

Hybrid Marriages and Phenotypic Heterosis in Offspring: Evidence from China

(Economics and Human Biology, Forthcoming)

Chen Zhu

¹ College of Economics and Management, China Agricultural University, Beijing 100083, China.

² Institute for Population and Labor Economics, Chinese Academy of Social Sciences, Beijing, China.

Xiaohui Zhang

School of Life Sciences, Nanjing University, Nanjing, Jiangsu, China.

Qiran Zhao

College of Economics and Management, China Agricultural University, Beijing 100083, China.

Qihui Chen

Center for Food and Health Economic Research, College of Economics and Management, China Agricultural University, Beijing 100083, China.

* Corresponding author. E-mail: chen1006@umn.edu. Phone: (+86) 150 1022 6027. Address: No. 17 Qinghuadonglu, Haidian, Beijing, 100081.

Abstract

In genetics, heterosis refers to the phenomenon that cross-breeding within species leads to offspring that are genetically fitter than their parents and exhibit improved phenotypic characteristics. Based on the theory of heterosis and existing genetic evidence, offspring of “hybrid” marriages (spouses originating from different states/provinces/countries/areas), though relatively rare due to physical boundaries, may exhibit greater genetic fitness in terms of intelligence, height, or physical attractiveness (“distance-performance” hypothesis). This study explores whether heterosis is a contributing factor to offspring’s educational attainment in China by applying a high-dimensional fixed effects (HDFE) modelling framework to the unique 0.1% micro-sample of the 2000 Chinese Population Census data. Concerning potential endogeneity of hybrid marriages, we conduct a series of robustness checks. Reassuringly, the estimated heterosis effect remains significantly positive across various measurements, after controlling for parental educational attainments/height, environmental influences, and over a thousand region and region-by-year fixed effects. The effects in male and higher-educated offspring are found to be stronger. Results are replicated when analyzing body height using data from the China Health and Nutrition Survey. Although endogeneity of “hybrid marriages” may not be completely ruled out, the current study sheds light on the potentially beneficial effects of interprovincial migration on population-level human capital accumulation, and we hope that this paper can intrigue future studies that further address endogeneity. The implied heterosis effect could, therefore, be profound for *Homo sapiens* as a species from an evolutionary point of view. An additional important implication is that, besides environmental factors, the overall genetic influences of parents on offspring’s performance may be further decomposed into a conventional heredity effect and a heterosis effect that has been

neglected previously.

Keywords: heterosis; hybrid marriage; genetic distance; educational attainment; height

JEL code: I00; I14; J16; J24

I. Introduction

Introduced by George Shull in 1914, heterosis (also known as hybrid vigour or heterozygote advantage) is a genetic phenomenon in which cross-breeding within species leads to offspring that are genetically fitter than their parents and exhibit improved phenotypic characteristics (Chen 2013; Lippman and Zamir 2007). Although biologists have documented a long list of heterosis in crossbred plants and animals, evidence for heterosis among human beings beyond the molecular level remains lacking. Yet given the substantial evidence of a positive correlation between geographic and genetic distances among different human populations (Manica et al. 2005; Jorde and Wooding 2004), offspring of “hybrid” marriages (in which spouses originated from different states/provinces/countries/areas) may exhibit better genetic fitness in terms of intelligence, height, or physical attractiveness (Prokosch et al. 2005; Lewis 2010), suggesting a potential heterosis effect among human populations.

A further belief in a general heterosis effect in human populations stems from three separate lines of biological evidence. First, there is an overwhelming amount of evidence on the heterozygote advantage for animals and plants found in both laboratory and wild environments (Chen 2013; Comings and MacMurray 2000). Second, the frequent evidence of inbreeding depression on human health indirectly suggests the existence of a heterosis effect among humans

(Bittles 2002; Polasek 2009). Third, there are accumulating examples of molecular heterosis in humans (Comings and MacMurray 2000; Aidoo et al. 2002). Despite these lines of evidence, however, only a limited number of studies have attempted to investigate heterosis in humans at a macro level. For example, Lewis (2010) proposed that the attractiveness of mixed-race individuals can be explained by heterosis. Mingroni (2007) argued that the tendency for IQ scores to increase in populations is a result of heterosis. And Koziel et al. (2011) found that marital distance is a key predictor of offspring height. Inspiring as they are, a common limitation of these studies is that they were mostly based on small-scale samples and failed to control for a large number of possible confounding factors.

In this study, we use the unique 1% sample of the 2000 Chinese Population Census Data to explore whether phenotypic heterosis exists among human population groups. China provides a good case study to achieve our research objective for three reasons. First, China contains a large geographical span and physical diversity. For example, the north-south distance of China is approximately 5,500 kilometres and the east-west distance 5,200 kilometres, which serve as the foundation for geographic isolation and genetic differentiation among populations (Manica et al. 2005). Second, there are 56 identified ethnic groups in China, and the existing literature has documented substantial evidence of distinct ancestors and genetic distances among different subpopulations¹ in the country (Chen et al. 2009; Du et al. 1997; Chu et al. 1998; Zhang et al. 2007; Wen et al. 2004; Zhang et al. 2010). For example, the Han ethnic group can be divided into two differentiated subgroups, northern Han and southern Han, who are approximately separated in

¹ Human populations in genetics are defined anthropologically. For example, both Han and Hui are ethnic groups in East Asia, and they can be treated as separate populations. On the other hand, Han is the largest ethnic group in China (accounted for about 91% of the total population), and Han people in different provinces/regions can be treated as separate populations, too. For example, in Du et al. (1997) and Xiao et al. (2002), they use terms such as Han-Shanghai, Han-Guangzhou and Han-Heilongjiang to distinguish subpopulations of Han people distributed in Shanghai, Guangzhou and Heilongjiang.

geographical territory by the Yangtze River (Chen et al. 2009; Wen et al. 2004). Third, as a politically unified nation for decades, China provides a context in which one can disentangle the potential heterosis effect relatively easier from certain socioeconomic, political and cultural factors than in the context of, say, Europe.

