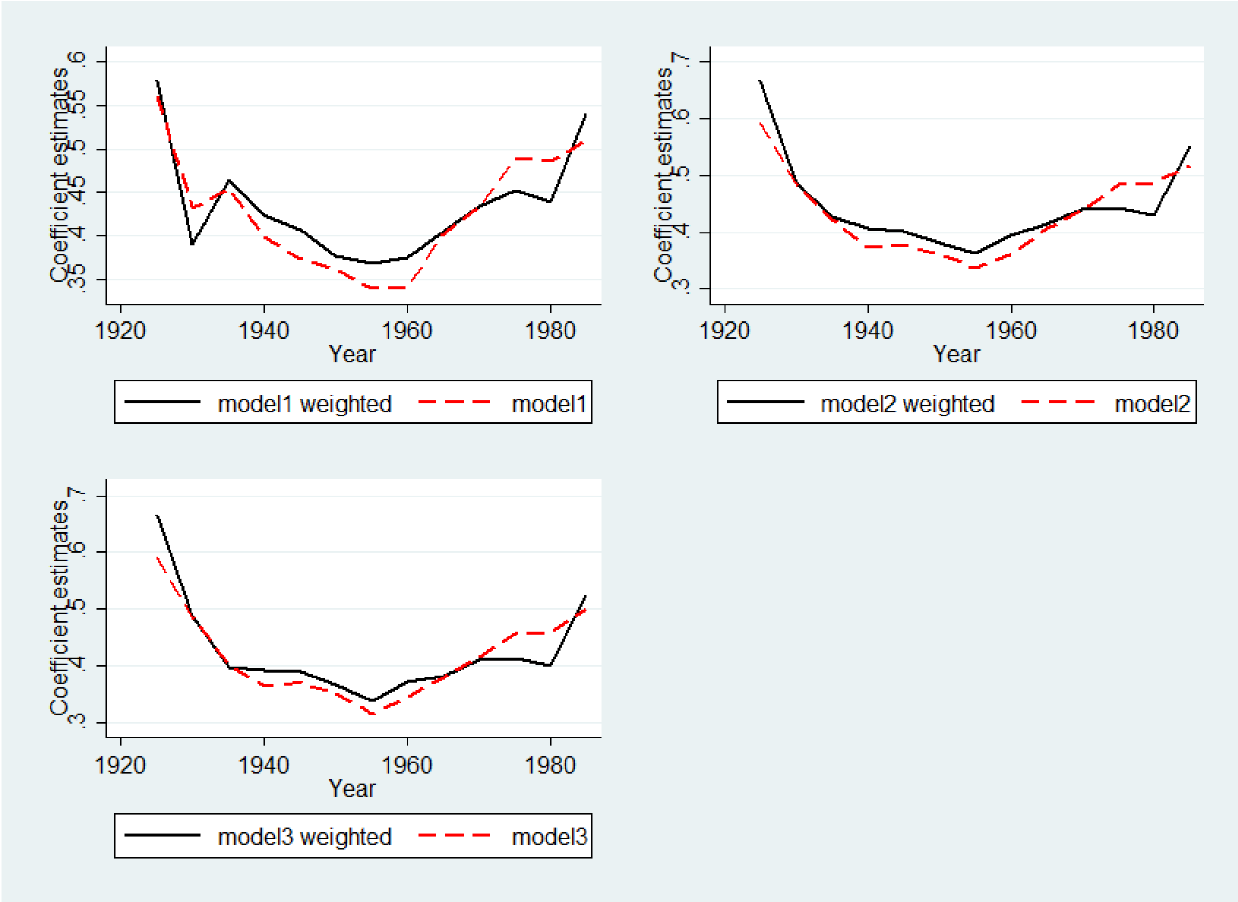
	High School			College		
	(1)	(2)	(3)	(4)	(5)	(6)
MS*YA	-0.0001	-	-0.0001	-0.0011**	-	-0.0004
	(0.0005)		(0.0001)	(0.0005)		(0.0005)
FS*YA	-	-0.0013***	-0.0009**	-	-0.0020***	-0.0015***
		(0.0004)	(0.0004)		(0.0003)	(0.003)
YA	0.012***	0.016***	-0.003	0.005***	0.010***	0.005***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
MS	0.044***	-	0.026***	0.031***	-	0.020***
	(0.001)		(0.001)	(0.001)		(0.001)
FS	-	0.042***	0.029***	-	0.030***	0.018***
		(0.001)	(0.001)		(0.001)	(0.001)
N	31,205	31,337	31,187	31,205	31,337	31,187
R^2	0.1748	0.1828	0.2065	0.1660	0.1658	0.1879

