More Unequal yet More Alike, the Changing Patterns of Family Formation, Generational Mobility and Household Income Inequality in China: A Counter-factual Analysis

Gordon Anderson

(University of Toronto, Canada)

Tongtong Hao

(University of Toronto, Canada)

Maria Grazia Pittau

(Sapienza University of Rome, Italy)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

PS13: Mobility, Opportunity and Longitudinal Data

WEDNESDAY, AUGUST 22, LATE AFTERNOON

More Unequal yet More Alike, the Changing Patterns of Family Formation, Generational Mobility and Household Income Inequality in China: A Counter-factual Analysis.

Gordon Anderson and Tongtong Hao

University of Toronto

Maria Grazia Pittau

Department of Statistics Sapienza University of Rome

May 18, 2018

Abstract

Chinas Household income inequality has grown steadily over the last 30 years. While many analyses focus on the effects of policies relating to urban-rural and inland-coastal distinctions, growth in inequality has prevailed on both sides of those respective divides suggesting something more fundamental is at play. Here, certain patterns of family formation and human capital transfer are shown to engender increases in household income inequality measures. A unique data set, linking grandparents, parents and children, yielded evidence of structural change toward such patterns over successive cohorts of households. Influenced by such events as the Cultural Revolution, the One Child Policy and the Economic Reforms, people intensified positive assortative matching behaviors and polarizing human capital transitions. Social class designations became less important and educational class designations became more important. A counterfactual analysis verified the impact of these changes on household income inequality in urban China, revealing increasing similarity between cohorts amidst growing inequality.

Keywords: Inequality, Intergenerational Mobility, Education, Social Classes

1 Introduction

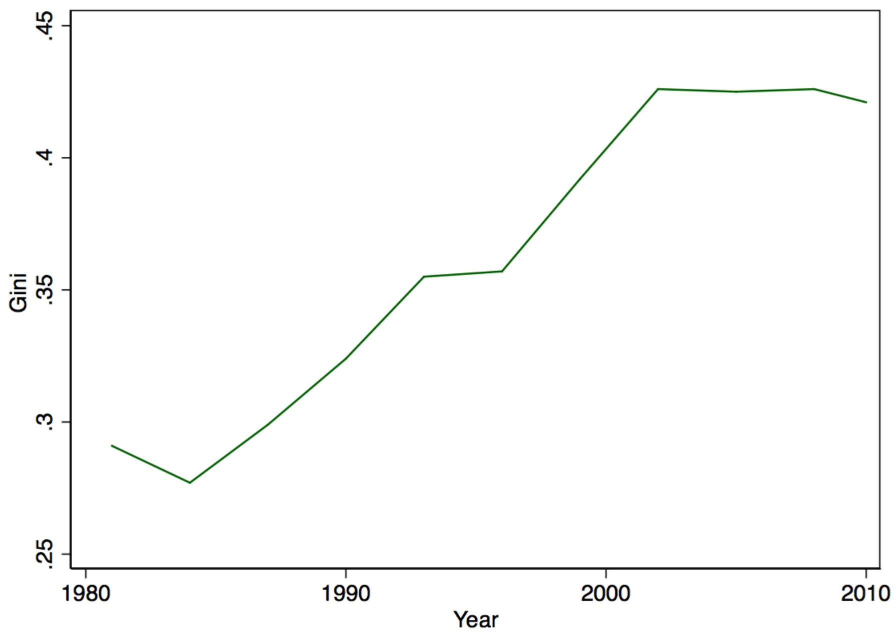
Household income distributions, and their concomitant inequality measures, can be based upon collections of household subgroups of different vintage cohorts where vintage designation is based upon the era in which the household was formed. The nature of a household, its attributes, its propensities for income generation, procreation, generational transmission of such attributes, are to some extent shaped by its vintage, a reflection of the fashions and constraints of the era in which it was formed. Had there been no vicissitude in such fashions and constraints, the income distribution could look very different from what it turned out to be. A counterfactual analysis of cohort vintage effects is employed in a subgroup decomposition of the Gini coefficient. Reflecting the fundamental distinction between polarization and inequality, that they can go in opposite directions, the results highlight paradoxically increasingly similar cohort distributions within an overall increase in inequality in Urban China, a consequence of changing patterns of family formation and generational transition.

Chinas' rapid economic growth since the Economic Reforms of the early 80's has been attended by an equally strident increase in inequality. With National Gini coefficients below 0.3 in the early 1980's rising to values above 0.5 in the first decade of the 21st century (Xie and Zhou, 2014). The rise has been persistent but uneven across many devides, Li (2012) reports rural Gini's of 0.24 and 0.37 in 1981 and 2011 respectively and urban Gini's of 0.15 to 0.34 in those same years. Rural-urban disparities, which are not a component of the "within" urban and "within" rural statistics[1], account for the National Gini being higher than its urban and rural counterparts, indeed they account for most of the overall inequality (Yang, 1999). As a consequence, researchers have looked to divide disparities in social and economic structure and policy treatment (Rozelle, 1994, Yang, 1999, Kanbur and Zhang, 1999, Gustafsson and Shi, 2002, Meng, Gregory, and Wang, 2005, Wu and Perloff, 2005, Hertel and Zhai, 2006, Ravallion and

[1]Gini is a mean normalized average of income differences within the urban society, within the rural society and between the urban and rural societies and the latter component only appears in the national measure.

Chen, 2007, Benjamin, Brandt, Giles, and Wang, 2008, Q. Deng, Gustafsson, and Li, 2013, Chen and Zhou, 2007, Cheng and Wu, 2017) as sources of increased inequality.

Diagram. 1: Gini Coefficient of China



Note: Gini coefficient reported by World Bank

However, Ravallion and Chen (2007) emphasize that, although relative inequality is higher in rural than urban areas, there has been steeper inequality gradient over time in urban areas. Moreover, after accounting for higher urban living costs, absolute inequality is higher in urban areas. In addition, due to the extensive urbanization process in China over the last 3 decades, the weight attached to the "rural" component has diminished substantially placing greater emphasis on the urban component of inequality (the 1981 urban population accounted for about 20% of 1.001 billion Chinese, by 2011 it had risen to about 51% of 1.347 billion[2]. This suggests that the respective divides have sources of increased inequality that are respectively unique. In this regard, increases in returns to education and the shifts in occupations have all been cited as sources of the rise in, and changing nature of, inequality (Meng, 2004, Wan, 2004, Zhang, Zhao, Park, and Song, 2005, Goh, Luo, and Zhu, 2009, Zhong, 2011, Meng, Shen, and Xue, 2013).

[2]National Bureau of Statistics of China 2014

While increasing inequality is everywhere and persistent, some of the roots of rising urban inequality and its changing nature are to be found in fundamental structural changes unique to that society and independent of urban-rural disparities.

Here, a rich data set of Urban Chinese households linking grandparents, parents and children, the 2002 Chinese Household Income Project (Li, Luo, Wei, and Yue, 2008) , is employed to explore structural changes in family formation and its intergenerational transmission processes as drivers of change in Urban inequality over family cohorts. Cohorts are defined by the age of the household head and their potential for being educated in historical eras before, during and after the Cultural Revolution (the latter period is associated with the economic reforms and the one child policy). The results indicate increased intensity of marriage partner matching by education and diminished or weakening matching by social class in post Cultural Revolution cohorts, also mobility diminished over the period (in other words dependency on circumstance increased). Ultimately this changed inequality relationships in and between the collection of cohorts making them more equal in the context of increasing overall inequality.

Section 2 provides an historical context for considering the effects of changing partner assortative matching and intergenerational transition patterns on household income generation and inequality. Some theoretical/algebraic relationships between these changing patterns and their effects on the Gini inequality measure together with some tools for measuring the extent of such changes and a decomposition of the Gini coefficient suitable for the purpose at hand are developed in Section 3. Section 4 examines the empirical existence of such changes and, after examining some empirical models of household size, household income and husband - wife educational relationships, section 5 explores their impact on the Gini coefficient counterfactually. Conclusions are drawn in section 6. In summary a source of increased urban inequality was found to be the increased dependency of household incomes on household human capital and diminished dependency on social class. Increased positive assortative matching in the Post Cultural Revolution Era increased disparities in household human capital which in turn increased the variation in household incomes and concomitantly the disparities in the circumstances of children whose educational outcomes were themselves highly dependent upon their parental circumstances. Somewhat paradoxically this made successive cohort income distributions more alike in the face of growing inequality.

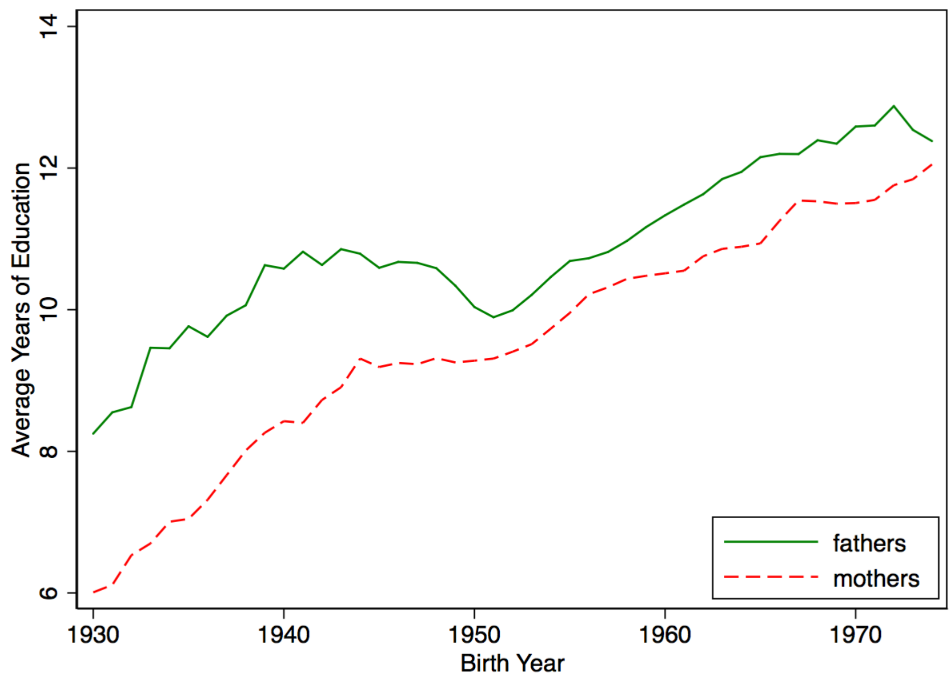
2 Background

The 1949 agrarian revolution in China saw the founding of a "new" social class system. In a society that was primarily agrarian, as much as half of the farmland was seized from landlords and redistributed to the formerly landless peasants (Walder and Hu, 2009, Clark, 2014). In this early stage of the revolution the entire population (the "grandparents" in this study) was formally classified into 12 ordered social classes according to family employment status, income sources and political loyalties at the time. The classes ranged from landless peasants through landlord classes to the aristocracy of the revolution, the revolutionary "fighters". An entire household was assigned a class label which would be inherited through the male line and remained with the offspring regardless of their political stance or behavior and became a primary criterion in their job search/promotion opportunities.

