Rising Intergenerational Income Persistence in China
Online Appendix

Yi Fan Junjian Yi Junsen Zhang [*]

November 13, 2019

Contents

Appendix A: Institutional Background **2**

Market-oriented Structural Changes 2

Economic Development 2

Public Expenditure and Expansion of Tertiary Education in China 3

Appendix B: Figures and Tables **6**

Appendix C: Missing Income, Selection Bias, and Computed Income **14**

Appendix D: Correlates of Changes in Intergenerational Income Persistence in China **25**

Appendix D1: Regression Specification and Variable Definitions 25

Appendix D2: Regression Results 26

[*]Yi Fan: Department of Real Estate, National University of Singapore, SDE Blk 1, 4 Architecture Drive, Singapore 117566; email: rstfyi@nus.edu.sg. Junjian Yi, Department of Economics, National University of Singapore, Faculty of Arts & Social Sciences, AS2 Level 6, 1 Arts Link, Singapore 117570; email: junjian@nus.edu.sg. Junsen Zhang (corresponding author), Department of Economics, 9/F&10/F, Esther Lee Building, Chinese University of Hong Kong, Shatin, N.T., Hong Kong; email: jszhang@cuhk.edu.hk. We thank Gary Becker, Steven Durlauf, David Figlio, James Heckman, Ivan Png, the Editor, and three referees for helpful comments and suggestions. The data used in this paper are from the China Family Panel Studies survey, funded by the 985 Program and carried out by the Institute of Social Science Survey of Peking University. Yi Fan acknowledges financial support from NUS Startup Grant No. R-297-000-134-133. Junjian Yi acknowledges financial support from the MOE Academic Research Fund No. FY2018-FRC3-006. A preliminary version of this paper began circulating in 2015:'The Great Gatsby Curve in China: Cross-Sectional Inequality and Intergenerational Mobility'.

Appendix A: Institutional Background

Market-oriented Structural Changes

China's market-oriented reforms, which started in the late 1970s, marked a shift from a planned economy to a market economy. These reforms have spurred China's economic growth, facilitated the transformation from agriculture to industry, and sparked rapid urbanization.[1] The household registration (*hukou*) system was gradually relaxed from the late 1980s onward, which resulted in an unprecedented increase in temporary domestic migration that amounted to 0.2 billion, according to the 1 percent mini-census in 2005. Meanwhile, private firms were legalized in 1997, and a considerable number of state-owned and collective enterprises were privatized (Zhu, 2012). Institutional reforms have adjusted the incentive structure, enhanced labor productivity, and increased private return to human capital (Ge & Yang, 2011, 2014). Figure B.1 shows that the return to 1 additional year of schooling increased fourfold, from 2 percent in 1988 to 10 percent in 2008. The return to college education underwent an even more drastic change, rising from 7 percent in 1988 to 49 percent in 2008 (Li *et al.*, 2012). Over the same period, income inequality also increased (Heckman & Yi, 2014). Figure A.1 shows that the annual wage for the high-education group was twice that of the low-education group and 1.5 times that of the medium-education group in the late 2000s, compared to almost no differences in 1988 (Li *et al.*, 2012).

Economic Development

The impact of economic development on intergenerational mobility is ambiguous from the perspective of intergenerational transmission of human capital (Becker & Tomes, 1986). On the one hand, poor families benefit more from economic development due to relaxed credit constraints on their children's education. In this sense, intergenerational income persistence is expected to decrease. On the other hand, the rising return to schooling incentivizes rich parents to invest in their children, and thus increases intergenerational persistence. In addition to the transmission channel of education, wealth plays an increasingly important role in intergenerational income persistence in China (Yuan & Chen, 2013). Parents who have increased their wealth as the economy has grown rapidly are able to bequeath more wealth to their children. Together with rising income inequality, as demonstrated in Figure 7 in Li *et al.* (2012) and Heckman & Yi (2014), the overall association between economic development and intergenerational income persistence remains an empirical question.

[1] See Zhu (2012) for a discussion of China's structural transformation and economic growth.

Public Expenditure and Expansion of Tertiary Education in China

China's economic reform has been accompanied by fiscal decentralization in primary and secondary education from the mid-1980s onward, and a tax reform of fiscal recentralization in 1994, which aggravated regional inequality in primary and junior secondary education. The central government makes partial transfers to finance local primary and secondary schools, and local governments are expected to fill the remaining gaps. As central-to-local transfers are insufficient, local governments, especially those in poor areas, are unable to fulfill their obligations. In rich areas, however, public expenditure is more generous. Figure A.2 suggests that the share of government expenditure for education in GDP doubled from 2 percent in 1992 to 4 percent in 2012, and Figure A.3 reveals that the expenditure is mainly borne by local governments. Heckman (2005) shows that in 2004, the per pupil government expenditure in Beijing was 16 times higher than in Guizhou (Table 7 in Heckman (2005)).[2]

In addition, the radical expansion of higher education, accompanied by the drastic rise in educational costs since the late 1990s, further exacerbates such inequality (Chow & Shen, 2006).The total number of fresh college graduates rose more than sixfold, to 7 million, between 2001 and 2013 (NBS, 2011). Figure A.4 displays the sharp increase in the share of college students in the 18-22 age cohort (Li & Xing, 2010). Average annual tuition fees surged from RMB 800 in 1995 to RMB 5,000 in 2004. Yearly expenditure per student reached an average of RMB 12,318, based on a national survey of college students conducted by Tsinghua University in 2010 (Li *et al.*, 2013). Furthermore, decentralization stratified higher education into two layers. The central government administers a small number of elite universities, whereas local governments administer most local colleges and universities. Rich parents are more able to send their children to elite universities, while poor families are increasingly subject to credit constraints. Li *et al.* (2013) note that the share of students in elite universities who come from rural and western regions has decreased. In 2010, 22 percent of college students are from families with annual income less than the average annual expenditure of college. Loans and scholarships account for less than 10 percent of the annual expenditure on college. Need-based aid targeted to low-income students is clearly misallocated (Li *et al.*, 2013).

[2] Knight *et al.* (2011) review the evolution of China's educational system.

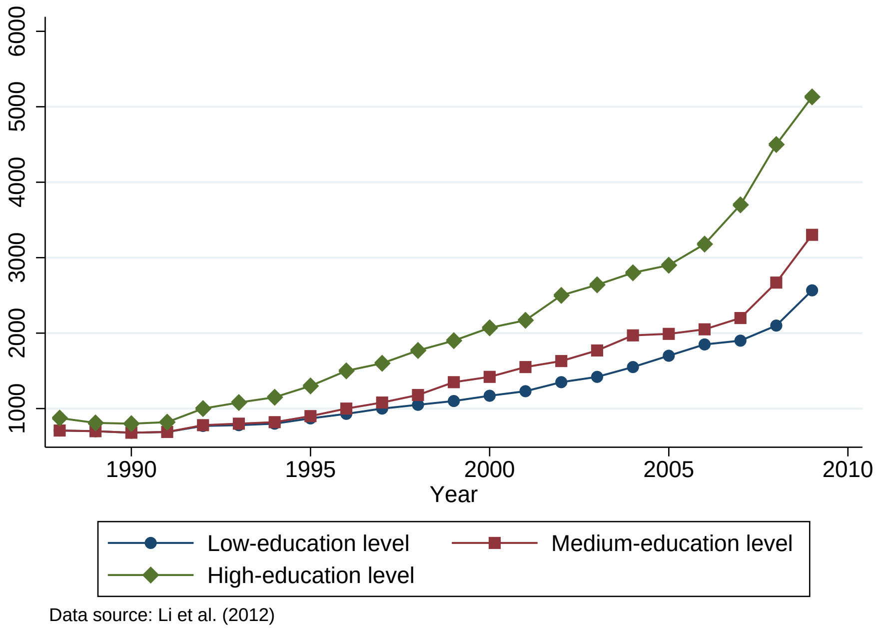


Figure A.1: Annual Wage of Urban Workers, 1988-2009

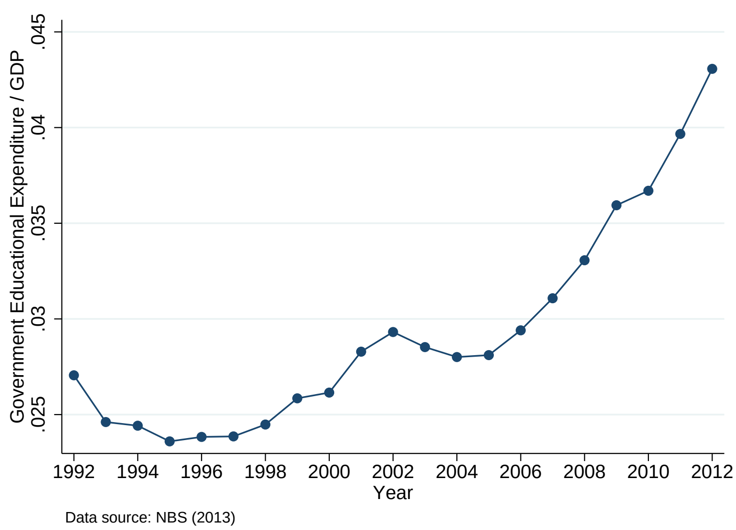


Figure A.2: Government Educational Expenditure/GDP, 1992-2012
Note: Data on government educational expenditure and GDP are from National Bureau of Statistics of China Statistical Yearbook (2013a).

