

DISCUSSION PAPER SERIES

IZA DP No. 12804

**Rising Intergenerational Income
Persistence in China**

Yi Fan
Junjian Yi
Junsen Zhang

NOVEMBER 2019

DISCUSSION PAPER SERIES

IZA DP No. 12804

Rising Intergenerational Income Persistence in China

Yi Fan

National University of Singapore

Junjian Yi

National University of Singapore and IZA

Junsen Zhang

*Chinese University of Hong Kong, Zhejiang
University and IZA*

NOVEMBER 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Rising Intergenerational Income Persistence in China*

This paper documents an increasing intergenerational income persistence in China since economic reforms were introduced in 1979. The intergenerational income elasticity increases from 0.390 for the 1970–1980 birth cohort to 0.442 for the 1981–1988 birth cohort; this increase is more evident among urban and coastal residents than rural and inland residents. We also explore how changes in intergenerational income persistence is correlated with market reforms, economic development, and policy changes.

JEL Classification: E24, J62, O15

Keywords: intergenerational income persistence, economic transition, great gatsby curve

Corresponding author:

Junsen Zhang
Department of Economics
Chinese University of Hong Kong
Shatin, NT, Hong Kong
Region of the People's Republic of China
E-mail: 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'.

China has experienced fast economic growth since the market-oriented reform in 1979. The real GDP per capita increased almost twenty-fold from 1980 to 2016. Over the same period, the Gini coefficient rose from 0.31 to 0.47 (Figure 1). Rising income inequality has become a major policy issue in China (Zhu, 2012; Xie & Zhou, 2014). Cross-sectional inequality, as measured by the Gini coefficient, portrays only a snapshot of inequality; a full analysis of the income distribution should include not only income inequality across different families in the same generation, but also across different generations of the same family, i.e., intergenerational income persistence (Becker & Tomes, 1979, 1986). To fully understand and address the rising income inequality during China's economic reform, we need to investigate how and why intergenerational income persistence has changed.

We study three sets of structural reforms that have likely impacted the evolution of intergenerational income persistence.¹ First, institutional reforms have reduced poverty, and transformed China from a planned and agricultural economy to a market and industrial economy (Zhu, 2012). Relaxation of the household registration (*hukou*) system has induced substantial domestic temporary migration.² The decline in the poverty rate and the agricultural sector share, along with the rise in the rural-to-urban migration rate, likely reduce intergenerational income persistence. Second, the economic reform has led to an increase in the returns to schooling (Figure B1 in Appendix B), accompanied by an increase in educational costs (Figure B2). In addition, the expansion of higher education, which was intended to improve intergenerational mobility, appears to disproportionately benefit the children of the rich (Li *et al.*, 2013). These factors likely enhance intergenerational income persistence. Third, several socioeconomic changes during the transition period may have ambiguous effects on intergenerational mobility. For example, the number of private enterprises has increased since the Open Door Policy was introduced in 1978.³ Thus, the evolution of intergenerational income persistence during China's transition period remains an empirical question.

We pursue two sets of empirical analyses. The first set is conducted at the individual level. We use the China Family Panel Studies (CFPS) survey data for 2010, 2012, 2014, and 2016 to estimate the intergenerational elasticity (IGE) in income by birth cohort at the national level. We find that the IGE increases from 0.390 for the 1970–1980 birth cohort to 0.442 for the 1981–1988 birth cohort. This pattern of increasing intergenerational income persistence is consistent under alternative measures, such as intergenerational log correlation, rank correlation, and transition matrix. Furthermore, increasing intergenerational income persistence is more evident among sons than daughters, and among urban and coastal residents than rural and inland residents.

The second set of analyses is conducted at the provincial level. We estimate the IGE in income by province and birth cohort, then correlate these estimates with proxy variables for structural changes, economic development, and policy changes. Although this set of analyses may shed light on the causes of rising intergenerational income persistence, we note that these factors may be associated with each other or with omitted variables. Thus our findings serve only as a starting point to understand the underlying mechanisms (Corak, 2013).

¹Appendix A provides a detailed description of the structural reforms in China.

²The *hukou* system is a governmental household registration system in China that officially identifies whether a person's status is rural or urban. Individuals born in rural areas are assigned agricultural or rural *hukou*, while their counterparts in urban areas are assigned nonagricultural or urban *hukou*.

³China has implemented the Open Door Policy since 1978. The policy includes decentralizing decision-making on exports and imports, establishing special economic zones and coastal open cities to attract foreign investment, replacing administrative restrictions on exports and imports with tariffs, quotas, and licensing requirements, and loosening foreign exchange controls (Wei, 1995).

We show the correlation between the IGE in income and the Gini coefficient; this correlation is termed the Great Gatsby Curve (Krueger, 2012; Corak, 2013).⁴ Different from mixed evidence for the U.S. (Chetty *et al.*, 2014b; Davis & Mazumder, 2017), we provide evidence on rising intergenerational income persistence in China since the economic reforms in 1979.⁵ We introduce a time-series dimension in addition to the cross-sectional one, and depict the Great Gatsby Curve across regions and cohorts. We find a positive correlation between the IGE and the Gini coefficient, which implies that rising intergenerational income persistence is associated with rising inequality, as has been documented in many developed countries.

Corak (2013) broadly ascribes variations in the IGE to family influence, labor markets, and public policy. Chetty *et al.* (2014a) focus on the geographic level of commuting zones, and outline five significant factors—residential segregation, income inequality, quality of the primary school attended, social capital, and family stability—that explain differences in intergenerational persistence across the U.S. In the context of China’s economic transition, we classify three categories of factors that are likely correlated with intergenerational income persistence—market-oriented structural changes, economic development, and public policies. We then correlate changes in these factors with changes in our estimates of intergenerational income persistence at the provincial level.

This paper appears to be the first to use the most recent panel data to document the rising intergenerational persistence across birth cohorts amid China’s economic transition. We also explore factors that are correlated with the transmission of economic status across generations. Our analyses suggest that removing migration restrictions and promoting means-tested policies may reduce the rising intergenerational income persistence. Understanding the changing intergenerational income persistence in China over the past 40 years of economic reform may provide insights, and yield policy implications for other countries at a similar stage of economic transition and development.

The rest of the paper is organized as follows. Section I defines four measures of intergenerational income persistence, and discusses strategies to address empirical challenges. Section II describes the data, and Section III presents the empirical results. Section IV presents the correlation between inequality and intergenerational income persistence. Section V discusses policy implications and concludes.

⁴Becker & Tomes (1986) develop a theory of intergenerational income mobility from a human capital perspective, in which inequality at the equilibrium is endogenously determined. Piketty (2000) summarizes theories of persistent inequality and intergenerational mobility. A positive correlation between inequality and intergenerational persistence has been documented across developed countries (Corak, 2004; Björklund & Jäntti, 2009; Blanden, 2013), as well as across areas within the U.S. (Chetty *et al.*, 2014a). Krueger (2012) first uses the term “the Great Gatsby Curve” to describe this positive relationship in a presentation at the Center for American Progress. To date, studies on the Great Gatsby Curve have largely focused on developed countries.

⁵Evidence on the recent trend of intergenerational income mobility in the U.S. is mixed. For example, Davis & Mazumder (2017) show declining intergenerational income mobility in the U.S., comparing cohorts that entered the labor market before and after 1980, when cross-sectional inequality increases. Chetty *et al.* (2014b) find no significant change in intergenerational mobility for children entering the labor market in the 2010s compared to those entering the market in the 1970s.

I Measures of Intergenerational Income Persistence

A Five Measures of Intergenerational Income Persistence

Intergenerational Elasticity The IGE is most commonly used to measure intergenerational income persistence (Solon, 1992; Mazumder, 2005; Corak, 2013; Chetty *et al.*, 2014a). Becker & Tomes (1986) first suggest a log-linear specification between a child’s income and his/her parents’ income. Solon (2004) later provides an economic justification. We start by regressing log child’s lifetime income ($\ln y_{it}$) on log parents lifetime income ($\ln y_{i,t-1}$):⁶

$$(1) \quad \ln y_{it} = \alpha_0 + \alpha_1 \ln y_{i,t-1} + \varepsilon_{it}.$$

The coefficient α_1 measures the percentage change in the child’s income with respect to the percentage change in parental income, i.e., α_1 represents the IGE. A larger α_1 indicates greater intergenerational income persistence, which implies less mobility across generations. We estimate the IGE separately for each birth cohort. Within each cohort, we also estimate the IGE by the child’s gender, *hukou* status (urban vs. rural areas), and region (coastal vs. inland).

Intergenerational Log Correlation The IGE does not take into account the difference in the dispersion of log income between two generations. For instance, elasticity may be low in one society simply because the variance in the log income of children is lower relative to the variance in the log income of parents.⁷ To account for the dispersions of log income, we investigate the intergenerational log correlation, which is defined as

$$(2) \quad \text{intergenerational log correlation} = \alpha_1 * \frac{\sigma_{t-1}}{\sigma_t},$$

where α_1 is the IGE estimated from Equation (1), and σ_{t-1} and σ_t are standard deviations of the log income of parents and children, respectively.

Figure 2 depicts the mean of log children’s income versus the mean of log parental income.⁸ For both the 1970–1980 and 1981–1988 birth cohorts, the slope is flat at the bottom, increases sharply in the middle, and is flat at the top; this pattern is similar to the U.S. pattern documented in Chetty *et al.* (2014a).

Intergenerational Rank Correlation Following the recent literature (Dahl & DeLeire, 2008; Chetty *et al.*, 2014a,b), we transform parents’ and child’s income into percentile ranks between 0 and 100, and present estimates of the rank-rank slope using the following equation:

$$(3) \quad \text{rank}_{it} = \beta_0 + \beta_1 \text{rank}_{i,t-1} + \epsilon_{it},$$

⁶As actual lifetime income is usually unobserved, the literature uses current income or income over various life stages as the proxy variable (Haider & Solon, 2006). Section C presents details of the construction of this variable.

⁷We plot the standard deviation of log income against age groups in Figure B3 for the parents’ and children’s generations.

⁸On the horizontal axis, we first rank log parental income, then sequentially average log parental income by every 100 observations. On the vertical axis, we calculate the mean of their children’s log income.

where $rank_{it}$ is the *national* percentile rank of children by income in each cohort, and $rank_{i,t-1}$ is the *national* percentile rank of parents by income in each cohort. When we investigate gender, *hukou*, and regional patterns, percentile ranks are also defined at the national level by cohort and generation.

Figure 3 shows the correlation between the percentile ranks of children and the percentile ranks of parents. The income rank of children is almost linearly dependent on parental rank, except in the upper tail for the 1981–1988 cohort.⁹

Intergenerational Transition Matrix of Relative Mobility To further explore this nonlinearity, we adopt an intergenerational transition matrix of relative mobility. This matrix displays the percentage of children in quintile i ($i=1,2,3,4,5$) conditional on their parents' income in quintile j ($j=1,2,3,4,5$), following Zimmerman (1992) and Nybom & Stuhler (2016a). This matrix is used to measure relative mobility, as it compares the income percentile of children from poor families with that of children from rich families.

In this transition matrix, we focus on children whose parents are in the bottom quintile. We examine the proportion of children who move into the top quintile, as well as the proportion of children who are trapped in the bottom quintile. We also look at children whose parents are in the top quintile, and measure the proportion who remain in the top quintile. The diagonal elements of the transition matrix may be underestimated while the off-diagonal elements, such as those that imply moving from the very bottom to the very top, may be overestimated (Nybom & Stuhler, 2016a). Since we are focusing on the *differences* between estimates from the two cohorts, such downward or upward biases are likely to be mitigated.

Intergenerational Transition Matrix of Absolute Mobility We also adopt an intergenerational transition matrix of absolute mobility, which displays the fraction of children earning more than 100 percent, 120 percent, and 150 percent of parents' income, conditional on their parents' income in quintile j ($j=1,2,3,4,5$). Different from the relative mobility, which examines the income distribution in the children's generation relative to that in their parents' generation, absolute mobility compares the absolute levels of income across generations. The measure of absolute mobility supplements that of relative mobility in describing intergenerational mobility, as the latter mainly reflects the effect from growth across income distribution, while the former captures additional impacts from growth in the overall size of the economy (Fields & Ok, 1999; Chetty *et al.*, 2017).

Chetty *et al.* (2017) use the fraction of children earning more than their parents as a measure of absolute mobility in the U.S. context. In China's transition period with rapid economic growth, we show proportions of children earning more than 120 percent and 150 percent of their parents' income, in addition to only 100 percent of parents' income. Similar to the analysis of relative mobility, we focus on the *differences* in estimates of absolute mobility between the two cohorts.

B Empirical Challenges

We discuss three empirical challenges in estimating intergenerational income persistence, and the strategies we employ to address these challenges.

⁹We also plot the rank-rank association across generations by cohort. We exclude parents in the top 10 percent of income in Figure B4.

Lifecycle Bias The bias most frequently discussed in the literature on intergenerational mobility is lifecycle bias, which was introduced by Jenkins (1987). Since the correlation between current earnings and lifetime earnings varies systematically over the lifecycle, using current earnings of children—especially at the early stages of life—produces inconsistent estimates of the IGE. The bias depends on the age at which the child’s current earnings are measured (Haider & Solon, 2006).¹⁰ Intergenerational income persistence is likely underestimated if earnings are measured at an early life stage, since the age-earnings profiles of individuals with high lifetime may have steep trajectories. Grawe (2006) shows that 40 percent of the variation in IGE estimates from previous studies is due to the estimation methodology and the father’s age during the observation period. Haider & Solon (2006) find that inclusion of polynomials in age does not fully address lifecycle bias, while Nybom & Stuhler (2016b) find that income measured at the mid-to-late life stage is least subject to lifecycle bias. In China’s context, Gong *et al.* (2012) discuss how the age-earnings profiles of various cohorts are differentially influenced by economic growth.

To address concerns associated with lifecycle bias, we first follow the literature to explicitly control for the change in average income across the lifecycle in Equations (1) and (3) by including age polynomials for children and parents. Second, we carefully choose the age cut-offs for children and parents. For children, we exclude those younger than 22 years, as they are likely to still be in school.¹¹ In addition, variation in children’s income is stable from their early twenties, as shown in Figure B3. To investigate the age-earnings profiles of high-skilled and low-skilled workers, we plot average earnings against age by educational groups in Figure B6, and find that the earnings of high-skilled and low-skilled workers diverge in their early twenties. For parents, we exclude those who are older than 64 years, as most individuals retire after 64. Third, we use parental average schooling as an instrument for parental lifetime income in one set of estimation to examine intergenerational income persistence, which is a common practice in the literature (Solon, 1992; Gong *et al.*, 2012). Fourth, we adopt the intergenerational rank correlation as our main measure since it is the most robust of the four measures with respect to the age at which income is measured (Nybom & Stuhler, 2016a). We also report the IGE and intergenerational log correlation.