Later, the Cultural Revolution 1966-76 (the educational period of many parents in this study) saw changes in the way human capital was generationally augmented within the family. An attempt at eliminating "the distinction between town and country, industry and agriculture, physical and mental labor", saw mass school closures in urban areas (Meng and Gregory, 2002, Z. Deng and Treiman, 1997) and a purge of intellectual "elites". The policies were designed to curtail the generational transmission of social status and educational advantage by social and educational elites, in essence an equal opportunity policy that levels down. Academics were ostracized and all levels of schools were closed (However, Meng and Gregory (2002) suggest that the largest negative impact was faced by children from lower educational achievement and lower social class families). When higher education institutions reopened after 1972, children from formerly lower social designations were given preference over those from higher social designations in educational and occupational opportunities. Higher education institutions did not resume recruiting based on merit until the Cultural Revolution ended (Clark, 2014).

The loss of schooling effects of the Cultural Revolution may be seen in the average number of years of schooling and average level of schooling profiles experienced by the birth cohorts who would have been educated in the period of the Cultural Revolution. Essentially the cohort born between 1948-1955 possibly missed senior high school due to the Cultural Revolution and the cohort born between 1956-1963 who missed part of primary school and junior high school or experienced a lower quality of school in the Cultural Revolution. From Diagram 2, the effects may be seen to have predominantly impinged upon educational growth trends in males, the growth trends in education for both genders diminished but for males it became negative over the 1945-1952 period so the male-female education gap was narrowed significantly. Over the same time period variations in educational attainment levels and education years across both genders diminished greatly, a consequence of the Cultural Revolution, it represents an equalization of circumstances for future generations.

Diagram. 2: Average Years of Education by Birth Year



1980 onwards saw the profound growth spurt precipitated by the Economic Reforms, which increased investment in child education, especially children born to parents who suffered the effects of the cultural revolution (Anderson and Leo, 2009). It also saw the effects of the One Child Policy which changed the way people chose partners. With procreation, child rearing and family income production each being part of household production, under a regime which constrains one or more of them (procreation and child rearing) relative to other outputs, potential partners with specialized procreation and child rearing skills become less attractive relative to partners with specialized income generating skills. Theoretically and empirically this resulted in an increase in the extent to which people chose partners similar to themselves in income generating dimensions relative to choosing partners on the basis of other dimensions such as social class (Becker, 1981, Anderson and Leo, 2009).

3 Relationships between Income inequality, family formation and human capital transmission

To understand the impact on inequality of paradigm shifts in marital matching and intergenerational transition behaviors that took place across cohorts, algebraic connections between matching intensity, generational dependencies and the Gini coefficient are outlined together with tools for measuring such intensities and dependencies and decomposition of the Gini coefficient that will illuminate the between cohort effects.

3.1 Family formation

When partners choose each other on the basis of similarity of their respective characteristics (for example pairing on the basis of similarity of education levels or social class) it is said to be a positive assortative match (see for example Chiappori, Salanié, and Weiss, 2017, Choo and Siow, 2006). It can be shown that Intensified positive assortative marital matching on any characteristic that is aggregative for the household and positively related to income, increases the household income Gini coefficient.

Increased (rank) correlation of spousal characteristics is frequently used to identify intensified assortative matching on a discretely (continuously) measure[3]. To demonstrate that increased spousal correlation increases the Gini, a simple mean preserving correlation increasing partner swap is contrived and shown to increase the Gini coefficient of household income. Let z be the ordered vector of husbands incomes (education

[3]It is important to note that, when opposite sides of the marriage market are unbalanced, in the sense that class sizes either side of the market are unequal, Spearmans Rank Correlation will not attain the value 1 even when everyone makes their best possible match. Given the well documented significant gender imbalance in China (approximately 118 boys are born for every 100 girls in China compared to a global average of 103 to 107) this is a likely prospect. To accommodate this a normalization of the coefficient is introduced which allows the statistic to attain 1 if everyone made their best possible match.

levels) and y be the associated wives incomes (education levels) so that the vector of household incomes (education levels) $x = z + y$. Let r_z and r_y be the vectors of corresponding ranks of z and y. Letting μ_w denote $E(w)$, note that $\mu_x = \mu_z + \mu_y$. Suppose the element $x_m = \mu_m + y_m$ i.e. the husband in the m'th household has the average husbands' income and, for convenience suppose $z_{m-1} < z_m < z_{m+1}$ so that $r_{zm+1} = r_{zm-1} + 2$, and suppose further $y_{m-1} = y_{m+1} + \delta$ with $\delta > 0$ so that $r_{ym-1} = r_{ym+1} + K$ where K is an integer ≥ 1. In essence assume spousal rankings are negatively correlated around the m'th observation. When husbands and wives in the $m - 1$ and $m + 1$ observations swap spouses, there will be increased positive assortative matching in terms of increased positive association in the correlation for continuously measured characteristics and rank correlation for discretely measured characteristics of husbands and wives. Consider RN, the numerator of correlation coefficient before and RN^*, the numerator of the correlation coefficient after the swap.

$$
\begin{aligned}
RN =& \sum_{i=1}^{n} (Z_i - \bar{Z}) Y_i \text{ and} \\
RN^* =& \sum_{i=1}^{n} (Z_i - \bar{Z}) Y_i - (Z_{m+1} - \bar{Z}) Y_{m+1} - (Z_{m-1} - \bar{Z}) Y_{m-1} \\
& \qquad + (Z_{m+1} - \bar{Z}) Y_{m-1} + (Z_{m-1} - \bar{Z}) Y_{m+1} \\
=& RN + \delta\{(Z_{m+1} - \bar{Z}) - \ (Z_{m-1} - \bar{Z})\} \\
& \text{where } \delta\{(Z_{m+1} - \bar{Z}) - \ (Z_{m-1} - \bar{Z})\} > 0
\end{aligned}
$$

For discretely measured characteristic assume for simplicity there are no ties in either husbands or wives' characteristics and consider Spearman's Rank Coefficient SR before (SR) and after (SR^*) the swap.

$$\text{Since } SR = 1 - \left(\frac{6 \sum_{i=1}^{n} (r_{zi} - r_{yi})^2}{n(n^2 - 1)} \right)$$

Note that

$$SR^* - SR = \frac{6}{n(n^2-1)} \Big((r_{zm+1} - r_{ym+1})^2 + (r_{zm-1} - r_{ym-1})^2 - (r_{zm+1} - r_{ym-1})^2 - (r_{zm-1} - r_{ym+1})^2 \Big)$$

Recall $(r_{zm+1} - r_{zm-1}) = 2$ and $(r_{ym-1} - r_{ym+1}) = K \geq 1$
substitution yields:

$$SR^* - SR = \frac{6}{n(n^2-1)} 4K > 0$$

For convenience write $GINI$ (before) and $GINI^*$ (after the swap) then the effect on household inequality in terms of the GINI may be seen as follows:

$$GINI = \frac{1}{\mu n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j| = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\frac{|x_i - x_j|}{\mu} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\frac{|nx_i - \sum_{j=1}^{n} x_j|}{\mu} = \frac{1}{n}\sum_{i=1}^{n}|\frac{x_i}{\mu} - 1|$$

Note that:

$$\begin{aligned} GINI^* - GINI &= |\frac{x^*_{m-1}}{\mu} - 1| + |\frac{x^*_{m+1}}{\mu} - 1| - |\frac{x_{m-1}}{\mu} - 1| - |\frac{x_{m+1}}{\mu} - 1| \\ &= |\frac{x_{m-1} - \delta}{\mu} - 1| + |\frac{x_{m+1} + \delta}{\mu} - 1| - |\frac{x_{m-1}}{\mu} - 1| - |\frac{x_{m+1}}{\mu} - 1| \\ &= \frac{2\delta}{\mu} > 0 \end{aligned}$$

Since $\frac{x_{m+1}}{\mu} > 1$ and $\frac{x_{m-1}}{\mu} < 1$.

Essentially intensified positive assortative matching on any variate that is positively associated with income will increase household income inequality.