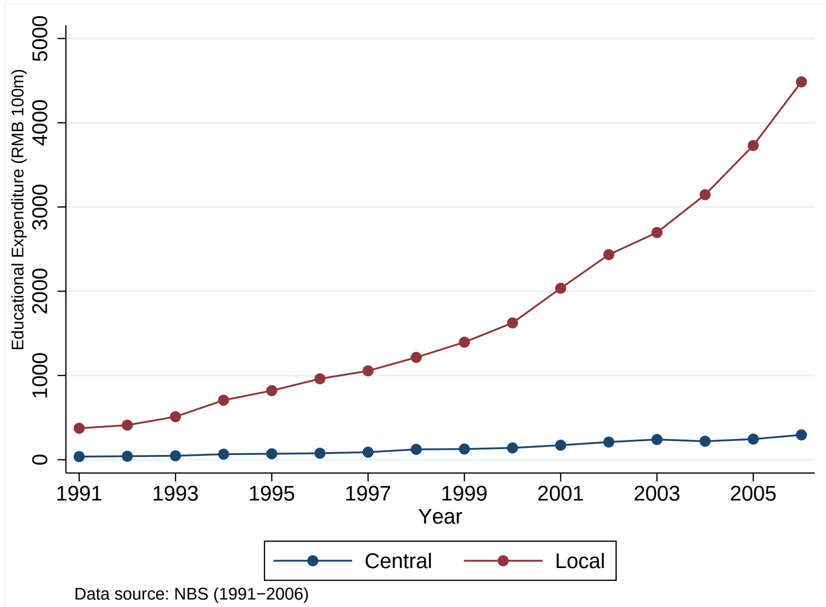


Figure A.3: Central and Local Governmental Expenditure on Education, 1991-2006

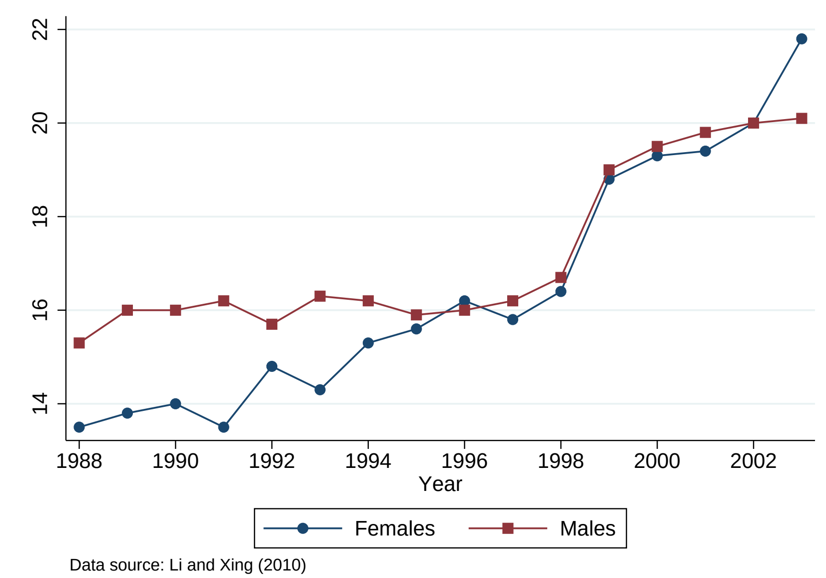


Figure A.4: Tertiary School Enrollment Rates, 1988-2003

Appendix B: Figures and Tables

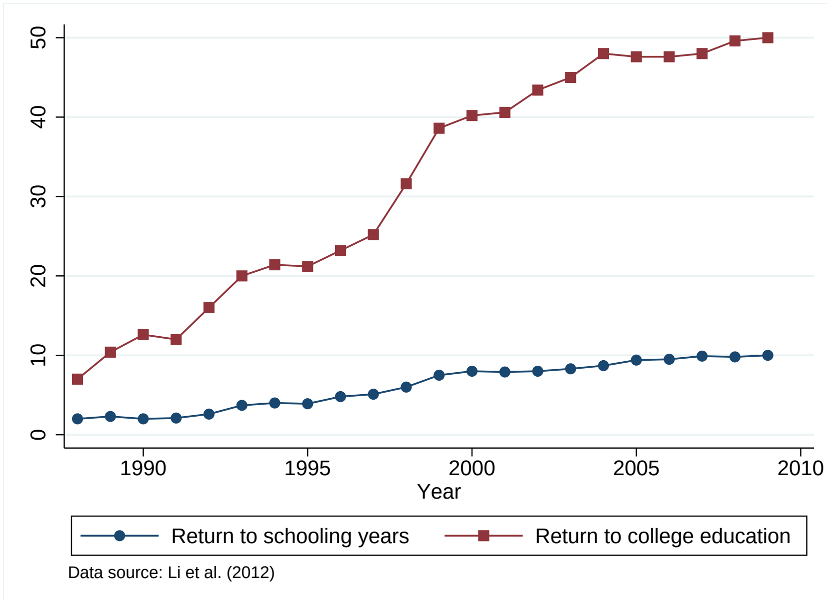


Figure B.1: Return to Education in Urban China, 1988-2009

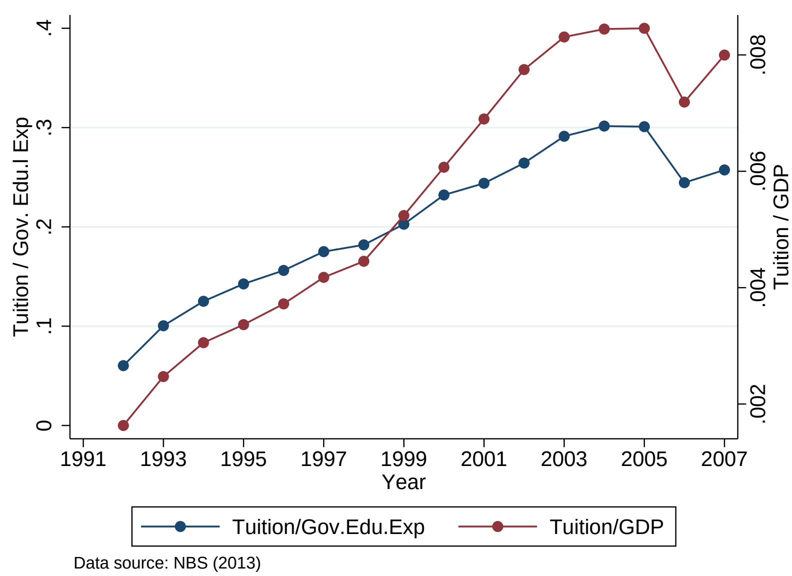


Figure B.2: Increase in Tuition in China, 1991-2007

Note: Data on tuition, government educational expenditure, and GDP are from National Bureau of Statistics of China Statistical Yearbook (2013b).

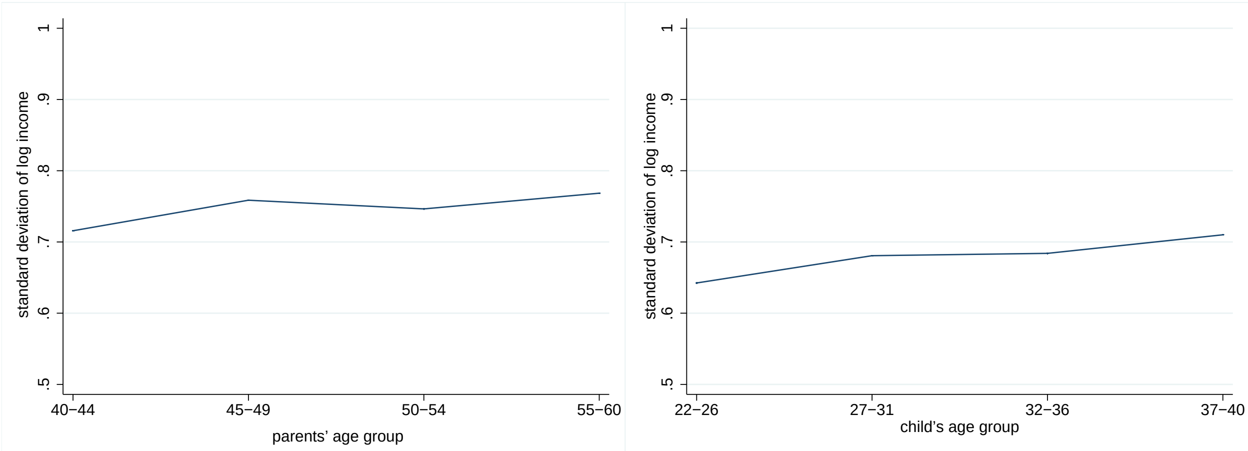


Figure B.3: Standard Deviation of Income against Age Groups
(Left: Parents, Right: Children)

Note: The cross-cohort comparison involves both age effect and cohort effect. These two effects may run in opposite direction and cancel out each other. Therefore the standard deviation of log income appears to be at a similar level across birth cohorts.