As a robustness check, we restrict children’s ages to be at least 24 years, and mothers’ ages to be at most 62 years, since females usually retire earlier than males. Moreover, following Chetty *et al.* (2014a), as educational attainment is more stable than income over the lifecycle, and is thus less subject to measurement errors, we replace children’s income with their educational attainment as the dependent variable. We also use the full sample of parents and children to predict each individual’s lifetime income to alleviate the concern that children in the late cohort are on average younger than children in the early cohort.

Attenuation Bias Attenuation bias comes from transitory fluctuation in income (Solon, 1989, 1992; Mazumder, 2005). Income in a specific year may not be a proper measurement of lifetime income, because it may contain transitory shocks or measurement errors. The estimate is downward biased by a factor of $\sigma_y^2/(\sigma_y^2 + \sigma_v^2)$, where σ_y^2 is the variance of income in either gen-

¹⁰As discussed in Grawe (2006) and Nybom & Stuhler (2016b), using the current earnings of parents may also produce inconsistent estimates of IGE.

¹¹According to the 2000 census, approximately 95 percent of China’s population who are at least 22 years old are not in school, as shown in Figure B5.

eration, and σ_v^2 is the variance of transitory fluctuations in lifetime income (Solon, 1989, 1992).¹² To address such attenuation bias, several studies take an average of income across different years (Solon, 1992; Mazumder, 2005; Lee & Solon, 2009). Mazumder (2005) argues that if substantial persistence in transitory fluctuations exists, even averaging income across four or five years may generate a poor measure of lifetime income. Gong *et al.* (2012) use demographic variables such as education as instruments for parental income when estimating the IGE in China.

Nybom & Stuhler (2016a) use long Swedish income series, and empirically evaluate the biases in standard measures of intergenerational dependence. They find that the intergenerational rank correlation and transition matrix are less subject to attenuation bias than the IGE and intergenerational log correlation. Following Nybom & Stuhler (2016a), we take an average of income across two to six years over the four waves using the longitudinal CFPS data, and report the intergenerational rank correlation and transition matrix.

Selection Bias Conventional household surveys, which interview individuals either living in the household or those who maintain close economic relationships with a household, are subject to two sources of selection bias. The first refers to co-residence bias if individuals choose to stay at home. When children get married, they usually leave their parents' household and form a new household. Household surveys interview either the parents' household or the children's household. The second selection bias arises from temporary migration. Household surveys usually do not record the income of temporary migrants. This type of selection bias might be severe in China, as the number of temporary migrants increased significantly during the economic reform. In our empirical analysis below, we use CFPS data and the Heckman model to address the selection bias.

II Data

A China Family Panel Studies

Studies on intergenerational income persistence in developed countries use long panel data or administrative data with tax information to address the biases discussed above (Lee & Solon, 2009; Chetty *et al.*, 2014a,b; Nybom & Stuhler, 2016b). In a developing country such as China, however, the tax system is less able to provide reliable information on income for intergenerational studies. Of all available household surveys in China, the China Family Panel Studies (2015) is the most appropriate for our research. The CFPS is a nationally representative and biennial longitudinal survey of Chinese individuals, families, and communities. It was launched in 2010 by the Institute of Social Science Survey of Peking University in China, and three follow-up surveys were conducted in 2012, 2014, and 2016. The baseline CFPS contains approximately 15,000 households and 30,000 individuals, with a response rate of 79 percent. Five provinces are chosen for initial oversampling, with 1,600 families drawn from each province to facilitate regional comparisons. The remainder of the sample is drawn from other provinces to render the overall sample nationally representative through weighting.¹³

¹²The basic assumptions are that variances in income y and transitory fluctuations v are the same for both generations. In addition, the v in each generation is correlated with neither each other nor income (Solon, 1989).

¹³This information is from the CFPS website: <http://www.issf.pku.edu.cn/cfps/EN/About/>.

We consider the CFPS to be the best available data set for our study for the following reasons. The CFPS is nationally representative; distributions of important demographic and socioeconomic variables, such as age, gender, schooling, and rural-urban stratification, are consistent with those from the census data (Xu & Xie, 2015). The survey covers urban and rural areas in 25 out of 34 provinces, municipalities, and autonomous regions in its baseline survey (Figure B7 in the appendix), and represents 95 percent of China’s population (Xie & Zhou, 2014). Furthermore, the CFPS employs a scientific stratified multistage probability sampling; the sample is drawn in three stages at the county, village, and household levels (Xie & Zhou, 2014). Similar to the Panel Study of Income Dynamics in the U.S., individuals included in the CFPS are tracked across waves. We are the first to use the panel structure of the CFPS to calculate lifetime income to study intergenerational income persistence. We exploit the CFPS unique survey design to overcome conventional selection bias in estimating intergenerational income persistence, as detailed in Section C.

The CFPS data are of high quality with solid technical and manpower support. First, computer-assisted personal interviewing technology, provided by the Survey Research Center at the University of Michigan, is employed. This technology tailors interviews to each household member, and reduces measurement error. Second, information on key variables, such as education, is collected through various channels and across survey waves to generate the most reliable values. Third, principal investigators are internationally recognized experts in economics, statistics, sociology, and public policy who work with an international advisory committee.¹⁴

B Sample Definition

Our sample contains 22,313 parent-child pairs. To estimate changes in intergenerational income persistence, children born between 1970 and 1980 are considered the early cohort, and children born between 1981 and 1988 are considered the late cohort.¹⁵ The cut-off point of 1980 denotes the beginning of market reforms in China. The early cohort comprises 10,980 parent-child pairs, and the late cohort comprises 11,333 parent-child pairs.

In addition to estimating intergenerational income persistence at the national level, we examine heterogeneous patterns by child’s gender, *hukou* status (urban vs. rural areas when the child was 3 years old), and region (coastal vs. inland). Prior to the economic reform, labor mobility between rural and urban areas was virtually illegal, and the *hukou* system effectively segregated China into two labor markets (Meng & Zhang, 2001). *Hukou* restrictions have gradually been relaxed, leading to an influx of temporary rural-to-urban migrants into cities. Because those temporary migrants do not have the same access to public education, health, pension, and welfare systems as urban citizens, they are included in the rural sample.

C Main Variables

Income Income is the key variable in this research. Annual income is the sum of five categories: wage, farming/self-employment, property (including rents from housing, land, or other family

¹⁴Principal investigators are Yu Xie from Princeton University, Xiaobo Zhang from Peking University and the International Food Policy Research Institute, and Ping Tu and Qiang Ren from Peking University. Detailed descriptions of the CFPS can be found at Xie *et al.* (2014).

¹⁵To facilitate calculation of lifetime income, the youngest children are restricted to be 22 years old in 2010 (born in 1988).

properties), transfers (including government subsidies and social transfers), and others (e.g., gifts in kind). We average individual income across at least two waves in the survey to generate a proxy variable for lifetime income. Income for 2012, 2014, and 2016 is adjusted by the Consumer Price Index to the 2010 price level. Parental income is the sum of the fathers and mothers income.

The income of parents or children might be missing in household surveys. Specifically, we categorize parents-children pairs in CFPS into three types:

(i) Parents and children live in the same household and work in the county of their registered *hukou*. The incomes of both parents and children are recorded by the CFPS.

(ii) Parents and children live in the same household but at least one party temporarily works outside the county of their registered *hukou*. The CFPS does not record the incomes of temporary migrants.

(iii) Parents and children live in different households. The CFPS surveys either the parents or the childrens household, but not both. Income is recorded for individuals in the surveyed household, but not for individuals in the non-surveyed household.

The missing income is likely to induce a standard incidental sample truncation problem (Wooldridge, 2010) when we use only the first type of parents-children pairs to estimate the intergenerational income persistence, as these parents-children pairs may be different from the other two types in many unobserved aspects. Moreover, the sample truncation problem may differentially impact the estimates of intergenerational income persistence across cohorts. Within the same survey year, parents and children in the 1970–1980 cohort are older than parents and children in the 1981–1988 cohort. The probability of living with one’s parents or one’s children is dependent on age. Likewise, the probability of being a temporary migrant changes over the life cycle. Consequently, the probability that the CFPS does not record a particular individual’s income depends on the individual’s age. Therefore, the sample truncation problem affects the two cohorts differently.

The CFPS has unique data characteristics that allow us to address this sample selection issue. In contrast to most conventional household surveys, the CFPS collects some information on temporary migrants and each household member’s relatives (including spouses, lineal relatives, and siblings) who are not living in the same household. The information collected includes demographic and socioeconomic characteristics such as education, employment status, and occupation, but not income. Therefore, even though the full sample of parents and children lacks income information, it has detailed demographic and socioeconomic information.

We take advantage of the CFPS’ unique survey design to correct for selection bias and predict the income for the full sample. First, we estimate the following probit model using the full sample of adult children with and without observed income:

$$(4) \quad I_i = \alpha_0 + \alpha_z z_i + X_i \alpha_X + \epsilon_i,$$

where I_i is a dummy variable that is equal to one if income information for adult child i is recorded in the CFPS, and zero otherwise; z is the number of the adult child’s siblings; X is a comprehensive set of demographic and socioeconomic variables, including education, gender, age, age squared, age cubed, and their full interactions with cohort, *hukou*, and regional dummies. Column (1) in Table C1 of Appendix C reports the estimation results of Equation (4). We also conduct a robustness check including additional interactions between the number of an adult child’s siblings with cohort, *hukou*, and coastal dummies; the results are presented in Column (2) of Table C1.

Second, using the estimates of Equation (4) for the full sample of adult children, we calculate

the inverse Mills ratio, λ_i , then include it in the income equation using the sample of adult children with observed income to correct for selection bias. Note that although Equation (4) is estimated using the full sample, Equation (5) below can only be estimated using the sample with observed income:

$$(5) \quad \ln y_i = \beta_0 + X_i\beta_X + \beta_\lambda\lambda_i + \delta_i,$$

where y_i is the observed income, and X_i is the same set of demographic and socioeconomic variables as in Equation (4).¹⁶ Because the CFPS records parents' and children's income for the two cohorts at different ages in the same survey years, we are unable to account for the possibility that returns to education may change over time. However, we account for the cohort, *hukou*, and regional variations in returns to education by including full interactions of education with cohort, *hukou*, and coastal dummies in X_i in Equation (5). The estimation results of Equation (5) are presented in Table C2 of Appendix C. Columns (1) and (2) present the estimates without and with correcting for selection bias for adult children, respectively. The R-squared in Column (2) is 0.235.

The variable z , the number of the adult child's siblings, appears in Equation (4) but not Equation (5). We use this variable as the excluded variable from the income equation to address the selection problem arising from missing income. First, the greater the number of siblings, the higher the probability that a sibling will take care of the parents, and therefore (i) the lower the probability that a child lives with his parents; and (ii) the higher the probability that the child works outside the home county. In both cases, the CFPS is less likely to record income information for adult children the greater the number of siblings (Figure B8). Thus, the variable z , the number of siblings, satisfies the monotonicity assumption in the two-stage estimation. We control for other variables, such as education, to mitigate the direct impact of the number of siblings on the child's income through the child quantity-quality trade-off (Becker & Lewis, 1973; Hanushek, 1992; Guo *et al.*, 2017).

Third, based on Equation (5), we compute income for the full sample of adult children with and without income using individual characteristics X_i , the calculated inverse Mills ratio λ_i , and the estimated coefficients β_0 , β_X and β_λ .

A similar procedure is applied to compute income for the full sample of parents.¹⁷ Here, z is the number of children. The number of schooling years is the average of the father's and mother's schooling years, and age is the average of the father's and mother's age. We also include two additional dummies to indicate whether the father is alive, and whether the mother is alive. Furthermore, we interact the father-alive dummy and the mother-alive dummy with cohort, *hukou*, and coastal dummies. The estimates of Equation (4) for the full sample of parents are reported in Column (3) of Table C1 in Appendix C. The result of the robustness check is shown in Column (4). Columns (3) and (4) of Table C2 present the estimates of Equation (5) without and with correcting for selection bias for parents, respectively. The R-squared in Column (4) is 0.250.

Table C3 summarizes the computed income for both cohorts. By correcting for the selection bias, α_1 in Equation (1) measures the elasticity of the child's computed income with respect to his

¹⁶Here, the observed income is defined as the average income across at least two waves of CFPS.

¹⁷We compute income for parents and children separately as both Equations (4) and (5) contain different covariates for the two samples. In a robustness analysis in Section 4.1.3, we use the full sample of parents and children, keep only the covariates that appear in both samples, and compute income for both parents and children using the same estimation procedure. Our main results on intergenerational income persistence remain robust.

parents' computed income. Standard errors are bootstrapped.

Education and other Demographic Characteristics Educational attainment is a key predictor of lifetime income. Through validation across family members and across survey waves, the CFPS survey team generates the most reliable measure of education for each adult. First, in the 2010 baseline survey, the team collects information on individuals' number of years of schooling and the highest level of education from (i) self-reports in the adult survey, and (ii) interviews with family representatives in the household survey, and/or (iii) interviews with spouses in the adult survey. Second, in the subsequent 2012 adult survey, individuals were again asked about their educational attainment. Hence, the 2010 dataset is further cross-checked against the 2012 self-reports.

In this study, we report the child's age in 2010. *Hukou* is a dummy variable that is equal to one if a child held an agricultural or rural *hukou* when he was three years old, and zero otherwise. The coastal dummy is equal to one if the household is living in any coastal provinces, which is the most developed area in China, and zero otherwise.¹⁸

D Summary Statistics

Table 1 summarizes statistics for the 1970-1980 and 1981-1988 birth cohorts. On average, parents and children in the 1970–1980 cohort are 59 years old and 34 years old, respectively, while parents and children in the 1981-1988 cohort are 52 years old and 25 years old, respectively. In both cohorts, around 70 percent of households are from rural areas and 38 percent are from coastal areas; these proportions are representative of the overall population in China. The number of years of schooling for both parents and children have increased across cohorts. Parents in the early cohort have an average of 4.2 years of schooling, and parents in the late cohort have an average of 5.7 years of schooling. Likewise, the children's number of years of schooling increased from an average of 8.3 years for the early cohort to 9.6 years for the late cohort. Likewise, computed income for both parents and children have increased across cohorts, from 17,979 yuan to 20,294 yuan for parents, and from 22,185 yuan to 23,761 yuan for children.

III Intertemporal Patterns in Intergenerational Income Persistence

Using individual-level data, we first estimate changes in intergenerational income persistence across birth cohorts at the national level, then analyze changes in intergenerational income persistence by gender, *hukou*, and region.

¹⁸The coastal provinces/municipalities are Liaoning, Hebei, Tianjin, Beijing, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, and Guangdong.