Note: This table reports the regression results for obtaining high school and college degrees based on equations (11)-(13). MS denotes years of mother's schooling; FS denotes years of father's schooling; YA denotes the years a person was affected by the CR. Controls in columns (1) and (4) include child's birth cohort dummies, gender, age and age squared, mother's age and mother's age squared; controls in columns (2) and (5) include birth cohort dummies, gender, age and age squared, father's age and father's age squared; controls in columns (3) and (6) include all of the controls mentioned above. The standard errors are clustered at the child's residential province level.

Table A18 Sample Descriptive Statistics by the Answer of "Whether Took CEE" in the CHIP

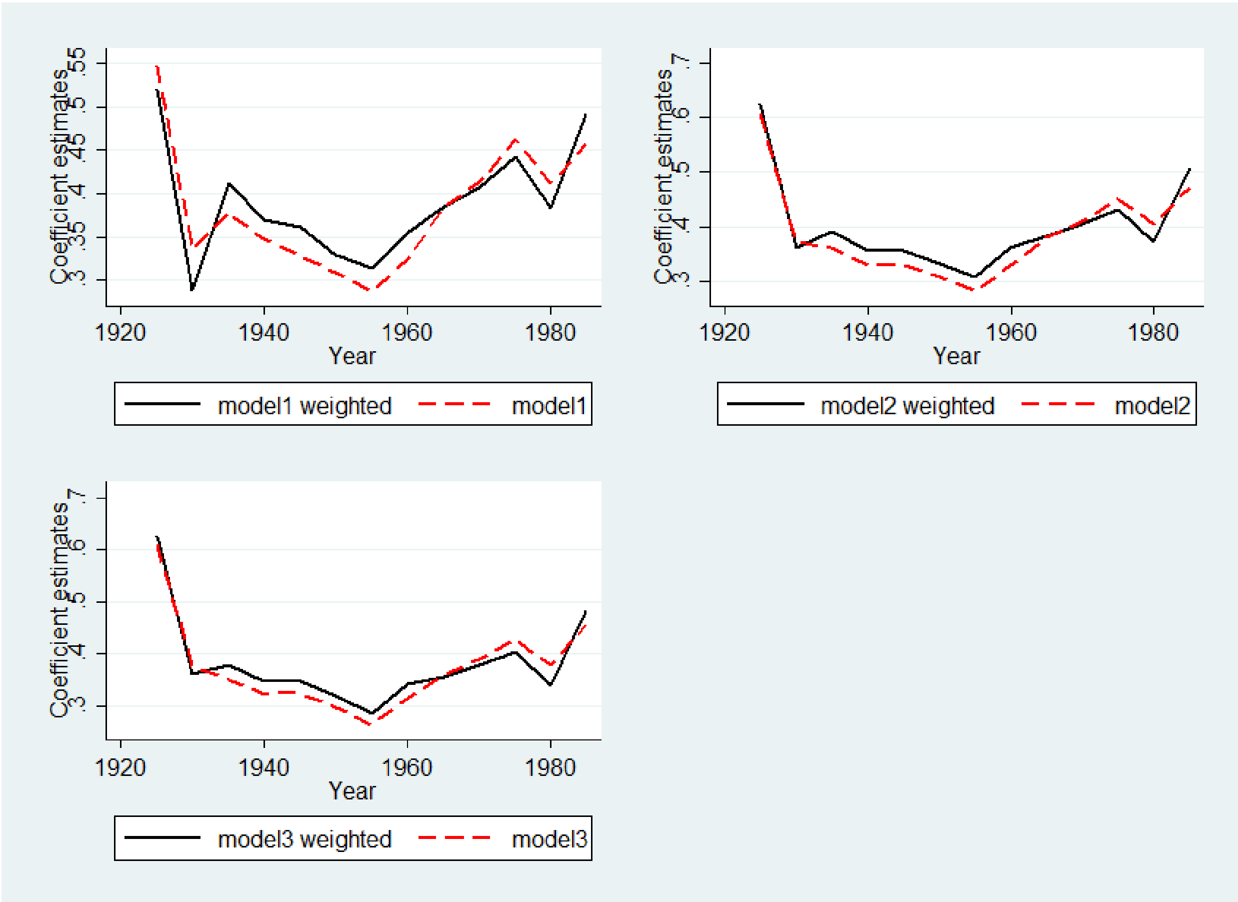
	Pre-CR	**CR**	**Post-CR**
	(1) Whether Took CEE		
	N (%)	N (%)	N (%)
No	616(29%)	4,098(31%)	5,247(30%)
Yes	56(3%)	778(6%)	2,499(14%)
Missing	1,427(68%)	8,508(63%)	9,747(55%)