3.2 Human capital transmission

With regard to the passing on of human capital, generational transition matrices may be construed as blueprints of the way in which human capital qualities are passed on through generations. When a society statically replicates itself, the Generational Transition matrix is said to be stationary (examples are the identity matrix or the perfect equality of opportunity matrix) and successive generations distributions will be identical. Transition matrices that change the anatomy of the arrival (inheritors) distribution from that of the departure (parents) distribution by moving inheritors into new positions relative to their ancestor's position in the departure distribution are not static matrices. Anderson (2017) characterized such transition matrices as polarizing or converging are not static matrices, when respectively the net transfer of mass is from the center of the departure distribution to the peripheries of the arrival distribution, or from the peripheries of the departure to the center of the arrival distribution. Based upon functions of cell values and initial class sizes, Anderson (2017) provides indexes on [0,1], which measure the extent to which a given transition matrix exhibits polarizing, converging, upward or downward transitional properties, Table 1 exemplifies matrices with such typologies. When incomes have a monotonic non-decreasing dependency upon human capital qualities, polarizing transitions can be seen to make future generations' outcomes more unequal and converging transition matrices can be seen to be making future generations' outcomes more equal, static transition matrices result in no change in the attainment distribution over time.

Table 1: Transition Types

Polarizing Transition	Converging Transition	Upward Transition	Static Dependent (Imobile) Transition	Static Independent (Mobile) transition
$\begin{bmatrix} 1 & 0.3 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0.3 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 1 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 & 0 \\ 0.3 & 0.5 & 0 \\ 0.2 & 0.5 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$

Explicit analysis of the effects of such transfers on inequality is facilitated by considering a rearrangement of the GINI coefficient interpreted as the average over all agents of a "relative to the mean" distance measure of each agent from all other agents. For grouped data, where π_i is the proportion of the population receiving income X_i $i = 1, K$, note the group GINI is written as:

$$\frac{1}{\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\pi_j|X_i-X_j| = \sum_{i=1}^{n}\pi_i\sum_{j=1}^{n}\pi_j\Big(\frac{|X_i-X_j|}{\mu}\Big) = \sum_{i=1}^{n}\pi_i\frac{|\sum_{j=1}^{n}\pi_jX_i-X_m|}{\mu} = \sum_{i=1}^{n}\pi_i|\frac{X_i}{\mu}-1| \tag{1}$$

Now consider the effect on these formulations of the GINI coefficients in the context of generational transition matrices with respect to educational attainments or income which are polarizing. It can be shown that any net transfer of mass from the center to the peripheries of a distribution will increase its GINI coefficient.

In terms of the grouped Gini for convenience suppose that n is odd and that μ is the mean of the distribution where $m = (n+1)/2$, thus $X_m = \mu$. Let's consider a shift of mass such that $\pi^*_m = \pi_m - \delta_{k1} - \delta_{k2}$, $\pi^*_{m+k1} = \pi_{m+k1} + \delta_{k1}$ and $\pi^*_{m-k2} = \pi_{m-k2} + \delta_{k2}$ for all δ_{k1}, δ_{k2} positive. Letting GINI* and GINI be the respective grouped Gini coefficients after and before the transfer, then from equation (1)

$$\begin{aligned} GINI^* - GINI &= \sum_{i=m-k2}^{m+k1} (\pi^*_i - \pi_i)|\frac{X_i}{\mu} - 1| \\ &= \delta_{k2}|\frac{X_{m-k2}}{\mu} - 1| + \delta_{k1}|\frac{X_{m+k1}}{\mu} - 1| > 0 \end{aligned}$$

Effectively the resultant income or educational attainment distributions become more unequal.

3.3 Gini Decomposition

The separate contributions of the cohorts to overall inequality will be examined via a cohort decomposition of the Gini coefficient. Following Mookherjee and Shorrocks (1982) and Anderson and Thomas (2017), when a population (with overall mean income μ) is composed of subgroups indexed $k = 1, \ldots, K$, with means μ_k, Ginis G_k, and population proportions w_k, the decomposition may be written as follows:

$$GINI = \sum_{k=1}^{K} w_k^2 \frac{\mu_k}{\mu} G_k + \frac{2}{\mu} \sum_{k=2}^{K} \sum_{j=1}^{k} w_k w_j |\mu_k - \mu_j| + NSF$$

where $NSF = \frac{2}{\mu} \sum_{k=2}^{K} \sum_{j=1}^{k-1} w_k w_j \int_0^\infty f_k(y) \int_0^\infty f_j(x)(x-y) dx dy$. NSF may be construed as a "Non-Segmentation Factor" measuring the extent to which distributions overlap or have elements in common. Coupled with the middle component, which measures the distance between subgroups they relate to the identification and alienation components in the Duclos, Esteban, and Ray, 2004 Polarization index, highlighting the distinction between polarization and inequality which sees the possibility of a society becoming more polarized yet more equal at the same time.

When there is no overlap between any subgroups so that $f_k(x) = 0$ for all $f_j(x) > 0$ and $f_j(x) = 0$ for all $f_k(x) > 0$ for all possible pairs j and k, $NSF = 0$. This is the perfect segmentation case of Mookherjee and Shorrocks (1982) wherein Gini is just the sum of the within and between subgroup inequality components i.e. $GINI =$

$\sum_{k=1}^{K} w_k^2 \frac{\mu_k}{\mu} G_k + \frac{2}{\mu} \sum_{k=2}^{K} \sum_{j=1}^{k} w_k w_j |\mu_k - \mu_j|$ making the Gini subgroup decomposable. Noting that all three components of $GINI$ are non-negative and that $0 \leq NSF \leq GINI$. $SI = 1 - NSF/GINI$ constitutes a Segmentation Index reflecting the lack of commonality amongst the subgroups.

4 Empirical Analysis

To evaluate the effects of intensified marital matching behaviour and increasingly polarizing human capital transitional patterns on inequality, the extent to which these trends have occurred is first examined. To see if positive assortative marriage matching patterns have intensified over the period, the rank correlations of partners educational and social class status of 3 marriage cohorts, those that took place, before, during and after the Cultural Revolution, are considered. For an analysis of changing generational transitional patterns household income to child educational achievements and household social class to child educational achievement transitions for the three cohorts are considered following Anderson, 2017. This is followed by a counterfactual study of the household income generation process.

A rich data set on Urban households, drawn from the 2002 Chinese Household Income Project (Lin, Wang, and Zhao, 2004), provides information on grandparent's social class designation given in the late 1940s, parent's educational status and child's (grandchildren's) educational status facilitating measurement of the transition from Grandparents Social class to parent's educational status and ultimately a child's educational status. Grandparent social classification (*Chengfen*) was C1: Poor Peasant or Landless (53.96%), C2: Lower Middle Peasant (14.14%), C3: Upper Middle Peasant (4.81%), C4 : Rich Peasant (2.01%), C5: Landlord (2.82%), C6: Manual Worker (8.21%), C7: Office Worker (3.30%), C8: Enterprise Owner (0.43%), C9 : Petty Proprietor (3.75%), C10: Revolutionary Cadre (1.38%), C11: Revolutionary Army Man (1.03%), C12: Other (4.16%). To simplify analysis, and because some cells were very small this categorization was condensed to 5 social classes. SC1 = {C1}, SC2 = {C2, C6}, SC3 = {C3, C9, C12}, SC4 = {C4, C7, C11}, SC5 = {C5, C8, C10}. The first group SC1 is poor peasant or landless persons, which accounts for roughly half of the population. SC2 is comprised of lower middle peasant and manual workers because they each have low social status. SC3 is made up of self-sufficient upper middle peasants and petty proprietors, also included in this group is the unidentified "other" because their education label is similar to the other 2 member classes. SC4 is comprised of rich peasant, office worker and revolutionary army man who have relatively more resources and typically has less manual labor obligations. SC5 is made up of Landlords, Enterprise owners and Revolutionary Cadres.

Based upon the highest category an individual attained the educational categories were 1 if never schooled, 2 if classes for eliminating illiteracy, 3 elementary school, 4 if junior middle school, 5 if senior middle school (including professional middle school), 6 if technical secondary school, 7 if junior college, 8 if college/university, 9 if graduate. Educational categories 1 through 9 were condensed to EDC1 = {1,2,3}, EDC2 = {4}, EDC3 = {5}, EDC4 = {6}, EDC5 = {7}, EDC6 = {8}, EDC7 = {9}. Information was available on 6610 parent - grandparent pairings and 1514 parent–child pairings (only children over 22 years old were used under the assumption they would have completed their education). Family cohort membership is determined by the age of the household head (father) at the time of the survey. Those whose household heads are born before 1948 are deemed to be the Pre Cultural Revolution Cohort of households (the education of these heads would not have been influenced by the vagaries of the Cultural Revolution). Those households whose heads are born between 1948 and 1963 are deemed the Cultural Revolution Cohort households and those born after 1963 are deemed the Post Cultural Revolution cohort, these household heads would have completed their education after the Cultural Revolution and made their marriage choices after the implementation of the one child policy.

