Intergenerational Earnings Mobility in China
Jimmy Chin

1. Introduction

China's three decades of economic growth has brought about an immense change in the structure of Chinese society. While economic reforms improved living standards and lifted millions out of poverty, the distribution of those gains were not shared equally as growth was accompanied by a large increase in income inequality. Both trends are well-documented. However, less attention has been paid to equality of opportunity in China. Equal opportunities that allow individuals, irrespective of background, to succeed and move up the economic ladder may partly compensate factor for the overall disparity in income. A key measure of equality of opportunity is the level of intergenerational mobility

A key measure of equality of opportunity is the level of intergenerational mobility. This study examines the heterogeneity in intergenerational income effects in the fast growing Chinese economy. Specifically, there are three questions I aim to answer. First, what is the level of intergenerational mobility of earnings in China? Second, how is mobility different across time and for different segments of the population? Finally, what are some of the mechanisms behind in the trends in intergenerational mobility? To answer the first question, I use the well-known Galton-Becker-Solon model to produce baseline estimates of intergenerational persistence. This allows for comparisons with other studies employing the same methods. In examining mechanisms, I plan to develop a model that focuses on the effects of one causal variable. This may be education, or the hukou system, or trade. I am currently working on the model.

There is a growing literature on intergenerational mobility in China. This study makes a number of contributions. First, this is the most up-to-date study on intergenerational mobility in China. Previous China studies use 2011 data at the latest, while this paper includes data from 2015. Second, this paper compensates for a number of methodological shortcomings in the literature on intergenerational mobility in China. Many previous studies fail to adequately correct for selection bias, lifecycle bias, endogeneity, and persistent transitory shocks to income. I use selection models to compensate for employment and co-residency bias, and employ college expansion at the municipal level as a novel instrument for parent income. A number of methods are used to correct for lifecycle bias, including a stochastic model for parent earnings that calculates the parent's fixed effect. The fixed effect represents the permanent component of parent earnings. This is also the first China study to estimate the level of intergenerational mobility using percentile ranks of parents and children. This avoids the problem that fast growth in overall earnings poses for using average earnings as a proxy for permanent income. Finally, this study uses a novel way to construct income from the CHNS dataset by including the individual's productivity-adjusted contribution to household income in the measurement of their individual income.

2 Literature Review

2.1 General Approaches

There exists an extensive literature examining the intergenerational mobility of income between parents and children.[1] Most empirical studies focus on the intergenerational dynamics of developed countries, largely owing to higher-quality data, but a sizable literature exists for developing countries as well. Moreover, most of these studies estimate the IGE for parent-child

[1] See Black and Devereux (2011) and Solon (1999) for surveys.

pairs. The canonical model originates in Becker and Tomes (1979) and is rationalized in Solon (2004). It takes the form:

$$Y_{it}^{c} = \alpha + \beta Y_{i}^{f} + \epsilon_{it},$$

where y_1 and $y_0$represent the log of child and parent income, respectively, and coefficient β is the IGE.

The first studies analyzing intergenerational mobility in the US estimate elasticity on the order of around 0.2. However, unrepresentative cross-sectional datasets used in these early papers biased estimates of the IGE downwards. In seminal work, Solon (1992) and Zimmerman (1992) average fathers' income over a period of three to five years and employ IVs to find US IGE to be approximately 0.4. Mazumber (2005) argues that these findings are still downward biased owing to persistent transitory shocks and estimates an elasticity of 0.6 using 16-year averages.

Contemporary research continues to use averages of parent income over a number of years, as well as IV or TS2SLS approaches, to estimate the IGE. Solon (1992) uses father's education and Zimmerman (1992) uses father's socioeconomic status as instruments for his permanent status. Instruments used in other studies include father's occupation (Björklund and Jäntti 1997), union status (Shea, 2000), and industry (Gong et al., 2012). Using an IV provides a reasonable upper limit on IGE if the instrument is a direct determinant of a child's income. The interval between OLS and IV estimates may therefore be interpreted as the bounds for the true value of income elasticity.

IGE estimates are also biased due to observations that occur during different parts of the lifecycle. Different methods have been employed to compensate for this lifecycle bias, as well as for persistent transitory shocks to income mentioned above. Hendricks (2007) computes the intergenerational persistence of income by simulating models of parent and child earnings. Mazumder (2003) replicates the results from Solon (1992) and uses an earnings model to calculate reliability ratios, or the degree of attenuation bias in the Solon's sample. Baker and Solon (2003) use a large dataset to estimate a model that incorporates both heterogeneous earnings effects and random walk components.

IGE in the US is generally large when compared to estimates of other OECD countries. For example, Bjorklund and Jantti (1997) find IGE to be 0.28 for Sweden; Corak and Heisz (1999) estimate an elasticity of 0.23 for Canada; and Couch and Dunn (1997) report an elasticity of 0.11 for Germany. There are, however, some exceptions within developed nations. For example, Britain has a low rate of intergenerational mobility comparable to the US. Ermisch and Francesconi (2004) find IGE for Britain as high as 0.45 to 0.75 for father-son pairs. For a survey of cross-country comparisons, see Solon (2002).

Fewer studies have examined trends in intergenerational mobility in developing countries—largely owing to a lack of quality data. However, the evidence thus far indicates a level of mobility considerably higher in developed than in developing states. Dunn (2007), for example, finds IGE in Brazil to be 0.80, and Piraino (2015) estimates elasticity for South Africa at over 0.60.