A Estimates at the National Level

A.1 Estimates from the Subsample with Observed Income

Our estimation addresses lifecycle bias, attenuation bias, and selection bias. To overcome the lifecycle bias and attenuation bias associated with household survey data, Solon (1992); Nicoletti & Ermisch (2007) use computed income to estimate intergenerational persistence in developed countries. Similarly, Gong *et al.* (2012) use the two-sample two-stage least squares method to compute lifetime income for parents and estimate IGE in urban China. As discussed in Section 2.2, selection bias arising from either non-coresidence or temporary migration is a serious issue with IGE estimates in developing countries such as China, but most studies on intergenerational mobility ignore selection bias (Chyi *et al.*, 2014; Chen *et al.*, 2015; Qin *et al.*, 2016). Deng *et al.* (2013), followed by Fan (2016), address selection bias arising from non-coresidence by using the regional co-residence rate as an instrumental variable; however, selection bias arising from temporary migration is not addressed.

To illustrate the three types of biases as discussed in Section 2.2, we first present the IGE estimates using raw income based on the sample of parents-children pairs with income information (Sample N_1) in Columns (1)–(3), Panel A of Table 2. The IGE estimates rise from 0.166 in the early cohort to 0.176 in the late cohort. Both estimates are statistically significant at the 1 percent level. The magnitudes of our IGE estimates from the raw income data are similar to the OLS estimates using parents' current income in Gong *et al.* (2012): 0.174 for mother-daughter pairs, 0.215 for father-daughter pairs, 0.241 for father-son pairs, and 0.302 for mother-son pairs. We are cautious in interpreting these results, as they may not fully address transitory shocks in income, and are subject to lifecycle bias and selection bias.

To address transitory shocks in income and lifecycle bias, researchers use parental education or occupation as an IV for parental lifetime income (Solon, 1992; Gong *et al.*, 2012). Following this practice, we use parental education and its interactions with cohort, *hukou*, and coastal dummies as instrumental variables for parental lifetime income.¹⁹ Since these instrumental variables do not vary much over the lifecycle, but are highly correlated with lifetime income, the lifecycle bias is significantly reduced. Additionally, using computed income (instead of raw income) in the IV estimation minimizes the attenuation bias arising from transitory income shocks. Columns (4)–(6) display IGE estimates under this IV specification. Compared with the IGE estimates from the raw income data (0.166 for the early cohort and 0.176 for the late cohort), the IV estimates are larger for both cohorts — 0.280 for the early cohort and 0.475 for the late cohort. The greater magnitude of the IV estimates indicate that estimates based on the raw income data suffer from attenuation bias, even though income was averaged across two to six years (Mazumder, 2005). Furthermore, the increase in the IV estimate is larger for the late cohort than the early cohort, implying that the estimate for the late cohort is more subject to lifecycle bias.

To evaluate the role of selection bias, we compare IGE estimates that use computed income with IV estimates that have corrected for selection bias. We follow the same procedure detailed in Section 3.3 to correct for the selection bias and compute the income for the full sample, but we use only Sample N_1 to estimate IGEs. The last three columns in Panel A of Table 2 report the results. Note that Columns (7)–(9) use the same instrumental variables as in Columns (4)–(6), but include λ to correct for the selection bias due to missing income (see Table C2 in the appendix). The

¹⁹First-stage results are presented in Table C2 of Appendix C.

IGE estimate increases to 0.353 for the early cohort (Column (7) vs. Column (4)), but decreases to 0.435 for the late cohort (Column (8) vs. Column (5)). This difference in the changes in the estimates is probably due to the selection bias, which has differential impacts on the two cohorts. We do not observe income for temporary migrants and individuals who are not living with their parents or children. The probability of being a temporary migrant and the probability of not living with parents or children vary across age. Since the CFPS records income for different cohorts (of different ages) in the same survey years, the incidence of missing income varies across cohorts. Thus, correcting the selection bias affects the estimates of intergenerational income persistence across the two cohorts differently. We consider the estimates in the last three columns to be robust to lifecycle bias, attenuation bias, and selection bias, even though the sample is based only on individuals with observed income.

A.2 Estimates from the Full Sample with and without Observed Income

To estimate intergenerational income persistence at the national level, we compute the income for all parents-children pairs with and without income records (defined as Sample N_2), following the method discussed in Section 3.3. Panel B of Table 2 presents the main results of this nationally representative sample. Columns (1)–(3) display IGE estimates from the OLS estimation of Equation (1). Columns (4)–(6) present log correlation estimates that have adjusted for different income distributions between generations based on Equation (2). Columns (7)–(9) present corresponding rank correlation estimates from the OLS estimation of Equation (3). The three sets of estimates reveal a consistent pattern of increasing intergenerational income persistence. Specifically, the IGE estimates are 0.390 and 0.442 for the early and late cohorts, respectively. These results suggest that a 1 percent increase in parental income increases the income of a child in the early cohort by 0.39 percent, and the income of a child in the late cohort by 0.44 percent. The increase of 0.052 across cohorts, as shown in Column (3), is statistically significant at the 1 percent level. These results indicate that intergenerational income persistence is stronger for the late cohort than for the early cohort.²⁰

It is worth mentioning that although children in the early cohort (with an average age of 34.1) are in midlife when lifecycle bias is considered small, this is still younger than the suggested ideal age range, from mid-to late life (Nybom & Stuhler, 2016b). Thus estimates for the early cohort are likely downward biased. Downward lifecycle bias is possibly more serious for children in the late cohort, with an average age of 25.5, which is an early stage of life (Haider & Solon, 2006; Nybom & Stuhler, 2016b). Hence our estimate of the difference between the two cohorts is likely a lower bound for the increase in intergenerational income persistence.²¹

The magnitudes of our main IGE estimates are similar to those from the sample of parents-children pairs with recorded income (Columns (7)–(9) in Panel A). The difference in the increase in IGE estimates between Sample N_1 and Sample N_2 is statistically insignificant at the 10 percent

²⁰In a separate iteration, we use the number of an adult child’s siblings and its interactions with cohort, *hukou*, and coastal dummies as excluded variables from the income equation in a robustness analysis, as shown in Table C4. The estimates remain consistent.

²¹This argument is conditional on two assumptions: that no bias arises from the parents side or that such bias can largely be cancelled by taking the difference across the two cohorts. As parents average age is 58.8 and 51.8 in early and late cohorts, respectively, they are approximately in the ideal age range of mid- to late life (Nybom & Stuhler, 2016b). Thus, we consider these assumptions to likely be valid.

level (Column (9) in Panel A vs. Column (3) in Panel B). The magnitudes of our main IGE estimates are similar to the IGE of 0.481 between fathers and children estimated by Qin *et al.* (2016), who estimate a simultaneous equations model using data from the 1989–2009 China Health and Nutrition Survey (CHNS). Our IGE estimates are smaller than the IV estimate of 0.556 from Chyi *et al.* (2014), who use father-son pairs from the 1989–2006 waves of the CHNS data, and employ father’s number of years of education as an instrumental variable for income.

Nybohm & Stuhler (2016a) find that the intergenerational log correlation and rank correlation are more robust to lifecycle bias than the IGE. Columns (4)–(6) present the intergenerational log correlation estimates, which show a similar increase from 0.434 for the early cohort to 0.519 for the late cohort. These estimates are adjusted for differences in the variance of income between generations, and corrected for selection bias. Rank correlation estimates (Columns (7)–(9)) show a similar increase from 0.443 for the early cohort to 0.494 for the late cohort. The increases in intergenerational log correlation estimates (Column (6)) and rank correlation estimates (Column (9)) are statistically significant at the 1 percent level.

To summarize the results in Panels A and B, we find that lifecycle bias, attenuation bias, and selection bias account for the differences among: (i) the estimates based on the raw data (i.e., the subsample with recorded income); (ii) the IV estimates; and (iii) the estimates based on computed income correcting for selection bias (drawing on the full sample with and without recorded income). When we use the raw data, and do not account for the lifecycle bias or attenuation bias, the estimated change in IGE across cohorts is 0.010. When these biases are addressed, the estimated change in IGE rises to 0.195; this estimate, however, is biased upward due to missing income. When we address the selection bias (still using the subsample with recorded income), the change in IGE is estimated to be 0.082. Finally, when we use the full sample (with and without recorded income), and address the selection bias, the change in IGE is estimated to be 0.052, which is approximately a 13 percent increase across cohorts; this estimate is five times as large as the estimate using the raw data. These findings convey three important results, namely: (i) the increase in intergenerational persistence is fairly robust, regardless of which estimate we consider; (ii) the downward lifecycle/attenuation bias is the largest bias; and (iii) the selection bias is not negligible and leads to an overestimate of the increase in IGE if it is not addressed.

A.3 Alternative Specifications

As discussed in Section B, estimates of intergenerational income persistence are sensitive to the life stage at which income is measured. Thus, our main results may suffer from lifecycle bias, because children’s income is measured at an early stage of the lifecycle, while parents’ income is measured at a late stage of the lifecycle (Table 1). To examine the robustness of the results, we first restrict children to be at least 24 years old, and mothers to be at most 62 years old. The results, shown in Panel C of Table 2, indicate a similar increase in intergenerational income persistence. The estimates of intergenerational income persistence, which are statistically significant, remain robust across different measures, with magnitudes varying within reasonable ranges.

Second, to alleviate the concern that children in the early cohort are systematically older than children in the late cohort, we use a full sample of parents and children to predict each individual’s lifetime income (Panel D). The magnitudes of the estimates, especially log correlation and rank correlation, remain similar to the main estimates in Panel B.

Third, we restrict our sample to the five oversampling provinces, and present the findings in

Panel E. The pattern of increasing intergenerational income persistence across cohorts is again evident. Estimates are similar to the main estimates (Panel B) in both magnitude and level of statistical significance.

In addition, we present intergenerational education persistence from Sample N_1 and Sample N_2 in Panels F and G, respectively. Consistent with the income pattern, there is a statistically significant increase in intergenerational education persistence across cohorts, as captured by the log correlation estimates in both Sample N_1 and Sample N_2 (Column (6)).²²

Finally, following Chetty *et al.* (2014a), we replace the dependent variable of children's income with children's educational attainment. Compared with income, schooling is more stable across the lifecycle and less subject to measurement error. Results are presented in Table B1 in the appendix. Specifically, we use three variables to measure a child's educational attainment: a dummy variable that indicates a minimum educational level of: (i) senior high school (Columns (1)–(3)); (ii) college (Columns (4)–(6)); or (iii) university (Columns (7)–(9)). Panel A presents the correlation between children's level of education and log parental income, and Panel B presents the correlation between children's level of education and the national rank of parents' income. Our results are robust under these three alternative measures of children's education, and further confirm that intergenerational persistence has increased across cohorts. The magnitudes of our estimates are comparable to those in Chetty *et al.* (2014a).

A.4 Transition Matrix of Relative Mobility

Table 3, depicting relative mobility as described in Section A, shows the percentage of children in quintile i ($i=1,2,3,4,5$), given parents in quintile j ($j=1,2,3,4,5$). Panels A and B present statistics for the early and late cohorts, respectively, while Panel C shows the differences between the two cohorts.

Consider the children of parents in the bottom quintile. The proportion of children who are trapped in the bottom quintile increases from 32.3 percent to 36.7 percent across cohorts (northwest corners in Panels A and B). This increase of 4.3 percentage points is statistically significant at the 1 percent level, and indicate an increasing intergenerational poverty trap. On the other hand, the proportion of children who grew up in the bottom quintile, but end up in the top quintile as adults, is 9.8 percent for the early cohort and 7.3 percent for the late cohort (northeast corners in Panels A and B). This decrease of 2.5 percentage points across cohorts is statistically significant at the 1 percent level, and suggest a diminishing likelihood of children from the most disadvantaged families rising to the top quintile.

We conclude that intergenerational income persistence for children from poor families has become increasingly strong across cohorts. This trend is possibly due to rising educational costs and skewed public educational expenditure, which does not benefit children of the poor. In addition, children whose parents migrate from rural to urban areas to seek employment are usually left in rural areas and lack educational prospects (Lu, 2012). These children are likely to remain trapped in poverty, as we shall discuss in Section 5.

Now consider the children of parents in the top quintile. The proportion of children who remain in the top quintile increases from 45.9 percent for the early cohort to 48.7 percent for the late cohort (southeast corners in Panels A and B). This increase of 2.8 percentage points is statistically

²²Since the distribution of years of schooling tends to be multimodal, i.e., individuals complete their schooling after passing a certain threshold, transforming this variable to percentile to estimate the rank specification is not appropriate.

significant at the 10 percent level, indicating that descendants of the rich are increasingly likely to stay in the same position as their parents.

A.5 Transition Matrix of Absolute Mobility

Now we turn to absolute mobility, measured by the percentage of children earning more than 100 percent, 120 percent, and 150 percent of parental income, given parents in quintile j ($j=1,2,3,4,5$), as shown in Table 4. The table format is the same as the one of Table 3. We first focus on the proportion of children earning more than 100 percent of their parents' income. In the early cohort, nearly all children (93.9 percent) from bottom-quintile families grow up to earn more than their parents. This extremely high proportion is possibly due to the fact that parents in the bottom quintile in the early cohort have very low incomes. However, the proportion of children from bottom-quintile families earning more than their parents decreases by 2.2 percentage points in the late cohort; this decrease is statistically significant at the 1 percent level. Children from the top quintile also experience lower absolute mobility across cohorts. Only 30.4 percent of children in the top quintile in the late cohort earn more than their parents, compared to 39.7 percent of children in the early cohort. Consistent with Chetty *et al.* (2017), the higher the parents income, the lower the rate of absolute mobility of children, as there is less scope for children to surpass their parents. Of greater relevance to China's transition period of rapid economic development, we present the share of children earning more than 120 percent and 150 percent of their parents' income. Figure 4 depicts the findings. Again, the rates of absolute mobility decrease from the early cohort to the late cohort; this decrease is especially acute for children earning more than 150 percent of their parents income. The steady fall in absolute mobility across cohorts is consistent with trends in the U.S. (Chetty *et al.*, 2017).

As discussed in Section 3.3, since income is recorded in a particular survey year, parents are significantly older than their children when their incomes are recorded. Hence parents' income tends to be more representative of lifetime income than children's income. Additionally, parents in the early cohort tend to be older than parents in the late cohort; likewise, children in the early cohort are older than children in the late cohort. Hence incomes in the early cohort tend to be more representative of lifetime income than incomes in the late cohort. Although we partially address this issue by averaging income across several survey years, and by using the full sample to adjust for age to predict lifetime income, we exercise caution in interpreting our results on absolute mobility.

B Heterogeneity in Intertemporal Patterns

We investigate how changes in intergenerational income persistence in China vary by child gender, urban/rural *hukou* status, and coastal/inland regions. Son preference has persisted for centuries in China. In addition, as documented by Xie & Zhou (2014), inequality in China is mainly driven by the urban-rural gap and the disparity between coastal and inland regions. Note that income ranks for both parents and children are defined at the national level, as in Chetty *et al.* (2014a).