4.1 Changing Marriage Matching Patterns

Differences in matching patterns in terms of social class and education class over the 3 Eras are compared by employing Spearman's Rank Correlation Coefficient (Spearman, 1904) of husbands and wives' education or social classes as a positive assortative matching index. In a balanced marriage market with effective market clearing under positive assortative matching, the rank correlation coefficient will be 1. However, there may be cause for concern with the use of the statistic as a matching index since it could understate the extent of positive assortative matching. If the marriage market was

Table 2: Positive Assortative Matching Indices

	Spearman Rank Correlation Coefficients			
	Education	Social Class	Scaled Education	Scaled Social Class
All Cohorts	0.5514	0.2800	0.5855	0.2823
(Variance)	(6.4629e-005)	(6.4629e-005)	(7.2874e-005)	(6.5697e-005)
[Maximal Value]	[0.9417]	[0.9918]		
Pre Cultural Revolution Cohort	0.5042	0.2854	0.5398	0.2880
(Variance)	(0.000267)	(0.000267)	(0.000306)	(0.000272)
[Maximal Value]	[0.9342]	[0.9910]		
Cultural Revolution Cohort	0.5058	0.2818	0.5341	0.2848
(Variance)	(0.000114)	(0.000114)	(0.000127)	(0.000116)
[Maximal Value]	[0.9470]	[0.9895]		
Post Cultural Revolution Cohort	0.6200	0.2350	0.6536	0.2377
(Variance)	(0.000338)	(0.000338)	(0.000376)	(0.000346)
[Maximal Value]	[0.9486]	[0.9886]		

Table 3: Spearman Rank Correlation difference analysis

Comparison	Matching Index	Spearman Difference	Standard Deviation	t-stat t-stat
PanelA: Spearman Rank Correlation difference analysis				
Post Cultural Revolution Cohort *vs*	Education	0.1158	0.0235	4.9277
Pre Cultural Revolution Cohort	Social Class	-0.0504	0.0237	5.8550
Post Cultural Revolution Cohort *vs*	Education	0.1142	0.0206	5.5437
Cultural Revolution Cohort	Social Class	-0.0468	0.0207	-2.2609
Cultural Revolution Cohort *vs*	Education	0.0016	0.0865	4.8962
Pre Cultural Revolution Cohort	Social Class	-0.0036	0.0187	-0.1925
PanelB: Scaled Spearman Rank Correlation difference analysis				
Post Cultural Revolution Cohort *vs*	Education	0.1138	0.0235	4.8426
Pre Cultural Revolution Cohort	Social Class	-0.0503	0.0237	-2.1224
Post Cultural Revolution Cohort *vs*	Education	0.1195	0.0206	5.8010
Cultural Revolution Cohort	Social Class	-0.0471	0.0207	-2.2754
Cultural Revolution Cohort *vs*	Education	-0.0057	0.0185	-0.3081
Pre Cultural Revolution Cohort	Social Class	-0.0032	0.0187	-0.1711

*The standard error for Spearman's Rank Correlation is $0.6325/\sqrt{n-1}$ and for the differences it is $\sqrt{0.4001 * (\frac{1}{n_1-1} + \frac{1}{n_2-1})}$ where nk is sample size for the k'th cohort. For the Scaled coefficient the standard error is scaled by the corresponding scaling factor.

unbalanced (i.e. there are insufficient numbers of a particular type on one side of the market to match with those on the other side of the market), then the correlation coefficient would record less than perfect matching even though the market cleared perfectly according to the positive assortative matching rule (Becker, 1981). This may be circumvented by rescaling the coefficient by its maximum value based on the assumption that everyone makes their best feasible match[4]. It would then record a value of 1 if market clearing was effective. Table 2 reports the corresponding matching indices.

Husband and wife scaled educational and social status correlations did not change significantly between the Pre Cultural Revolution and Cultural Revolution eras. The significance of the unscaled Spearman statistic and non-significance of the scaled Spearman statistic suggests that the Pre Cultural Revolution-Cultural Revolution change in educational matching had more to do with the increased capacity for matching as evident in diagram 2. However, both scaled and non-scaled the educational class correlations increased substantially in the post Cultural Revolution period whereas the corresponding social class correlations diminished significantly suggesting education matching and social class matching behaviors reflect different objectives or responses in the Post Cultural Revolution era. This is consistent with the theoretical reasoning in Anderson and Leo, 2013 which predicts intensified positive assortative matching on education relative to social status when household production of children is rationed, as was the case in the Post Cultural Revolution era.

4.2 Changing Generational Transition Patterns

Changes in transitional structures affect the income distribution both indirectly and directly. While social class may affect incomes both directly and through its effect on educational classification, educational classification cannot affect exogenously determined social class but it can influence Income status. Study of social and educational class transitions to income classes is facilitated by a semi-parametric decomposition of the household income distribution which produces individual household income class membership probabilities (details in the appendix) in a 5 income class model. Details of various transition matrix typologies, their mobility and polarizing properties and

[4]The maximum value can be obtained by separately sorting husbands and wives matching index, pair husbands and wives according to rank and calculate the Spearman rank correlation index for such a pairing.

associated indices are outlined and discussed in Table 1 and Anderson (2017) and reported in the following. The indices all lay on the unit interval and are asymptotically normally distributed. On the null hypothesis that mobility does not favor a direction or polarizing/converging trend, it can be shown that both Upward Mobility and Polarizing/Converging indexes are N(0.5,(0.25/n)) where n is the sample size yielding standard errors of 0.01223, 0.00824 and 0.01410 for Pre CR, CR and Post CR cohorts respectively for these indices.

Table 4: Parent Social Class-Educational Transitions

Pre CR Fathers(N=1672)	SocClass1	SocClass2	SocClass3	SocClass4	SocClass5
Education Class 1	0.2040	0.1311	0.1056	0.0891	0.0899
Education Class 2	0.3511	0.2674	0.3310	0.2574	0.2472
Education Class 3	0.1508	0.1568	0.1690	0.1485	0.2135
Education Class 4	0.1248	0.1799	0.1549	0.1386	0.1685
Education Class 5	0.1088	0.1671	0.1127	0.2079	0.1011
Education Class 6	0.0581	0.0951	0.1232	0.1485	0.1798
Education Class 7	0.0025	0.0026	0.0035	0.0099	0.0000
CR Fathers(N=3680)					
Education Class 1	0.0439	0.0481	0.0557	0.0119	0.0229
Education Class 2	0.3366	0.2861	0.3559	0.2460	0.2114
Education Class 3	0.2937	0.3058	0.1864	0.2579	0.2743
Education Class 4	0.0828	0.0937	0.1162	0.0913	0.1200
Education Class 5	0.1740	0.1936	0.1816	0.2579	0.2457
Education Class 6	0.0611	0.0641	0.1017	0.1190	0.0857
Education Class 7	0.0079	0.0086	0.0024	0.0159	0.0400
Post CR Fathers(N=1258)					
Education Class 1	0.0099	0.0081	0.0238	0.0000	0.0192
Education Class 2	0.2133	0.1707	0.1190	0.2000	0.0385
Education Class 3	0.2219	0.2358	0.2024	0.2308	0.2115
Education Class 4	0.1110	0.0894	0.1667	0.0923	0.1154
Education Class 5	0.2552	0.2724	0.3333	0.2769	0.3077
Education Class 6	0.1800	0.2154	0.1310	0.1692	0.2885
Education Class 7	0.0086	0.0081	0.0238	0.0308	0.0192

Observe from Table 4 and 3a, that social class to education class mobility was at its highest for the cultural revolution cohort, a direct effect of the Cultural Revolution,

Table 3a: Social Class-Education Transition Indices

Cohort	Mobility	Upward	Polarize
Pre CR	0.7853967	0.4892344	0.3560291
CR	0.9073940	0.5875000	0.3319542
Post CR	0.6947007	0.7428458	0.3893200

mobility was significantly progressively upward over the 3 cohorts but the transitions were never polarizing indeed they were significantly convergent or equalizing. Turning to the social class–income class transitions, Table 4 and 4a indicate that mobility was invariably quite high implying that income distributions of the various social classes was very similar, put another way social class had little impact on the shape of the income distribution over all cohorts. Transitions were invariably upward and progressively so over the cohorts, though they were never polarizing, and none of the differences were profoundly significant.

A very different story emerges for education class to income class transitions reported in Table 5 and 5a. Transition matrices characterize a very immobile society (and increasingly so over the cohorts) suggesting that a household place in the income distribution is very much governed by its educational status and increasingly so. Transitions are typically upward but to a diminishing extent. Most significantly for present purposes transitions are always polarizing and increasingly so over more recent cohorts. In effect social class appears to have a weaker direct effect on household incomes than does educational classifications. However educational outcomes are dependent on social class and changes in the way social class translates to educational class influences the income distribution indirectly.

Lefranc, Pistolesi, and Trannoy (2008; 2009) propose evaluating the presence of equality of opportunity by evaluating the extent of second order dominance relationships between the various conditional outcome distributions with absence of dominance supporting the equality of opportunity hypothesis. Strictly speaking this is not possible here because only outcome classes are being considered and only first order dominance comparisons can be made. However, some insight on the differences across regimes can be gleaned from examining the first order comparisons and noting that dominance at the first order implies dominance at the second order. Turning to the cumulative household distributions conditioned on social class and education class of the household in Tables

Table 4: Social Class-Income Transition of father

	SocClass1	SocClass2	SocClass3	SocClass4	SocClass5
Overall					
IncomeClass1	0.00742	0.00544	0.00693	0.00324	0.00670
IncomeClass2	0.0160	0.0147	0.0110	0.0146	0.00876
IncomeClass3	0.256	0.231	0.236	0.199	0.198
IncomeClass4	0.354	0.344	0.349	0.340	0.335
IncomeClass5	0.367	0.404	0.398	0.444	0.451
PreCR					
IncomeClass1	0.0120	0.00350	0.00815	0.00312	0.0217
IncomeClass2	0.0246	0.0116	0.0171	0.0296	0.00442
IncomeClass3	0.268	0.223	0.242	0.231	0.201
IncomeClass4	0.349	0.352	0.341	0.335	0.349
IncomeClass5	0.346	0.410	0.392	0.402	0.424
CR					
IncomeClass1	0.00722	0.00505	0.00574	0.00343	0.00127
IncomeClass2	0.0143	0.0175	0.00872	0.00628	0.0126
IncomeClass3	0.252	0.230	0.233	0.178	0.207
IncomeClass4	0.355	0.340	0.351	0.337	0.330
IncomeClass5	0.372	0.407	0.402	0.475	0.449
PostCR					
IncomeClass1	0.00359	0.00959	0.00902	0.00274	0.000330
IncomeClass2	0.0120	0.00985	0.00273	0.0230	0.00316
IncomeClass3	0.253	0.248	0.227	0.229	0.163
IncomeClass4	0.358	0.347	0.366	0.356	0.331
IncomeClass5	0.373	0.386	0.396	0.390	0.503