There is also another strand of literature that investigates the mechanisms behind intergenerational mobility. Mechanisms that transmit income or status between parents and children include parental influences like the provision of social connections and genetic transmission of ability, as well as non-parental factors like education policy (Nybom and Stuhler, 2013) and political participation and public spending on education (Ichino et al., 2011). Chetty et al. (2014) estimate correlations between mobility and a number of other variables, including race, neighborhoods, and family structure. Corak (2013) surveys the literature on the relationship between intergenerational mobility and income inequality.

2.2 China

There is a growing literature on intergenerational mobility in China, and a consensus has yet to emerge. Researchers find different levels of mobility using different data sources, models, and methods of constructing income. Guo and Min (2008) estimate IGE in urban China at 0.32, but only use single-year observations for income. Gong et al. (2012) build on their study by averaging parent incomes and using TS2SLS methods to obtain an elasticity of 0.63. Deng et al. (2013) and Fan (2016) both find IGEs at around 0.5 using China Household Income Project (CHIP) surveys for the years of 1995 and 2002.

Several studies utilize the China Health and Nutrition Survey (CHNS) to estimate income elasticity in China. Zhang and Eriksson (2010) find an elasticity of 0.45 for survey waves from 1989 to 2006. Up to 2009, Qin et al. (2016) obtains an elasticity of 0.481 using a simultaneous equations model that accounts for human capital transmission. For the same years, Yuan (2017) estimates the IGE at around 0.5 to 0.6, which varies considerably for different income groups as well as for urban and rural areas.

In the China literature, there has been less attention on the transmission mechanisms that drive intergenerational mobility. One study by Yuan and Chen (2013) decomposes income elasticity and find that human capital, social capital, and wealth explain more than 60 percent of variation. Fan (2016) examines education, social capital, and ownership of the child's work unit and finds that effects vary for different income groups. The mechanisms of intergenerational mobility in China deserve closer attention.

3. Data

This study uses longitudinal data from the China Health and Nutrition Survey (CHNS). The survey is a long-standing collaborative project between the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention. Its primary goal is to collect data that allow researchers to examine the effects of health, nutrition, and family planning policies during the past several decades of Chinese social and economic transformation.

Thus far the survey includes 10 waves (1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011, and 2015) and covers 9 to 15 provinces and direct-controlled municipalities per wave.[2] The average number of households surveyed per wave ranges from 3,500 to 4,500 during the years 1989 to 2009. That range increases to 5,900 to 7,300 with significant expansions of the survey in the next two waves. In addition to the household survey, members currently living in the household participate in the individual survey. In 2015, 20,694 household members responded to the individual questionnaire. These respondents reside in 361 communities (or cities, townships and villages), which represents a large increase from about 200 communities surveyed before 2011.

CHNS uses a multistage, random cluster sampling process for sample selection and aims to be diverse in various socioeconomic characteristics (e.g., income, health, education, modernization). Counties are stratified by income and selected using a weighted sampling method. Villages and townships within the counties are selected randomly. For municipal cities, urban and suburban neighborhoods are randomly sampled.

[2] The CHNS survey covers the following provinces: Beijing (2011, 2015), Chongqing (2011, 2015), Guangxi, Guizhou, Heilongjiang (1997-on), Henan, Hubei, Hunan, Jiangsu, Liaoning (except for 1997), Shaanxi (2015), Shandong, Shanghai (2011, 2015), Yunnan (2015), and Zhejiang (2015).

One advantage of CHNS is that it is a large longitudinal dataset spanning 26 years. The length and design of the survey is a benefit given the vulnerability of short panels to attenuation bias produced by persistent transitory income shocks (see Mazumder, 2005). Life-course surveys that occur every few years may yield estimates less exposed to this bias than more frequent surveys with the same number of waves.

CHNS also controls partially for co-residency bias. Other surveys that provide incomes for parent-child pairs only when they are living in the same household produce estimates that are vulnerable to this form of selection bias. The advantage of CHNS is that it links parents and children through identification numbers as long as they are observed living together in at least one wave. Thus children are connected to their parents even after moving out. However, this does not account for children who are never observed living with parents in the survey.[3] I use a Heckman selection model to correct for this bias. A last advantage of the dataset used in this study is its provision of new data. To my knowledge, studies using CHNS to estimate intergenerational mobility have only analyzed waves up to 2009. Other studies using different datasets use data up to 2005 at the latest. Thus this study uses a greater number of observations than the most recent papers, and it is the first to examine intergenerational mobility under the Xi Jinping government.

One disadvantage of the dataset is the attrition of individuals and households in the survey. Households may exit the survey for unknown reasons, and they are replaced in subsequent surveys in order to maintain a similar number of individuals and households in each wave. Individuals who leave the household and move to a municipality or province not covered in the survey are also lost.

I define an individual's annual labor earnings as an individual's total wage from primary and secondary jobs, plus their labor earnings from household farming, livestock and poultry production, fishing, and household business.[4] Earnings from household activities are individualized using hours engaged in work and the estimated productivity of the individual which is computed based on age, education, and year and province fixed effects. Wages are imputed if an individual participates in a labor activity but does not report any earnings.[5] Two CHNS-constructed measures of income are used in the sensitivity analysis. The first contains annual wages, bonuses, and other cash and non-cash non-labor income. The second represents total net individual income. In addition to wages, bonuses, and non-labor income, it adds income derived from the individual's contribution to household income based on hours worked. All measures of earnings and income are adjusted for inflation using provincial CPI.