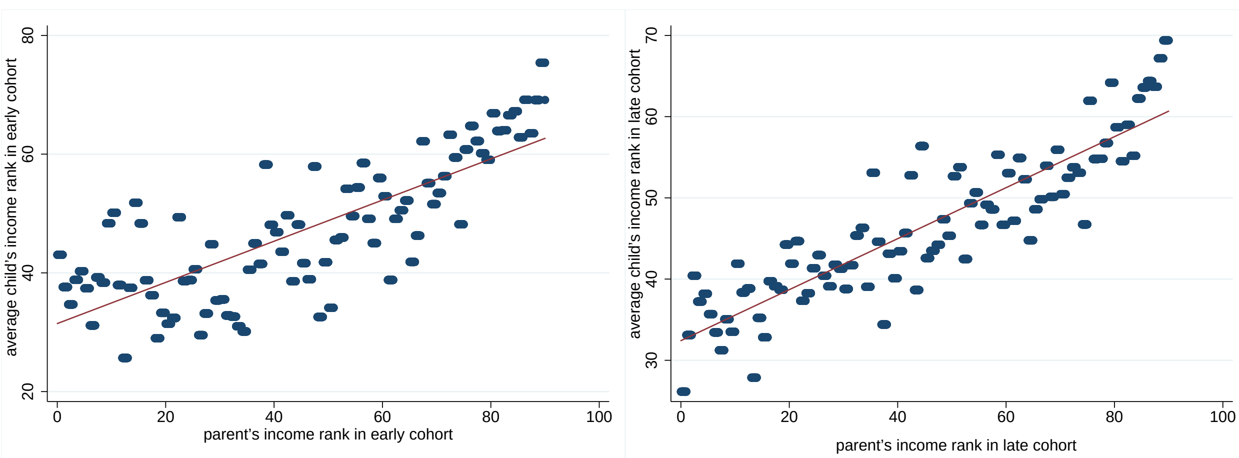


Figure B.4: Income Rank of Children versus Income Rank of Parents in the Early Cohort (Left) and Late Cohort (Right), with Top 10 percent Dropped

Note: The sample excludes parents in the top 10 percent.

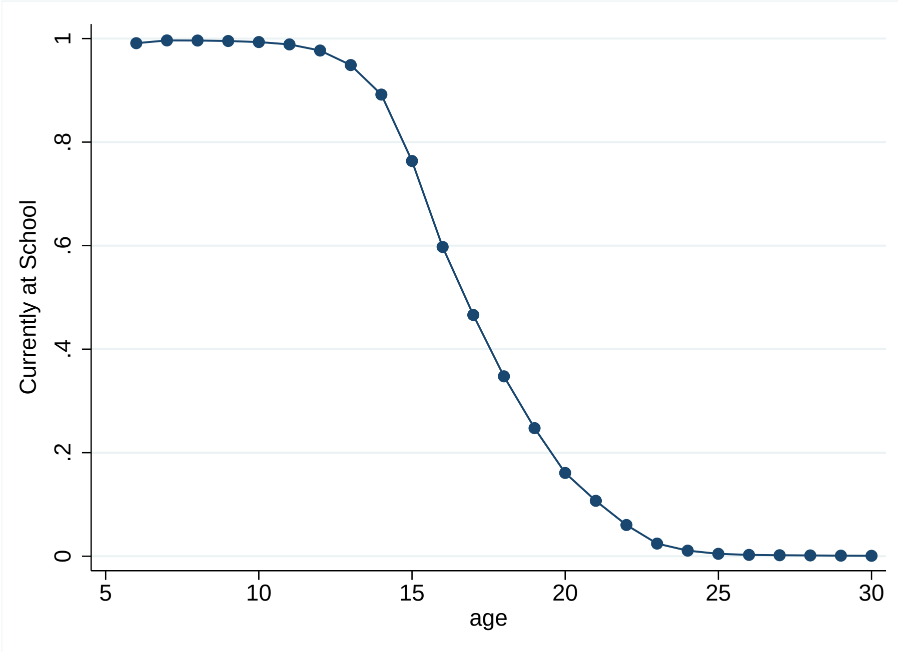


Figure B.5: Probability of Currently Being at School against Age

Note: Data from census in 2000.

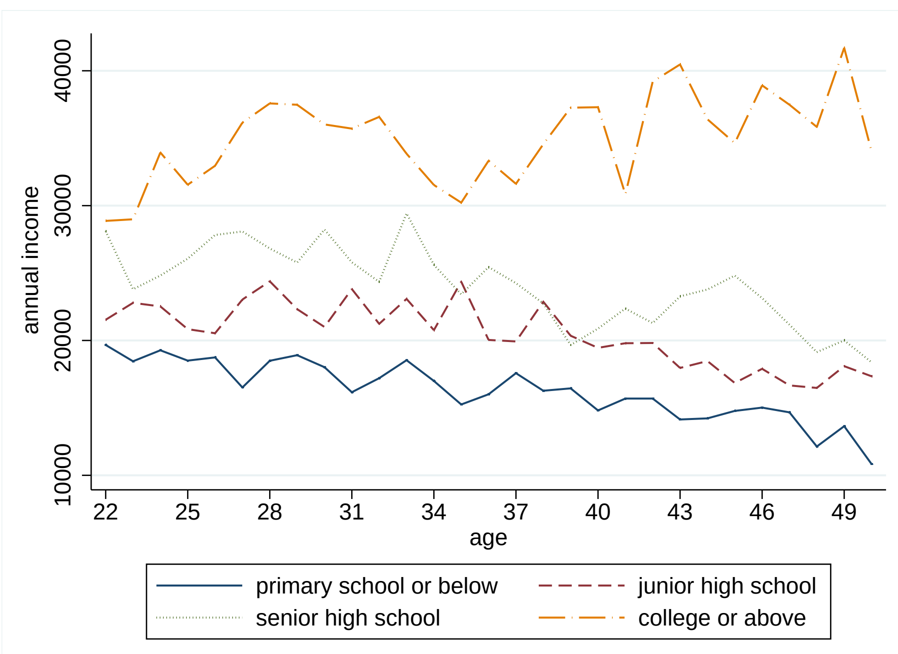


Figure B.6: Income against Age by Educational Groups for Children

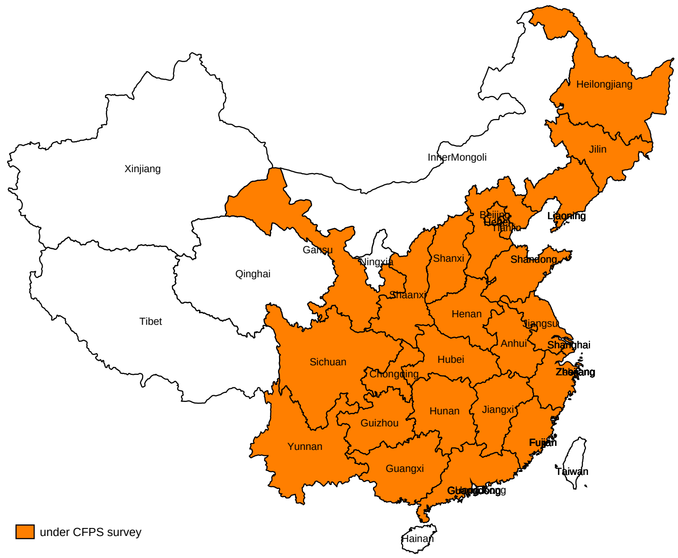


Figure B.7: The 25 Provinces and Municipalities under the China Family Panel Studies 2010
Note: Geographic data are from China's Official Geographic Data (2015).

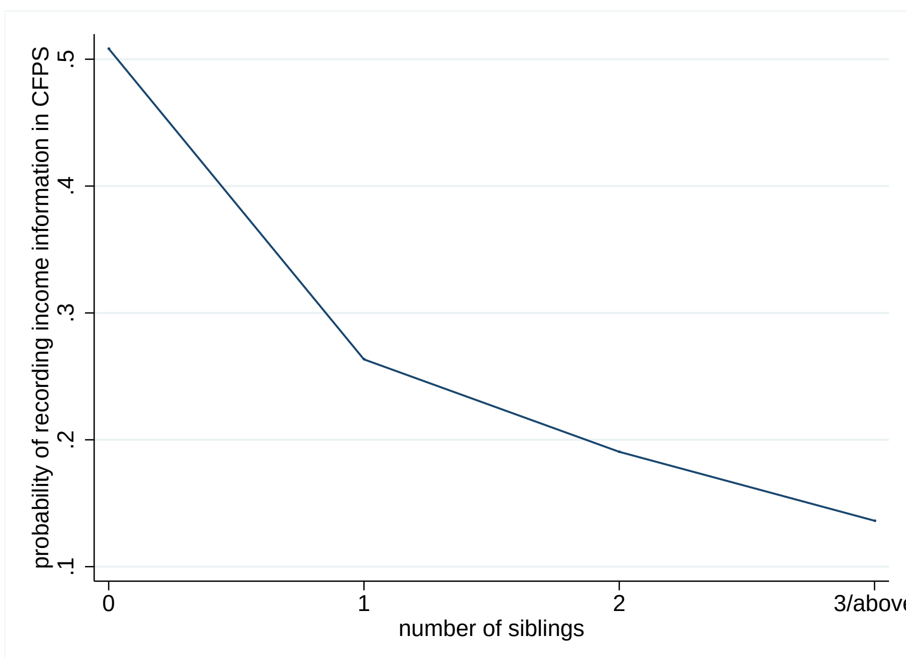


Figure B.8: Probability of Recording Income Information in CFPS against Number of Siblings