Table 4a: Social Class-Income Transition indices

Cohort	Mobility	Upward	Polarize
Pre CR	0.8793059	0.6654964	0.3145048
CR	0.8280157	0.6948723	0.3072642
Post CR	0.8475474	0.7265366	0.2834459

Table 5: Educational-Income Transition of Fathers

	EduClass1	EduClass2	EduClass3	EduClass4	EduClass5	EduClass6	EduClass7
Overall							
IncomeClass1	0.0276	0.0111	0.00581	0.00145	0.000802	6.68e-05	1.06e-05
IncomeClass2	0.0588	0.0223	0.0142	0.0100	0.00181	9.86e-06	0
IncomeClass3	0.368	0.310	0.253	0.220	0.170	0.121	0.0663
IncomeClass4	0.331	0.362	0.361	0.358	0.343	0.310	0.256
IncomeClass5	0.214	0.295	0.366	0.410	0.484	0.569	0.678
PreCR							
IncomeClass1	0.0359	0.00799	0.00668	0.00111	0.000268	7.89e-05	2.08e-06
IncomeClass2	0.0683	0.0222	0.00772	0.00790	0.00237	7.69e-06	0
IncomeClass3	0.366	0.286	0.231	0.217	0.159	0.126	0.0381
IncomeClass4	0.323	0.367	0.361	0.361	0.329	0.308	0.222
IncomeClass5	0.207	0.317	0.394	0.413	0.509	0.567	0.739
CR							
IncomeClass1	0.0158	0.0119	0.00476	0.00122	0.00122	6.45e-05	9.94e-06
IncomeClass2	0.0466	0.0203	0.0153	0.0113	0.00200	1.72e-05	5.22e-11
IncomeClass3	0.362	0.304	0.252	0.207	0.160	0.111	0.0553
IncomeClass4	0.342	0.362	0.358	0.351	0.337	0.298	0.231
IncomeClass5	0.233	0.302	0.370	0.429	0.499	0.592	0.714
PostCR							
IncomeClass1	0.00225	0.0139	0.00887	0.00263	0.000278	6.18e-05	1.54e-05
IncomeClass2	0.0155	0.0323	0.0165	0.0104	0.00109	3.10e-06	0
IncomeClass3	0.485	0.386	0.281	0.260	0.198	0.129	0.104
IncomeClass4	0.377	0.355	0.368	0.370	0.363	0.325	0.330
IncomeClass5	0.120	0.213	0.326	0.357	0.438	0.546	0.566

Table 5a: Educational-Income Transition Indices

Cohort	Mobility	Upward	Polarize
Pre CR	0.4276802	0.7327413	0.6351860
CR	0.4189770	0.6771935	0.6824907
Post CR	0.3871865	0.5284033	0.7040519

6 and 7 respectively, note that income distributions for higher social classes do not always dominate those of lower social classes both overall and across the three cohorts. Indeed, the high value of the overlap measure of the extreme distribution comparison indicates small differences between the income distributions of various social classes. On the other hand, income distributions for higher education classes always dominate lower education classes for all conditional distributions in all cohorts (except for the lowest educational class in the Post Cultural Revolution cohort), that is to say there is a strict ordering of income class outcomes by educational class. Furthermore, the overlap between the extreme income distributions conditional on educational classes is much lower indicating greater variation in the conditional income distributions by educational class. This reflects the lack of mobility indicated in Table 4 which is characteristic of a society where educational rather than social status governs income status.

5 Household Income Generation and Inequality

To examine the effect of these phenomena on inequality, a counterfactual analysis of the income generating process is performed. A sense of the influence on household income production of the nature of the family is provided by simple regression equations for household size, parental educational differences and Adult Equivalized Household Income[5] on a variety of factors, the equations are respectively reported in Tables 8, 9 and 10.

The size of a household turned out to be a concave function of vintage (age of household head) and negative in the relevant range, it switched to a convex function for vintages in the range affected by the Cultural Revolution, so generally older households were larger. Higher social class families were significantly smaller with an implied elasticity of -0.01. The overall effect of education is to engender slightly smaller families though the larger the father-mother educational gap the larger the family size, an effect which outweighs the positive effect in the income equation so the net effect is negative, consistent with the idea of parental complementarity in family production which would predict positive assortative matching in income. Although the household income equation suggests some substitutability in household income production, positive as-

[5] Adult Equivalization uses the square root rule (Brady and Barber 1948) essentially it is household income divided by the square root of the number of people in the household.

Table 6: Income Class Cumulative densities conditional on social class

	SocClass1	SocClass2	SocClass3	SocClass4	SocClass5
Overall Social Class CDF, overlap of extreme pdf's=0.91558					
IncomeClass1	0.00742	0.00544	0.00693	0.00324	0.00670
IncomeClass2	0.02339	0.02012	0.01792	0.01785	0.01546
IncomeClass3	0.27918	0.25156	0.25346	0.21674	0.21362
IncomeClass4	0.63340	0.59580	0.60248	0.55624	0.54897
IncomeClass5	1.00000	1.00000	1.00000	1.00000	1.00000
Pre CR Social Class CDF, overlap of extreme pdf's= 0.91292					
IncomeClass1	0.01196	0.00350	0.00815	0.00312	0.02170
IncomeClass2	0.03654	0.01510	0.02523	0.03268	0.02612
IncomeClass3	0.30473	0.23841	0.26691	0.26354	0.22754
IncomeClass4	0.65379	0.59005	0.60832	0.59835	0.57645
IncomeClass5	1.00000	1.00000	1.00000	1.00000	1.00000
CR Social Class CDF, overlap of extreme pdf's=0.92304					
IncomeClass1	0.00722	0.00505	0.00574	0.00343	0.00127
IncomeClass2	0.02150	0.02257	0.01446	0.00971	0.01386
IncomeClass3	0.27360	0.25271	0.24772	0.18765	0.22107
IncomeClass4	0.62811	0.59284	0.59839	0.52483	0.55115
IncomeClass5	1.00000	1.00000	1.00000	1.00000	1.00000
Post CR Social Class CDF, overlap of extreme pdf's=0.87024					
IncomeClass1	0.00359	0.00959	0.00902	0.00274	0.00033
IncomeClass2	0.01561	0.01944	0.01175	0.02576	0.00349
IncomeClass3	0.26880	0.26725	0.23881	0.25457	0.16665
IncomeClass4	0.62720	0.61414	0.60447	0.61010	0.49743
IncomeClass5	1.00000	1.00000	1.00000	1.00000	1.00000

Table 7: Income Class Cumulative densities conditional on education class

	EduClass1	EduClass2	EduClass3	EduClass4	EduClass5	EduClass6	EduClass7
Overall Education Class CDF, overlap of extreme pdf's=0.53582							
IncomeClass1	0.02761	0.01112	0.00581	0.00145	0.00080	0.00007	0.00001
IncomeClass2	0.08638	0.03340	0.02004	0.01149	0.00262	0.00008	0.00001
IncomeClass3	0.45476	0.34297	0.27345	0.23179	0.17293	0.12105	0.06631
IncomeClass4	0.78617	0.70510	0.63396	0.58970	0.51581	0.43084	0.32199
IncomeClass5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
Pre CR Education Class CDF, overlap of extreme pdf's=0.46773							
IncomeClass1	0.03594	0.00799	0.00668	0.00111	0.00027	0.00008	0.00000
IncomeClass2	0.10423	0.03022	0.01440	0.00901	0.00263	0.00009	0.00000
IncomeClass3	0.47020	0.31607	0.24551	0.22636	0.16152	0.12582	0.03815
IncomeClass4	0.79279	0.68277	0.60617	0.58718	0.49091	0.43336	0.26052
IncomeClass5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
CR Education Class CDF, overlap of extreme pdf's=0.51889							
IncomeClass1	0.01585	0.01194	0.00476	0.00122	0.00122	0.00006	0.00001
IncomeClass2	0.06242	0.03220	0.02005	0.01256	0.00321	0.00008	0.00001
IncomeClass3	0.42486	0.33665	0.27177	0.21929	0.16361	0.11070	0.05532
IncomeClass4	0.76706	0.69823	0.63024	0.57064	0.50095	0.40841	0.28595
IncomeClass5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
Post CR Education Class CDF, overlap of extreme pdf's=0.55406							
IncomeClass1	0.00225	0.01390	0.00887	0.00263	0.00028	0.00006	0.00002
IncomeClass2	0.01773	0.04618	0.02532	0.01299	0.00137	0.00006	0.00002
IncomeClass3	0.50266	0.43227	0.30591	0.27287	0.19946	0.12954	0.10387
IncomeClass4	0.87995	0.78718	0.67380	0.64254	0.56234	0.45409	0.43402
IncomeClass5	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Table 8: Household size equation reparametrized (dependent variable: $\ln(\sqrt{\text{household size}})$)

VARIABLES	Coefficient	t-statistics	Coefficient	t-statistics
vintage	0.0221**	(2.158)	0.0224**	(2.193)
vintage2	-0.000296***	(-2.904)	-0.000299***	(-2.935)
Father edu	-0.0412**	(-2.035)	-0.0443**	(-2.196)
Father edu*Mother edu	-0.00180	(-0.844)	-0.00180	(-0.845)
edu difference	0.0415***	(3.746)	0.0435***	(3.928)
Social Class	-0.0391***	(-3.086)	-0.0322**	(-2.536)
CR	3.846***	(2.649)	3.871***	(2.673)
Father edu*CR	0.0152	(0.606)	0.0148	(0.593)
Father edu*Mother edu*CR	-0.000747	(-0.271)	-0.000632	(-0.230)
Social Class*CR	0.0342**	(2.034)	0.0289*	(1.724)
vintage*CR	-0.165***	(-2.625)	-0.165***	(-2.635)
vintage2*CR	0.00166**	(2.449)	0.00165**	(2.448)
Constant	3.030***	(12.42)	3.043***	(12.37)
Prov FE			Yes	
Observations	6,599		6,599	
R-squared	0.022		0.032	

*** p<0.01, ** p<0.05, * p<0.1

Note: CR is Cultural Revolution Dummy, postCR is post Cultural Revolution Dummy sortative matching appears to prevail and increases in extent for younger cohorts. A simple regression reported in Table 9 reflects the extent to which positive assortative matching intensified over the period in question with older vintage families (vintage is age of household head) exhibiting larger educational differences on average.