Summary statistics for the raw sample and the father-son estimation sample are found in Tables 1 and 2 below. The raw sample includes all individuals, including those who do not report parents in the sample. The estimation sample includes only father-son pairs that satisfy the age restriction. In the estimation sample, 18.5 percent of respondents report urban status and 85.3 percent report being Han Chinese.

Table 3 contains a father-son intergenerational transition matrix. It shows that children born into lowest quintile of the earnings distribution are less mobile than those born at the top. There seems to be significant mobility in the middle quintiles. Note, however, that the transition

[3] Co-residency rates vary by age. 83 percent of 25 year old males are linked to fathers. That number decreases to 49 and 22 percent for 35 and 45 year old males.

[4] See Appendices 1 and 3 for detailed variable descriptions.

[5] Some STATA code for CHNS data cleaning, including construction of annual labor earnings, was provided by Professor Peter.

matrix presents data on relative earnings. In a fast-growing economy, a child may lose their relative position but earn much more than their parents in absolute terms. There are also floor and ceiling effects which make children born into the bottom and top of the earnings distribution less mobile (since there is one less direction to move). Roemer (2004) discusses the normative significance of the transition matrix, and points out equal entries in each cell of the matrix should not be identified with perfect equality of opportunity. It may be undesirable to completely compensate children for the influence of parents (e.g., for the formation of preferences or the genetic transmission of ability).

4. Empirical Model

4.1 Baseline model

First, I focus on a baseline model that takes the form:

$$Y_{it}^{c} = \alpha + \beta_1 Y_i^f + \beta_2 X_{it}^c + \beta_3 X_i^f + \beta_4 \theta_t + \epsilon_{it}, \quad (1)$$

where Y_{it}^{c} represents the log of child labor earnings in year *t* and Y_i^f represents the log of permanent labor earnings of child's father. Both variables are in real terms. X_{it}^c is a vector that contains characteristics of the child: a quadratic function of age, , ethnicity, current province of residence, and urban status of residence. X_i^f is a vector that contains father's birth province and a quadratic function of the age measured at the time of observed earnings. θ_t represents wave or year dummies, and ϵ_{it} is an error term. β_1 is the key parameter of interest that captures the intergenerational persistence of earnings between fathers and children.

I estimate three different baseline models that address a number of issues and biases in estimation. The first uses averages obtained over a minimum of nine years to proxy for the father's permanent earnings. This is the baseline model that yields an estimate of the IGE. The second uses percentile ranks in the earnings distribution to compute a rank-rank slope as a measure of intergenerational persistence. The third estimates a stochastic model for father's earnings and predicts and employs the fixed effect as a measure of father's permanent earnings. This method also produces an estimate for the IGE. In sections 4.2 to 4.6 that follow, I describe the estimation issues and methods that are used to correct for them. Section 4.7 describes a model for a causal mechanism of intergenerational mobility.

4.2 Persistent transitory shocks

Persistent transitory shocks to income are one source of attenuation bias in estimating the IGE. Using averages instead of single-year observations for earnings is one way to compensate for this bias. Solon (1992) and Zimmerman (1992) average fathers' income over a period of three to five years and estimate an IGE of around 0.2, which is significantly greater than previously thought. Mazumder (2005) averages incomes over 16 years and finds an even higher IGE of around 0.6. He also shows using simulations that earnings averaged over a smaller number of years incur significant attenuation bias. Notice, however, that the studies described above all focus on the US. Black and Devereux (2011) note that cross-country differences in transitory earnings variation may be significant. For example, Nilsen et al. (2008) finds smaller effects on the IGE using averages of father's earnings.

Following the literature convention, I average parent earnings to proxy for permanent income. I compute averages using a minimum time span of nine years. . This requirement ensures averages are obtained using a minimum of four observations. Mean ages for the time period in which averages are computed are used when controlling for quadratics of age.

A drawback of this approach is that increasing the time span from which averages are computed significantly reduces the sample size. As an alternate approach, I generate a stochastic model for earnings and compute each parent's fixed effect in order to estimate permanent income.

The fixed effect model for parent earnings is as follows:

$$Y_{it}^{f} = \alpha + \beta_1 X_{it}^{f} + +\beta_2 \theta_t + \beta_3 \mu_i^{f} + \epsilon_{it}. \quad (2)$$

Y_{it}^{f} represents the log of father's annual labor earnings. X_{it}^{f} includes time-variant characteristics relevant to the father like (household size, marital status, hours, type of activity). θ_t represents year dummies. μ_i^{f} represents a fixed effect component where time-invariant characteristics like urban status, ethnicity, birth and current province, and so on are captured. The permanent component of the log of father's labor earnings is computed using the following equation:

$$Y_i^{f} = \hat{\mu}_i^{f} + \beta_1 \alpha + \beta_2 \varphi_{it}^{f} * 40 + \beta_3 \tau_{it}^{f} * 40^2, \quad (3)$$

where Y_i^{f} is the estimated permanent component for fathers at age 40 in 1991. $\hat{\mu}_i^{f}$ is the estimated fixed effect, α is the constant term, φ_{it}^{f} is age, and τ_{it}^{f} is age squared. Y_i^{f} computed in equation (3) is then used in equation (1).