Total	2,099(100%)	13,384(100%)	17,493(100%)
	(2) Whether Took CEE among People with Years of Schooling>12		
	N (%)	N (%)	N (%)
No	231(75%)	1,398(65%)	2,564(48%)
Yes	55(18%)	584(27%)	2,499(47%)
Missing	22 (7%)	178(8%)	297(6%)
Total	308(100%)	2,160(100%)	5,323(100%)
	(3) Descriptive Statistics by Whether Took CEE among People with Years of Schooling>12		
	Mean (Std)	Mean (Std)	Mean (Std)
	(i) Age		
CEE taker	73.38(4.43)	54.70(3.82)	28.25(6.87)
No taker	73.41(4.23)	57.21(4.21)	40.13(6.36)
Missing	73.25(5.03)	56.42(3.96)	40.78(6.35)
	(ii) Schooling		
CEE taker	15.45(1.80)	13.49(1.83)	14.78(1.93)
No taker	12.90(1.70)	12.53(1.44)	12.80(1.70)
Missing	12.60(1.67)	12.29(1.17)	12.40(1.35)
	(iii) Father's Schooling		
CEE taker	5.45(4.87)	6.00(4.50)	8.59(4.16)
No taker	4.00(4.12)	4.70(4.30)	7.39(4.00)
Missing	3.76(4.16)	4.03(3.89)	6.82(3.82)
	(iv) Mother's Schooling		
CEE taker	2.92(4.05)	4.05(4.38)	6.96(4.45)
No taker	1.59(3.10)	2.68(3.79)	5.56(4.11)
Missing	1.42(2.71)	2.37(3.37)	5.55(3.83)