Existing evidence in genetics and experimental observations offer three important predictions about the heterosis effect in humans – the “distance-performance” hypothesis. First, the offspring of individuals from separate populations may show heterosis and perform better in a number of quantitative traits, such as intelligence, height and physical attractiveness (Emmrich et al. 2015; Prokosch et al. 2005; Lewis 2010). Second, the heterosis effect may rise with increasing genetic differences of the parents (Bikard et al. 2009; Moll et al. 1965). Third, given the substantial evidence that geographic and genetic distances among different populations are positively correlated (Manica et al. 2005; Jorde and Wooding 2004), offspring of “hybrid” marriages (with spouses originating from different states/provinces/countries/areas) may exhibit better genetic fitness and greater phenotypic heterosis.

The major phenotype analyzed in this study is the educational attainment (EA) of offspring. It is selected on the basis of a broad agreement that the educational achievement is at least moderately associated with intelligence levels (Plug and Vijverberg, 2003; Deary et al. 2007), and thus can reflect a general biological/genetic fitness to some extent (Beauchamp 2016; Prokosch et al. 2005). To reduce selection bias and the risk of confounding environmental factors, we employ the high-dimensional fixed effects (HDFE) model proposed by Guimaraes and Portugal (2010), which allows us to include a large number of fixed effects (FE) in estimation. In addition, we exploit exogenous migration decisions of respondents’ parents and province-year level average marital distance as instrumental variables (IV) for measures of genetic diversity to estimate high-

dimensional fixed effects-two stage least squares (HDFE-2SLS) models. We further replicate the results by analyzing body height, an alternative indicator of genetic fitness that is considered less likely to be affected by socioeconomic environments (Koziel et al. 2011; Mascie-Taylor 1991), using data from the China Health and Nutrition Survey.

Our empirical results find strong evidence of the existence of a heterosis effect among human beings. Consistent with predictions of the “distance-performance” hypothesis, a 1000-kilometre increase in geographic distance between parents’ birthplaces would increase the offspring’s total years of schooling and height by 0.21 years and 0.94 centimetres, respectively. The estimated heterosis effect is robust across various measurements of genetic diversity and remains similar after we simultaneously control for parental educational attainments/height, environmental influences, and over a thousand province and province-by-year FEs. And the impacts in male and higher-educated offspring are found to be larger in magnitude. Although endogeneity of “hybrid marriages” may not be completely ruled out, our study serves among the first major attempts toward understanding heterosis in humans beyond the molecular level, and we hope that this paper can intrigue future studies that further address endogeneity issues.

The remainder of this paper is structured as follows. Section II summarizes the data and describes variables used in the analysis. Section III details our empirical framework, discusses potential sources of endogeneity problems, and develops strategies to solve these problems. Section IV presents estimation results and supplemental robustness checks. Section V concludes and points out a number of directions for future research.

II. Data

The main proprietary dataset we use is the unique 0.1% micro-sample of population census data

($N = 1,180,111$) collected by the National Bureau of Statistics of China (NBS) in 2000 (henceforth, 2000 census data of China), which covered 31 provinces in mainland China.² The dataset explicitly documents both the location of current (registered) residence and birthplace (province of origin) of all sampled individuals.

We use three complementary variables to measure the genetic diversity between the two parents of an investigated individual. First, we use a binary variable to identify whether the parents of a given respondent are originated from two different provinces in China, i.e., engaged in a hybrid marriage. Second, we calculate the geographic (geodesic) distance of the couple's provinces of origin in each hybrid marriage (based on the coordinate of capital city for each province). Third, we collect information on the genetic distance (measured by F_{ST}) among Chinese populations from pairwise provinces based on gene frequencies (Du et al. 1997; Cavalli-Sforza et al. 1994).³

Excluding individuals who did not have a spouse at the time of survey yields a subsample with 553,538 observations (276,769 identified couples) from the original dataset. Table 1 provides summary statistics of hybrid marriages by province of residence for each province. As expected, Beijing, Tianjin and Shanghai, three large and most urbanized cities in China, have the highest rates of hybrid marriages. It is worth noting that the rate in Beijing is near 1/3, almost double that of Tianjin and Shanghai. The corresponding average geographic distance and genetic difference of couples in hybrid marriages for each province are also provided in columns 3-4. According to the subsample, the total rate of hybrid marriages by province of residence in China was 5.88% in 2000.⁴

² In China, both "municipalities" and "autonomous regions" are provincial-level administration units. For simplicity, we use the unified term "provinces" to represent "provinces, municipalities and autonomous regions" in the current context.

³ F_{ST} is the standard parameter used to describe the genetic structure and amount of differentiation among subpopulations in population genetics, first introduced by Wright (1949).

⁴ We also calculate the total hybrid marriage rate by province of origin (5.78%) and for each province (results not shown).