4.3 Fast overall growth in earnings

Averaging earnings has a drawback when estimating intergenerational mobility in fast-growing economies. In the case of post-reform China, high rates of growth and the associated increases in earnings swamps the effect of transitory shocks. Average earnings are highly sensitive to the years in which the observations of earnings are drawn. To illustrate, in my sample, average annual labor earnings is 4,000 yuan in 1991; 8,000 yuan in 2004; and 30,000 yuan in 2015 (all in real terms). Average earnings are therefore not an adequate proxy for permanent income.

To compensate for earnings growth, I estimate persistence of earnings using the percentile ranks of individuals in the yearly earnings distribution. The resulting estimate for intergenerational mobility is the rank-rank slope between parents and children. The slope and the IGE are two different but related measures. Chetty et al. (2014) note that both measures capture the joint distribution of parent and child percentile ranks, but only the IGE is affected by the distribution of parent and child incomes. Previous studies employed the rank-rank specification (Chetty et al., 2014; Dahl and DeLeire, 2008), but this article is the first to do so for China. Importantly, percentile ranks are still exposed to transitory earnings shocks. I use average percentile ranks to mitigate these effects. I follow the same procedure for averaging earnings described above.

4.3 Life-cycle bias

Estimates of the IGE are biased when observations of earnings at different points in the life-cycle are not representative of permanent earnings. For example, Grawe (2006) finds a negative correlation between the size of the IGE and the father's age at time of measurement. For men, attenuation bias due to the lifecycle is high in their twenties and after their late 40s (Haider and Solon, 2006; Böhlmark and Lindquist, 2006). Most studies include quadratics of parent and child age in their models, but this only controls for lifecycle bias on the level of earnings, and not heteroskedastic measurement error (Mazumder, 2003). I follow convention in the literature by controlling for quadratics of age. When using average earnings or percentile ranks, I employ the mean age of the individual for the range of years in which the average was taken. Also following convention, I restrict age ranges and exclude children younger than 25 and parents older than 65 in the sample. Furthermore, I use percentile ranks adjusted for age in order to account for life-cycle bias. Finally, the stochastic model approach accounts for this bias by estimating the parent's fixed effect.

4.4 Sample selection bias

As noted above, CHNS partially compensates for co-residency bias because children need not reside with their parents in order to be identified as a child. As long as parents and children have lived together at some point in the survey, they will be linked. However, this does not account for parents and children who are never observed living together in the survey. Indeed, sons in our sample who are linked to parents have lower earnings than those who are not. Therefore, estimates that do not control for this form of selection are biased. Moreover, bias occurs when only fathers who report employment are selected into the sample. I use a selection model to correct for these two biases. The dependent variable in the probit model equals one if the individual is in the estimation sample (i.e., has a father whose earnings are averaged over at least nine years) and equals zero if the individual is not linked to any father. The independent variables are a quadratic of age, ethnicity, urban status, education, and current province. Following Deng et al. (2013) and Fan (2016), I use municipal co-residency rates as the exclusion restriction. The inverse propensity weight is then computed and added into baseline equation (1).

4.5 Endogeneity

Parent education is a frequently used instrument in the literature, but it is at best an upper-bound on the IGE because of its endogeneous relationship with child earnings. I use college expansion, as measured by the number of campuses opened in each municipality, as an instrument for parent earnings. The number of campuses is measured at age 17 of the parent. College expansion proxies for parent income and should be uncorrelated with the error term.

4.7 Mechanism analysis

The second part of this paper moves beyond estimating the level of intergenerational mobility and examines a mechanism of mobility. One possible causal variable to examine is education. Most models of this kind are static. I may develop a dynamic simultaneous equations model following Nybom and Stuhler (2014). The authors exploit their model to analyze how educational reforms in the distant past affect intergenerational mobility in the multiple generations that follow. Ichino et al. (2011) look at the political economy of intergenerational mobility. The mechanisms in their model are political participation and public spending on education. Ermisch et al. (2006) examine assortative mating as a mechanism of intergenerational mobility. I will develop a model that similarly examines a mechanism of mobility. Some variables of interest include education, hukou status, corruption, and trade.

5. Results

5.1 Baseline results

Table 4 contains the results for the baseline estimates. Each specification averages father's earnings over a minimum nine year time span. The models presented proxy for father's permanent income in three different ways. (1) uses the log of average earnings; (2) uses average percentile ranks; and (3) uses the permanent component of earnings derived from the log of average earnings. (1) and (3) estimate the IGE, while (2) estimates the rank-rank slope. The three specifications yield estimates of 0.49, 0.45, and 0.45 for intergenerational persistence of earnings, respectively. The estimates are nearly similar. Ethnicity is barely significant in only one of the models, while urban residence significantly increases intergenerational persistence in all three specifications.

Figure 1 shows the results of the quantile regression with a 95 percent confidence interval. Individuals in the upper end of the earnings distribution are highly mobile with estimated elasticities of around 0.2 to 0.3. On the other hand, those with lower earnings show extremely high levels of persistence at approximately 0.8. These results are consistent with findings from the transition matrix. One explanation advanced by Yuan (2017) for the disparity in persistence is that the variance in economic backgrounds for the rich is significantly greater than it is for the poor. This decreases the explanatory power of the father's income for the outcomes of the child.

5.2 Sensitivity analysis

Table 5 presents the results for the sensitivity analysis. (4) and (5) present estimates of the IGE using 3 year and 15 year minimum time spans for average father's earnings, respectively. Consistent with findings from the literature, averaging earnings over a longer period of time produces larger estimates of persistence.

(6) and (7) present two specifications with alternate constructions of percentile rank. Percentile ranks are unadjusted for age in (6), and ranks are adjusted for both age and location in (7). Nine year minimum averages are used, and the results are similar to the age-adjusted measure.