Panel A in Table 5 presents the gender pattern. Increasing intergenerational income persistence is slightly more evident for sons than daughters under the three measures. The IGE for sons increases from 0.34 to 0.40. This difference, as well as the estimates for each cohort, is statistically

significant at the 1 percent level (Columns (1)–(3)). Log correlation estimates (Columns (4)–(6)) and rank correlation estimates (Columns (7)–(9)) exhibit a similar pattern. For daughters, intergenerational income persistence also rises across cohorts, although the magnitudes are slightly smaller than those for sons for all three measures.²³

Panel B presents estimates of intergenerational income persistence by *hukou*. We find that the increase in intergenerational persistence is larger in urban areas than rural areas. Intergenerational income persistence increases in urban areas by 0.10, 0.18, and 0.09 under the measures of IGE, log correlation, and rank correlation, respectively. The three estimates are statistically significant at the 1 percent level (Columns (3), (6), and (9)). Intergenerational income persistence also increases in rural areas, although the magnitudes are smaller than those documented in urban areas. The urban-rural difference in the increase in intergenerational income persistence seems to be mainly driven by temporary outflows of migrants during the economic transition; we provide suggestive evidence in Appendix D2. With the relaxation of the *hukou* restriction, adult children from disadvantaged rural areas in the late cohort find it easier to migrate to cities, relative to their counterparts in the early cohort. Therefore, relative to urban residents, rural residents experience a smaller increase in intergenerational income persistence.

We then examine the disparity between coastal and inland regions. Market reforms were initiated in coastal areas, and later expanded inland. We present the patterns of intergenerational income persistence in coastal and inland regions in Panel C. Both regions show an increasing trend in intergenerational income persistence across cohorts; the exception is the IGE in the inland region, which is statistically insignificant. In the early cohort, estimates for coastal regions are smaller than estimates for inland regions. In the late cohort, however, estimates for coastal regions are larger than estimates for inland regions. Consequently, the increase in intergenerational income persistence is far greater in coastal regions than in inland regions. The increase in coastal regions is 0.2 for the IGE (Columns (1)–(3)); 0.15 for the log correlation (Columns (4)–(6)); and 0.24 for the rank correlation (Columns (7)–(9)). These estimates are statistically significant at the 1 percent level. Conversely, changes in intergenerational income persistence in inland regions are either statistically insignificant (Columns (3) and (9)) or of small magnitude (Column (6)), possibly due to temporary migration from inland to coastal regions. The results suggest that the increase in intergenerational income persistence is mainly driven by trends in coastal regions.

IV Inequality and Intergenerational Mobility: The Great Gatsby Curve in China

We present the Great Gatsby Curve in China, which plots the IGE against the Gini coefficient at the provincial level. The Chinese Statistical Yearbooks report Gini coefficients by year, province, and *hukou*. According to Krueger (2012) and Corak (2013), the Gini coefficient of parental income is ideally measured when children are growing up. Thus, we use the Gini coefficients in 1990 and 1999 from the Chinese Statistical Yearbooks as proxies for inequality during the teen years of children in the early and late cohorts, respectively.

²³Figure B9 shows the density distribution of log income for fathers and sons separately for the early cohort and the late cohort.

Figure 5(a) plots the IGE against the Gini coefficient by cohort, province, and *hukou*. We observe a positive correlation between the IGE and the Gini coefficient, which is similar to the pattern in developed countries. The OLS estimate is 0.83, which is statistically significant at the 1 percent level. This estimate indicates that the IGE increases by 0.083 when the Gini coefficient increases by 0.1. One potential reason for the large magnitude of the estimate is autocorrelation, as we pool the IGE and Gini coefficient for two cohorts in one estimation. Another possible reason is that we measure the Gini coefficient when children are on average 15–16 years old.²⁴ Blanden (2013) finds that the older the children when the Gini coefficient is recorded, the stronger the IGE-Gini correlation.

We further investigate the association between the intergenerational log correlation and the Gini coefficient. The pattern remains robust, as shown in Figure 5(b). Figure 5(c) plots the expected income rank of children born to parents at the bottom 20th national percentile rank against the Gini coefficient. As expected, the slope is negative. The higher the level of income inequality, the smaller the possibility that children from poor families can climb up the socioeconomic ladder, and the stronger the intergenerational income persistence. The estimate is -41.13, which is statistically significant at the 5 percent level. This estimate implies that the expected income rank of children born to parents at the bottom 20th percentile rank falls by 4 when the Gini coefficient increases by 0.1.²⁵

The correlation between intergenerational income persistence and inequality presented in Figure 5 might be driven by unobserved cross-province heterogeneity. We conduct a regression analysis to examine the association between the change in inequality and the change in intergenerational income persistence across cohorts by province and *hukou*. Appendix D details the regression specification. Estimation results are presented in Panel A of Table 6. Consistent with our findings from Figure 5, the change in the Gini coefficient is positively correlated with the change in IGE, and negatively correlated with the change in the expected income rank of children born to parents at the bottom 20th percentile rank. Both estimates are statistically insignificant.

The pattern of the Great Gatsby Curve may change over time, either because the children’s cohorts grew up under different conditions, as is the focus of our paper, or because the parents grew up under different conditions (Nyblom & Stuhler, 2016c). In China, most parents in the early cohort were born before China’s “Great Leap Forward,” while most parents in the late cohort were born after that period. Selection into fertility may have changed as well. Although the descriptive pattern of the Great Gatsby Curve in China cannot address all of these questions, it sheds light on the interaction between inequality and intergenerational mobility during China’s transition era.

We also make a preliminary attempt to associate the change in intergenerational income persistence with changes in socioeconomic factors during China’s transition period. Our purpose is merely to provide stylized facts to shed light on future research on causal determinants and new development of economic theories of intergenerational mobility for transition economies. We classify the socioeconomic changes during China’s economic transition into market-oriented structural changes, economic development, and public policies. Appendix A details the socioeconomic changes and their expected correlation with the change in intergenerational persistence. Appendix D1 specifies the regression equation and defines the variables. Appendix D2 presents and discusses the regression results. We find that the rising outflow of migrants is negatively associated with intergenerational income persistence, and (i) increases in the share of private enterprises; (ii)

²⁴The Chinese Statistical Yearbooks first report Gini coefficients by province and *hukou* in 1990.

²⁵Figures B10–B11 plot the Great Gatsby Curve in urban and rural areas, respectively.

public expenditure on education, science, culture, and public health per capita; and (iii) university enrollment rates are positively associated with intergenerational income persistence. Our evidence is weak possibly due to the small sample size at the province level.

V Policy Implications and Conclusion

We investigate the changing transmission of economic status across generations since China began economic reforms in 1979. Using household data from the CFPS survey, we first show intergenerational income persistence has increased across cohorts at the national level. We also present the Great Gatsby Curve across provinces and cohorts in China. The positive correlation between the IGE and Gini coefficient is similar to the pattern documented in developed countries. We then examine factors that may be correlated with the rising intergenerational income persistence across cohorts at the provincial level.

Urban areas have experienced greater increases in intergenerational income persistence than rural areas, possibly due to the relaxation of the *hukou* restriction, which facilitates temporary migration from rural areas to urban areas. The results suggest that the Chinese government should continue to remove rural-urban migration barriers. On the other hand, the misallocation of public resources during structural reforms may exert negative effects on intergenerational mobility. For instance, intergenerational income persistence has increased with the rise in government educational expenditure, possibly because the allocation of such expenditure is not means-tested (Li *et al.*, 2013). The expansion of tertiary education, with misallocated loans and scholarships, is negatively correlated with the upward mobility of children from poor families. The Chinese government should initiate various programs to subsidize the education of children from disadvantaged families, such as “left-behind” children. In addition, the efficacy of loan and scholarship programs at the tertiary level should be improved.

It should be noted that we have used the most appropriate methods to estimate intergenerational income persistence in China. Given the data limitations, however, the estimates are likely biased downward and thus interpretation of the magnitude of the levels or changes in the estimates requires great caution. Nevertheless, as discussed earlier, we believe that our estimates are likely a lower bound for the increase in intergenerational income persistence, which implies that the worsening intergenerational mobility may be more serious than what the reported estimates suggest. Further study of such mobility will have to await better panel data in China.

References

- Becker, Gary S., & Lewis, H. Gregg. 1973. On the Interaction between the Quantity and Quality of Children. *Journal of Political Economy*, **81**(2, Part 2), S279–S288.
- Becker, Gary S., & Tomes, Nigel. 1979. An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, **87**(6), 1153–1189.
- Becker, Gary S., & Tomes, Nigel. 1986. Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, **4**(3), S1–S39.
- Björklund, Anders, & Jäntti, Markus. 2009. Intergenerational Income Mobility and the Role of Family Background. *Oxford Handbook of Economic Inequality*, Oxford University Press, Oxford, 491–521.
- Blanden, Jo. 2013. Cross-Country Rankings in Intergenerational Mobility: A Comparison of Approaches from Economics and Sociology. *Journal of Economic Surveys*, **27**(1), 38–73.
- Chen, Yuyu, Naidu, Suresh, Yu, Tinghua, & Yuchtman, Noam. 2015. Intergenerational Mobility and Institutional Change in 20th Century China. *Explorations in Economic History*, **58**, 44–73.
- Chetty, Raj, Hendren, Nathaniel, Kline, Patrick, & Saez, Emmanuel. 2014a. Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *Quarterly Journal of Economics*, **129**(4), 1553–1623.
- Chetty, Raj, Hendren, Nathaniel, Kline, Patrick, Saez, Emmanuel, & Turner, Nicholas. 2014b. Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review: Papers & Proceedings*, **104**(5), 141–147.
- Chetty, Raj, Grusky, David, Hell, Maximilian, Hendren, Nathaniel, Manduca, Robert, & Narang, Jimmy. 2017. The Fading American Dream: Trends in Absolute Income Mobility since 1940. *Science*, **356**(6336), 398–406.
- China Family Panel Studies, CFPS. 2015. *Institute of Social Science Survey, Peking University*.
- Chyi, Hau, Zhou, Bo, Jiang, Shenyi, & Sun, Wei. 2014. An Estimation of the Intergenerational Income Elasticity of China. *Emerging Markets Finance and Trade*, **50**(6), 122–136.
- Corak, Miles. 2004. Generational Income Mobility in North America and Europe: An Introduction. *Generational Income Mobility in North America and Europe*. Cambridge, 1–37.
- Corak, Miles. 2013. Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, **27**(3), 79–102.
- Dahl, Molly W., & DeLeire, Thomas. 2008. *The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates using Administrative Data*. University of Wisconsin-Madison, Institute for Research on Poverty.
- Davis, Jonathan, & Mazumder, Bhashkar. 2017. The Decline in Intergenerational Mobility after 1980. *FRB of Chicago Working Paper No. WP-2017-5*.

- Deng, Quheng, Gustafsson, Björn, & Li, Shi. 2013. Intergenerational Income Persistence in Urban China. *Review of Income and Wealth*, **59**(3), 416–436.
- Fan, Yi. 2016. Intergenerational Income Persistence and Transmission Mechanism: Evidence from China. *China Economic Review*, **41**, 299–314.
- Fields, Gary S, & Ok, Efe A. 1999. The measurement of income mobility: an introduction to the literature. *Pages 557–598 of: Handbook of income inequality measurement*. Springer.
- Gong, Hongge, Leigh, Andrew, & Meng, Xin. 2012. Intergenerational Income Mobility in Urban China. *Review of Income and Wealth*, **58**(3), 481–503.
- Grawe, Nathan D. 2006. Lifecycle Bias in Estimates of Intergenerational Earnings Persistence. *Labour Economics*, **13**(5), 551–570.
- Guo, Rufei, Yi, Junjian, & Zhang, Junsen. 2017. Family size, birth order, and tests of the quantity–quality model. *Journal of Comparative Economics*, **45**(2), 219–224.
- Haider, Steven J., & Solon, Gary. 2006. Life-Cycle Variation in the Association between Current and Lifetime Earnings. *American Economic Review*, **96**(4), 1308–1320.
- Hanushek, Eric A. 1992. The Trade-off between Child Quantity and Quality. *Journal of Political Economy*, **100**(1), 84–117.
- Jenkins, Stephen. 1987. Snapshots versus Movies: Lifecycle Biases and the Estimation of Intergenerational Earnings Inheritance. *European Economic Review*, **31**(5), 1149–1158.
- Krueger, Alan. 2012. The Rise and Consequences of Inequality. *Presentation made to the Center for American Progress*.
- Lee, Chul-In, & Solon, Gary. 2009. Trends in Intergenerational Income Mobility. *Review of Economics and Statistics*, **91**(4), 766–772.
- 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. 2018. The Current Income Distribution in China (in Chinese). *Academics*, **3**, 5–19.
- Lu, Yao. 2012. Education of Children Left behind in Rural China. *Journal of Marriage and Family*, **74**(2), 328–341.
- Mazumder, Bhashkar. 2005. Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data. *Review of Economics and Statistics*, **87**(2), 235–255.
- Meng, Xin, & Zhang, Junsen. 2001. The Two-tier Labor Market in Urban China: Occupational Segregation and Wage Differentials between Urban Residents and Rural Migrants in Shanghai. *Journal of Comparative Economics*, **29**(3), 485–504.

- National Bureau of Statistics of China Statistical Yearbook. 2003-2016. *Gini coefficients*. Available at <http://www.scio.gov.cn/zhzc/2/32764/Document/1421797/1421797.htm>; <http://www.stats.gov.cn/ztjc/ztfx/18fzcyj/201802/P020180212572614710685.pdf>.
- Nicoletti, Cheti, & Ermisch, John F. 2007. Intergenerational earnings mobility: changes across cohorts in Britain. *The BE Journal of Economic Analysis & Policy*, **7**(2).
- Nybom, Martin, & Stuhler, Jan. 2016a. Biases in Standard Measures of Intergenerational Income Dependence. *Journal of Human Resources*, 0715–7290.
- Nybom, Martin, & Stuhler, Jan. 2016b. Heterogeneous Income Profiles and Lifecycle Bias in Intergenerational Mobility Estimation. *Journal of Human Resources*, **51**(1), 239–268.
- Nybom, Martin, & Stuhler, Jan. 2016c. Interpreting Trends in Intergenerational Mobility. *Working Paper*.
- Piketty, Thomas. 2000. Theories of Persistent Inequality and Intergenerational Mobility. In A.B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, **Chapter 6**, 429–476. Amsterdam: North Holland.
- Qin, Xuezheng, Wang, Tianyu, & Zhuang, Castiel Chen. 2016. Intergenerational Transfer of Human Capital and Its Impact on Income Mobility: Evidence from China. *China Economic Review*, **38**(C), 306–321.
- Ravallion, Martin, & Chen, Shaohua. 2007. China's (Uneven) Progress against Poverty. *Journal of Development Economics*, **82**, 1–42.
- Solon, Gary. 1989. Biases in the Estimation of Intergenerational Earnings Correlations. *Review of Economics and Statistics*, **71**(1), 172–174.
- Solon, Gary. 1992. Intergenerational Income Mobility in the United States. *American Economic Review*, **82**(3), 393–408.
- Solon, Gary. 2004. A Model of Intergenerational Mobility Variation Over Time and Place. *Generational Income Mobility in North America and Europe*, 38–47.
- Wei, Shang-Jin. 1995. The open door policy and China's rapid growth: Evidence from city-level data. *Pages 73–104 of: Growth Theories in Light of the East Asian Experience, NBER-EASE Volume 4*. University of Chicago Press.
- Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT press.
- World Bank. 1981-2016. *GDP per capita*. Available at <https://data.worldbank.org/indicator/NY.GDP.PCAP.KD?end=2016&locations=CN&start=1981>.
- Xie, Yu, & Zhou, Xiang. 2014. Income Inequality in Today's China. *Proceedings of the National Academy of Sciences*, **111**(19), 6928–6933.
- Xie, Yu, Hu, Jingwei, & Zhang, Chunni. 2014. The China Family Panel Studies: Design and Practice. *Society: Chinese Journal of Sociology*, **34**(2).