Figure A1 Estimated Coefficients of the Intergenerational Coefficients of Table A2 (Rank of Schooling)



Note: These graphs were plotted based on the estimated results of the three model specifications in Table A2. Model 1 is the specification with no controls, model 2 has basic controls, and model 3 has extensive controls. The solid lines are estimated intergenerational coefficients using the weights described in Appendix A1, and the dash lines are estimated coefficients without weighting.

Figure A2 Estimated Coefficients of the Intergenerational Correlations of Table A3 (Years of Schooling)

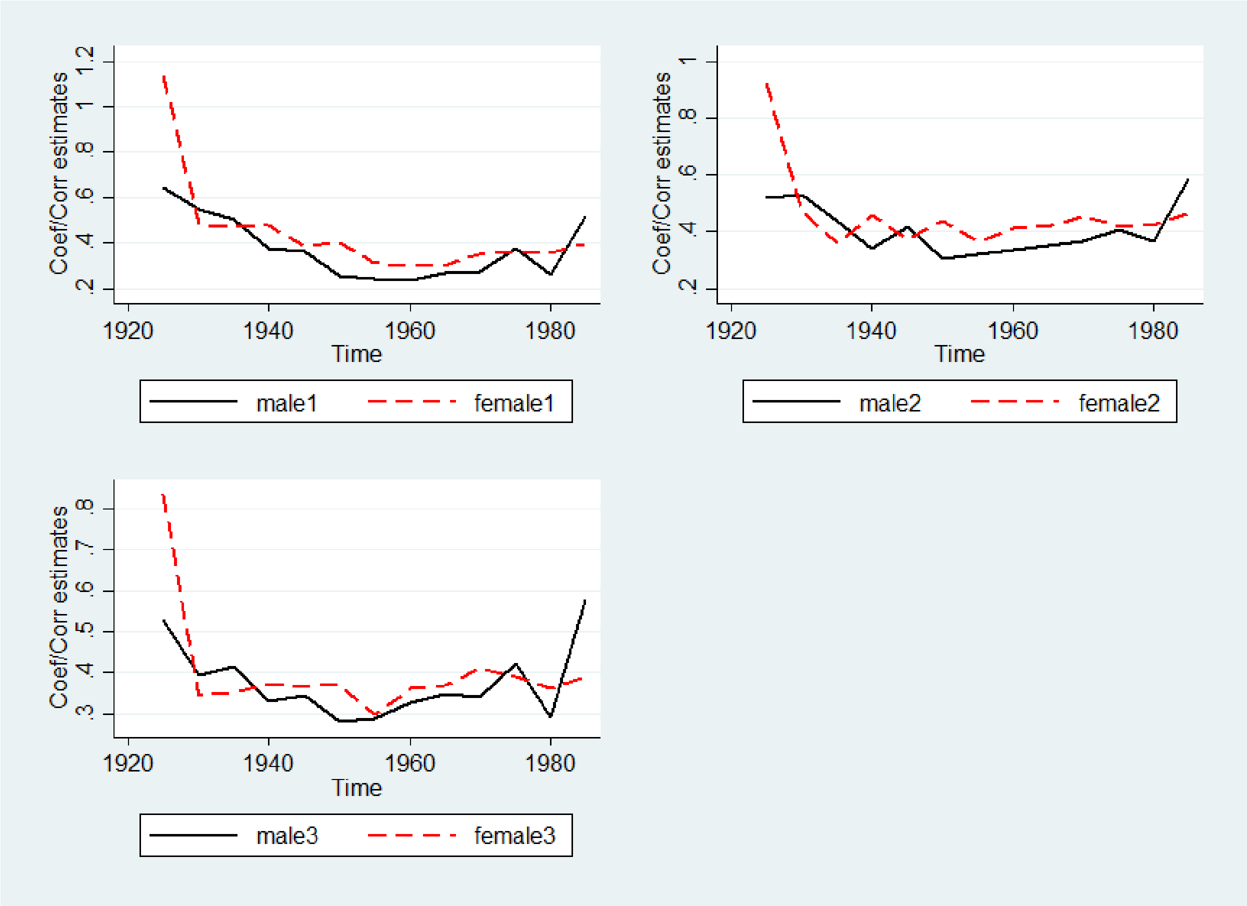


Note: These graphs were plotted based on the estimated results of the three model specifications in Table A3. Model 1 is the specification with no controls, model 2 has basic controls, and model 3 has extensive controls. The solid lines are estimated intergenerational coefficients using the weights described in Appendix A1, and the dash lines are estimated coefficients without weighting.

Figure A3 Estimated Coefficients of the Intergenerational Schooling Coefficients and Correlations for Urban and Rural Residents