(8) and (9) employ the CHNS-constructed measures of earnings. In (8), using the total net individual income produces an estimate of 0.36. (9) uses individual wage, bonus, and other income and estimates the IGE at 0.149. Both are significantly lower than the results of all of the other specifications. One reason for this discrepancy may be that the CHNS variables contain transitory components of income like bonuses and cash and non-cash gifts. This could attenuate estimates of the IGE. Previous studies using these variables estimate the IGE at around 0.4-0.5 (Qin et al. 2016; Yuan, 2017). Their results are close to my result using annual labor earnings. However, both studies use a limited number of controls. Yuan (2017) includes only age controls in his model, while Qin et al. (2016) controls for age, working in a state-owned sector, urban status, and living in a coastal region. When I include only age controls in models using the CHNS variables, I am able to obtain results closer to those found in the literature.

Bibliography

Baker, M. & Solon, G. (2003). Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records. Journal of Labor Economics, 21(2), 289-321. http://www.jstor.org/stable/10.1086/345559

Becker, G. S., & Tomes, N. (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, *87*(6), 1153–1189. https://doi.org/10.2307/1833328

Black, S. & Devereux, P.J. (2011). Recent developments in intergenerational mobility. In Orley Ashenfelter and David Card (Eds.), *Handbook of Labor Economics Volume 4B* (pp. 1487–1541). Elsevier.

Björklund, A., & Jäntti, M. (1997). Intergenerational Income Mobility in Sweden Compared to the United States. *The American Economic Review*, *87*(5), 1009–1018.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553–1623. https://doi.org/10.1093/qje/qju022

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, *27*(3), 79–102. https://doi.org/10.1257/jep.27.3.79

Corak, M., & Heisz, A. (1999). The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data. *The Journal of Human Resources*, *34*(3), 504–533. https://doi.org/10.2307/146378

Couch, K.A., & Dunn, T.A. (1997). Intergenerational Correlations in Labor Market Status: A Comparison of the United States and Germany, *The Journal of Human Resources, 32*(1), 210-232. http://www.jstor.org/stable/146246

Deng, Q., Gustafsson, B., & Shi, L. (2013). Intergenerational Income Persistence in Urban China. *Review of Income and Wealth, 59*(3), 416-436.

Dunn, C. (2007). The Intergenerational Transmission of Lifetime Earnings: Evidence from Brazil. *The B.E. Journal of Economic Analysis & Policy*, *7*(2). https://doi:10.2202/1935-1682.1782

Ermisch, J & Francesconi, M. (2004). Intergenerational mobility in Britain: new evidence from the British Household Panel Survey. In M. Corak (Eds.), *Generational Income Mobility in North America and Europe* (pp. 147-189). Cambridge, UK: Cambridge University Press.

Fan, Y. (2016). Intergenerational income persistence and transmission mechanism: Evidence from urban China. *China Economic Review*, *41*, 299–314. https://doi.org/10.1016/j.chieco.2016.10.005

Gong, H., Leigh, A., & Meng, X. (2012). Intergenerational Income Mobility in Urban China. *Review of Income and Wealth, 58*(3), 481–503. https://doi.org/10.1111/j.1475-4991.2012.00495.x

Hendricks, L. (2007). The intergenerational persistence of lifetime earnings. *European Economic Review, 51*(1), 125-144. https://doi.org/10.1016/j.euroecorev.2005.11.009

Ichino, A., Karabarbounis, L., & Moretti, E. (2011). The Political Economy of Intergenerational Income Mobility. *Economic Inquiry*, *49*(1), 47–69. https://doi.org/10.1111/j.1465-7295.2010.00320.x

Mazumder, B. (2003). Revised Estimates of Intergenerational Income Mobility in the United States. (FRB of Chicago Working Paper No. 2003-16). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=473188.

Mazumder, B. (2005). Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data. *The Review of Economics and Statistics*, *87*(2), 235–255. https://doi.org/10.1162/0034653053970249

Nilsen, O.A., Vaage, K., Arkvik, A., & Jacobsen, K. (2008). Estimates of intergenerational elasticities based on lifetime earnings. (Discussion paper no. 3709, Institute for the Study of Labor (IZA)).

Nybom, M. & Stuhler, J. (2014). Interpreting Trends in Intergenerational Income Mobility. (Swedish Institute for Social Research Working Paper 3/2014). Retrieved from https://econpapers.repec.org/paper/hhssofiwp/2014_5f003.htm.

Piraino, P. (2015). Intergenerational Earnings Mobility and Equality of Opportunity in South Africa. *World Development*, *67*, 396–405. https://doi.org/10.1016/j.worlddev.2014.10.027

Qin, X., Wang, T., & Zhuang, C. C. (2016). Intergenerational transfer of human capital and its impact on income mobility: Evidence from China. *China Economic Review*, *38*, 306–321. https://doi.org/10.1016/j.chieco.2014.10.005

Shea, J. (2000). Does parents' money matter? *Journal of Public Economics, 77*(2), 155-184. https://doi.org/10.1016/S0047-2727(99)00087-0

Roemer, J. (2004). Equal opportunity and intergenerational mobility: going beyond intergenerational mobility: going beyond intergenerational income transition matrices. In M. Corak (Eds.), *Generational Income Mobility in North America and Europe* (pp. 48-57). Cambridge, UK: Cambridge University Press.