Table 1. Spatial description of hybrid marriages in China (by province of residence)

	(1)	(2)	(3)	(4)
Province	Sample size	Hybrid marriage rates	Geographic distance (1,000 kilometers), hybrid marriages only	Genetic distance (F_{ST}), hybrid marriages only
Beijing	5,948	28.38%	1.597	0.01718
Tianjin	4,976	16.92%	0.947	0.01904
Shanghai	7,512	14.35%	1.422	0.00792
Inner Mongolia	11,564	14.23%	1.105	0.00695
Heilongjiang	17,853	13.57%	0.640	0.00359
Xinjiang	8,260	12.62%	0.701	0.02159
Ningxia	2,464	11.69%	1.231	0.01058
Qinghai	1,900	10.42%	1.349	0.01988
Liaoning	21,766	8.50%	1.310	0.01349
Jilin	12,782	7.70%	0.605	0.00405
Shaanxi	14,952	6.26%	0.674	0.01853
Hebei	30,948	6.11%	1.075	0.00503
Chongqing	12,480	5.79%	1.712	0.02258
Jiangsu	34,512	5.68%	0.422	0.00774
Gansu	9,940	5.67%	0.818	0.01819
Anhui	26,040	5.48%	2.061	0.05540
Shanxi	14,594	5.39%	0.615	0.02208
Guangdong	26,380	5.34%	0.675	0.01015
Zhejiang	21,824	5.09%	1.286	0.01366
Hubei	23,902	4.51%	0.584	0.00856
Hainan	2,808	4.42%	0.747	0.00555
Jiangxi	15,300	4.31%	1.028	0.02587
Guizhou	14,348	4.04%	0.645	0.00728
Fujian	13,020	3.90%	0.647	0.00310
Henan	40,272	3.60%	1.877	0.02244
Shandong	45,470	3.54%	1.283	0.02575
Xizang	658	3.34%	0.597	0.02970
Yunnan	17,120	3.14%	0.784	0.01045
Sichuan	33,232	3.02%	1.096	0.01766
Guangxi	16,356	2.63%	0.753	0.01789
Hunan	26,418	2.45%	0.885	0.01580

Notes: Summary statistics for hybrid marriages (by province of residence) are based on information of 276,769 couples identified from the 2000 census data of China. The

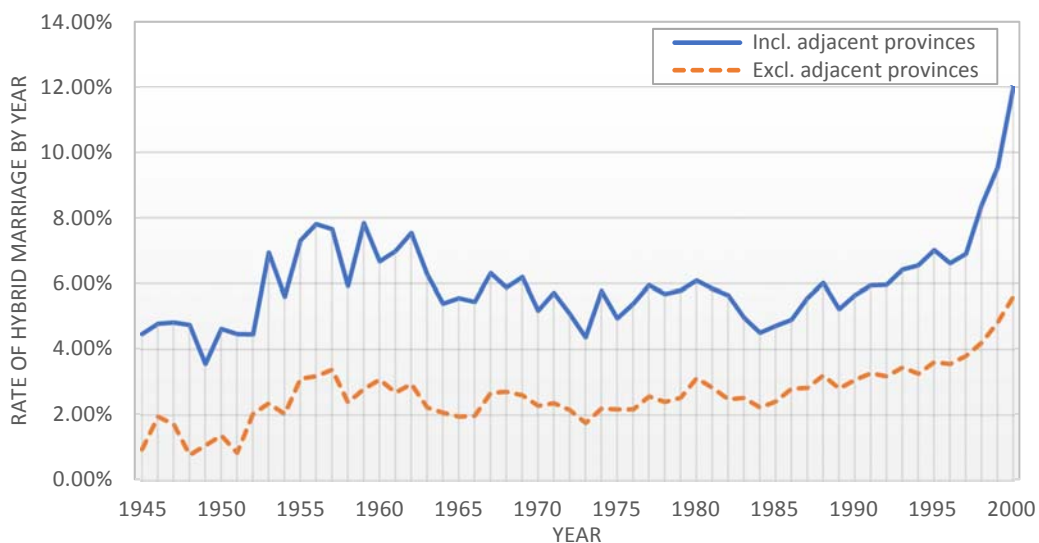
Although slightly different in magnitude, the general pattern of the spatial distribution of hybrid marriages in China remains similar (results are available from the authors upon request).

geographic distance is calculated by the coordinates of capital cities of the couple's provinces of origin in each hybrid marriage. The genetic distance information is obtained from Du et al. 1997 and Cavalli-Sforza et al. 1994. The subsample used are individuals with spouses at the time of survey (N=553,538). In 2000, the total rate of hybrid marriages (by province of residence) was 5.88% in China.