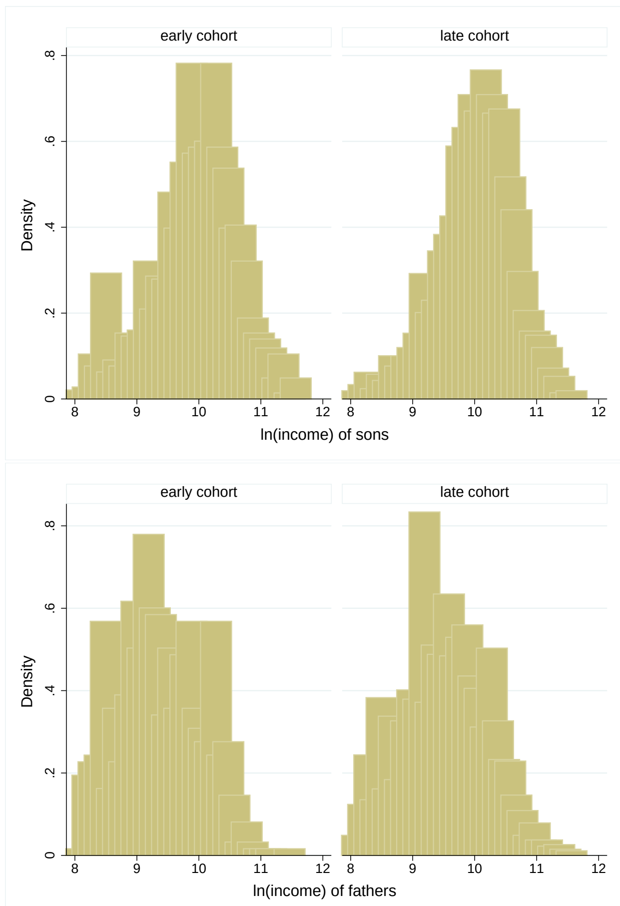
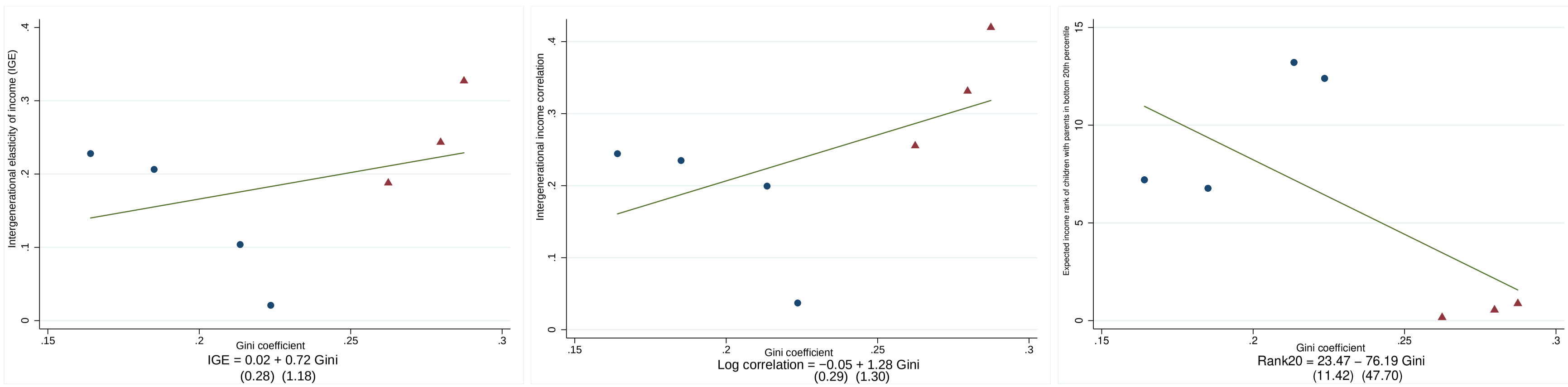


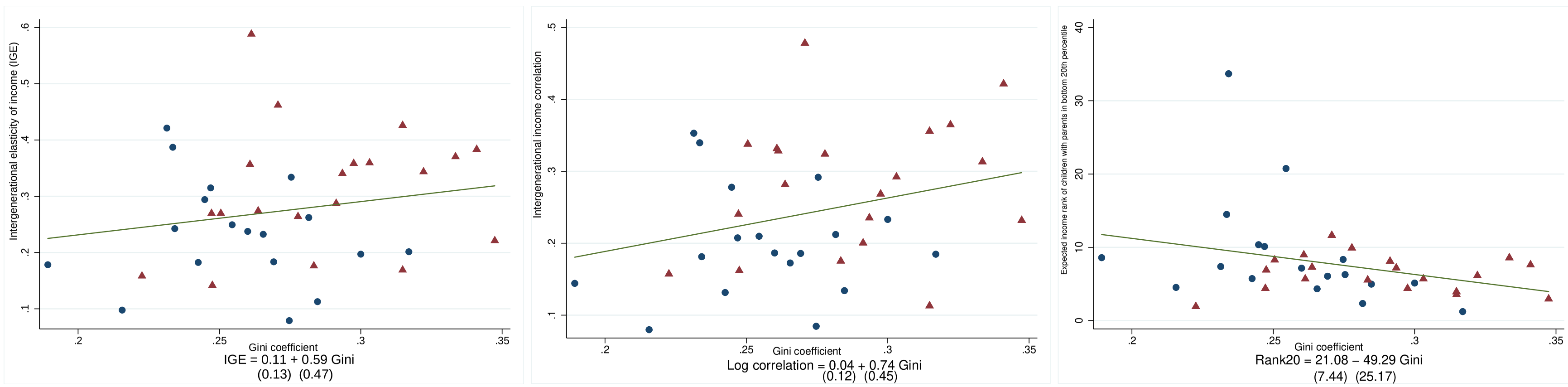
Figure B.9: Density Distribution of Log of Income by Fathers and Sons in Early and Late Cohorts



(a) **(b)** **(c)**

Figure B.10: The Great Gatsby Curve in Urban China

Note: Bootstrapped standard errors are in parentheses. Circles represent estimates from the 1970–1980 (early) cohort, and triangles represent estimates from the 1981–1988 (late) cohort. Lines are linearly fitted lines. The Gini coefficient is calculated by province and cohort. Cells with fewer than 100 observations are excluded. Rank20 refers to the income rank of children born to parents at the bottom 20 percent national percentile rank.



(a) **(b)** **(c)**

Figure B.11: The Great Gatsby Curve in Rural China

Bootstrapped standard errors are in parentheses. Circles represent estimates from the 1970–1980 (early) cohort, and triangles represent estimates from the 1981–1988 (late) cohort. Lines are linearly fitted lines. The Gini coefficient is calculated by province and cohort. Cells with fewer than 100 observations are excluded. Rank20 refers to the income rank of children born to parents at the bottom 20 percent national percentile rank.

Table B1: Robustness Checks on Intergenerational Mobility Using Child's Schooling as Alternative Outcome

	Early cohort (1)	Late cohort (2)	Change (3)	Early cohort (4)	Late cohort (5)	Change (6)	Early cohort (7)	Late cohort (8)	Change (9)
Panel A. Robustness Check 1: Correlation between Child's Schooling and Parental Logarithm Income									
	Child's being at least senior high school			Child's being at least college			Child's being at least university		
Estimate	0.423	0.617	0.194	0.270	0.446	0.176	0.131	0.255	0.124
	(0.012)	(0.010)	(0.015)	(0.011)	(0.012)	(0.016)	(0.008)	(0.010)	(0.013)
Obs.	10,980	11,333	-	10,980	11,333	-	10,980	11,333	-
Panel B. Robustness Check 2: Correlation between Child's Schooling and Parental Income Rank									
	Child's being at least senior high school			Child's being at least college			Child's being at least university		
Estimate	0.529	0.761	0.232	0.321	0.522	0.201	0.154	0.283	0.128
	(0.015)	(0.014)	(0.020)	(0.011)	(0.014)	(0.017)	(0.009)	(0.012)	(0.016)
Obs.	10,980	11,333	-	10,980	11,333	-	10,980	11,333	-

Note: Bootstrapped standard errors are in parentheses. The data are from the China Family Panel Studies (CFPS) for 2010, 2012, 2014, and 2016. Additional regressors include child's age and age squared, and father's age and age squared. The sample includes children who are at least 22 years old and parents who are at most 64 years old. The early cohort comprises children born between 1970 and 1980. The late cohort comprises children born between 1981 and 1988. Both the child's income and the parents' income refer to annual income averaged across at least two waves of the CFPS in 2010, 2012, 2014, and 2016. Income is adjusted to 2010 prices using the CPI.