Table 9: Absolute Education Class Difference

VARIABLES	Coefficient	t-statistics	Coefficient	t-statistics
vintage	0.229**	(2.408)	0.0250***	(2.591)
vintage2	-0.000163*	(-1.711)	-0.000181*	(-1.871)
Constant	0.215	(0.936)	0.154	(0.650)
Prov FE			Yes	
Observations	6,684		6,684	
R-squared	0.005		0.006	

*** p<0.01, ** p<0.05, * p<0.1

The household income regression reported in Table 10 reveals strong cohort effects (F test for no cohort effects 2.9157, $P(f > 2.9157) = 0.00013$) and a strong dependence on the educational status of both parents throughout the eras. In the Pre-Cultural Revolution cohort mother's educational status has a bigger impact than fathers' educational status on household income. The difference disappears in the Cultural Revolution era and is re-established in the post Cultural Revolution era. There does appear to be some substitutability of parental education in income production with respect to education with a significantly negative cross partial derivative which, following Becker, 1981 suggests that the propensity for positive assortative matching is not as strong as would otherwise be the case (but recall income production is not the only household objective). Absolute differences in mother father education levels, reflecting the positive assortative matching effect, appears to have little impact on income generation in this era.

Household income is a weakly increasing concave function of household vintage (head of household's age) a life cycle income pattern which is positive for all households whose head is < 75. Equivalized Household income is decreasing in household size, (not surprising given adult equivalization) however in the Cultural Revolution and Post Cultural Revolution eras the value of the parameter diminishes somewhat to the point where its effect is eliminated for the youngest households. Having a head who was

Table 10: The Structure of Household Income Generation

VARIABLES	Coefficient	t-statistics	Coefficient	t-statistics
vintage	0.0265***	(3.376)	0.0122*	(1.687)
vintage2	-0.000170**	(-2.296)	-4.35e-05	(-0.635)
Mother edu	0.194***	(7.353)	0.183***	(6.924)
Father edu	0.154***	(5.441)	0.149***	(5.502)
family size	-0.181***	(-14.38)	-0.189***	(-16.30)
Father edu*Mother edu	-0.0131***	(-2.646)	-0.0124**	(-2.552)
Father-Mother edu difference	0.0341**	(2.082)	0.0339**	(2.239)
Social Class	-0.00901	(-0.784)	-0.00997	(-0.937)
CR	0.106	(0.500)	0.246	(1.228)
postCR	-0.555***	(-3.150)	-0.149	(-0.429)
Mother edu*CR	-0.0664	(-1.636)	-0.0949**	(-2.436)
Father edu*CR	-0.0401	(-0.956)	-0.0663*	(-1.682)
family size*CR	0.0505***	(2.669)	0.0455***	(2.604)
Father edu*Mother edu*CR	0.0111	(1.519)	0.0162**	(2.334)
Father-Mother edu difference*CR	-0.00526	(-0.264)	0.00420	(0.228)
Social Class*CR	0.0391***	(2.833)	0.0332***	(2.611)
Mother edu*postCR	0.00504	(0.553)	-0.0636	(-1.017)
Father edu*postCR	0.000408	(0.0109)	-0.0495	(-0.789)
family size*postCR	0.156***	(5.414)	0.133***	(5.034)
Father edu*Mother edu*postCR	0.00184	(0.423)	0.0121	(1.159)
Father-Mother edu difference*postCR	-0.0283	(-1.133)	-0.0156	(-0.622)
Social Class*postCR	0.0365**	(2.027)	0.0363**	(-2.186)
Constant	7.451***	(31.30)	7.880***	(34.97)
Prov FE			Yes	
Observations	6,137		6,137	
R-squared	0.265		0.378	

*** p<0.01, ** p<0.05, * p<0.1

CR is Cultural Revolution Dummy, postCR is post Cultural Revolution Dummy potentially affected by the cultural revolutions educational exigencies and the social class of the family does not appear to significantly affect household income except through the fathers' social class. The interaction of class and the Cultural Revolution dummy is significantly positive indicating that the higher social class of a family head (who potentially missed years of education), the higher would household income be. In a similar fashion the post Cultural Revolution dummy and social class interaction appears to enhance the income generation prospects of a household.

Since it is evident that marriage matching patterns and generational transition patterns differed significantly across the cohorts, a counterfactual study of cohort Gini coefficients was performed. Matching and generational transition models were estimated for the Pre Cultural Revolution cohorts and matches and educational endowments projected for households in Cultural Revolution and Post Cultural Revolution cohorts as though they were made in the Pre Cultural Revolution fashion. The consequent "counterfactual" Gini Coefficients were then computed and compared with the true Gini coefficients for those eras.

5.1 Marital Matching Counterfactual Analysis

The Pre Cultural Revolution matching model was based upon writing a wives' educational status as a quadratic function of wives age, social class and husbands' educational status and reported in Table 11. Then wives' educational status and income was projected for Cultural Revolution and Post Cultural Revolution cohorts under the assumption that matching patterns were the same in those cohorts as in the Pre Cultural Revolution Cohort. Household incomes were reconstituted using projected wives' incomes and Counterfactual Gini coefficients recalculated for the Cultural Revolution and the Post Cultural Revolution Cohorts and compared with the original "True" cohort Gini coefficients. As may be observed the counterfactual analysis generates a significant reduction in the Gini coefficients as predicted.

5.2 Intergenerational Transition Analysis

The impact of changes in the structure of intergenerational transition across cohorts was studied in a similar fashion where the grandparent – parent transmission mechanism for the Pre Cultural Revolution was assumed to prevail in the Cultural Revolution

Table 11: Wife's education year – Pre Cultural Revolution cohort

VARIABLES	Coefficient	t-statistics
Wife's age	0.7622***	(4.17)
Wife's age^2	-0.0071***	(-4.63)
Husband Education Years	0.3117***	(3.58)
Husband Education Years^2	0.0093***	(2.2)
Wives Social Class	0.7577***	(2.19)
Wives Social Class ^{2}	-0.0848	(-1.48)
Husband Education Years & Wives Social Class	0.0084	(0.43)
Constant	-16.6073***	(-3.04)
Province FE	Yes	
Observations	1519	
R-squared	0.37	

*** p<0.01, ** p<0.05, * p<0.1

Table 11a: Matching Counterfactual Gini

cohort	Actual Gini	Actual Gini S.D.	Counterfactual Gini	z-stat
CR(n=3344)	0.3140	(0.0094)	0.2569	-350.234
postCR(n=1131)	0.3037	(0.0162)	0.2690	-71.583

and Post Cultural Revolution eras. A grandparent – parent generational regression for both genders that was quadratic in both grandparents' educational class and family social class with cohort and provincial fixed effects was estimated and reported in Table 12 with the following results. Again the counterfactual Gini analysis indicates a significant reduction in inequality had transition patterns remained the same in the Cultural Revolution and Post Cultural Revolution cohorts as they were in the Pre Cultural Revolution cohorts. The estimated counterfactual Gini's are reported in Table 12a.

Table 12: Intergenerational Tramsition across cohorts

Dependent Variable: Father's Education Years	Coefficient.	t-stat	Dependent Variable: Mother's Education Years	Coefficient.	t-stat
Father's age	-0.0772*	(-1.7)	Mother's age	0.1434***	(3.6)
Father's age^2	-0.0002	(-0.36)	Mother's age^2	-0.0022***	(-5.45)
Paternal GF edu Years	0.1701***	(5.34)	Maternal GF edu Years	0.2202***	(6.93)
Paternal GF edu Years^2	-0.0003	(-0.16)	Maternal GF edu Years^2	-0.0019	(-0.93)
Paternal GM edu Years	0.0486	(1.2)	Maternal GM edu Years	0.1398***	(3.63)
Paternal GM eduYears^2	0.0047*	(1.66)	Maternal GM eduYears^2	-0.0007	(-0.26)
Social Class	0.6044***	(3.64)	Social Class	0.5255***	(3.28)
Social Class^2	-0.0666**	(-2.31)	Social Class^2	-0.0387	(-1.39)
CR Dummy (CRD)	-0.5828**	(-2.5)	CR Dummy (CRD)	1.0554	(4.96)
Paternal GF edu Years*CRD	-0.0568**	(-2.09)	Maternal GF edu Years*CRD	-0.1088***	(-3.92)
Paternal GM edu Years*CRD	-0.0341	(-0.94)	Maternal GM edu Years*CRD	0.0022	(0.06)
Social Class*CRD	-0.1476*	(-1.71)	Social Class*CRD	-0.2300	(-2.71)
Post CR Dummy (PCRD)	-0.1993	(-0.56)	Post CR Dummy (PCRD)	1.5800***	(4.86)
Paternal GF edu Years*PCRD	-0.0599*	(-1.72)	Maternal GF edu Years*PCRD	-0.0847**	(-2.36)
Paternal GM edu Years*PCRD	-0.0777*	(-1.88)	Maternal GM edu Years*PCRD	-0.0026	(-0.06)
Social Class*PCRD	-0.0679	(-0.62)	Social Class*PCRD	-0.1124	(-1.06)
constant	14.1580***	(11.32)	constant	5.5103***	(5.3)
Province FE	Yes			Yes	
Observations	6231			6237	
R-squared	0.14			0.23	

*** p<0.01, ** p<0.05, * p<0.1
Note: GF stands for Grandfather, GM stands for Grandmother
Note: CRD stands for CR cohort dummy, PCRD stands for post-CR cohort dummy

5.3 Decomposition Analysis

To see the overall effect on inequality of intensified positive assortative marriage matching and polarizing generational transmissions, the counterfactual income distribution

Table 12a: Transmission Counterfactual Gini

cohort	Actual Gini	Actual Gini S.D.	Counterfactual Gini	z-stat
CR(n=3344)	0.3140	(0.0094)	0.196147	-720.312
postCR(n=1131)	0.3037	(0.0162)	0.195139	-224.939

over all 3 groups can be constructed and the corresponding overall Gini computed and compared with the true Gini. Furthermore, Gini coefficients can be decomposed into a sum of 3 components representing within cohort inequality, between cohort inequality and a component representing the extent to which the cohorts are not distinct or segmented facilitating a more detailed comparison. Details of these measures are reported in Table 13.