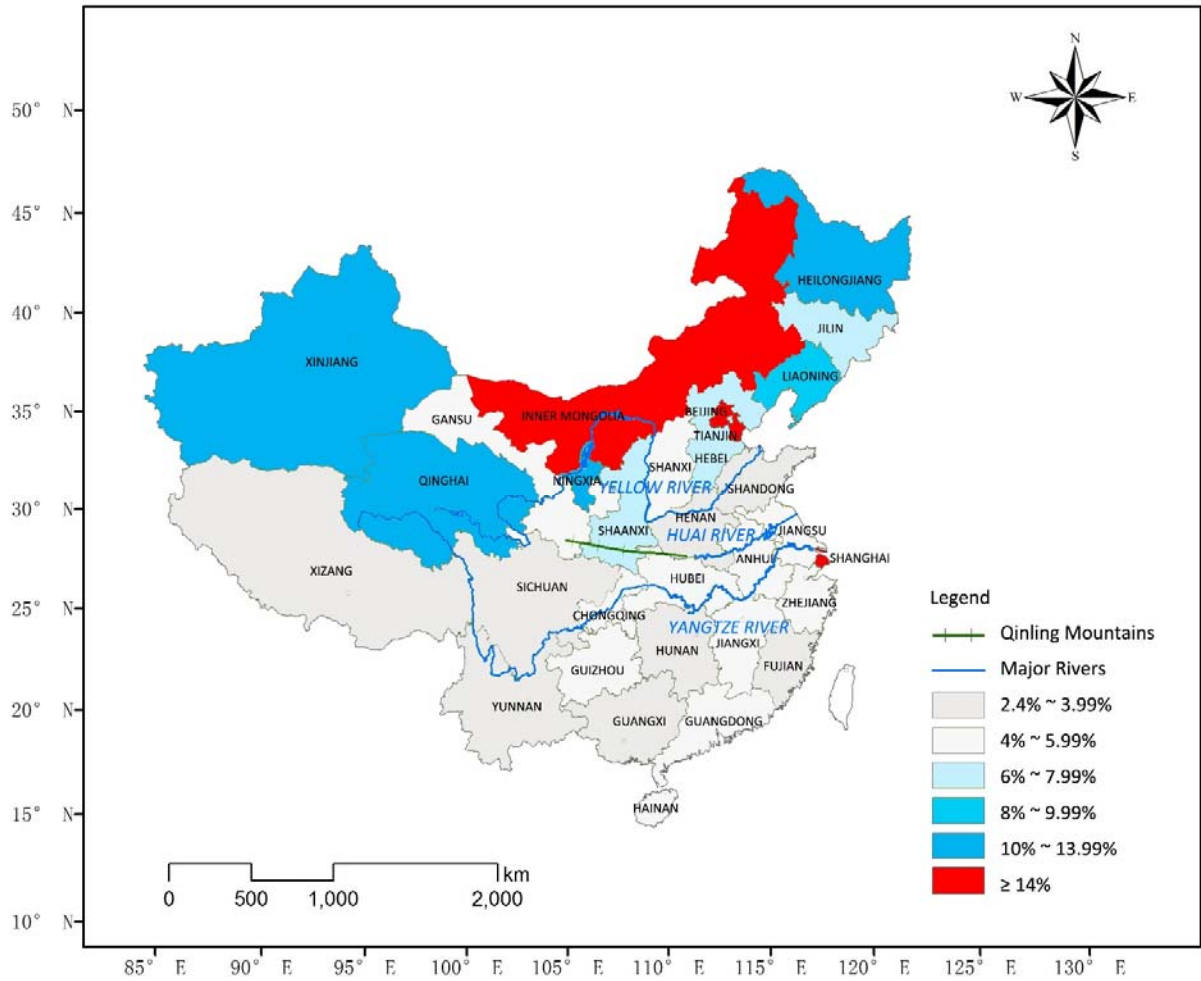
Figure 1 depicts the distribution of hybrid marriage rates by marriage year (panel A) and province of residence (panel B). As shown in panel A, the rate of hybrid marriages was relatively stable at around 6% (2.5% if excluding marriages involving adjacent provinces) before the 1990s, while it rose sharply from 1996 on, likely as a result of the massive labor migration since the early-1990s (Liang 2001). Seen from panel B, provinces in the northern part of China generally have higher rates of hybrid marriage (e.g., Beijing 28.38%, Tianjin 16.92%, Inner Mongolia 14.23%, Heilongjiang 13.57%). In contrast, people living in the southern part, specifically south of the Yangtze River, are more likely to have intra-provincial marriages and thus lower rates of hybrid marriages (e.g., Hunan 2.45%, Guangxi 2.63%).

Figure 1. Hybrid marriage rates in China

A. by marriage year



B. by province of residence



Notes: The distribution of hybrid marriage rates by marriage year (panel A) and province of residence (panel B) are calculated based on information of 276,769 couples identified from the 2000 census data of China.

III. Empirical Strategy

In the empirical analysis, we restrict our sample to offspring who were older than 18, had completed their education by the time of survey, and with full information on their parents' provinces of origin (N = 85,972). Table 2 reports summary statistics of key variables in this sample. Approximately 4.6% of the respondents are offspring from families of hybrid marriages. The average geographic distance for hybrid marriages, 889.48 kilometres (standard deviation = 583.61 kilometres), is 26 times that for all couples, 34.31 kilometres (standard deviation = 185.19 kilometres). The mean parental genetic distance is 0.0000544 (standard deviation = 0.0012919). Regarding socio-demographic characteristics, 67.01% of the respondents are male, and 10.14% are ethnic minorities. The average respondent is 26.21 years old (standard deviation = 6.30 years), completed 9.33 years of education (standard deviation = 2.36 years), and has 0.88 siblings (standard deviation = 0.95). Their parents, on average, have 6.45 years of education (standard deviation = 2.95 years).

Table 2. Descriptive information of the estimated sample

Variable	Mean/Percentage	Range
Years of education (years)	Mean = 9.33, S.D. = 2.36	Range: 0-19
Hybrid marriage	4.6% (N = 3,955)	
Genetic distance (F_{ST})	Mean = 0.0000544, S.D. = 0.0012919	Range: 0-0.1036
Geographic distance (1,000 kilometers)	Mean = 0.03431, S.D. = 0.18519	Range: 0-3.26240
Age (years)	Mean = 26.21, S.D. = 6.30	Range: 18-65
Male	67.01% (N = 57,612)	
Ethnicity	10.14% (N = 8,718)	
Urban	18.05% (N = 15,518)	
Married	33.96% (N = 29,192)	
Migration history (ever migrated = 1)	11.71% (N = 10,068)	
Number of siblings	Mean = 0.88, S.D. = 0.95	Range: 0-9
Parents: average years of education (years)	Mean = 6.45, S.D. = 2.95	Range: 0-19

Parents: number of migrated parents	Mean = 0.10, S.D. = 0.38	Range: 0-2
Parents: migration reason (ever migrated due to exogenous job assignments = 1)	0.22% (N = 189)	

Notes: The estimated sample includes individuals that were older than 18, had completed their education by the time of survey, and with full information on their parents' provinces of origin (N = 85,972) from the original 2000 census data of China.