Solon, G. (1992). Intergenerational Income Mobility in the United States. *The American Economic Review*, *82*(3), 393–408. https://doi.org/10.2307/2117312

Solon, G. (1999). Intergenerational Mobility in the Labor Market. In Orley Ashenfelter and David Card (Eds.), *Handbook of Labor Economics Volume 3A* (pp. 1761-1800). Elsevier.

Solon, G. (2002). Cross-Country Differences in Intergenerational Earnings Mobility. *The Journal of Economic Perspectives*, *16*(3), 59–66. Retrieved from http://www.jstor.org/stable/3216950

Solon, G. (2004). A model of intergenerational mobility variation over time and place. In M. Corak (Eds.), *Generational Income Mobility in North America and Europe* (pp. 38-47). Cambridge, UK: Cambridge University Press.

Yuan, W. (2017). The Sins of the Fathers: Intergenerational Income Mobility in China. *Review of Income and Wealth*, *63*(2), 219–233. https://doi.org/10.1111/roiw.12222

Yuan, Z., & Chen, L. (2013). The Trend and Mechanism of Intergenerational Income Mobility in China: An Analysis from the Perspective of Human Capital, Social Capital and Wealth. *The World Economy*, *36*(7), 880–898. https://doi.org/10.1111/twec.12043

Zhang, Y., & Eriksson, T. (2010). Inequality of opportunity and income inequality in nine Chinese provinces, 1989–2006. *China Economic Review*, *21*(4), 607–616. https://doi.org/10.1016/j.chieco.2010.06.008

Zimmerman, D. J. (1992). Regression Toward Mediocrity in Economic Stature. *American Economic Review*, *82*(3). Retrieved from http://www.jstor.org/stable/2117313

Table 1: Raw Sample Summary Statistics

Variables	Variable Definition	Observations	Mean	Std. Dev.	Min	Max
	Son's variables					
annearni	Annual labor earnings	72,861	9503.814	25312.81	-32680.45	1200000
age	Age (years)	185,283	36.14311	20.46881	0	111
urban	=1 if urban	185,365	.3026299	.4593976	0	1
ethhan	=1 if Han Chinese	185,365	.8505165	.3565653	0	1
	Father's variables					
annearni		30,940	11019.62	27259.56	-32680.45	882000
age		53,425	42.97024	16.76148	0	111

Table 2: Estimation Sample Summary Statistics

Variables	Variable Definition	Observations	Mean	Std. Dev.	Min	Max
	Son's variables					
annearni	Annual labor earnings from individual jobs and household and business income (using different individual productivity)	2,957	16260.21	36548.31	2.812697	882000
age	Age (years)	2,957	32.1116	5.950778	25	65
urban	=1 if urban	2,957	.0940142	.2918978	0	1
ethhan	=1 if Han Chinese	2,957	.8454515	.3615349	0	1
	Father's variables					
av_annearni_f	Average annual labor earnings with minimum time span of nine years	2,957	6687.206	7280.126	603.2895	153126.6
av_age_f	Age (years)	2,957	53.00005	4.963354	34.5	60.75

Table 3: Father-Son Intergenerational Transition Matrix

Fathers Earnings Quintile (Averages)	Son Earnings Quintile 1	2	3	4	5	Total	
1	194	145	128	87	43	597	obs
	32.5	24.29	21.44	14.57	7.2	100	%
2	154	163	134	98	71	620	obs
	24.84	26.29	21.61	15.81	11.45	100	%
3	101	119	126	139	140	625	obs
	16.16	19.04	20.16	22.24	22.4	100	%
4	91	85	98	141	135	550	obs
	16.55	15.45	17.82	25.64	24.55	100	%
5	58	81	103	124	199	565	obs
	10.27	14.34	18.23	21.95	35.22	100	%
Total	598	593	589	589	588	2,957	obs
	20.22	20.05	19.92	19.92	19.89	100	%

Table 4: Baseline IGE Estimates

	(1)	(2)	(3)
	Log average earnings	Average percentile rank (age-adjusted)	Log Permanent earnings component
Father's permanent earnings measure	0.491***	0.455***	0.452***
	(0.045)	(0.035)	(0.041)
Age of father	-0.11	1.352	-0.124
	(0.08)	(1.618)	(0.079)
Age of father squared	0.001	-0.02	0.002**
	(0.001)	(0.016)	(0.001)
Age of son	0.081*	-0.964	0.102**
	(0.042)	(0.818)	(0.042)
Age of son squared	-0.001*	0.013	-0.001**
	(0.001)	(0.011)	(0.001)
Ethnicity	0.13	3.075	0.167*
	(0.095)	(1.908)	(0.094)
Urban	0.378***	6.551***	0.314***
	(0.079)	(1.707)	(0.080)
Observations	2,957	2,987	2,957