Appendix C: Missing Income, Selection Bias, and Computed Income

Table C1: Determinants of Having Income Information

Outcome Variable: Having Income Information (=1)				
	Children Probit		Parents Probit	
	(1)	(2)	(3)	(4)
number of live siblings	-0.284***	-0.300***		
	(0.009)	(0.027)		
number of live siblings × cohort		-0.161***		
		(0.041)		
number of live siblings × coastal		-0.015		
		(0.028)		
number of live siblings × Hukou		0.091**		
		(0.029)		
number of live siblings × cohort × coastal		0.006		
		(0.039)		
number of live siblings × cohort × Hukou		0.086*		
		(0.043)		
number of live children			-0.435***	-0.384***
			(0.013)	(0.059)
number of live children × cohort				-0.238**
				(0.080)
number of live children × coastal				-0.006
				(0.047)
number of live children × Hukou				-0.010
				(0.059)
number of live children × cohort × coastal				0.053
				(0.057)
number of live children × cohort × Hukou				0.176*
				(0.080)
education	0.035***	0.035***	-0.026	-0.024
	(0.009)	(0.009)	(0.018)	(0.018)

cohort (late = 1)	114.345	117.678	-7.262	-7.208
	(87.484)	(88.267)	(17.204)	(17.218)
coastal region	55.882	58.578	-27.076	-27.384
	(87.928)	(88.142)	(27.454)	(27.421)
Hukou (rural = 1)	7.338	7.510	-20.164	-19.449
	(71.976)	(71.910)	(18.167)	(18.122)
age	4.201	4.263	0.070	0.107
	(5.341)	(5.350)	(0.562)	(0.561)
age squared/100	-11.303	-11.480	-0.067	-0.101
	(15.453)	(15.477)	(0.491)	(0.490)
age cubed/1000	1.000	1.017	-0.021**	-0.021**
	(1.484)	(1.487)	(0.008)	(0.008)
male	0.033	0.028		
	(0.066)	(0.066)		
live father			-0.107	-0.114
			(0.211)	(0.209)
live mother			0.965*	0.957*
			(0.456)	(0.451)
education × cohort	-0.002	-0.009	0.051*	0.041
	(0.013)	(0.013)	(0.022)	(0.023)
education × Hukou	-0.031**	-0.029**	0.042*	0.040*
	(0.011)	(0.011)	(0.020)	(0.020)
cohort × Hukou	-51.721	-54.522	26.125	26.343
	(102.431)	(102.996)	(19.273)	(19.264)
education × cohort × Hukou	0.020	0.026	-0.032	-0.023
	(0.015)	(0.015)	(0.025)	(0.025)
education × coastal	-0.013	-0.014	-0.004	-0.004
	(0.013)	(0.013)	(0.024)	(0.024)
cohort × coastal	-79.626	-84.563	24.610	26.013
	(123.680)	(124.850)	(28.676)	(28.677)
education × cohort × coastal	0.021	0.017	-0.044	-0.043
	(0.019)	(0.019)	(0.030)	(0.030)
Hukou × coastal	10.652	11.825	12.630	13.045
	(109.204)	(109.193)	(31.493)	(31.415)
education × Hukou × coastal	-0.004	-0.002	-0.042	-0.041
	(0.017)	(0.017)	(0.028)	(0.028)
cohort × Hukou × coastal	68.331	69.381	-10.921	-12.236

	(153.143)	(153.968)	(32.770)	(32.723)
education × cohort × Hukou × coastal	-0.015	-0.012	0.041	0.041
	(0.024)	(0.024)	(0.035)	(0.035)
age × cohort	-11.791	-12.206	0.291	0.293
	(9.126)	(9.227)	(0.611)	(0.612)
age × Hukou	-0.208	-0.238	0.719	0.694
	(6.266)	(6.260)	(0.635)	(0.634)
age × cohort × Hukou	5.798	6.180	-0.926	-0.938
	(10.678)	(10.757)	(0.684)	(0.684)
age × coastal	-4.499	-4.736	0.982	0.993
	(7.663)	(7.682)	(0.949)	(0.948)
age × cohort × coastal	7.561	8.072	-0.852	-0.903
	(12.843)	(12.992)	(1.003)	(1.003)
age × Hukou × coastal	-1.310	-1.402	-0.467	-0.482
	(9.508)	(9.507)	(1.092)	(1.089)
age × cohort × Hukou × coastal	-8.437	-8.618	0.407	0.454
	(15.893)	(16.005)	(1.147)	(1.146)
age squared/100 × cohort	41.688	43.493	-0.272	-0.267
	(32.966)	(33.390)	(0.543)	(0.544)
age squared/100 × Hukou	-0.635	-0.540	-0.636	-0.615
	(18.111)	(18.095)	(0.555)	(0.554)
age squared/100 × cohort × Hukou	-22.465	-24.188	0.822	0.830
	(38.572)	(38.917)	(0.607)	(0.608)
age squared/100 × coastal	11.886	12.581	-0.888	-0.897
	(22.174)	(22.228)	(0.821)	(0.820)
age squared/100 × cohort × coastal	-24.894	-26.653	0.770	0.814
	(46.206)	(46.824)	(0.878)	(0.879)
age squared/100 × Hukou × coastal	4.988	5.227	0.421	0.435
	(27.483)	(27.482)	(0.947)	(0.945)
age squared/100 × cohort × Hukou × coastal	34.895	35.724	-0.397	-0.441
	(57.180)	(57.664)	(1.006)	(1.005)
age cubed/1000 × cohort	-5.048	-5.303	-0.039**	-0.041**
	(4.084)	(4.142)	(0.015)	(0.015)
age cubed/1000 × Hukou	0.186	0.174	0.000	0.000
	(1.738)	(1.737)	(0.009)	(0.009)
age cubed/1000 × cohort × Hukou	2.965	3.215	0.002	0.004
	(4.780)	(4.829)	(0.016)	(0.016)

age cubed/1000 × coastal	-1.027	-1.094	0.014	0.014
	(2.130)	(2.135)	(0.011)	(0.011)
age cubed/1000 × cohort × coastal	2.851	3.052	-0.022	-0.022
	(5.707)	(5.790)	(0.020)	(0.020)
age cubed/1000 × Hukou × coastal	-0.597	-0.618	-0.011	-0.011
	(2.637)	(2.637)	(0.012)	(0.012)
age cubed/1000 × cohort × Hukou × coastal	-4.800	-4.914	0.032	0.033
	(7.066)	(7.132)	(0.022)	(0.022)
male × cohort	0.329***	0.316***		
	(0.092)	(0.093)		
male × Hukou	0.373***	0.392***		
	(0.078)	(0.079)		
male × cohort × Hukou	-0.065	-0.066		
	(0.109)	(0.110)		
male × coastal	0.063	0.062		
	(0.094)	(0.095)		
male × cohort × coastal	-0.250	-0.240		
	(0.131)	(0.133)		
male × Hukou × coastal	-0.205	-0.196		
	(0.119)	(0.119)		
male × cohort × Hukou × coastal	0.226	0.206		
	(0.164)	(0.165)		
live father × cohort			0.460	0.489
			(0.330)	(0.332)
live father × Hukou			0.611*	0.605*
			(0.250)	(0.248)
live father × cohort × Hukou			-0.462	-0.473
			(0.369)	(0.371)
live father × coastal			-0.001	0.003
			(0.314)	(0.312)
live father × cohort × coastal			-0.306	-0.328
			(0.448)	(0.450)
live father × Hukou × coastal			0.173	0.172
			(0.403)	(0.400)
live father × cohort × Hukou × coastal			-0.196	-0.185
			(0.537)	(0.537)
live mother × cohort			-0.391	-0.397

			(0.566)	(0.564)
live mother × Hukou			-0.846	-0.851
			(0.475)	(0.470)
live mother × cohort × Hukou			0.425	0.445
			(0.597)	(0.595)
live mother × coastal			-0.281	-0.261
			(0.653)	(0.650)
live mother × cohort × coastal			0.070	0.051
			(0.781)	(0.781)
live mother × Hukou × coastal			0.438	0.420
			(0.705)	(0.701)
live mother × cohort × Hukou × coastal			-0.183	-0.166
			(0.861)	(0.861)
Observations	22,313	22,313	22,313	22,313

Note: Standard errors in parentheses. *** significant at 0.01, ** significant at 0.05, * significant at 0.1.
The data are from the China Family Panel Studies (CFPS) for 2010, 2012, 2014, and 2016. The sample includes children who are at least 22 years old and parents who are at most 64 years old. The early cohort comprises children born between 1970 and 1980. The late cohort comprises children born between 1981 and 1988. Both the child's income and the parents' income refer to annual income averaged across at least two waves of the CFPS in 2010, 2012, 2014, and 2016. Income is adjusted to 2010 prices using the CPI.

Table C2: Income Equation with and without Bias Corrected

Outcome Variable: Ln (Income)				
	Children		Parents	
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
Inverse Mills Ratio (lambda)		-0.141***		-0.182***
		(0.037)		(0.039)
education	0.052***	0.048***	0.070**	0.068**
	(0.009)	(0.009)	(0.022)	(0.022)
cohort (late = 1)	-92.859	-98.740	31.603	28.469
	(77.332)	(77.255)	(20.182)	(20.139)
coastal region	-34.026	-28.877	15.499	20.517
	(79.958)	(79.874)	(33.141)	(33.069)
Hukou (rural = 1)	-80.855	-75.380	3.338	5.130
	(67.023)	(66.958)	(22.662)	(22.604)
age	-3.301	-3.286	0.819	0.675
	(5.090)	(5.084)	(0.681)	(0.680)
age squared/100	9.427	9.299	-0.745	-0.614
	(14.773)	(14.756)	(0.603)	(0.602)
age cubed/1000	-0.901	-0.880	0.004	0.006
	(1.424)	(1.422)	(0.010)	(0.010)
male	0.285***	0.270***		
	(0.059)	(0.059)		
live father			0.263	0.292
			(0.241)	(0.240)
live mother			-0.655	-0.760
			(0.681)	(0.680)
education × cohort	-0.021	-0.021	0.007	0.003
	(0.012)	(0.012)	(0.025)	(0.025)
education × Hukou	-0.013	-0.010	-0.017	-0.020
	(0.010)	(0.010)	(0.025)	(0.025)
cohort × Hukou	182.468*	181.681*	-13.451	-15.400
	(89.165)	(89.059)	(23.449)	(23.389)
education × cohort × Hukou	0.038**	0.036**	-0.009	-0.006
	(0.014)	(0.014)	(0.028)	(0.028)
education × coastal	-0.001	-0.001	-0.016	-0.011
	(0.012)	(0.012)	(0.028)	(0.028)