One crucial threat to our analysis is the potential selection bias, which arises from the possibility that the genetic or geographic distance between partners of a couple may be correlated with their ability. For example, a man with higher intelligence and greater potential is more likely to migrate and/or find a partner who originates from a different province. Heredity, another genetic factor as opposed to heterosis, thus plays a more vital role in offspring's intellectual development. Another example is the case of assortative mating, where people tend to find partners who share similar traits with them (Robinson et al. 2017; Guo et al. 2014). Moreover, if parents migrated from poorer areas to more prosperous areas due to socioeconomic reasons, the difference detected between offspring of hybrid and from non-hybrid marriages may be driven by unobserved regional socioeconomic attributes, rather than heterosis *per se*. As reported in Table 3, the rates of respondents being offspring of hybrid marriages increase with educational attainments, which agitate the endogeneity suspicion.

Table 3. Offspring EA and rates from hybrid marriages

Respondents' EA	N	Rate from hybrid marriages
No school	1,462	1.33%
Primary School	16,686	2.41%
Secondary School	50,113	3.22%
High School	13,704	8.11%
College	3,981	20.45%
Graduate School	26	37.93%

We adopt three empirical strategies to address potential endogeneity issues. Our first strategy is to employ the high-dimensional fixed effects model proposed by Guimaraes and Portugal (2010), which enables us to include a large number of province and province-by-year fixed effects, to partially allay the selection bias. The general form of our HDFE model is:

$$EA = \lambda\Psi + X\beta + \theta_1\textit{Birthplace} + \theta_2\textit{Birthplace} \times \textit{Birthyear} + \theta_3\textit{Province} + \theta_4\textit{City} + \theta_5\textit{Parent1_Birthplace} + \theta_6\textit{Parent2_Birthplace} + \varepsilon \quad (1)$$

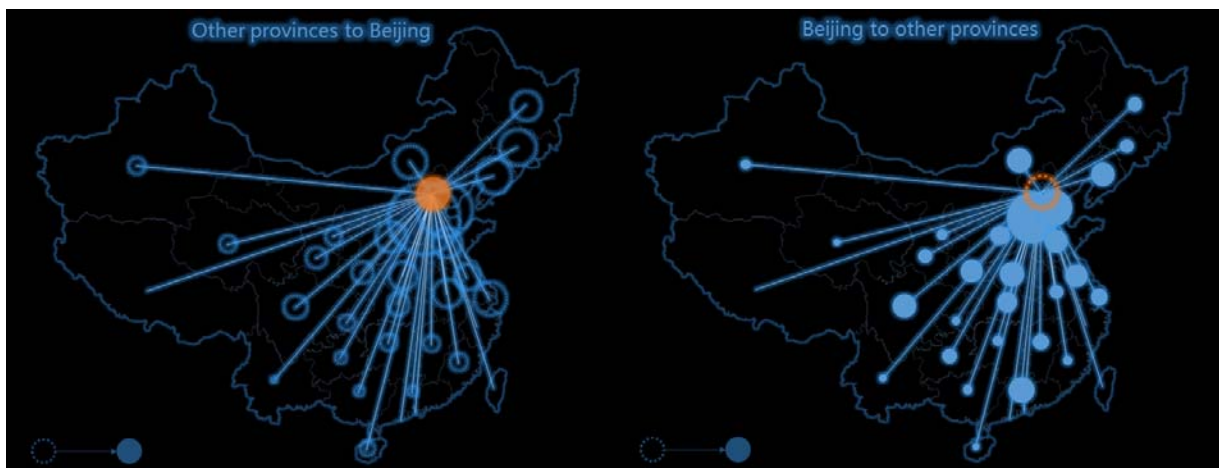
where the dependent variable is a respondent's total years of schooling; Ψ represents different measures of parental genetic diversity and λ is the coefficient vector of primary interest; X contains a set of observed individual socio-demographic characteristics, including age, gender, ethnicity, urban residence, marital status, migration history, the number of siblings, and the number of parents who migrated. Also included is parents' average years of schooling, considering the vital role heredity plays in offspring's EA development. There are a total of 1,537 fixed effects included in the model: 30 province of origin FEs, 67 birth year FEs, 1,350 province of origin \times birth year FEs, 30 province of residence FEs, 30 father's province of origin FEs, and 30 mother's province of origin FEs. Finally, ε is a stochastic error term assumed to be normally distributed, i.e., $\varepsilon \sim N(0, \sigma^2)$.

Figure 2 provides migration flow maps of major provinces from spatial analysis of migration patterns. Two popular migration destinations (Beijing and Shanghai) and two large migrant-sender provinces (Anhui and Henan) are selected. For each investigated province, there are two flow maps: one shows the percentage of migrants from other provinces to the investigated province (left panel),

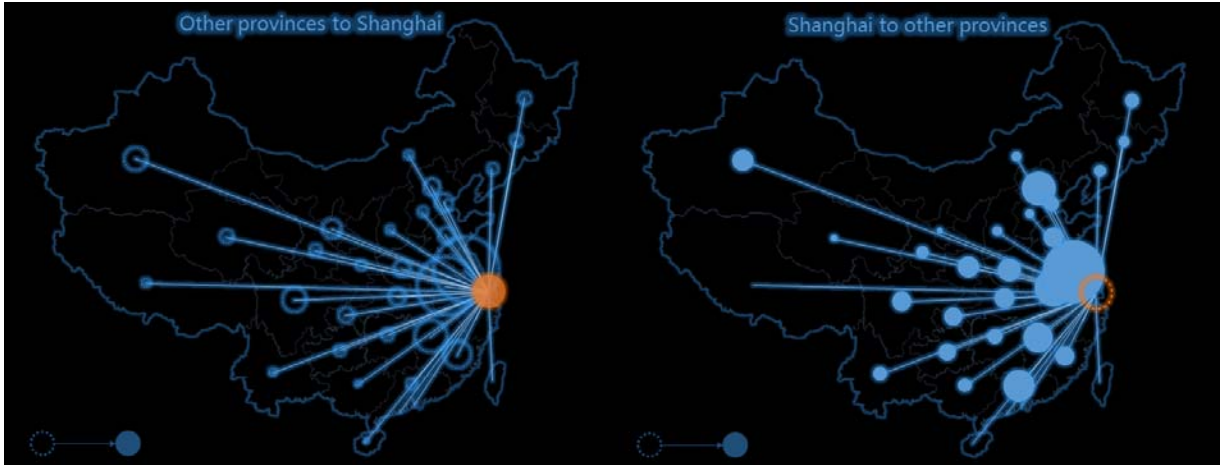
and the other one shows the reverse direction (right panel), with the size of the bubbles indicating the percentage of individuals migrated from the sender-province (hollow circle) to the destination-province (solid circle). As demonstrated by the flow maps, the migration paths are almost always bidirectional for all provinces investigated. For example, 0.28% of Beijing-born individuals end up migrating to Henan, whereas 0.35% of Henan-born individuals migrated to Beijing. It indicates that the inclusion of province-level FEs in our empirical model can provide control for the spatial distribution of migrations to a large extent.