- Xu, Hongwei, & Xie, Yu. 2015. The causal effects of rural-to-urban migration on children's well-being in China. *European Sociological Review*, **31**(4), 502–519.
- Zhu, Xiaodong. 2012. Understanding China's Growth: Past, Present, and Future. *Journal of Economic Perspectives*, **26**(4), 103–124.
- Zimmerman, David John. 1992. Regression toward Mediocrity in Economic Stature. *American Economic Review*, **82**(3), 409–29.

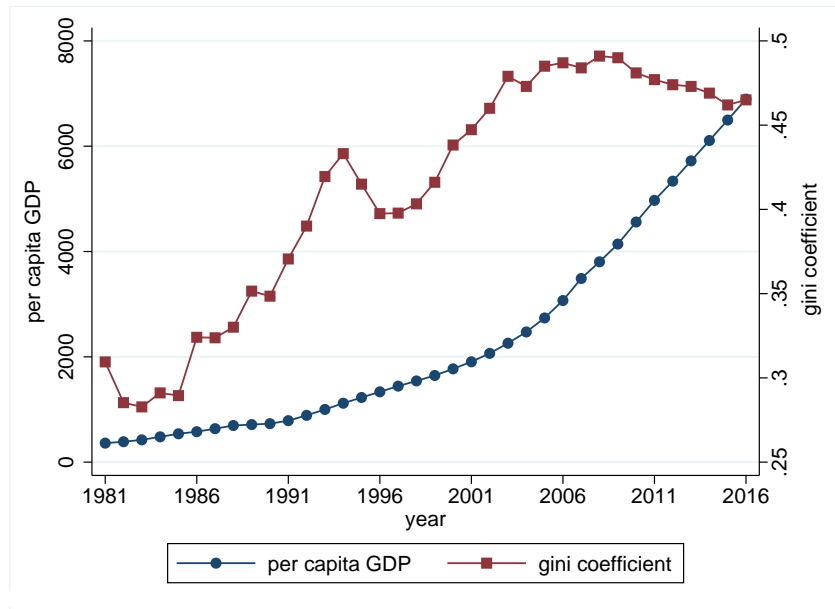


Figure 1: Per Capita GDP and Gini Coefficient in China, 1981–2016

Note: Data on GDP per capita are from World Bank (1981-2016). Gini coefficients in 1981–2000 are from Ravallion & Chen (2007); the Gini coefficient in 2002 is from Li (2018); the Gini coefficients in 2003-2016 are from National Bureau of Statistics of China Statistical Yearbook (2003-2016). Per capita GDP is measured in 2010 USD.

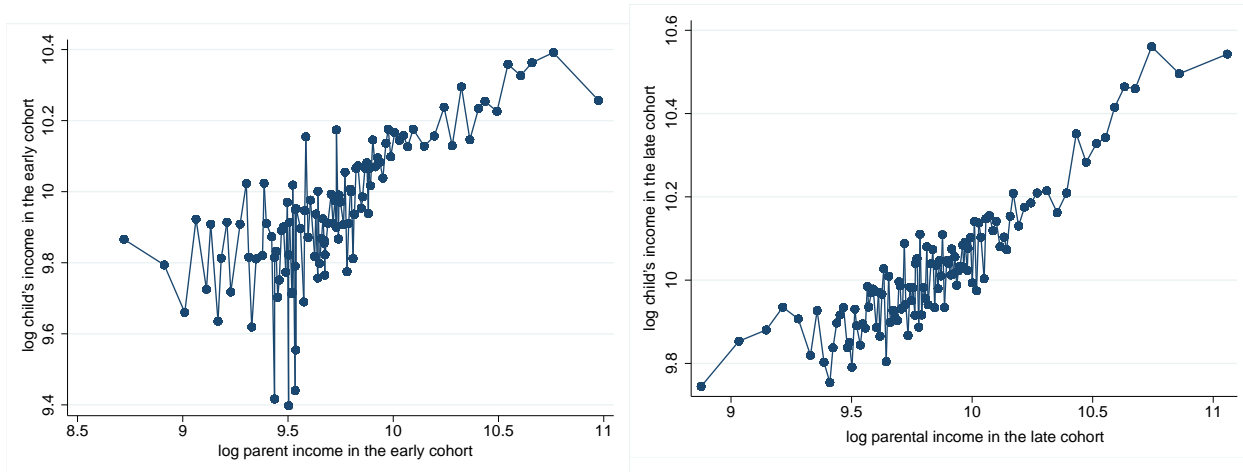


Figure 2: Log Child's Income versus Log Parental Income in the 1970–1980 (Early) Cohort (Left) and the 1981–1988 (Late) Cohort (Right)

Note: On the horizontal axis, we first rank log parental income, and then sequentially average log parental income by every 100 observations. On the vertical axis, we calculate the mean of the corresponding log child's income.

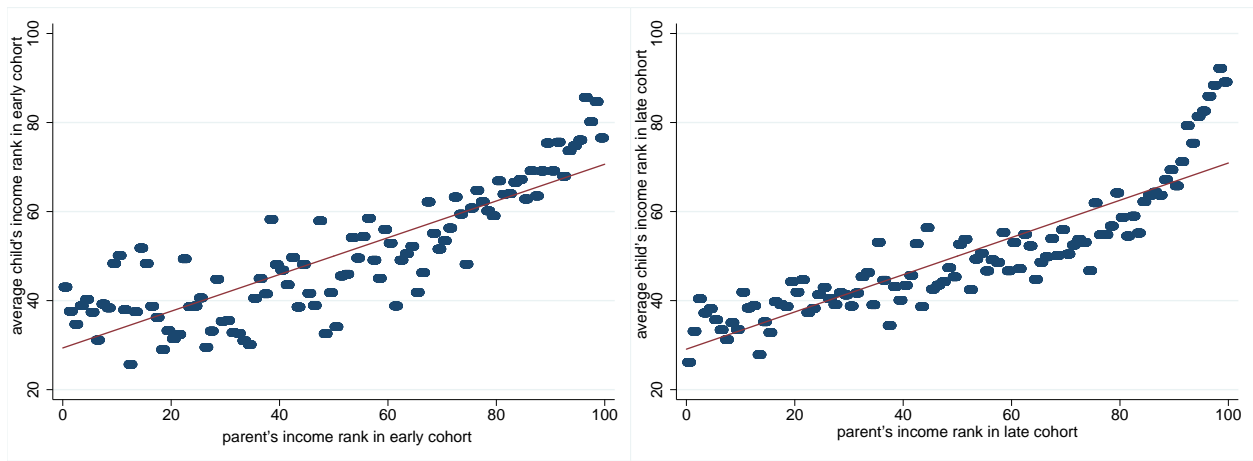
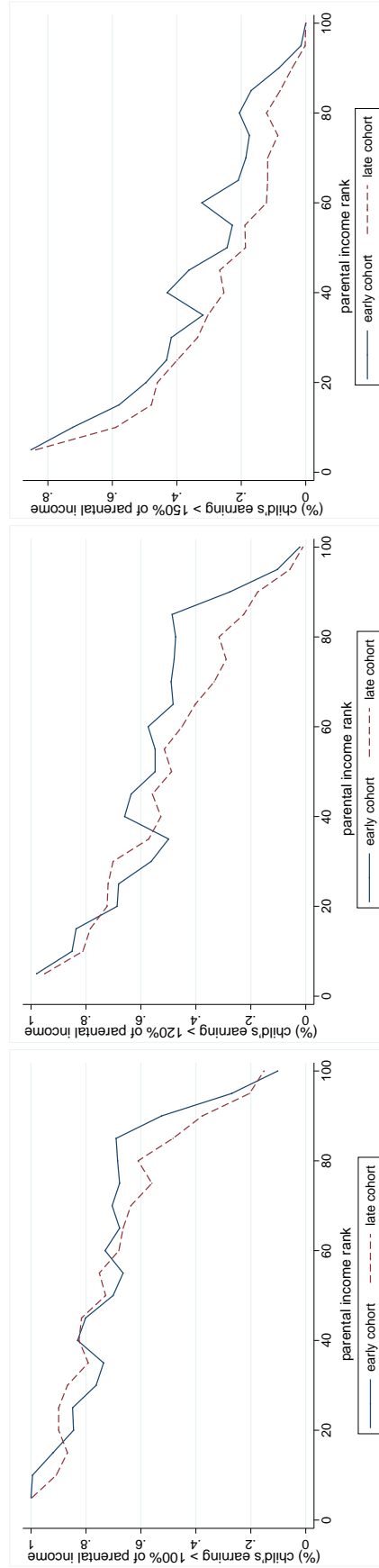


Figure 3: Income Rank of Children versus Income Rank of Parents in the 1970–1980 (Early) Cohort (Left) and the 1981–1988 (Late) Cohort (Right)

Note: Income rank is calculated at the national level by cohort and generation.



(a)

(b)

(c)

Figure 4: Percentage of Children Earning More than 100 Percent, 120 Percent, and 150 Percent of Parental Income in 1970–1980 (Early) Cohort and the 1981–1988 (Late) Cohort

Note: Figures (a), (b), and (c) show the percentage of children earning more than 100 percent, 120 percent, and 150 percent, respectively, of parental income by cohort. Graphs are drawn by taking averages across five ranks of parental income in each cohort.

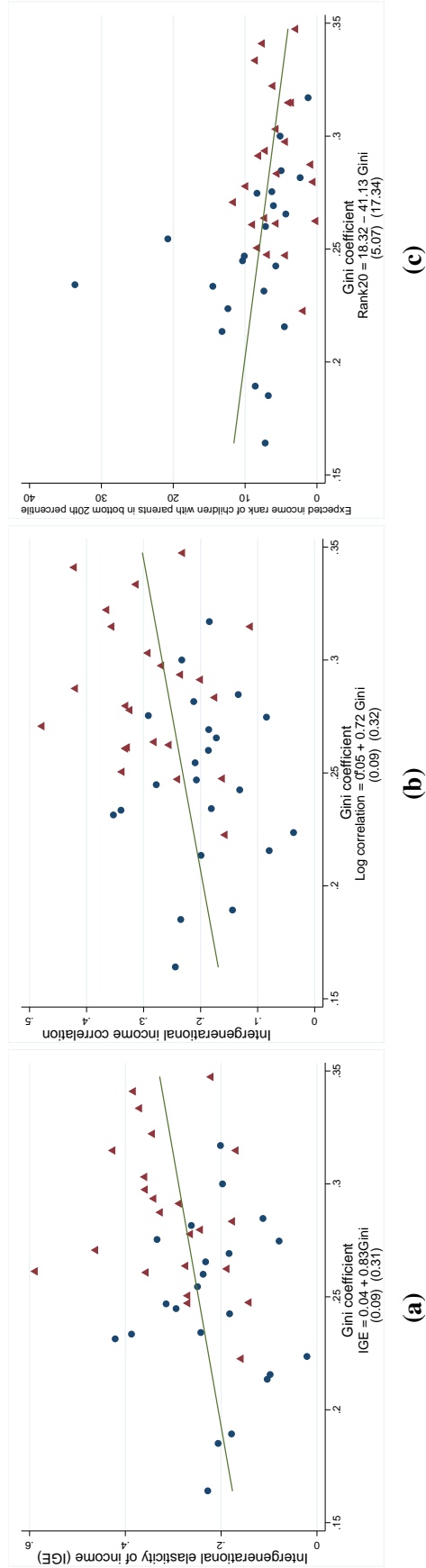


Figure 5: The Great Gatsby Curve in China

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, cohort, and *hukou*. 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 1: Summary Statistics

	Mean (Standard deviation)	
	Early cohort	Late cohort
Child's income (computed)	22,185.44 (7,674.02)	23,761.27 (7,380.33)
Ln child's income (computed)	9.95 (0.34)	10.03 (0.30)
Parents' income (computed)	17,978.72 (8,161.70)	20,293.59 (8,223.27)
Ln parental income (computed)	9.72 (0.38)	9.85 (0.36)
Child's schooling years	8.32 (4.30)	9.57 (4.18)
Parents' average schooling years	4.22 (3.42)	5.72 (3.76)
Child's age	34.13 (2.98)	25.46 (2.30)
Parents' average age	58.78 (2.96)	51.83 (4.68)
Child's gender (male=1)	0.48 (0.50)	0.47 (0.50)
Child's <i>hukou</i> status at 3 years old (agricultural=1)	0.69 (0.46)	0.71 (0.45)
Coastal area (=1)	0.37 (0.48)	0.39 (0.49)
Observation	10,980	11,333

Note: The data are from the China Family Panel Studies (CFPS) in 2010, 2012, 2014, and 2016. Children are at least 22 years old, and parents 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 2: Estimates of Intergenerational Income Persistence

Early cohort	Late cohort	Change	Early cohort	Late cohort	Change	Early cohort	Late cohort	Change
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Intergenerational Elasticity of Family Members with Income (Sample N_1)								
Raw income data			Computed income without selection bias correction			Computed income with selection bias correction		
0.166 (0.029)	0.176 (0.018)	0.010 (0.034)	0.280 (0.131)	0.475 (0.077)	0.195 (0.151)	0.353 (0.020)	0.435 (0.014)	0.082 (0.024)
Panel B. Main Results (Family Members with and without Income & Selection Bias Corrected: Sample N_2)								
Intergenerational elasticity					Log correlation		Rank correlation	
0.390 (0.008)	0.442 (0.007)	0.052 (0.011)	0.434 (0.009)	0.519 (0.008)	0.085 (0.012)	0.443 (0.008)	0.494 (0.009)	0.051 (0.013)
Panel C. Robustness Check 1: Children at Least 24 Years Old and Mothers at Most 62 Years Old (Sample N_2)								
0.403 (0.008)	0.431 (0.008)	0.028 (0.012)	0.452 (0.009)	0.517 (0.010)	0.064 (0.013)	0.457 (0.009)	0.499 (0.010)	0.042 (0.013)
Panel D. Robustness Check 2: Use Full Sample to Predict Lifetime Income (Sample N_2)								
0.236 (0.011)	0.287 (0.011)	0.051 (0.017)	0.402 (0.016)	0.454 (0.015)	0.052 (0.022)	0.419 (0.016)	0.475 (0.014)	0.055 (0.022)
Panel E. Robustness Check 3: Restrict to Five Oversampling Provinces (Sample N_2)								
0.434 (0.011)	0.482 (0.011)	0.048 (0.015)	0.480 (0.012)	0.552 (0.012)	0.072 (0.017)	0.487 (0.013)	0.536 (0.012)	0.049 (0.017)
Panel F. Intergenerational Education Persistence^a (Sample N_1)								
0.452 (0.038)	0.474 (0.020)	0.022 (0.043)	0.393 (0.033)	0.458 (0.019)	0.064 (0.039)	-	-	-
Panel G. Intergenerational Education Persistence^a (Sample N_2)								
0.561 (0.010)	0.555 (0.010)	-0.006 (0.014)	0.445 (0.008)	0.500 (0.009)	0.054 (0.012)	-	-	-

Note: Robust standard errors are in parentheses in Columns (1)–(6) in Panel A. Bootstrapped standard errors are in parentheses in Columns (7)–(9) in Panel A and all columns in Panels B–E. The numbers of observations in Panels A–E are 3,074, 22,313, 18,261, 5,451, and 10,271, respectively. Annual incomes of less than 3,000 yuan or more than 1,000,000 yuan are excluded as outliers. The data are from the China Family Panel Studies (CFPS) in 2010, 2012, 2014, and 2016. Additional regressors include child's age and age squared, and father's age and age squared. Panels AB and D–E include children who are at least 22 years old and parents who are at most 64 years old. Panel C includes children who are at least 24 years old and mothers who are at most 62 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. Intergenerational log correlation = intergenerational income coefficient * σ_p/σ_c , where σ_p and σ_c are the standard deviations of log income of parents and of children, respectively.