cohort × coastal	98.910	96.694	-23.249	-27.748
	(105.867)	(105.743)	(34.007)	(33.929)
education × cohort × coastal	0.035*	0.033	0.012	0.012
	(0.017)	(0.017)	(0.032)	(0.032)
Hukou × coastal	118.599	102.112	-49.341	-54.417
	(97.639)	(97.617)	(39.047)	(38.958)
education × Hukou × coastal	0.017	0.017	-0.009	-0.008
	(0.015)	(0.015)	(0.034)	(0.034)
cohort × Hukou × coastal	-146.481	-140.318	60.774	64.788
	(129.227)	(129.083)	(39.937)	(39.839)
education × cohort × Hukou × coastal	-0.045*	-0.043*	0.008	0.006
	(0.021)	(0.021)	(0.038)	(0.038)
male × cohort	0.002	-0.023		
	(0.081)	(0.081)		
male × Hukou	0.174*	0.135		
	(0.071)	(0.072)		
male × cohort × Hukou	-0.125	-0.122		
	(0.096)	(0.096)		
male × coastal	0.082	0.078		
	(0.085)	(0.085)		
male × cohort × coastal	-0.151	-0.125		
	(0.114)	(0.114)		
male × Hukou × coastal	-0.116	-0.099		
	(0.106)	(0.106)		
male × cohort × Hukou × coastal	0.310*	0.290*		
	(0.140)	(0.140)		
age × cohort	9.738	10.402	-1.206	-1.095
	(7.899)	(7.892)	(0.716)	(0.715)
age × Hukou	6.950	6.431	-0.179	-0.245
	(5.850)	(5.844)	(0.798)	(0.796)
age × cohort × Hukou	-18.968*	-19.038*	0.581	0.647
	(9.113)	(9.103)	(0.832)	(0.830)
age × coastal	2.917	2.422	-0.592	-0.768
	(6.980)	(6.973)	(1.154)	(1.152)
age × cohort × coastal	-10.823	-10.701	0.871	1.023
	(10.793)	(10.780)	(1.192)	(1.189)
age × Hukou × coastal	-10.225	-8.739	1.833	2.009

	(8.513)	(8.512)	(1.364)	(1.361)
age × cohort × Hukou × coastal	13.698	13.485	-2.247	-2.381
	(13.176)	(13.160)	(1.403)	(1.399)
age squared/100 × cohort	-34.699	-37.244	1.133	1.034
	(28.021)	(27.996)	(0.641)	(0.639)
age squared/100 × Hukou	-19.889	-18.261	0.170	0.231
	(16.953)	(16.938)	(0.706)	(0.704)
age squared/100 × cohort × Hukou	66.860*	67.684*	-0.597	-0.653
	(32.365)	(32.327)	(0.743)	(0.741)
age squared/100 × coastal	-8.334	-6.764	0.544	0.701
	(20.232)	(20.212)	(1.008)	(1.006)
age squared/100 × cohort × coastal	40.194	40.204	-0.825	-0.959
	(38.216)	(38.170)	(1.048)	(1.045)
age squared/100 × Hukou × coastal	29.261	24.815	-1.648	-1.805
	(24.644)	(24.642)	(1.193)	(1.190)
age squared/100 × cohort × Hukou × coastal	-43.527	-44.294	2.055	2.174
	(46.690)	(46.635)	(1.234)	(1.231)
age cubed/1000 × cohort	4.194	4.525	-0.003	0.001
	(3.426)	(3.423)	(0.015)	(0.015)
age cubed/1000 × Hukou	1.893	1.724	-0.001	-0.001
	(1.631)	(1.630)	(0.011)	(0.011)
age cubed/1000 × cohort × Hukou	-7.979*	-8.147*	0.001	0.000
	(3.963)	(3.958)	(0.016)	(0.016)
age cubed/1000 × coastal	0.798	0.634	-0.011	-0.012
	(1.947)	(1.945)	(0.014)	(0.014)
age cubed/1000 × cohort × coastal	-5.051	-5.109	0.014	0.015
	(4.665)	(4.660)	(0.020)	(0.020)
age cubed/1000 × Hukou × coastal	-2.783	-2.342	0.009	0.010
	(2.369)	(2.368)	(0.016)	(0.016)
age cubed/1000 × cohort × Hukou × coastal	4.716	4.983	-0.011	-0.015
	(5.707)	(5.701)	(0.023)	(0.023)
live father × cohort			0.095	0.030
			(0.340)	(0.340)
live father × Hukou			0.092	-0.003
			(0.313)	(0.313)
live father × cohort × Hukou			0.101	0.164
			(0.409)	(0.408)

live father × coastal			0.044	0.049
			(0.348)	(0.347)
live father × cohort × coastal			0.262	0.286
			(0.455)	(0.454)
live father × Hukou × coastal			-0.744	-0.733
			(0.503)	(0.501)
live father × cohort × Hukou × coastal			0.344	0.360
			(0.602)	(0.600)
live mother × cohort			0.328	0.342
			(0.785)	(0.783)
live mother × Hukou			0.941	1.061
			(0.701)	(0.700)
live mother × cohort × Hukou			-0.460	-0.498
			(0.814)	(0.812)
live mother × coastal			1.276	1.163
			(0.959)	(0.957)
live mother × cohort × coastal			-0.413	-0.253
			(1.066)	(1.064)
live mother × Hukou × coastal			-1.169	-1.050
			(1.012)	(1.010)
live mother × cohort × Hukou × coastal			0.470	0.289
			(1.141)	(1.139)
Observations	5,820	5,820	3,923	3,923
R squared	0.233	0.235	0.244	0.248

Note: Standard errors in parentheses. *** significant at 0.01, ** significant at 0.05, * significant at 0.1.
The data are from the China Family Panel Studies (CFPS) for 2010, 2012, 2014, and 2016. The sample includes children who are at least 22 years old and parents who are at most 64 years old. The early cohort comprises children born between 1970 and 1980. The late cohort comprises children born between 1981 and 1988. Both the child's income and the parents' income refer to annual income averaged across at least two waves of the CFPS in 2010, 2012, 2014, and 2016. Income is adjusted to 2010 prices using the CPI.

Table C3: Summary Statistics for Computed Income

	Mean (Standard Deviation)	
	Early Cohort	**Late Cohort**
Income of children	22,185.44	23,761.27
	(7,674.016)	(7,380.334)
Income of parents	17,978.72	20,293.59
	(8,161.698)	(8,223.27)
Log(income of children)	9.949	10.030
	(0.344)	(0.304)
Log(income of parents)	9.717	9.851
	(0.383)	(0.357)
Observations	10,980	11,333

Note: The data are from the China Family Panel Studies (CFPS) for 2010, 2012, 2014, and 2016. The sample includes children who are at least 22 years old and parents who are at most 64 years old. The early cohort comprises children born between 1970 and 1980. The late cohort comprises children born between 1981 and 1988. Both the child's income and the parents' income refer to annual income averaged across at least two waves of the CFPS in 2010, 2012, 2014, and 2016. Income is adjusted to 2010 prices using the CPI.

Table C4: Robustness of the Main IGE Estimates

Outcome Variable: Ln (Predicted Income of Children)				
	Early Cohort		Late Cohort	
	(1)	(2)	(3)	(4)
ln (predicted income of parents)	0.390***	0.391***	0.442***	0.448***
	(0.009)	(0.010)	(0.008)	(0.008)
child's age	-0.051	-0.054*	0.022	0.012
	(0.028)	(0.024)	(0.027)	(0.033)
child's age squared/100	0.051	0.055	-0.028	-0.013
	(0.041)	(0.035)	(0.053)	(0.065)
father's age	-0.217***	-0.219***	0.042***	0.043***
	(0.037)	(0.035)	(0.008)	(0.010)
father's age squared/100	0.197***	0.199***	-0.030***	-0.031***
	(0.031)	(0.029)	(0.008)	(0.009)
Observations	10,980	10,980	11,333	11,333
R-squared	0.198	0.195	0.255	0.245

Note: Bootstrapped standard errors in parentheses. *** significant at 0.01, ** significant at 0.05, * significant at 0.1. Columns (1) and (3) present estimates using number of child's live siblings to address selection bias. Columns (2) and (4) show robustness checks, using the number of child's live siblings and its interactions with cohort, coastal, and *hukou* dummies to address selection bias. The data are from the China Family Panel Studies (CFPS) for 2010, 2012, 2014, and 2016. The sample includes children who are at least 22 years old and parents who are at most 64 years old. The early cohort comprises children born between 1970 and 1980. The late cohort comprises children born between 1981 and 1988. Both the child's income and the parents' income refer to annual income averaged across at least two waves of the CFPS in 2010, 2012, 2014, and 2016. Income is adjusted to 2010 prices using the CPI.

Appendix D: Correlates of Changes in Intergenerational Income Persistence in China

Appendix D1: Regression Specification and Variable Definitions

We investigate correlation between the change in correlates and the change in IGE using the following regression equation:

$$(D1) \qquad \hat{IGE}_{it} - \hat{IGE}_{i,t-1} = \gamma_0 + \gamma_1(Correlate_{it} - Correlate_{i,t-1}) + \nu_i,$$

where $\hat{IGE}_{it}$ ($\hat{IGE}_{i,t-1}$) is the IGE estimate for the late (early) cohort in province i, based on the CFPS data. Table D1 reports IGE estimates. CFPS surveys 25 provinces/municipalities, but since Beijing and Tianjin contain fewer than 100 observations in our sample, we exclude these two municipalities and in total have 23 provinces/municipalities across two cohorts.[3] $Correlate_{it}$ ($Correlate_{i,t-1}$) is the specific correlate for the late (early) birth cohort, which is described in the section below. We use correlates around the years of 1990 and 2000 from the China Compendium of Statistics and other datasets to measure the socioeconomic environment when children in the early and late cohorts, respectively, were growing up.[4] Standard errors are bootstrapped.