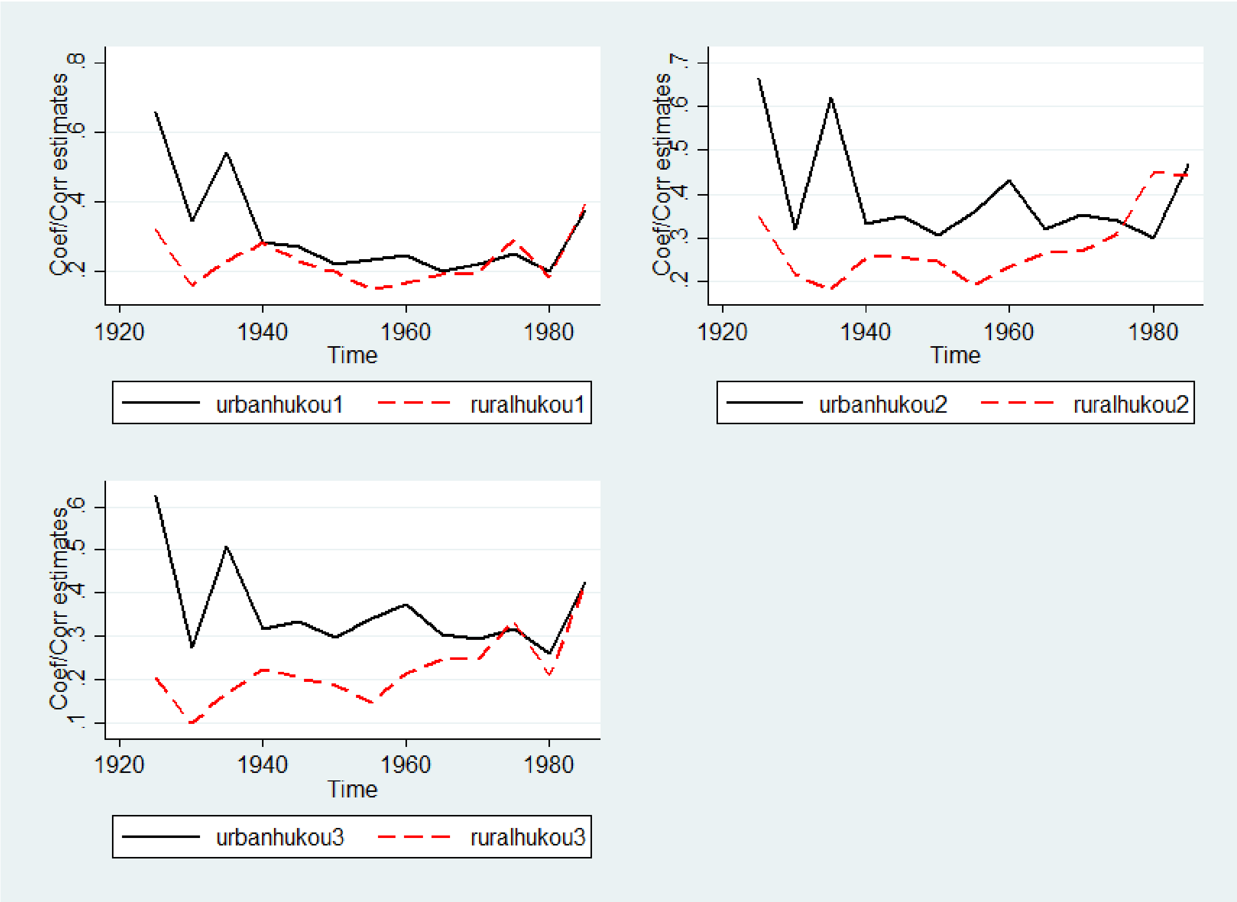
Note: These graphs were plotted based on the results of the third model specification of equation (1). The top-left subgraph plots the estimates of the intergenerational coefficients using years of schooling, the top-right subgraph plots the estimates of the intergenerational coefficients using rank of schooling, and the bottom-left subgraph plots intergenerational correlations using years of schooling. The solid lines are the estimates for urban sample; the dotted lines are the estimates for rural sample.

Figure A4 Estimates of the Intergenerational Schooling Coefficients and Correlations for Male and Female