^a Since the distribution of years of schooling tends to be multimodal, i.e., individuals complete their schooling after passing a certain threshold, transforming this variable to percentile to estimate the rank specification is not appropriate.

Table 3: Quintile Transition Matrix of Intergenerational Income Persistence

Panel A. Early Cohort: Percentage of Children in Each Quintile Conditional on Parents' Quintile						
		Child's Income Quintile				
		1	2	3	4	5
Parents' Income Quintile	1	0.323 (0.468)	0.223 (0.416)	0.204 (0.403)	0.151 (0.358)	0.098 (0.298)
	2	0.331 (0.471)	0.205 (0.404)	0.230 (0.421)	0.138 (0.345)	0.096 (0.294)
	3	0.219 (0.414)	0.214 (0.410)	0.214 (0.410)	0.213 (0.409)	0.140 (0.347)
	4	0.109 (0.312)	0.239 (0.427)	0.194 (0.395)	0.250 (0.433)	0.207 (0.405)
	5	0.018 (0.132)	0.119 (0.324)	0.157 (0.364)	0.248 (0.432)	0.459 (0.498)
Panel B. Late Cohort: Percentage of Children in Each Quintile Conditional on Parents' Quintile						
		Child's Income Quintile				
		1	2	3	4	5
Parents' Income Quintile	1	0.367 (0.482)	0.220 (0.414)	0.179 (0.384)	0.161 (0.368)	0.073 (0.261)
	2	0.260 (0.439)	0.248 (0.432)	0.204 (0.403)	0.179 (0.383)	0.109 (0.312)
	3	0.184 (0.388)	0.245 (0.430)	0.217 (0.412)	0.210 (0.407)	0.144 (0.351)
	4	0.135 (0.342)	0.206 (0.405)	0.236 (0.424)	0.236 (0.425)	0.187 (0.390)
	5	0.054 (0.226)	0.081 (0.272)	0.164 (0.370)	0.215 (0.411)	0.487 (0.500)
Panel C. Difference between Early and Late Cohorts						
		Child's Income Quintile				
		1	2	3	4	5
Parents' Income Quintile	1	0.043 (0.014)	-0.003 (0.012)	-0.025 (0.012)	0.010 (0.011)	-0.025 (0.008)
	2	-0.071 (0.014)	0.043 (0.013)	-0.026 (0.012)	0.040 (0.011)	0.013 (0.009)
	3	-0.035 (0.012)	0.031 (0.013)	0.003 (0.012)	-0.003 (0.012)	0.004 (0.010)
	4	0.026 (0.010)	-0.033 (0.012)	0.042 (0.012)	-0.014 (0.013)	-0.021 (0.012)
	5	0.036 (0.006)	-0.038 (0.009)	0.007 (0.011)	-0.033 (0.013)	0.028 (0.015)

Note: Each cell in Panels A and B reports the percentage of children in the quintile (as given by the column), conditional on parental income in the quintile (as given by the row). Standard deviations are reported in parentheses in Panels A and B. Panel C displays the difference between Panel A and Panel B, with standard errors reported in parentheses.

Table 4: Absolute Intergenerational Income Mobility

Panel A. Early Cohort: Percentage of Children Earning 100 Percent, 120 Percent, and 150 Percent of Parents' Income Conditional on Parents' Quintile				
		100 Percent	120 Percent	150 Percent
Parents' Income Quintile	1	0.939 (0.239)	0.838 (0.369)	0.661 (0.474)
	2	0.795 (0.404)	0.601 (0.490)	0.401 (0.490)
	3	0.723 (0.447)	0.576 (0.494)	0.288 (0.453)
	4	0.687 (0.464)	0.482 (0.500)	0.194 (0.396)
	5	0.397 (0.489)	0.222 (0.416)	0.067 (0.250)
Panel B. Late Cohort: Percentage of Children Earning 100 Percent, 120 Percent, and 150 Percent of Parents' Income Conditional on Parents' Quintile				
		100 Percent	120 Percent	150 Percent
Parents' Income Quintile	1	0.918 (0.275)	0.817 (0.387)	0.591 (0.492)
	2	0.845 (0.362)	0.629 (0.483)	0.325 (0.469)
	3	0.745 (0.436)	0.504 (0.500)	0.189 (0.391)
	4	0.618 (0.486)	0.334 (0.472)	0.111 (0.314)
	5	0.304 (0.460)	0.118 (0.323)	0.031 (0.173)
Panel C. Difference between Early and Late Cohorts				
		100 Percent	120 Percent	150 Percent
Parents' Income Quintile	1	-0.022 (0.008)	-0.021 (0.011)	-0.069 (0.014)
	2	0.050 (0.011)	0.028 (0.015)	-0.076 (0.014)
	3	0.022 (0.013)	-0.072 (0.015)	-0.100 (0.013)
	4	-0.069 (0.014)	-0.147 (0.015)	-0.083 (0.011)
	5	-0.092 (0.014)	-0.104 (0.011)	-0.036 (0.006)

Note: Each cell in Panels A and B reports the percentage of children earning more than 100 percent, 120 percent, and 150 percent of their parents income (as given by the column), conditional on parental income in the quintile (as given by the row). Standard deviations are reported in parentheses in Panels A and B. Panel C displays the difference between Panel A and Panel B, with standard errors reported in parentheses.

Table 5: Heterogeneity in Intergenerational Income Persistence

Intergenerational elasticity		Intergenerational log correlation			Intergenerational rank correlation			
Early cohort	Late cohort	Change	Early cohort	Late cohort	Change	Early cohort	Late cohort	Change
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Sons versus Daughters								
Sons								
0.343 (0.008)	0.403 (0.008)	0.060 (0.011)	0.513 (0.012)	0.594 (0.012)	0.080 (0.017)	0.390 (0.009)	0.453 (0.009)	0.063 (0.013)
Daughters								
0.429 (0.010)	0.479 (0.010)	0.050 (0.013)	0.550 (0.012)	0.605 (0.012)	0.056 (0.017)	0.491 (0.011)	0.538 (0.009)	0.047 (0.014)
Panel B. Urban versus Rural Areas								
Urban Areas								
0.184 (0.009)	0.288 (0.009)	0.104 (0.013)	0.297 (0.017)	0.477 (0.015)	0.179 (0.023)	0.225 (0.014)	0.317 (0.012)	0.093 (0.018)
Rural Areas								
0.342 (0.014)	0.391 (0.014)	0.049 (0.020)	0.328 (0.013)	0.391 (0.014)	0.063 (0.019)	0.354 (0.011)	0.392 (0.011)	0.037 (0.016)
Panel C. Coastal versus Inland Regions								
Coastal Regions								
0.250 (0.012)	0.451 (0.012)	0.201 (0.018)	0.328 (0.017)	0.482 (0.013)	0.153 (0.022)	0.252 (0.015)	0.490 (0.019)	0.238 (0.024)
Inland Regions								
0.330 (0.010)	0.317 (0.010)	-0.012 (0.016)	0.357 (0.012)	0.394 (0.013)	0.036 (0.017)	0.369 (0.011)	0.382 (0.012)	0.013 (0.016)

Note: Bootstrapped standard errors are in parentheses. The numbers of observations for sons, daughters, urban areas, rural areas, coastal regions and inland regions are 10,603, 11,710, 6,647, 15,666, 8,469 and 13,844, respectively. 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. Intergenerational log correlation = intergenerational income coefficient * σ_p/σ_c , where σ_p and σ_c are the standard deviations of log income of parents and of children, respectively.

Rising Intergenerational Income Persistence in China

Online Appendix

Yi Fan Junjian Yi Junsen Zhang *

November 7, 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.

¹See 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).