To study intergenerational income persistence for children from poor or rich families, we further examine the association between change in the correlates and change in the expected income rank if parental income is at the bottom 20$^{\text{th}}$ or top 20$^{\text{th}}$ national percentile rank. Although our regression analysis is at the provincial level, the calculation of children's expected rank is based on a common scale at the national level, which validates cross-province comparison. For a child born in the late (early) cohort in province i, we use $rank^{20}_{it}$ ($rank^{20}_{i,t-1}$) to denote the expected percentile rank of the child in his/her national income distribution if his/her parents are at the 20$^{\text{th}}$ rank in the national distribution of their generation. Following Chetty *et al.* (2014),

$$(D2) \qquad rank^{20}_{it} = \hat{\beta_{0,it}} + 20\hat{\beta_{1,it}},$$

where $\hat{\beta_{0,it}}$ and $\hat{\beta_{1,it}}$ are estimates of intercept and rank correlation from Equation (3) for the late cohort in province i based on CFPS. A large value of $rank^{20}_{it}$ indicates a high expected rank for a child from a poor family. To study intergenerational persistence for rich families, we define

[3] We exert caution in interpreting cross-province comparisons, since the CFPS oversamples five provinces (Shanghai, Liaoning, Henan, Gansu, and Guangdong) to ensure scientific comparison across these five provinces. The rest is drawn through weighting to ensure that the overall sample is nationally representative (see detailed description in Section 3.1). Thus, strictly speaking, provinces/municipalities other than the five oversampled ones cannot be used to draw precise estimates for provincial comparison. To the best of our knowledge, however, currently there is no other way to overcome this data limitation. Therefore, our cross-province comparison provides suggestive evidence only.

[4] Detailed data sources are described in Table D2.

$rank^{80}$ in a similar way as $rank^{20}$. By replacing the dependent variable with $rank^{20}_{it} - rank^{20}_{i,t-1}$ ($rank^{80}_{it} - rank^{80}_{i,t-1}$) in Equation ($D1$), we investigate the correlation between change in correlates and change in $rank^{20}$ ($rank^{80}$).

Following Chetty *et al.* (2014), we have standardized both measures of cohort-province-specific intergenerational income persistence and their correlates. Correlates of changes in intergenerational income persistence, which are detailed in the next section, are measured in different units. Similarly, we use three different measures for intergenerational persistence. Standardizing these variables facilitates comparison of magnitudes of univariate regression coefficients estimated by Equation ($D1$). Table D3 presents summary statistics for both non-standardized and standardized measures of intergenerational persistence and correlates at the provincial level.

Appendix D2: Regression Results

Table D4 presents estimation results of Equation ($D1$). Column (1) shows the results using the IGE as the measure for intergenerational income persistence. Columns (2) and (3) present results with the ranks of children born to parents at the bottom and top 20^{th} percentile national ranks, respectively, as outcome variables. We note that province-level variables in Equation ($D1$) could be associated with each other or with omitted variables. Our estimates are thus interpreted as correlation rather than causality. Perhaps because of the small sample size at the provincial level, all but four estimates are statistically insignificant.

Market-oriented Structural Changes Panel B presents the association between changes in intergenerational income persistence and structural changes. Specifically, we use changes in the share of primary, secondary, and tertiary industry; outflow migration rate; urbanization rate; and share of private enterprises as proxy variables for structural changes. We find that the IGE in source regions falls by 1.1 standard deviations when the migration rate increases by 1 standard deviation. The estimate is statistically significant at the 10 percent level. This result is consistent with the findings in Table 5 that increasing intergenerational persistence is less evident in rural and noncoastal areas. With a 1-standard-deviation increase in the share of private enterprises, the rise in the IGE is as large as 4 standard deviations, and is statistically significant at the 5 percent level. Other estimates generally have the expected signs, but are statistically insignificant.

Economic Development Panel C presents the association between economic development, captured by changes in gross regional product (GRP) per capita and poverty rate, and the change in intergenerational persistence.[5] Our results show that the IGE is positively correlated with both

[5] Because of data availability, the poverty rate is by rural and urban areas with at least 100 observations in each province-rural/urban cell.

GRP and poverty rates. With the rising GRP and declining poverty rate, the income rank at the national level for children of the poor (bottom 20 percent) decreases, while that for children of the rich (top 20 percent) increases. All estimates are statistically insignificant.

Public Expenditure and Expansion of Tertiary Education Panel D presents the association between changes in intergenerational persistence and changes in government expenditure on education, science, culture, and public health per capita and university students per 10,000 people. The estimation result shows that intergenerational income persistence rises with increasing expenditure on education and university enrollment rate. Specifically, with a one-standard-deviation increase in public expenditure per capita, the IGE rises by 2.4 standard deviations. The expected income rank of children born to parents at the bottom 20^{th} percentile rank decreases by 1.2 standard deviations with a one-standard-deviation increase in the university enrollment rate. The two estimates are statistically significant at the 5 percent and 10 percent levels, respectively. The evidence echoes Lai *et al.* (2011), who report that children from low-income families find it increasingly difficult to attend elite schools.[6]

We exercise caution in interpreting the bivariate correlation analysis results in this section. When we apply the Bonferroni correction for the multiple hypothesis testing for each measure of intergenerational income persistence, all estimates in Table D4 are statistically insignificant. For each measure of intergenerational income persistence, we also carry out multivariate correlation analyses by including all factors in one regression.[7] We find that none of the variables is statistically significant; in fact, they are jointly statistically insignificant. The p-values of the F-statistics vary between 0.188 and 0.798.[8]

[6]Li *et al.* (2013) show that the share of students in elite universities from rural and western regions has decreased. In 2010, 22 percent of college students come from families with an annual income that is less than the average annual expenditure of college students. Loans and scholarships account for less than 10 percent of the annual expenditure for college students.

[7]The variable "share of tertiary industry is excluded because of collinearity with "share of primary industry" and "share of secondary industry".

[8]These results are available upon request.

Table D1: IGE Estimates by Cohort in 23 Provinces/Municipalities

	IGE	
Province	Early cohort	Late cohort
(1)	(2)	(3)
Hebei	0.246	0.386
Shanxi	0.239	0.223
Liaoning	0.275	0.507
Jilin	0.341	0.192
Heilongjiang	0.263	0.248
Shanghai	0.218	0.341
Jiangsu	0.061	0.359
Zhejiang	0.123	0.067
Anhui	0.136	0.197
Fujian	0.254	0.636
Jiangxi	0.266	0.275
Shandong	0.208	0.343
Henan	0.271	0.257
Hubei	0.323	0.213
Hunan	0.274	0.19
Guangdong	0.212	0.419
Guangxi	0.309	0.234
Chongqing	0.364	0.27
Sichuan	0.412	0.486
Guizhou	0.257	0.331
Yunnan	0.234	0.387
Shannxi	0.168	0.283
Gansu	0.398	0.334

Note: Beijing and Tianjin in both cohorts are excluded, as the observations in each cohort are fewer than 100.

Table D2: Data Sources for Correlates with Intergenerational Income Persistence

Variable	**Data**	**Notes**
Gini coefficient	Chinese Statistical Yearbooks in 1990 and 1999	Data are by rural and urban areas, as reported in the statistical yearbooks. Observations in cohort-hukou-province cells with fewer than 100 observations are excluded.
Share of primary industry	China Compendium of Statistics 1949-2008 in 1990 and 2000	Data cover all 25 provinces in CFPS.
Share of secondary industry	China Compendium of Statistics 1949-2008 in 1990 and 2000	Data cover all 25 provinces in CFPS.
Share of tertiary industry	China Compendium of Statistics 1949-2008 in 1990 and 2000	Data cover all 25 provinces in CFPS.
Outflow migration rate	Census data in 1990 and 2000	The outflow migration rate is calculated for the age cohorts of 16 to 26 with high rates of migration. Data cover 23 provinces in the CFPS, with missing values in Sichuan and Chongqing. The latter was separated as a municipality from the former in 1998.
Urbanization rate	China Compendium of Statistics 1949-2008 in 1990 and 2000	Data cover 20 provinces in the CFPS, with missing values in Zhejiang, Fujian, Chongqing, Sichuan, and Shannxi. The value in Guangdong in 1990 is replaced by the one for 1989 because of data availability.

Share of private enterprises	China Compendium of Statistics 1949-2008 in 1990 and 1997	Statistics for 2000 are not available, and are thus replaced by those for 1997. Data cover all 25 provinces in the CFPS.
Per capita GRP (1,000 yuan)	China Compendium of Statistics 1949-2008 in 1990 and 2000	Data cover all 25 provinces in the CFPS.
Urban poverty rate	China Compendium of Statistics 1949-2008 in 2000 and 2008	Statistics for 1990 and 2000 are not available, and thus are replaced by those for 2000 and 2008 sequentially. Data cover 24 provinces in CFPS, with missing values for Jilin. The values for Hebei, Jiangsu, Sichuan, and Shannxi in 2000 are replaced by those for 2003, 2002, 2005, and 2001, sequentially, due to data availability. The value for Hubei in 2008 is replaced by the one for 2004.
Rural poverty rate	China Compendium of Statistics 1949-2008 in 2000 and 2008	Statistics for 1990 and 2000 are not available, and are thus replaced by those for 2000 and 2008 sequentially. Data cover 17 provinces in the CFPS, with missing values for Tianjin, Shanxi, Jilin, Jiangsu, Shandong, Henan, Hubei, and Chongqing. Values for Heilongjiang and Guizhou for 2000 are replaced by those for 2002 and 2005 sequentially.
Educational expenditure per capita	China Compendium of Statistics 1949-2008 in 1990 and 2000	Data cover all 25 provinces in the CFPS. The value for Sichuan in 1990 is replaced by the one for 1994 because of data availability.