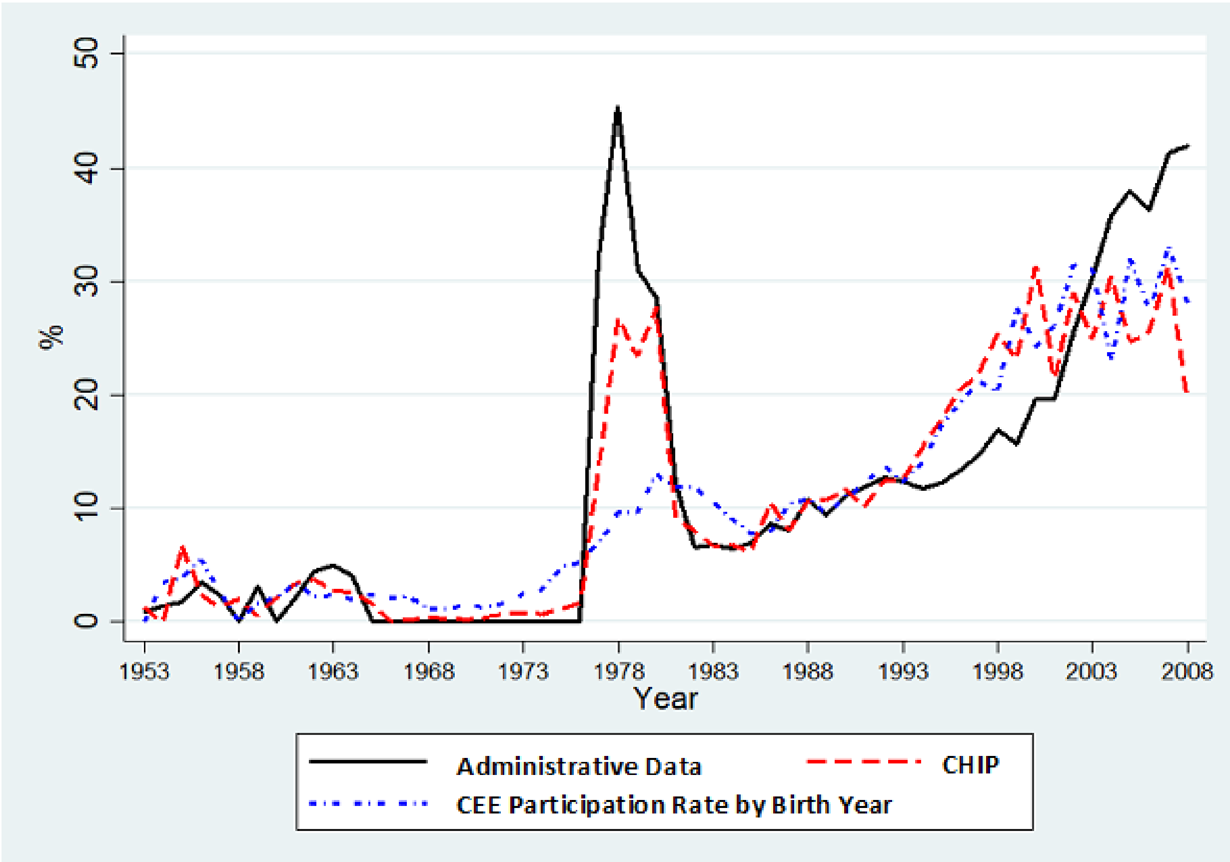
Note: These graphs were plotted based on the results of the third model specification (without controlling gender) of equation (1). The top-left subgraph plots the estimates of the intergenerational coefficients using years of schooling, the top-right subgraph plots the estimates of the intergenerational coefficients using rank of schooling, and the bottom-left subgraph plots intergenerational correlations using years of schooling. The solid lines are the estimates for male sample; the dotted lines are the estimates for female sample.

Figure A5 Estimates of the Intergenerational Schooling Coefficients and Correlations for Urban-born and Rural-born Samples



Note: These graphs were plotted based on the results of the third model specification of equation (1). The top-left subgraph plots the estimates of the intergenerational coefficients using years of schooling, the top-right subgraph plots the estimates of the intergenerational coefficients using rank of schooling, and the bottom-left subgraph plots intergenerational correlations using years of schooling. The solid lines are the estimates for urban-born sample (proxied by people who had an Urban Hukou at birth), the dotted lines are the estimates for rural-born sample (proxied by people who had an Agricultural Hukou at birth).

Figure A6 Number of People Who Reported Taking the CEE as Proportions of 19-year-old Population and the Proportion of 19-year-old Population Registered for the CEE by Year: Comparison between the CHIP and Administrative Data



Note: This graph plots the fraction of CEE takers as a proportion of 19-year-olds, comparing the CHIP and administrative data by year. The solid line shows the fractions of CEE takers using Census 2000 and an administrative file of CEE takers by CEE-taking year, the short-dashed line shows the fractions of CEE takers reported by CHIP respondents by CEE-taking year, and the dotted line shows the proportion of 19-year-olds who took the CEE by their birth year (the horizontal line denotes the year when they were 19 years old) using the CHIP