Figure 2. Migration flow maps of Beijing, Shanghai, Anhui, and Henan

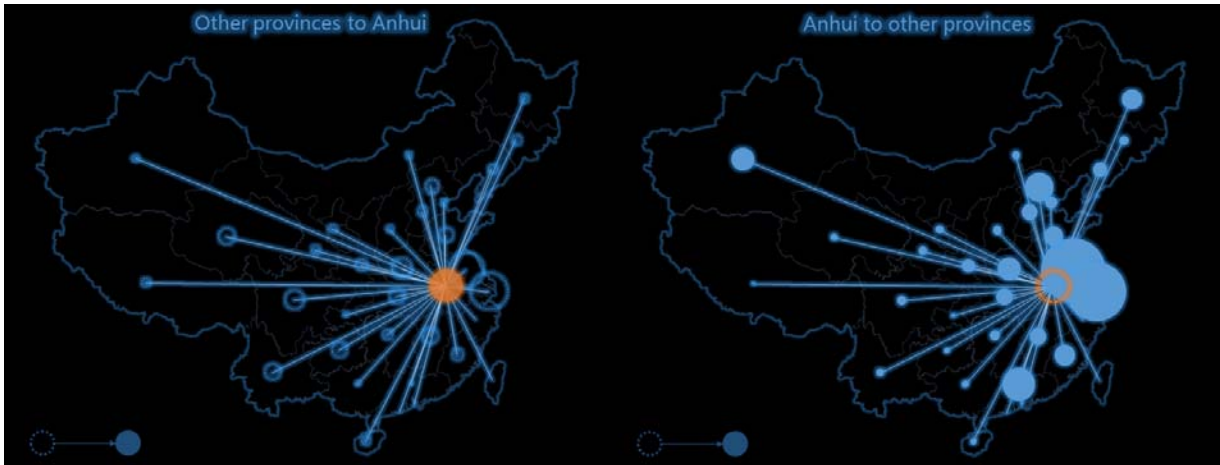
(A) Beijing



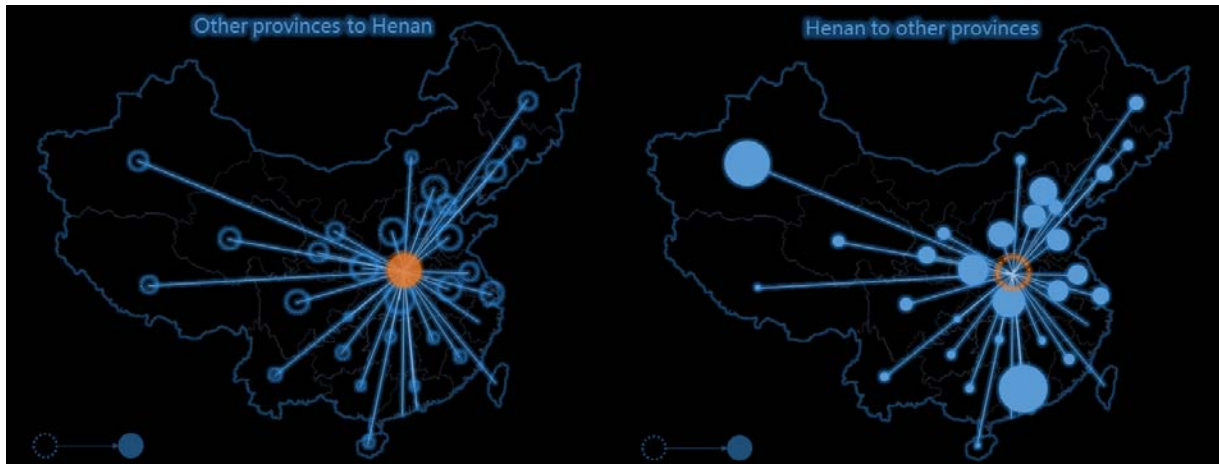
(B) Shanghai



(C) Anhui



(D) Henan



Note: The size of each bubble indicates the percentage of individuals migrated from the sender-province (hollow circle) to the destination-province (solid circle) based on the 2000 census data of China. These maps indicate that migration paths are normally bidirectional between provinces.

As a second strategy, we exploit exogenous shocks to parental migration decision to create instrument variables (IVs) for measures of genetic diversity. Since our dataset documents different migration reasons, we are able to construct a variable that identifies parents' migration that is due to exogenous job assignment and transfer. The job assignment system in urban China in the "planned economy" era was exogenous in the sense that at least between 1952 and 1984 (and may well extend to the mid-1990s), virtually all jobs were assigned by the government, according to a plan set up together with secondary schools and universities. During that time, all urban employees were guaranteed a lifelong employment (Ngok, 2008). In such a system, it is very difficult for employees to change their job voluntarily (Unger, 1982). Usually, job changes were done by the employer. Thus, our argument here is that conditional on an individual's education level, job assignment was usually made by his or her employer, not based on this individual's genetic endowment but rather the need of the employer (especially for people who entered the labor market before the mid-1980s). In addition, as area-based measures are commonly used as instruments for individual level variables (Card, 1993; Currie and Cole, 1993; Morris, 2006), we include province-year level average marital distance as a second instrumental variable. We then test whether hybrid

marriages and the associated geographic/genetic distances remain significant predictors of offspring EA, by estimating high-dimensional fixed effects-two stage least-squares models.