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 1: Quantile Regression

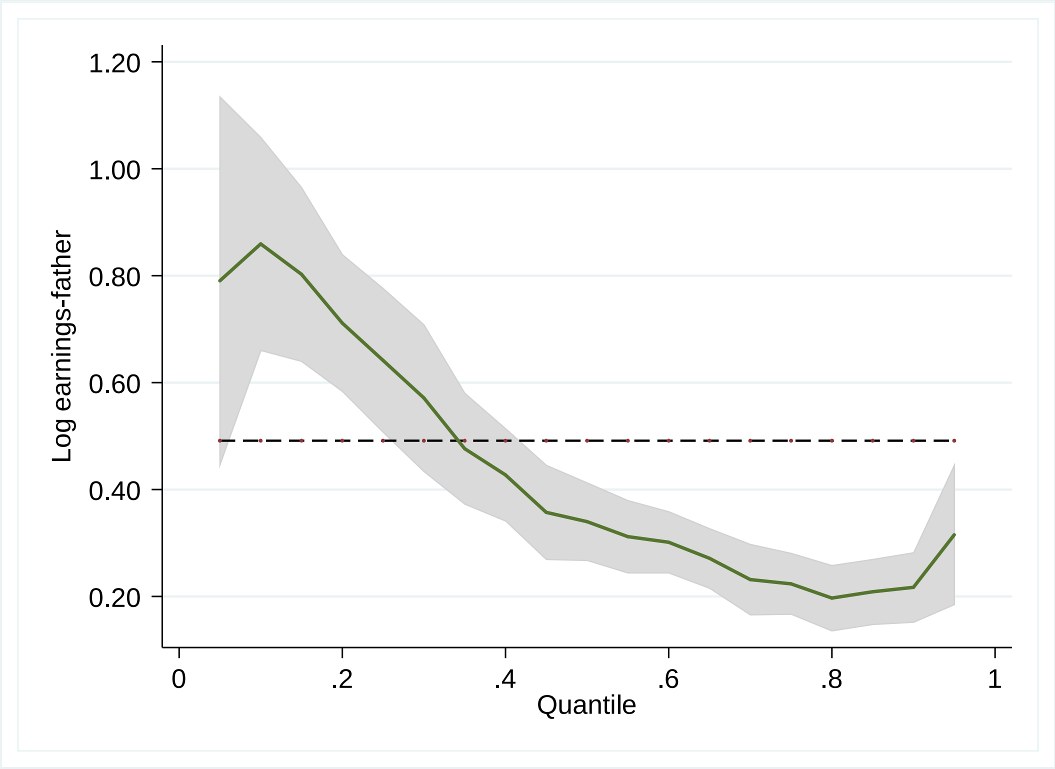

Table 5: Sensitivity Analysis

	(4) Log average earnings (3 year)	(5) Log average earnings (15 year)	(6) Average percentile rank (unadjusted)	(7) Average percentile rank (age, location adjusted)	(8) Log individual net income	(9) Log individual wage, other income
Father's permanent earnings measure	0.424***	0.537***	0.427***	0.404***	0.363***	0.149***
	(0.037)	(0.059)	(0.030)	(0.034)	(0.033)	(0.047)
Age of father	-0.161***	-0.242**	-1.937	1.819	-0.044	0.007
	(0.059)	(0.111)	(1.442)	(1.676)	(0.056)	(0.099)
Age of father squared	0.002***	0.003**	0.017	-0.022	0.000	-0.000
	(0.001)	(0.001)	(0.015)	(0.017)	(0.001)	(0.001)
Age of son	0.087***	0.060	1.762**	-0.703	0.056*	0.041
	(0.033)	(0.078)	(0.722)	(0.806)	(0.032)	(0.068)
Age of son squared	-0.001***	-0.001	-0.025**	0.009	-0.001	-0.000
	(0.000)	(0.001)	(0.010)	(0.011)	(0.000)	(0.001)
Ethnicity	0.109	0.089	2.620	3.277*	0.107	-0.060
	(0.081)	(0.120)	(1.774)	(1.885)	(0.075)	(0.131)
Urban	0.362***	0.238**	4.971***	-11.869***	0.217***	0.104
	(0.056)	(0.110)	(1.535)	(1.994)	(0.050)	(0.070)
Observations	3,714	1,829	2,987	2,987	2,981	751

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix

A. Variable Definitions

Annual labor earnings

Earnings are constructed from an individual's earnings from primary and secondary jobs, collective and household farming, and household business. Wage labor income is computed using the reported average monthly wage and number of months worked last year. Values for months worked are imputed with the median value if a respondent reports working but months worked is zero or missing. Values for monthly wage are imputed using regression imputation. if the wage is missing but the individual reports working.

Earnings from collective and household activities are obtained by using an individual's number of hours worked in each activity, along with the assumption of different levels of productivity. For each individual, CHNS provides data on how many hours they worked in each collective and household activity. These activities include household business, farming, fishing, gardening, livestock production, and collective farming. Hours are imputed if an individual reports working. Hours are also aggregated and scaled down proportionately if they exceed 7000. An individual's contribution to household labor is computed by dividing their share of hours worked in household activities.

Different components of household income are then aggregated. Quantities and prices of crops are surveyed from 1989 to 1997. Gross income from crop production is computed as the quantity of crops produced multiplied by the free market price, plus the quantity crops of sold to the government multiplied by the difference between government and free market prices. Gross income from livestock production, fishing, gardening, and small business is obtained by adding the value of sold, consumed, and given away goods last year.

Household income is imputed if respondents indicate that their household engaged in an activity (production, sales, consumptions, gifting), but the values for production are missing. There are two income elements to household farming and business: sales and consumption. If either or both elements are missing, values are imputed in a multi-step regression sequence.

In the case of missing consumption, logs of consumption are regressed on the log of sales and the log of costs of production, dummies for household activity, year, province, urban status, log of the total household hours worked by all household members in a given activity, average years of schooling of household members engaged in a given activity, and the number of household members engaged in a given activity (total, over 16, and over 60). The next model takes the same form but excludes the costs of production. The final model includes sales, costs, and only the dummy variables specified above. In all three models, predicted values are obtained and imputed for either consumption or sales. The same procedure is followed for missing sales data.