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

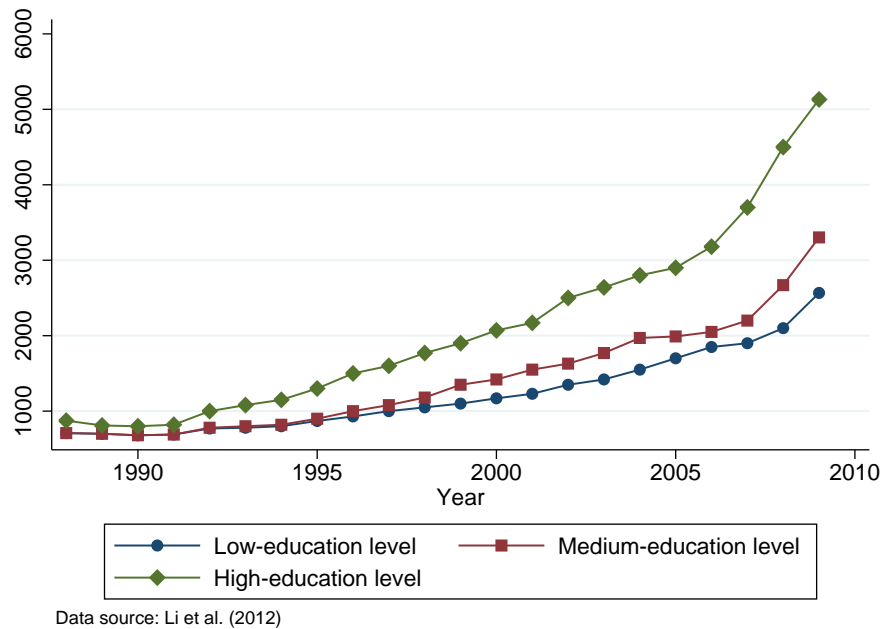


Figure A.1: Annual Wage of Urban Workers, 1988-2009

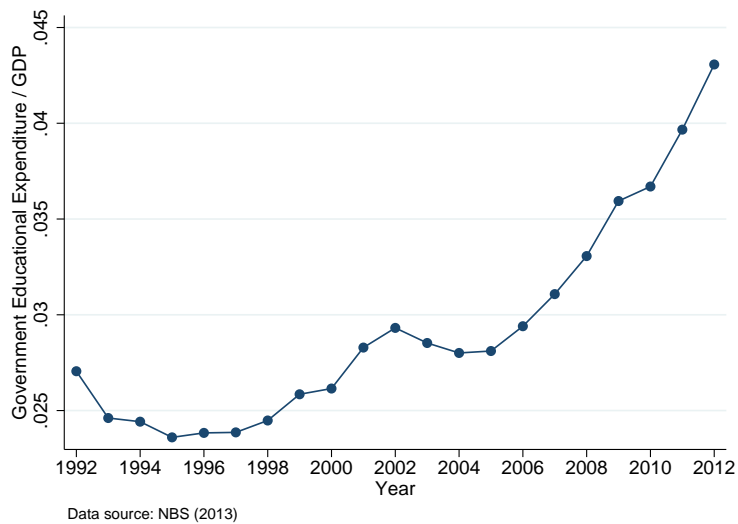


Figure A.2: Government Educational Expenditure/GDP, 1992-2012

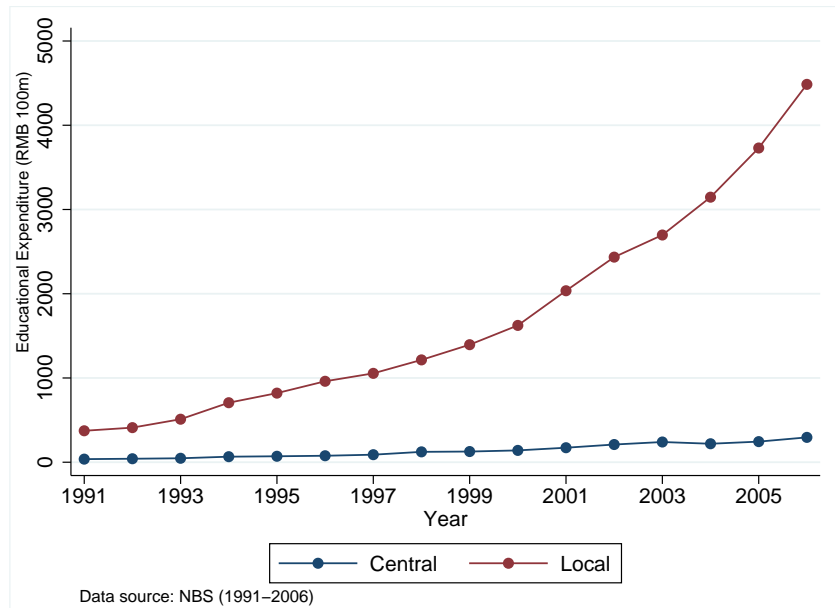


Figure A.3: Central and Local Governmental Expenditure on Education, 1991-2006

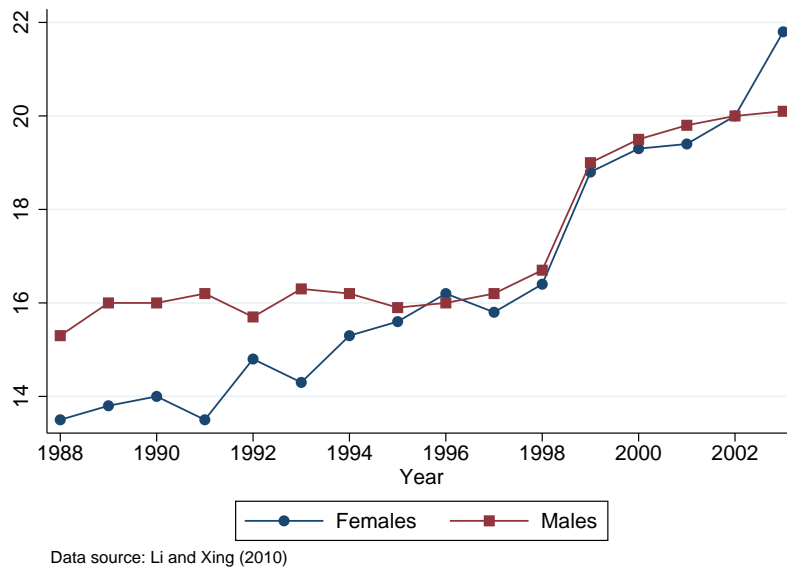


Figure A.4: Tertiary School Enrollment Rates, 1988-2003

Appendix B: Figures and Tables

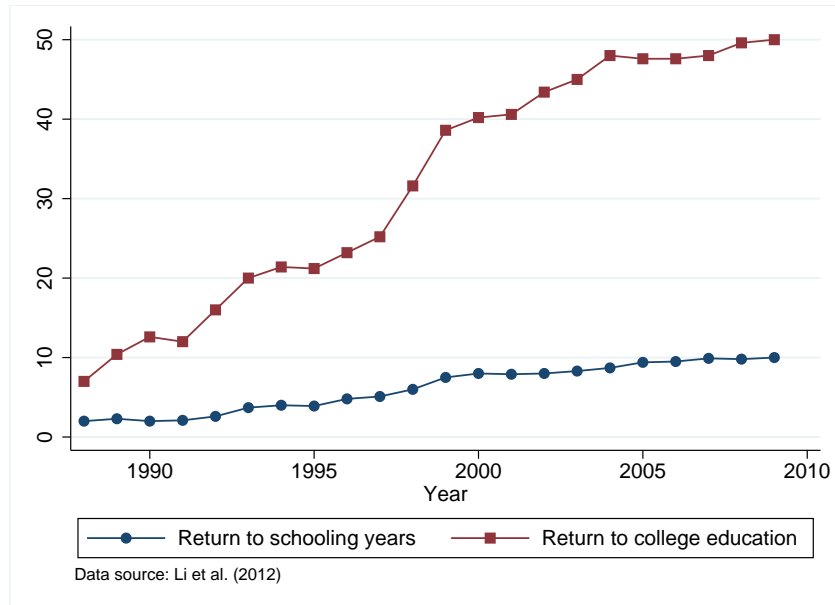


Figure B.1: Return to Education in Urban China, 1988-2009

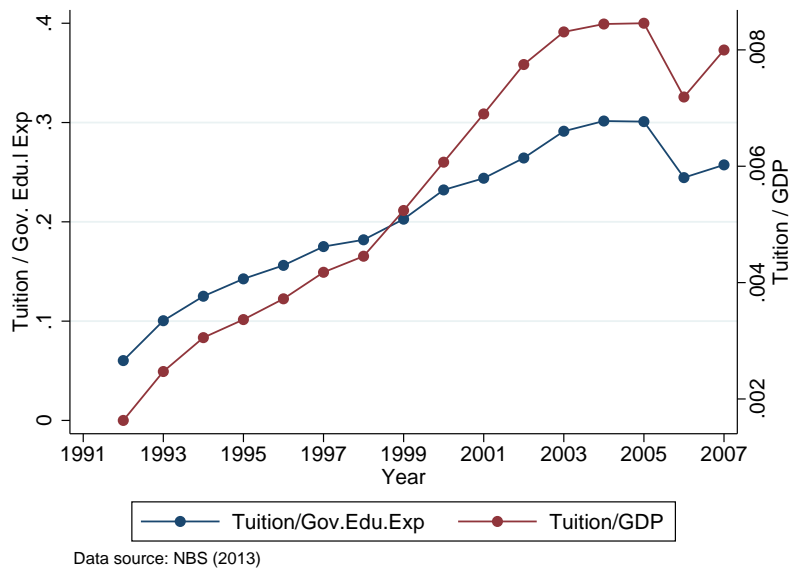


Figure B.2: Increase in Tuition in China, 1991-2007

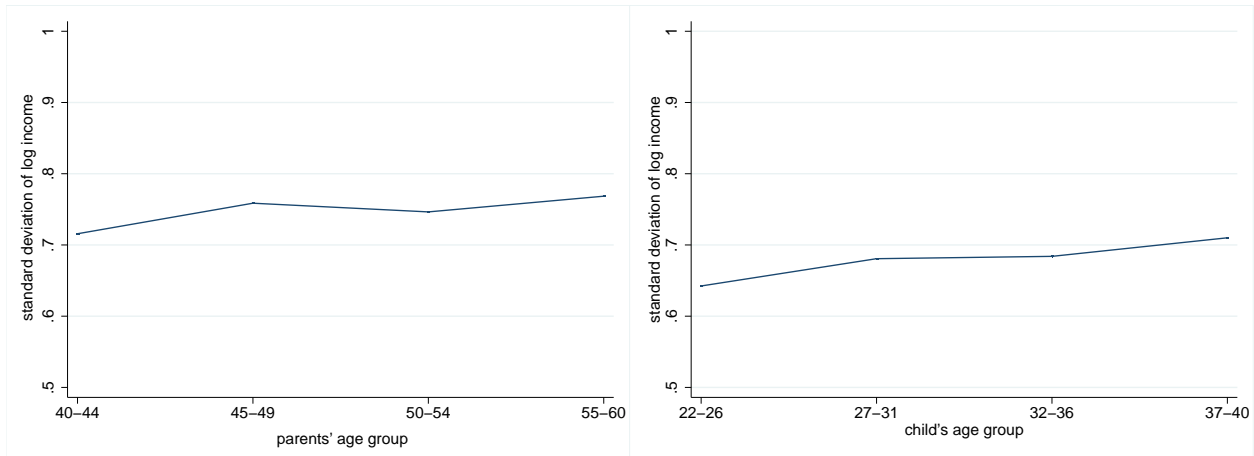


Figure B.3: Standard Deviation of Income against Age Groups
(Left: Parents, Right: Children)

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.

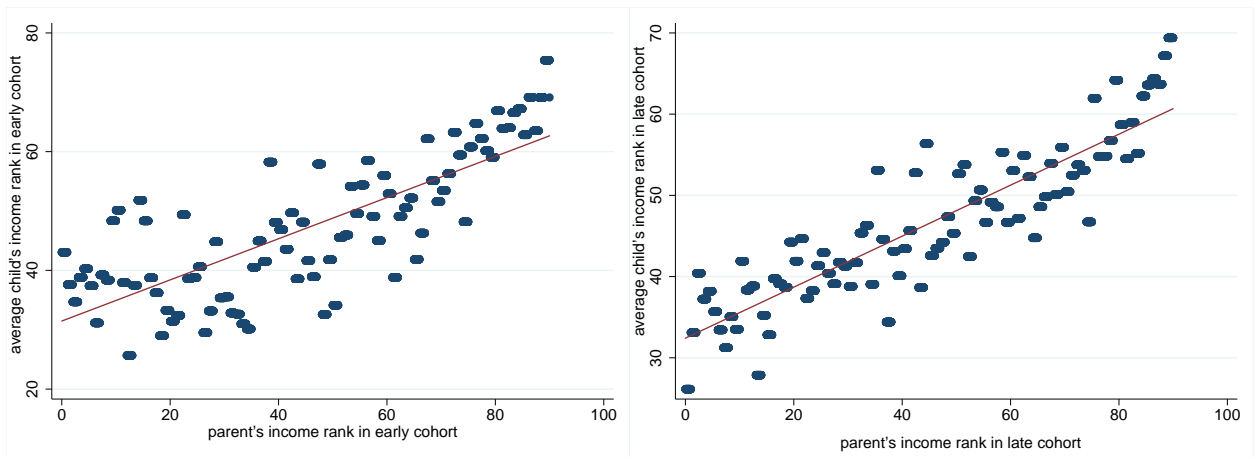


Figure B.4: Income Rank of Children versus Income Rank of Parents in the Early Cohort (Left) and Late Cohort (Right), with Top 10 percent Dropped

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

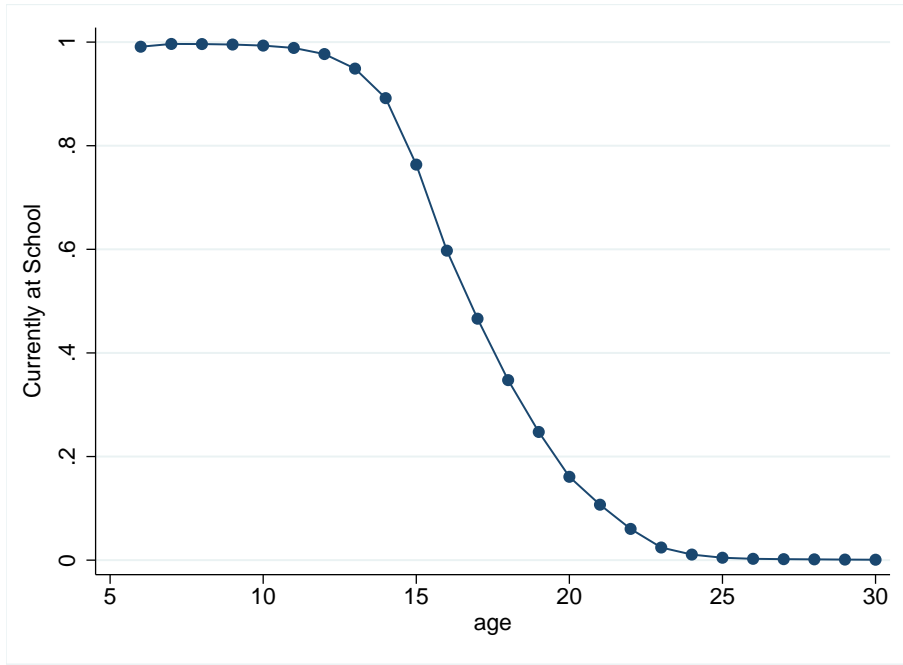


Figure B.5: Probability of Currently Being at School against Age

Note: Data from census in 2000.

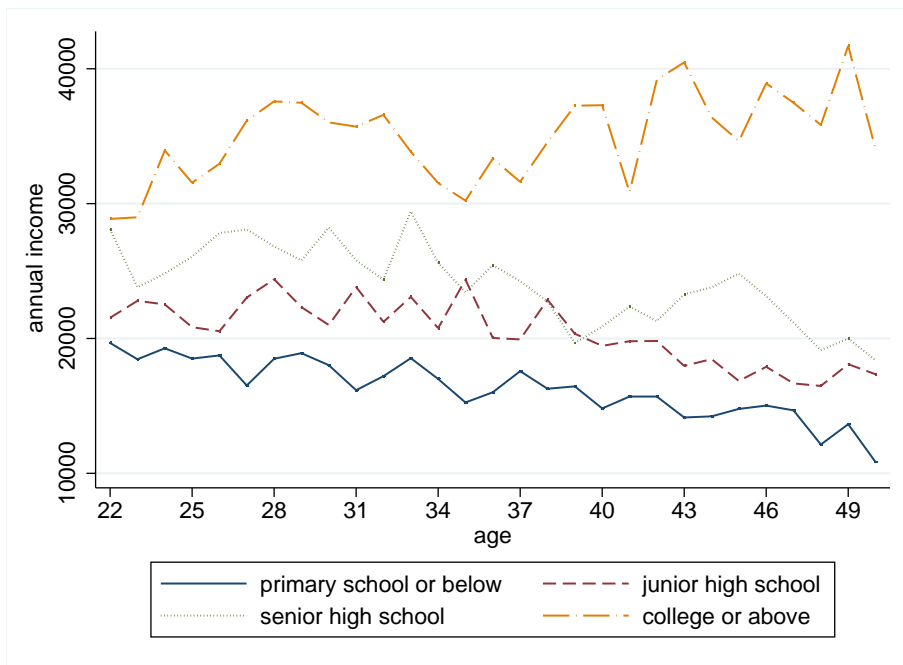


Figure B.6: Income against Age by Educational Groups for Children

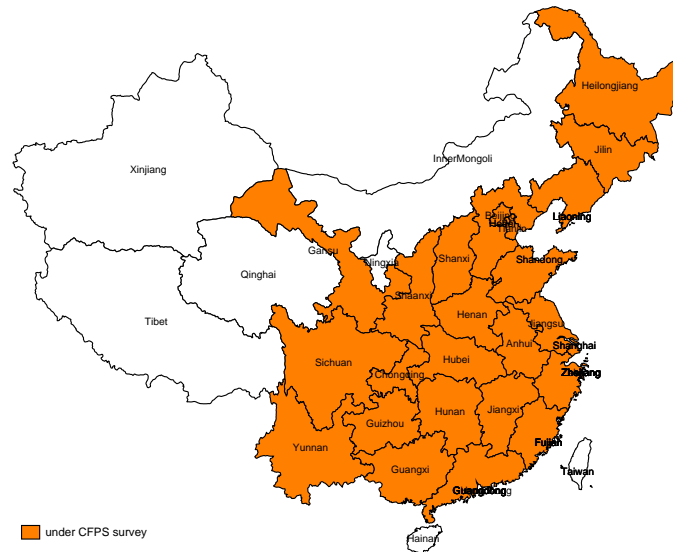


Figure B.7: The 25 Provinces and Municipalities under the China Family Panel Studies 2010

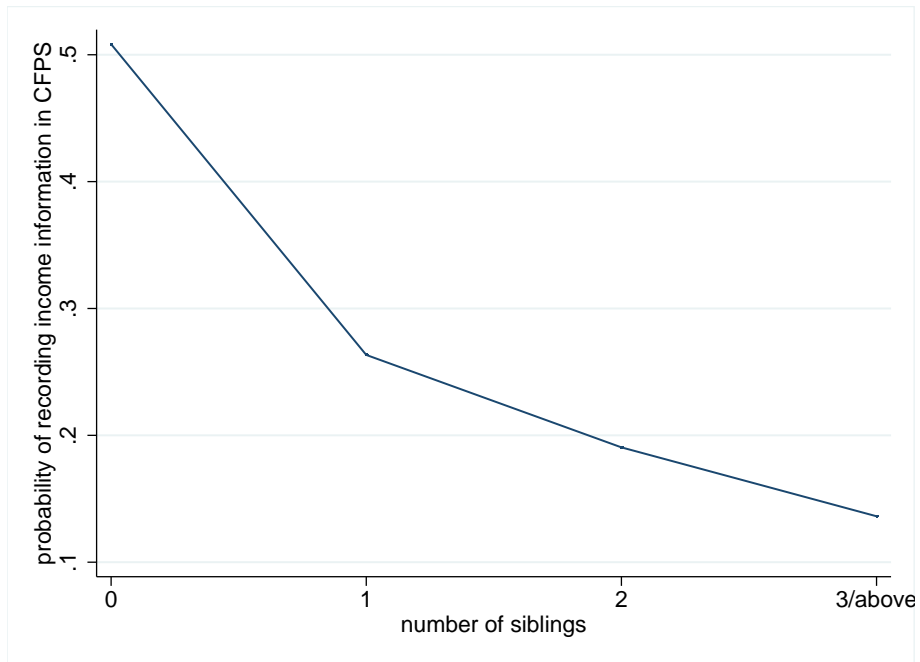


Figure B.8: Probability of Recording Income Information in CFPS against Number of Siblings