Our third strategy is to examine an alternative outcome variable of human capital attainment. As previously noted, height is another potential indicator of biological fitness, and is considered to be less influenced by one's socioeconomic background (Koziel et al., 2011; Mascie-Taylor, 1991). Thus, one could estimate the potential heterosis effect on respondents' height as a robust check. Unfortunately, the population census data do not contain information on respondents' morphology. As an alternative, we supplement the analysis with an additional dataset from the China Health and Nutrition Survey (CHNS), which contains information on respondents' height, as an additional strategy for robustness check.

CHNS is an international collaborative project between the Carolina Population Centre at the University of North Carolina at Chapel Hill and the Chinese Center for Disease Control and Prevention. It is a longitudinal survey with detailed information on socio-demographics, anthropometrics, and health-related conditions of the respondents and their family members (Popkin et al., 2010). Like the census data, CHNS also records information of respondents' province-of-origin, allowing us to construct the same parental genetic diversity variables as described in section II. One distinct feature is that the height of each household member was measured by a health professional during each survey interview (Liu and Zhu, 2014). Hence compared with self-reported data, the measurement error of height in CHNS is considerably lower. After excluding respondents under age of 18 as in Baten and Murray (2000), the resulting sample contains 10,412 observations from 9 provinces.

With regard to body height across generations, it has been shown that protein production, milk production in particular, is an important determinant of regional heights – the “milk hypothesis”

dubbed by Baten and Murray (2000) and Baten (2009). An attractive strategy is thus to control for milk production conditions in the analysis. However, as historical data on regional milk and dairy products production/consumption in China when most of the respondents were still in their early childhood are either incomplete or unavailable, we instead include a large set of region and region \times birth year dummy variables to help control for regional milk production conditions. Specifically, we estimate the following HDFE model to detect the potential heterosis effect on body height:

$$Height = \lambda\Psi + X\beta + \theta_1 Birthplace + \theta_2 Birthplace \times Birthyear + \theta_3 Province + \theta_4 Community + \theta_5 Parent1_Birthplace + \theta_6 Parent2_Birthplace + \theta_7 Wave + \varepsilon \quad (2)$$

where the dependent variable is respondents' height in centimetres measured by health professionals; Ψ denotes parental genetic diversity, so that λ captures the potential heterosis effect on body height; X includes age, age squared, gender, ethnicity, and the average height of respondents' parents (also measured by health professionals during the survey) that captures heritability of body height.⁵ Also included are a total of 814 fixed effects: 20 province of origin FEs, 504 province of origin \times birth year FEs, 9 province of residence FEs, 228 community of residence FEs, 22 province of origin of parent 1 FEs, 24 province of origin of parent 2 FEs, and 7 wave FEs. ε is a stochastic error term.

⁵ The inclusion of parental EA (in equation 1) and height (in equation 2) can help reduce the endogeneity problem, but it may not be able to fully purge out the influence of genetic heritability (because offspring EA and height are not *entirely* determined by genetic inheritance).

IV. Estimation Results

4.1 Results on EA from census data

Table 4 reports HDFE estimates for three different models, where the dependent variables are respondents' total years of schooling for all three models. Panel A reports results of the standard HDFE model (equation 1) for the pooled sample. Panel B presents HDFE-2SLS regression results using exogenous parental migration decisions and province-year level average marital distance as instruments. When examining these regression results, one should keep in mind that respondents' heterogeneity at birthplace / birthplace \times birth year / province of residence levels and birthplace heterogeneity of both parents have been captured by the large number of fixed effects described in Section III.

Table 4. Estimation results from HDFE and HDFE-2SLS regressions of parental genetic diversity on offspring EA

	(A) HDFE			(B) HDFE-2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Hybrid Marriage	0.088*			0.113*		
	(0.046)			(0.063)		
Geographic Distance		0.160***			0.208**	
		(0.048)			(0.098)	
Genetic Distance			20.235***			22.492**
			(5.998)			(9.398)
Age	0.001	0.001	0.000	-0.002	-0.003	-0.000
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Male	0.020	0.020	0.033**	0.015	0.026	0.035*
	(0.015)	(0.015)	(0.015)	(0.018)	(0.019)	(0.020)
Ethnicity	-0.171***	-0.170***	-0.180***	-0.135***	-0.111**	-0.192***
	(0.033)	(0.033)	(0.034)	(0.042)	(0.045)	(0.045)
Urban	2.053***	2.053***	2.042***	2.045***	2.043***	2.075***
	(0.021)	(0.021)	(0.022)	(0.025)	(0.027)	(0.032)
Married	-0.025	-0.025	-0.015	-0.003	-0.000	0.010