University students per 10,000 people	China Compendium of Statistics 1949-2008 in 1990 and 2000	Data cover all 24 provinces in the CFPS, with missing value for Beijing. The value for Sichuan for 1990 is replaced by the one for 1994 because of data availability.

Table D3: Summary Statistics for Intergenerational Persistence and Correlates at Provincial Level

	Mean (Standard Deviation)			
	Early Cohort		Late Cohort	
	Non-standardized	Standardized	Non-standardized	Standardized
Intergenerational income elasticity	0.254	-0.265	0.312	0.265
	(0.084)	(0.773)	(0.124)	(1.141)
Rank of child with parents from 20th national rank	12.096	0.494	6.060	-0.494
	(7.106)	(1.164)	(2.588)	(0.424)
Rank of child with parents from 80th national rank	29.095	0.073	27.616	-0.073
	(10.750)	(1.065)	(9.581)	(0.949)
Gini coefficient	0.246	-0.507	0.286	0.485
	(0.038)	(0.940)	(0.033)	(0.809)
Share of primary industry	0.291	0.609	0.173	-0.609
	(0.088)	(0.903)	(0.066)	(0.674)
Share of secondary industry	0.408	-0.179	0.434	0.179
	(0.082)	(1.124)	(0.062)	(0.845)
Share of tertiary industry	0.300	-0.754	0.393	0.754
	(0.035)	(0.566)	(0.045)	(0.732)
Outflow migration rate [a]	0.042	-0.929	0.178	0.929
	(0.012)	(0.164)	(0.033)	(0.460)
Urbanization rate	0.317	-0.221	0.397	0.181
	(0.191)	(0.959)	(0.203)	(1.018)
Share of private enterprises	0.105	-0.215	0.212	0.215
	(0.237)	(0.950)	(0.256)	(1.023)
Per capita GRP (1,000 yuan)	1.718	-0.622	8.152	0.622
	(1.040)	(0.201)	(5.659)	(1.094)
Poverty rate	0.130	-0.130	0.271	0.137
	(0.174)	(0.330)	(0.740)	(1.404)
Expenditure on education, science, culture, & public health per capita	50.876	-0.523	211.809	0.523
	(23.375)	(0.152)	(185.115)	(1.204)
University students per 10,000 people	20.426	-0.517	46.451	0.517
	(16.739)	(0.665)	(25.741)	(1.022)

Note: Beijing and Tianjin in both cohorts are excluded, as the observations in each cohort are fewer than 100. The data are from the China Compendium of Statistics, 1949–2008 in 1990 and 2000, except for the outflow migration rate, which is from the 1990 and 2000 censuses. The data on the share of private enterprise are from 1990 and 1997. Poverty rates are recorded separately for urban and rural areas, and the data are from 2000 and 2008. Gini coefficients are recorded separately for urban and rural areas, and the data are from 1990 and 1999. If data are missing in a specified year, the closest alternative within a five-year window is used instead. In total, there are 12 replacements for all data.

[a] The outflow migration rate is calculated for the 16- to 26-year-olds, who have high rates of migration.

Table D4: Correlates of Changes in Intergenerational Income Persistence

	Outcome Variable		
Correlates	Intergenerational income elasticity	Rank of child born to parents at the bottom 20th percentile rank	Rank of child born to parents at the top 20th percentile rank
(1)	(2)	(3)	(4)
Panel A. Inequality			
Gini coefficient	0.350	-0.932	-0.109
	(0.578)	(0.499)	(0.341)
Panel B. Structural Changes			
Share of primary industry	0.671	-0.712	-0.083
	(0.723)	(0.771)	(0.677)
Share of secondary industry	0.595	-0.025	0.116
	(0.592)	(0.519)	(0.300)
Share of tertiary industry	-0.493	0.077	-0.074
	(0.383)	(0.406)	(0.243)
Outflow migration rate[a]	-1.084	0.299	-0.120
	(0.648)	(0.749)	(0.466)
Urbanization rate	-0.054	0.131	0.080
	(0.599)	(0.747)	(0.920)
Share of private enterprises	4.012	-0.682	1.296
	(1.778)	(1.735)	(1.395)
Panel C. Economic Development			
Per capita GRP	1.307	-0.350	0.310
	(0.819)	(0.706)	(0.557)
Poverty rate	0.559	0.063	-0.474
	(0.835)	(1.632)	(1.411)
Panel D. Redistribution and Education Policies			
Expenditure on education, science, culture, & public health per capita	2.379	-1.297	0.104
	(1.136)	(1.259)	(0.894)
University students per 10,000 people	0.556	-1.230	-0.434
	(0.713)	(0.741)	(0.637)

Note:All variables have been normalized. Beijing and Tianjin in both cohorts are excluded, as the number of observations in each cohort are fewer than 100.Column (1) lists the correlates of changes in intergenerational income persistence. Each cell in Columns (2)–(4) presents the OLS estimate of γ_1 in Equation (D1) in Appendix D, with bootstrapped standard errors in parentheses. Specifically, Columns (2), (3), and (4) show, respectively, the OLS estimates of the changes in the correlates with the changes in intergenerational income elasticity, the rank of a child born to parents at the bottom 20^{th} national percentile rank, and the rank of a child born to parents at the top 20^{th} national percentile rank. The data are from the China Compendium of Statistics, 1949–2008 in 1990 and 2000, except for the outflow migration rate, which is from the 1990 and 2000 censuses. The data on the share of private enterprise are from 1990 and 1997. Poverty rates are recorded separately for urban and rural areas, and the data are from 2000 and 2008. Gini coefficients are recorded separately for urban and rural areas, and the data are from 1990 and 1999. If data are missing in a specified year, the closest alternative within a five-year window is used instead. In total, there are 12 replacements for all data.

[a] The outflow migration rate is calculated for the 16- to 26-year-olds, who have high rates of migration.

References

Becker, Gary S., & Tomes, Nigel. 1986. Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, **4**(3), S1–S39.

Chetty, Raj, Hendren, Nathaniel, Kline, Patrick, & Saez, Emmanuel. 2014. Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *Quarterly Journal of Economics*, **129**(4), 1553–1623.

China's Official Geographic Data. 2015. *China's Official Geographic Data*. Available at `http://www.resdc.cn/data.aspx?DATAID=200` by registration.

Chow, Gregory C., & Shen, Yan. 2006. Demand for Education in China. *International Economic Journal*, **20**(2), 129–147.

Ge, Suqin, & Yang, Dennis Tao. 2011. Labor Market Developments in China: A Neoclassical View. *China Economic Review*, **22**(4), 611–625.

Ge, Suqin, & Yang, Dennis Tao. 2014. Changes in China's Wage Structure. *Journal of the European Economic Association*, **12**(2), 300–336.

Heckman, James J. 2005. China's Human Capital Investment. *China Economic Review*, **16**(1), 50–70.

Heckman, James J., & Yi, Junjian. 2014. Human Capital, Economic Growth, and Inequality in China. *The Oxford Companion to the Economics of China*, 459.

Knight, John, Sicular, Terry, & Yue, Ximing. 2011. *Educational Inequality in China: The Intergenerational Dimension*. Tech. rept. CIBC Working Paper No. 201113, University of Western Ontario.

Lai, Fang, Sadoulet, Elisabeth, & De Janvry, Alain. 2011. The Contributions of School Quality and Teacher Qualifications to Student: Performance Evidence from a Natural Experiment in Beijing Middle Schools. *Journal of Human Resources*, **46**(1), 123–153.

Li, Hongbin, Li, Lei, Wu, Binzhen, & Xiong, Yanyan. 2012. The End of Cheap Chinese Labor. *Journal of Economic Perspectives*, **26**(4), 57–74.

Li, Hongbin, Meng, Lingsheng, Shi, Xinzheng, & Wu, Binzhen. 2013. Poverty in China's Colleges and the Targeting of Financial Aid. *China Quarterly*, **216**, 970–992.

Li, Shi, & Xing, Chunbing. 2010. China's Higher Education Expansion and its Labor Market Consequences. *IZA Discussion Paper*, **No. 4974**.

National Bureau of Statistics of China Statistical Yearbook. 2013a. *Government educational expenditure and GDP*. Available at `http://www.stats.gov.cn/tjsj/ndsj/2013/indexeh.htm`.

National Bureau of Statistics of China Statistical Yearbook. 2013b. *Tuition, government educational expenditure, and GDP*. Available at http://www.stats.gov.cn/tjsj/ndsj/2013/indexeh.htm.

NBS. 2011. *China Statistics Yearbook*. Beijing: China Statistics Press.

Yuan, Zhigang, & Chen, Lin. 2013. The Trend and Mechanism of Intergenerational Income Mobility in China: An Analysis from the Perspective of Human Capital, Social Capital and Wealth. *World Economy*, **36**(7), 880–898.

Zhu, Xiaodong. 2012. Understanding China's Growth: Past, Present, and Future. *Journal of Economic Perspectives*, **26**(4), 103–124.