Table 13: Gini Decomposition

Indices	Actual	Counterfactual	
		Marriage Matching	Intergenerational
Gini over all cohorts	0.30741580	0.27334450	0.22143736
Within cohort component	0.12657076	0.11394637	0.08827589
Between cohort component	0.01426836	0.08726255	0.02677773
Non segmentation factor	0.16657668	0.07213557	0.10638373
Segmentation Index	0.45813884	0.73610014	0.51957640

Note the Counterfactual overall Gini's are significantly lower (Gini standard error = .0073512), signaling the inequality increasing effect of intensified positive assortative matching and polarizing intergenerational transfers. These structural changes increased the within cohort inequality component (especially with respect to intergenerational transitional patterns) but decreased the between group inequality component (especially with respect to changed marriage matching patterns). The diminished segmentation index reveals that, while individually the cohorts experienced increasing inequality, as a collection of groups they were experiencing increasing income commonality. Since the cohorts are associated with vintages it seems that the changes resulted in an increasing overlap of older family and younger family income distributions, i.e. they became more similar.

6 Conclusion

The strident growth in Chinese household income inequality has been ubiquitous in the last 35 years. Here the changing nature of family formation and changes in the way that human capital is passed on through the generations, are examined as sources of growing urban household income disparities. Shaped by historical events, the Cultural Revolution, The One Child Policy and the Economic Reforms, people changed the way they chose partners and invested in children, consequently changing the structure of generational relationships and the social order.

After demonstrating that ceteris paribus certain types of intergenerational transition structure and intensified marital matching behavior engender increases in income inequality, a three cohort study of social class to education, social class to income and education to income transition patterns and marital matching patterns was performed. An urban data set linking grandparents, parents and children across cohorts determined by age of head of household and potential time of marriage in Pre Cultural Revolution, Cultural Revolution Post Cultural Revolution Eras revealed that such matching and transitional patterns prevailed in each Era though they changed over the eras in a fashion that could increase household income inequality. Positive assortative partner matching on education intensified and intergenerational educational transitions were polarizing over the Eras.

In essence a source of increased urban inequality was an increased dependency of household incomes on household human capital, diminished dependency on social class and increased positive assortative matching which increased the disparities in household human capital and concomitantly increased the disparities in household incomes. An interesting sidebar was that, although educational polarization persists throughout the time, there was a substantial narrowing of the educational status in the Cultural Revolution equalizing the circumstances of later generations. In addition the middle social class is elevated after the Cultural Revolution and ends up dominating both the lower and upper social classes in its education and income outcome distributions.

To examine the ultimate impact of these phenomena on inequality a "counterfactual" analysis was performed wherein matching and transitional patterns that prevailed in the Pre Cultural Revolution Eras were assumed to also prevail in the Cultural Revolution and Post Cultural Revolution Eras. Counterfactual household income distributions were constructed together with their corresponding Gini coefficients and compared with the actual Household Income Gini coefficients that arose. In all cases the counterfactual Gini coefficients were significantly lower than the actual Gini coefficients providing evidence that a source of the ubiquitous increase in inequality was the intensified positive assortative partner choice and polarizing intergenerational transition patterns. Decomposition of the Gini coefficient in terms of the vintage cohorts revealed that, while they individually became more unequal as a consequence of the changes, collectively they were becoming more alike, there was indeed increasing generational similarity amidst growing within cohort inequality.

References

Anderson, G. (2017). Measuring Aspects of Mobility, Polarization and Convergence in the Absence of Cardinality: Indices Based Upon Transitional Typology. *Social Indicators Research*, 1-21. doi: 10.1007/s11205-017-1767-1

Anderson, G., and Leo, T. W. (2009). Child Poverty, Investment in Children and Generational Mobility: The Short and Long Term Wellbeing of Children in Urban China After the one Child Policy. *Review of Income and Wealth*, *55*, 607-629. doi: 10.1111/j.1475-4991.2009.00333.x

Anderson, G., and Leo, T. W. (2013). An empirical examination of matching theories: The one child policy, partner choice and matching intensity in urban China. *Journal of Comparative Economics*, *41*(2), 468-489. doi: 10.1016/j.jce.2012.12.005

Anderson, G., and Thomas, J. (2017). More Unequal Yet More Alike: The Changing Anatomy of Constituent Canadian Income Distributions in the 21st Century. *Mimeo University of Toronto*.

Becker, G. S. (1981). *A Treatise on the Family.* Harvard University Press.

Benjamin, D., Brandt, L., Giles, J., and Wang, S. (2008). Income Inequality during China's Economic Transition. In L. Brandt and T. G. Rawski (Eds.), *China's Great Economic Transformation* (p. 729-775). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511754234.019

Chen, Y., and Zhou, L.-A. (2007). The long-term health and economic consequences of the 1959–1961 famine in China. *Journal of Health Economics*, *26*(4), 659-681. doi: 10.1016/j.jhealeco.2006.12.006

Cheng, W., and Wu, Y. (2017). Understanding the Kuznets Process—An Empirical Investigation of Income Inequality in China: 1978–2011. *Social Indicators Research*, *134*(2), 631-650. doi: 10.1007/s11205-016-1435-x

Chiappori, P.-A., Salanié, B., and Weiss, Y. (2017). Partner Choice, Investment in Children, and the Marital College Premium. *American Economic Review*, *107*(8), 2109-2167. doi: 10.1257/aer.20150154

Choo, E., and Siow, A. (2006). Who Marries Whom and Why. *Journal of Political Economy*, *114*(1), 175-201. doi: 10.1086/498585

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility - Gregory Clark - Google Books.*

Deng, Q., Gustafsson, B., and Li, S. (2013). Intergenerational Income Persistence in Urban China. *Review of Income and Wealth*, *59*(3), 416-436. doi: 10.1111/roiw.12034

Deng, Z., and Treiman, D. J. (1997). The Impact of the Cultural Revolution on Trends in Educational Attainment in the People's Republic of China. *American Journal of Sociology*, *103*(2), 391-428. doi: 10.1086/231212

Duclos, J.-Y., Esteban, J., and Ray, D. (2004). Polarization: Concepts, Measurement, Estimation. *Econometrica*, *72*(6), 1737-1772. doi: 10.1111/j.1468-0262.2004.00552.x

Goh, C.-c., Luo, X., and Zhu, N. (2009). Income growth, inequality and poverty reduction: A case study of eight provinces in China. *China Economic Review*, *20*(3), 485-496. doi: 10.1016/j.chieco.2008.10.008

Gustafsson, B., and Shi, L. (2002). Income inequality within and across counties in rural China 1988 and 1995. *Journal of Development Economics*, *69*(1), 179-204. doi: 10.1016/S0304-3878(02)00058-5

Hertel, T., and Zhai, F. (2006). Labor market distortions, rural–urban inequality and the opening of China's economy. *Economic Modelling*, *23*(1), 76-109. doi: 10.1016/j.econmod.2005.08.004

Kanbur, R., and Zhang, X. (1999). Which Regional Inequality? The Evolution of Rural–Urban and Inland–Coastal Inequality in China from 1983 to 1995. *Journal of Comparative Economics*, *27*(4), 686-701. doi: 10.1006/jcec.1999.1612

Lefranc, A., Pistolesi, N., and Trannoy, A. (2008). Inequality of Opportunities Vs. Inequality of Outcomes: Are Western Societies All Alike? *Review of Income and Wealth*, *54*(4), 513-546. doi: 10.1111/j.1475-4991.2008.00289.x

Lefranc, A., Pistolesi, N., and Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France - ScienceDirect. *Journal of Public Economics*, *93*(11-12), 1189-1207.

Li, S. (2012). *Changes in income inequality in China in the past three decades Mimeo.* Institute of Income Distribution, Beijing Normal Univeristy, Beijing, China.

Li, S., Luo, C., Wei, Z., and Yue, X. (2008). The 1995 and 2002 household surveys: Sampling methods and data description. *Inequality and Public Policy in China*, 337-353.

Lin, J. Y., Wang, G., and Zhao, Y. (2004). Regional Inequality and Labor Transfers in China. *Economic Development and Cultural Change*, *52*(3), 587-603. doi: 10.1086/421481

Meng, X. (2004). Economic Restructuring and Income Inequality in Urban China. *Review of Income and Wealth*, *50*(3), 357-379. doi: 10.1111/j.0034-6586.2004.00130.x

Meng, X., Gregory, R., and Wang, Y. (2005). Poverty, inequality, and growth in urban China, 1986–2000. *Journal of Comparative Economics*, *33*(4), 710-729. doi: 10.1016/j.jce.2005.08.006

Meng, X., and Gregory, R. G. (2002). The Impact of Interrupted Education on Subsequent Educational Attainment: A Cost of the Chinese Cultural Revolution. *Economic Development and Cultural Change*, *50*(4), 935-959. doi: 10.1086/342761

Meng, X., Shen, K., and Xue, S. (2013). Economic reform, education expansion, and earnings inequality for urban males in China, 1988–2009. *Journal of Comparative Economics*, *41*(1), 227-244.