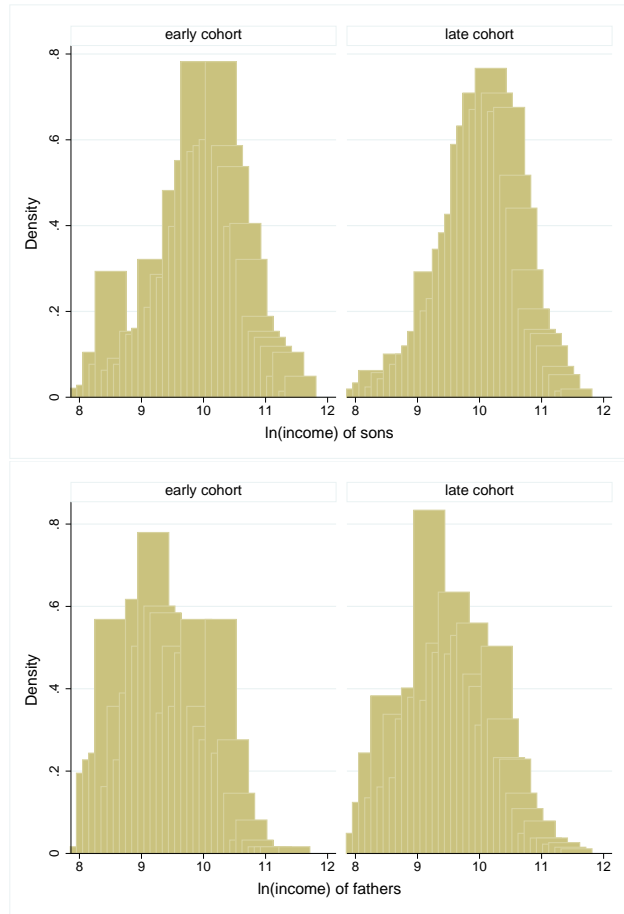


Figure B.9: Density Distribution of Log of Income by Fathers and Sons in Early and Late Cohorts

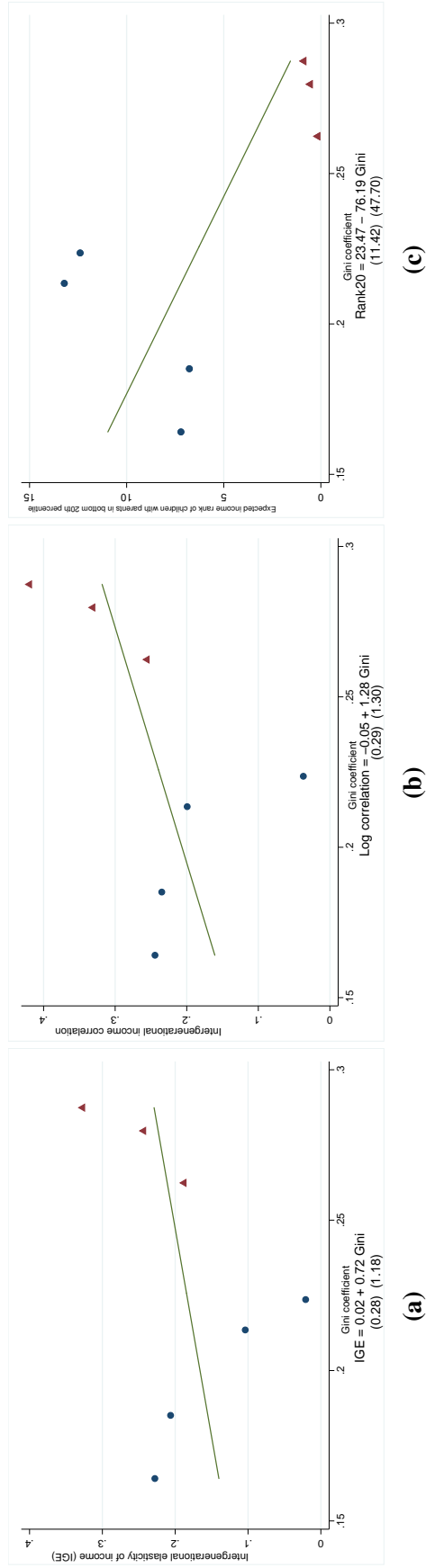


Figure B.10: The Great Gatsby Curve in Urban China

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.

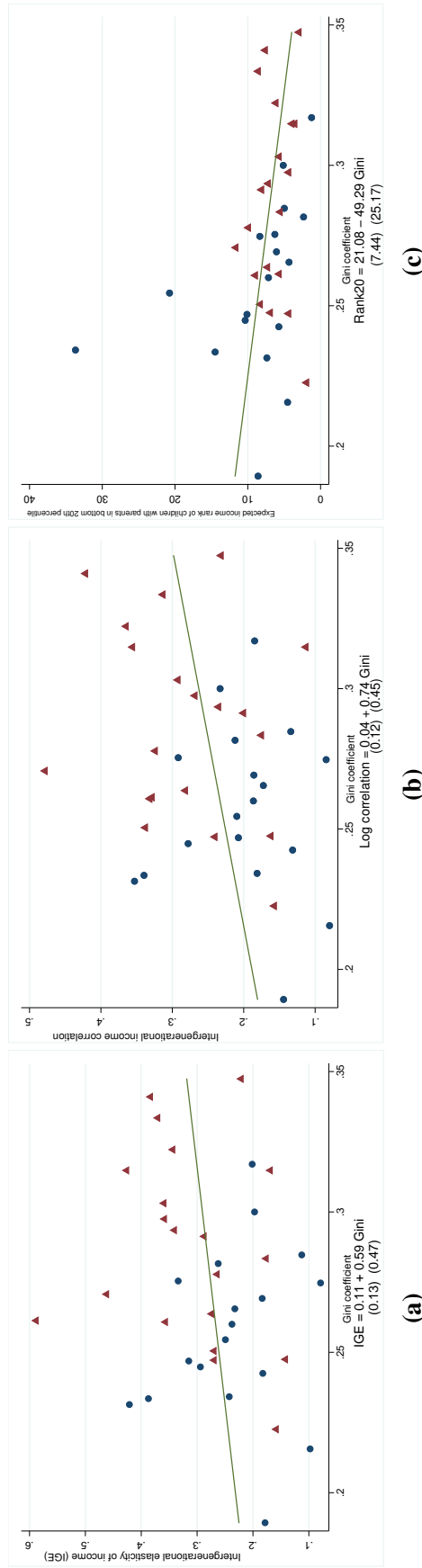


Figure B.11: The Great Gatsby Curve in Rural China

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	Late cohort	Change	Early cohort	Late cohort	Change	Early cohort	Late cohort	Change
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(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 (0.012)	0.617 (0.010)	0.194 (0.015)	0.270 (0.011)	0.446 (0.012)	0.176 (0.016)	0.131 (0.008)	0.255 (0.010)
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 (0.015)	0.761 (0.014)	0.232 (0.020)	0.321 (0.011)	0.522 (0.014)	0.201 (0.017)	0.154 (0.009)	0.283 (0.012)
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*** (0.009)	-0.300*** (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*** (0.013)	-0.384*** (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*** (0.009)	0.035*** (0.009)	-0.026 (0.018)	-0.024 (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 (2.130)	-1.094 (2.135)	0.014 (0.011)	0.014 (0.011)
age cubed/1000 × cohort × coastal	2.851 (5.707)	3.052 (5.790)	-0.022 (0.020)	-0.022 (0.020)
age cubed/1000 × Hukou × coastal	-0.597 (2.637)	-0.618 (2.637)	-0.011 (0.012)	-0.011 (0.012)
age cubed/1000 × cohort × Hukou × coastal	-4.800 (7.066)	-4.914 (7.132)	0.032 (0.022)	0.033 (0.022)
male × cohort	0.329*** (0.092)	0.316*** (0.093)		
male × Hukou	0.373*** (0.078)	0.392*** (0.079)		
male × cohort × Hukou	-0.065 (0.109)	-0.066 (0.110)		
male × coastal	0.063 (0.094)	0.062 (0.095)		
male × cohort × coastal	-0.250 (0.131)	-0.240 (0.133)		
male × Hukou × coastal	-0.205 (0.119)	-0.196 (0.119)		
male × cohort × Hukou × coastal	0.226 (0.164)	0.206 (0.165)		
live father × cohort			0.460 (0.330)	0.489 (0.332)
live father × Hukou			0.611* (0.250)	0.605* (0.248)
live father × cohort × Hukou			-0.462 (0.369)	-0.473 (0.371)
live father × coastal			-0.001 (0.314)	0.003 (0.312)
live father × cohort × coastal			-0.306 (0.448)	-0.328 (0.450)
live father × Hukou × coastal			0.173 (0.403)	0.172 (0.400)
live father × cohort × Hukou × coastal			-0.196 (0.537)	-0.185 (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)
<hr/>				
Observations	22,313	22,313	22,313	22,313
<hr/>				

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)
<hr/>				
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 (7,674.016)	23,761.27 (7,380.334)
Income of parents	17,978.72 (8,161.698)	20,293.59 (8,223.27)
Log(income of children)	9.949 (0.344)	10.030 (0.304)
Log(income of parents)	9.717 (0.383)	9.851 (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*** (0.009)	0.391*** (0.010)	0.442*** (0.008)	0.448*** (0.008)
child's age	-0.051 (0.028)	-0.054* (0.024)	0.022 (0.027)	0.012 (0.033)
child's age squared/100	0.051 (0.041)	0.055 (0.035)	-0.028 (0.053)	-0.013 (0.065)
father's age	-0.217*** (0.037)	-0.219*** (0.035)	0.042*** (0.008)	0.043*** (0.010)
father's age squared/100	0.197*** (0.031)	0.199*** (0.029)	-0.030*** (0.008)	-0.031*** (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:

$$(D1) \quad I\hat{G}E_{it} - I\hat{G}E_{i,t-1} = \gamma_0 + \gamma_1(\text{Correlate}_{it} - \text{Correlate}_{i,t-1}) + v_i,$$

where $I\hat{G}E_{it}$ ($I\hat{G}E_{i,t-1}$) 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} ($\text{Correlate}_{i,t-1}$) 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 20th or top 20th 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_{it}^{20} ($\text{rank}_{i,t-1}^{20}$) to denote the expected percentile rank of the child in his/her national income distribution if his/her parents are at the 20th rank in the national distribution of their generation. Following Chetty *et al.* (2014),

$$(D2) \quad \text{rank}_{it}^{20} = \hat{\beta}_{0,it} + 20\hat{\beta}_{1,it},$$

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_{it}^{20} indicates a high expected rank for a child from a poor family. To study intergenerational persistence for rich families, we define

³We 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.

⁴Detailed data sources are described in Table D2.

$rank^{80}$ in a similar way as $rank^{20}$. By replacing the dependent variable with $rank_{it}^{20} - rank_{i,t-1}^{20}$ ($rank_{it}^{80} - rank_{i,t-1}^{80}$) 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 20th 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

⁵Because 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 20th 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

⁶Li *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.

⁷The variable “share of tertiary industry is excluded because of collinearity with “share of primary industry” and “share of secondary industry”.

⁸These results are available upon request.

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

Province	IGE	
	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 (0.084)	-0.265 (0.773)	0.312 (0.124)	0.265 (1.141)
Rank of child with parents from 20 th national rank	12.096 (7.106)	0.494 (1.164)	6.060 (2.588)	-0.494 (0.424)
Rank of child with parents from 80 th national rank	29.095 (10.750)	0.073 (1.065)	27.616 (9.581)	-0.073 (0.949)
Gini coefficient	0.246 (0.038)	-0.507 (0.940)	0.286 (0.033)	0.485 (0.809)
Share of primary industry	0.291 (0.088)	0.609 (0.903)	0.173 (0.066)	-0.609 (0.674)
Share of secondary industry	0.408 (0.082)	-0.179 (1.124)	0.434 (0.062)	0.179 (0.845)
Share of tertiary industry	0.300 (0.035)	-0.754 (0.566)	0.393 (0.045)	0.754 (0.732)
Outflow migration rate ^a	0.042 (0.012)	-0.929 (0.164)	0.178 (0.033)	0.929 (0.460)
Urbanization rate	0.317 (0.191)	-0.221 (0.959)	0.397 (0.203)	0.181 (1.018)
Share of private enterprises	0.105 (0.237)	-0.215 (0.950)	0.212 (0.256)	0.215 (1.023)
Per capita GRP (1,000 yuan)	1.718 (1.040)	-0.622 (0.201)	8.152 (5.659)	0.622 (1.094)
Poverty rate	0.130 (0.174)	-0.130 (0.330)	0.271 (0.740)	0.137 (1.404)
Expenditure on education, science, culture, & public health per capita	50.876 (23.375)	-0.523 (0.152)	211.809 (185.115)	0.523 (1.204)
University students per 10,000 people	20.426 (16.739)	-0.517 (0.665)	46.451 (25.741)	0.517 (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 20 th percentile rank	Rank of child born to parents at the top 20 th percentile rank
(1)	(2)	(3)	(4)
Panel A. Inequality			
Gini coefficient	0.350 (0.578)	-0.932 (0.499)	-0.109 (0.341)
Panel B. Structural Changes			
Share of primary industry	0.671 (0.723)	-0.712 (0.771)	-0.083 (0.677)
Share of secondary industry	0.595 (0.592)	-0.025 (0.519)	0.116 (0.300)
Share of tertiary industry	-0.493 (0.383)	0.077 (0.406)	-0.074 (0.243)
Outflow migration rate ^a	-1.084 (0.648)	0.299 (0.749)	-0.120 (0.466)
Urbanization rate	-0.054 (0.599)	0.131 (0.747)	0.080 (0.920)
Share of private enterprises	4.012 (1.778)	-0.682 (1.735)	1.296 (1.395)
Panel C. Economic Development			
Per capita GRP	1.307 (0.819)	-0.350 (0.706)	0.310 (0.557)
Poverty rate	0.559 (0.835)	0.063 (1.632)	-0.474 (1.411)
Panel D. Redistribution and Education Policies			
Expenditure on education, science, culture, & public health per capita	2.379 (1.136)	-1.297 (1.259)	0.104 (0.894)
University students per 10,000 people	0.556 (0.713)	-1.230 (0.741)	-0.434 (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 20th national percentile rank, and the rank of a child born to parents at the top 20th 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.
- 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**.
- 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.