If both income elements are missing, values are imputed in four steps. First, log of production is regressed on costs, and the dummy and household variables listed above. Then, log of production is regressed on dummy and household variables alone. Third, log of production is regressed on costs and dummy variables, and finally, log of production is regressed on dummy variables alone. Values are predicted and imputed at each step.

Costs of production are imputed in a two-step sequence if missing. First, log of costs are regressed on log of production, dummy, and household variables. Then, log of costs are regressed on household variables only. Again, values are predicted and imputed at each step.

Household expenses are treated as investment if expenses exceed gross income by more than 50 percent. Investment made from previous savings are not considered as a component of gross labor income

The individual contribution to household labor income is computed by assuming different productivity levels based on schooling and age. Log of wages are regressed on schooling years, a quadratic age, and year and province fixed effects. An index is generated from the product of schooling years, age, age squared, their coefficients, and the individual contribution to household labor. The adjustment weight is generated by dividing the value of the index by its sum within every activity in a given household and year. The sum of the adjustment weight represents the contribution of each member in hourly household income based on different levels of productivity. The adjustment weight is multiplied by the unadjusted individual contributions to household income.

Finally, income from collective farming is measured as the total cash and in-kind payments from collective farming outside a regular job.[6] Annual labor earnings are generated by adding the total wage earned from primary and secondary jobs, income from collective farming, and net individual income in household farming and business. Respondents report earnings obtained from the year preceding the survey year. Therefore, earnings are inflated to 2014 price levels using provincial CPI obtained from China Data Online and price levels from each wave's preceding year.

Appendix 2. Sample constraints

The raw sample contains 185,365 observations and includes all individuals surveyed from 1989 to 2015 regardless of whether they are linked to any parents. There are a number of constraints I impose to obtain the estimation samples. First, I drop the 1989 wave because of a different survey design. This leaves 169,442 observations. Since there are few working mothers in the sample, I focus on father-son and father-daughter pairs in my estimation samples. Dropping individuals not linked to fathers eliminates over half of the sample and leaves 72,106 observations. (At this point, observations are unique father-child-year combinations.) Next, I restrict ages so children are no younger than 25 years old and fathers are no older than 65 years old. The sample reduces to 25,484 observations. I also require that averages of father's earnings are computed over a minimum of nine years. This constraint leaves 14,095 observations. Note that observations from provinces added in 2011and 2015 are excluded from the sample because earnings averages for fathers in these provinces cannot be attained over a period of at least nine years.

There are 8,075 total father-son-year combinations. Dropping sons with zero or missing earnings leaves a final estimation sample size of 2,957. For fathers and daughters, there are 6,020 total combinations and 634 combinations without zero or missing daughter's earnings.

Appendix 3. Variables

Individual wage, bonus, other income (CHNS variable)

Includes annual wages and bonuses from primary and secondary jobs, as well as other cash and non-cash non-labor income. Cash income excludes cash received from other household members.

[6] In the entire survey, only 369 observations received cash or in-kind payments from collective farms. The number of observations is not sufficient to perform meaningful imputations. 48 missing values are treated as zeroes.

Non-cash income like clothing and food also excludes non-cash items given by household members. Incomes are inflated to 2014 price levels using provincial CPI. .

Total net individual income (CHNS variable)

The sum of total individual income and the individual contribution to household revenue minus expenditures. The individual contribution to household income is computed from each person's share of household hours worked. Sources of household income include business, farming, fishing, gardening, and livestock, with missing values imputed for each activity. Household income is combined with wages, bonuses, and other non-labor income ("Individual wage, bonus, other income" described above) to generate total net individual income.[7] Wages are similarly adjusted for inflation.

Total wage earnings

The sum of wage earnings from primary and secondary jobs. CHNS asks respondents for their average monthly wage for each job last year. Total wage earnings are computed using the product of average monthly wage and the number of months worked last year. Wages are similarly adjusted for inflation.

Percentile rank

The position of individuals in the earnings distribution of a given year. I compute three measures of percentile ranks using a regression-based method. Annual labor earnings are regressed on a vector X in year t. Residuals are predicted and each individual is ranked based on their corresponding percentile rank in the residual distribution. The first measure is unconditional where X includes only the intercept. The second measure is conditional on age and includes a quartic of age in X. The third is conditional on age, urban status, and current province location, which are all included in X.

Ethnicity

A dummy variable that equals one if the respondent is Han Chinese and zero if otherwise.

Urban area

A dummy variable that equals one if the respondent's household is registered in an urban area and zero otherwise.

Residence area

This variable takes on four values: one for city, two for suburb, three for town or county capital city, and four for rural village.

Job number

A dummy variable that equals one to indicate a primary job and two to indicate a secondary job.

Province

The survey covers 12 provinces (Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shaanxi, Shandong, Yunnan, and Zhejiang) and three direct-controlled

[7] A detailed description is found on the CHNS website: http://www.cpc.unc.edu/projects/china/data/datasets/Individual%20Income%20Variable%20Construction.pdf

municipalities (Beijing, Shanghai, Chongqing). From 1989 to 1997, CHNS sampled eight provinces. Heilongjiang replaced Liaoning in 1997, but Liaoning returned in the following wave 2000. CHNS surveyed nine provinces until 2011, when Beijing, Shanghai, and Chongqing were added that year. Zhejiang, Yunnan, and Shaanxi entered the survey in 2015.

Female, age, year of survey
Self-explanatory.