Mookherjee, D., and Shorrocks, A. (1982). A decomposition analysis of the trend in UK income inequality. *The Economic Journal*.

Ravallion, M., and Chen, S. (2007). China's (uneven) progress against poverty. *Journal of Development Economics*, *82*(1), 1-42. doi: 10.1016/j.jdeveco.2005.07.003

Rozelle, S. (1994). Rural Industrialization and Increasing Inequality: Emerging Patterns in China′s Reforming Economy. *Journal of Comparative Economics*, *19*(3), 362-391. doi: 10.1006/jcec.1994.1108

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *15*(1), 72-101. doi: 10.2307/1412159

Walder, A. G., and Hu, S. (2009). Revolution, Reform, and Status Inheritance: Urban China, 1949–1996. *American Journal of Sociology*, *114*(5), 1395-1427. doi: 10.1086/595949

Wan, G. (2004). Accounting for income inequality in rural China: A regression-based approach. *Journal of Comparative Economics*, *32*(2), 348-363. doi: 10.1016/j.jce.2004.02.005

Wu, X., and Perloff, J. M. (2005). China's Income Distribution, 1985–2001. *The Review of Economics and Statistics*, *87*(4), 763-775. doi: 10.1162/003465305775098206

Xie, Y., and Zhou, X. (2014). Income inequality in today's China. *Proceedings of the National Academy of Sciences*, *111*(19), 6928-6933. doi: 10.1073/pnas.1403158111

Yang, D. T. (1999). Urban-Biased Policies and Rising Income Inequality in China. *The American Economic Review*, *89*(2,).

Zhang, J., Zhao, Y., Park, A., and Song, X. (2005). Economic returns to schooling in urban China, 1988 to 2001. *Journal of Comparative Economics*, *33*(4), 730-752. doi: 10.1016/j.jce.2005.05.008

Zhong, H. (2011). The impact of population aging on income inequality in develop- ing countries: Evidence from rural China. *China Economic Review*, *22*(1), 98-107.

Appendix

A Data Summary

Table A1: Household Adult Equivalized Income data

N=6226	Mean	Std. Dev.	Min	Max
Household Equivalized Income(1)	12911.67	8406.592	0.00	113968.90
Household vintage (2)	48.23863	10.62279	27.00	75.00
Household vintage^2	2439.792	1079.848	729.00	5625.00
Father edu	5.375504	1.538859	1.00	9.00
Mother edu	4.964889	1.445191	1.00	9.00
Social Class (3)	1.836103	1.151939	1.00	5.00
Family Size	3.023266	0.7778607	1.00	9.00
CR Dummy (4)	0.5551081	0.4969909	0.00	1.00
vintage*CR	25.7648	23.30947	0.00	54.00
vintage^2*CR	1207.075	1124.387	0.00	2916.00
Father edu*CR	2.952801	2.851391	0.00	9.00
Mother edu*CR	2.789631	2.671376	0.00	9.00
Social Class*CR	1.025227	1.258468	0.00	5.00
Family Size*CR	1.680089	1.578344	0.00	8.00

(1) Brady Barber square root rule.
(2) Age of Household Head.
(3) Sum of fathers parents social class ranks and mothers parents social class ranks)/4.
(4) Household heads between ages 38 and 52 at the time of the survey would have been affected by the shut-down of schools in the Cultural Revolution, D is an indicator of heads of this age.

B Human Capital Transmission Continuous Gini

This section present the continuous Gini in Section 3.2.

The GINI coefficient can be interpreted as the average over all agents of a “relative to the mean” distance measure of each agent from all other agents. For grouped data where $f(x)$ is continuous income distribution, GINI may be written as:

$$\frac{1}{\mu}\int_a^b f(x)\int_a^b f(y)|x-y|dxdy = \int_a^b f(y)|\frac{E_{f(x)}(x)-y}{\mu}|dy = \int_a^b f(y)|1-\frac{y}{\mu}|dy = E_{f(y)}\Big(|1-\frac{y}{\mu}|\Big) \tag{1'}$$

Now consider the effect on these formulations of the GINI coefficients in the context of generational transition matrices with respect to educational attainments or income which are polarizing. Consider a distribution $f(x)$ for $x \in [a,b]$ where $0 < a < b$ and $E(x) = \int xf(x)dx = \mu$ and contemplate another distribution $f^*(x)$ where mass has been transferred in $f(x)$ from the center to the peripheries in the following fashion:

$$f^* \leq f(x) \text{ for } x \in [\mu \pm \delta] \text{ (stritly} < \text{somewhere)}$$
$$f^* \geq f(x) \text{ for } x \notin [\mu \pm \delta] \text{ (stritly} > \text{somewhere)}$$

Noting that:

$$\begin{aligned}&-\int_{\mu-\delta}^{\mu+\delta}(f^*(x)-f(x))dx \\ &= \Big\{\int_a^{\mu-\delta}(f^*(x)-f(x))dx + \int_{\mu+\delta}^b(f^*(x)-f(x))dx\Big\} > 0\end{aligned}$$

and

$$0 < |1-\frac{x}{\mu}|_{x\in[\mu\pm\delta]} < |1-\frac{x}{\mu}|_{x\notin[\mu\pm\delta]}$$

so that

$$\begin{aligned}&-\int_{\mu-\delta}^{\mu+\delta}(f^*(x)-f(x))|1-\frac{x}{\mu}|dx \\ &= \Big\{\int_a^{\mu-\delta}(f^*(x)-f(x))|1-\frac{x}{\mu}|dx + \int_{\mu+\delta}^b(f^*(x)-f(x))|1-\frac{x}{\mu}|dx\Big\}\end{aligned}$$

From equation (1) GINI*-GINI is given by:

$$\int_{\mu-\delta}^{\mu+\delta}(f^*(x)-f(x))|1-\frac{x}{\mu}|dx+\int_{a}^{\mu-\delta}(f^*(x)-f(x))|1-\frac{x}{\mu}|dx \\ +\int_{\mu+\delta}^{b}(f^*(x)-f(x))|1-\frac{x}{\mu}|dx>0$$

C Determination of Mixture Components

To study the various direct transition effects on the income distribution, following Anderson et. al (2016), the distribution of adult equivalized household income y was modeled as a K component mixture distribution of $ln(y)$ of the form:

$$f(lny)=\sum_{k=1}^{K}w_k f_k(\ln\,y,\mu_k,\delta_k^2)$$

$$\text{where } f_k(\ln\,y,\mu_k,\delta_k^2)=\frac{1}{\sqrt{2\pi\delta_k^2}}e^{-\frac{(\ln\ y-\mu_k)^2}{2\delta^2}}$$

The preferred specification had 5 components details of which are reported in Table 4. The structure can be seen to be primarily a 3 class model of roughly similar sizes with 2 components representing the poorest 2% of the sample.

Table A2: component parameters

Component	μ_k	δ_k	w_k
1	7.491525332	0.595191821	0.004187388
2	7.970963519	0.177103563	0.018043197
3	8.859922538	0.457431763	0.257995935
4	9.220069174	0.457684671	0.352934957
5	9.617802324	0.43377088	0.366838523

Transitions to an income distribution class can be explored by computing $P(I\in \text{Class}k|y_i)$ the probability that a household with ln income y_i is in class k by using the formula:

$$P(\text{household } i\in \text{ Class } k|y_i)\frac{w_k f_k(\ln\,y,\mu_k,\delta_k^2)}{\sum_{k=1}^{K}w_k f_k(\ln\,y,\mu_k,\delta_k^2)}\text{ for } k=1,...,K \qquad (1)$$

A regression of these probabilities on social or educational class membership dummies will yield the corresponding transition matrix.

To determine the optimal number of mixture components in the mixture distribution the comparison of each mixture with a kernel estimate of the distribution using versions of Gini's Transvariation Statistic with and without importance weighting and with and without parsimony penalization. Closer proximity of the mixture distribution $fM(x)$ to the kernel estimate of the distribution $fK(x)$ is the objective function here and GINI's Transvariation and importance weighted measures measure that proximity in terms of:

$$\begin{aligned}
GINIIT = &\int_0^\infty |f_M(x) - f_K(x)|dx = \int_0^\infty |\frac{f_M(x)}{f_K(x)} - 1|f_K(x)dx \\
=&E_{f_x}\big(|\frac{f_M(x)}{f_K(x)} - 1|\big) = \frac{1}{n}\sum_{i=1}^{n}|\frac{f_M(x_i)}{f_K(x_i)} - 1| \\
GINITIM = &\int_0^\infty |f_M(x) - f_K(x)|f_k(x)^{-0.5}dx = \int_0^\infty |\frac{f_M(x)}{f_K(x)} - 1|f_k(x)^{-0.5}f_K(x)dx \\
= &E_{f_x}\big(|\frac{f_M(x)}{f_K(x)} - 1|f_k(x)^{-0.5}\big) = \frac{1}{n}\sum_{i=1}^{n}|\frac{f_M(x_i)}{f_K(x_i)} - 1|f_k(x)^{-0.5}
\end{aligned}$$

The parsimony penalization factor was $+\ 3n(k)$ where $n(k)$ is the number of parameters estimated in the k component mixture. The results are reported in Table A2. In all cases the 5 component mixture minimized the statistic.

Table A3: Transvariation statistics for the mixture–kernel distribution comparisons.

Num of Components	GINIT	GINIT Penalized	GINITIMP	GINITIMP Penalized
1	0.10904065	0.10993805	65.379380	65.380278
2	0.022838910	0.024633705	1.5981538	1.5999486
3	0.030588285	0.033280477	3.5404273	3.5431195
4	0.022758709	0.026348299	0.15000316	0.15359275
5	0.018226032	0.022713020	0.061861663	0.066348651
6	0.018332401	0.023716786	0.50103731	0.50642169