	(0.019)	(0.019)	(0.019)	(0.024)	(0.024)	(0.027)
Ever Migrated	0.550***	0.549***	0.579***	0.537***	0.495***	0.560***
	(0.024)	(0.024)	(0.025)	(0.029)	(0.034)	(0.034)
No. of Siblings	-0.132***	-0.132***	-0.133***	-0.123***	-0.118***	-0.127***
	(0.008)	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)
No. of Migrated Parents	0.110***	0.102***	0.118***	-0.428	-0.919	-0.779
	(0.035)	(0.031)	(0.041)	(1.561)	(0.887)	(0.772)
Average Years of Schooling of Parents	-0.013*	-0.013*	-0.012	0.014	0.038**	0.018
	(0.007)	(0.007)	(0.007)	(0.014)	(0.017)	(0.015)
Average Years of Schooling of Parents squared	0.018***	0.018***	0.018***	0.015***	0.012***	0.014***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Province of Residence FE	Yes	Yes	Yes	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes
Birthplace × Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes
Father's Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes
Mother's Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes
Total No. of Fixed Effects	1,537	1,537	1,537	1,537	1,537	1,537
First-stage F statistic	-	-	-	19.797***	16.284***	14.454***
p-value	-	-	-	0.000	0.000	0.000
Hansen-Sargan statistic	-	-	-	2.641	0.284	1.475
p-value	-	-	-	0.104	0.594	0.225
N	85,972	85,972	85,972	85,972	85,972	85,972

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. First-stage estimation results of HDFE-2SLS are reported in Appendix Table A1.

Results of the benchmark model presented in panel A reveal a generally positive and statistically significant correlation between different measures of parental genetic diversity (Ψ) and offspring's EA (panel A, columns 1-3). These results suggest that, even after purging out the influence of parental migration motivated by socioeconomic reasons (besides the inclusion of regional fixed effects), the genetic diversity of parents still has a significantly positive effect on offspring's EA.

Results of HDFE-2SLS presented in panel B are also consistent with the findings from HDFE models reported in panels A, while an intensified heterosis effect (columns 4-6). Weak-IV tests and overidentification tests also do not provide any evidence against the validity of the IVs. Results of HDFE-2SLS regressions suggest that hybrid marriage adds 0.11 years of schooling to the

offspring (panel B, column 4), and an additional 1000-kilometre increase in geographic distance between parents' birthplaces increases offspring's total years of schooling by 0.21 years (panel B, column 5). To put these results into perspective, consider a hybrid marriage involving Beijing and Shanghai, the distance between which is approximately 1000 km. If a Beijing-born man goes to college in Shanghai and marries a Shanghai-born woman, then years later their children will receive 2.5 months more education on average than Beijing-born men who do not migrate. In 1950s, many young people from Hunan province, the homeland of Mao Zedong, migrated to the Xinjiang province for political ideals and border defense obligations (Joniak-Lüthi, 2013). The distance between Hunan and Xinjiang is about 3,000 km. Thus our results imply that the children of these Hunan-born individuals and the local residents of Xinjiang are likely to have 7.5 more months of schooling compared to children whose parents were originated from Hunan or Xinjiang but did not migrate.

Simply put, the positive parameter estimates of hybrid marriage/parental geographic distance/parental genetic distance are robust in all cases.⁶ They hold conditional on parental EA, environmental influences, and a large number of fixed effects. These results imply that heterosis plays an important role in determining offspring's EA, lending strong empirical support to the "distance-performance" hypothesis proposed above.

The estimated impacts of other explanatory variables are in general as expected and are mostly

⁶ We also split the sample by different breakpoints of offspring EA and run separate HDFE-2SLS regressions. Key estimates from 12 HDFE-2SLS regressions are reported in Appendix Table A2. When splitting by either finishing primary school (panel A) or finishing secondary school (panel B), measurements on genetic diversity are significant only in the higher-educated group (columns 2 and 4). Combining the fact that parental hybrid marriage rates are lower in less educated offspring groups as shown in Table 3 (e.g. 2.41% in primary school, 3.22% in secondary school, 8.11% in high school), these results suggest that the potential gains from heterosis may be heterogeneous depending on offspring's levels of education. Full estimation results are available from the authors upon request.

in line with previous findings. Parental average years of schooling, which is usually used to control for “genetic influences from parents”, also has a positive nonlinear impact on offspring EA.⁷ An implication here is that, besides environmental factors, parents’ genetic influences of on offspring’s EA may be further decomposed into genetic inheritance and a heterosis effect. Urban residence has a positive effect on respondents’ EA, consistent with findings of Knight and Li (1996) and Connelly and Zheng (2003). A significantly negative effect is detected for the number of siblings on respondents’ total years of schooling, implying a general resource dilution effect at the family level in China, which is consistent with the findings of Li et al. (2008).

4.2 Results on body height from CHNS data

While we find strong evidence that indicators of parental genetic diversity are positively associated with increased offspring EA, we still cannot be sure that such an association represents a causal relationship that is free from selection on unobserved characteristics. We thus extend the analysis to the phenotype of offspring’s height. Our presumption is that individuals could be sorted into hybrid marriages due to higher ability and intelligence, but the selection should be much less endogenous with respect to height, especially when parental height has been controlled for. We are mainly interested in determining whether or not the estimated heterosis effect persists for offspring height as an alternative indicator of genetic fitness. Table 5 shows that this is indeed the case.

Echoing the results on EA, the estimated heterosis effect on body height is significantly positive, and remains robust across various measurements of parental genetic diversity, even after

⁷ To check for robustness, we replace parental average years of schooling and its quadratic form with a set of dummy variables indicating parental EA. The usage of dummy variables is less restrictive for the nonlinear relationship between parental and offspring EA. Key estimates of genetic diversity measurements from HDPE-2SLS results reported in Appendix Table A3 mostly resemble what have been reported in Table 4. It is worth noting that the magnitude of impact from hybrid marriages on offspring EA tends to be closest to that of father’s/mother’s having completed primary school (column 1).

controlling for the average height of parents and a total of 814 fixed effects (including respondents' province of residence FEs of respondents, community of residence FEs, province of origin FEs, province of origin \times birth year FEs, two parents' province of origin FEs, and wave FEs) accounting for individual/family heterogeneity. On average, offspring of hybrid marriages are 0.88 centimetres taller than those of with-province marriages (Table 5, column 1). A 1000-kilometre increase in the geographic distance between parents' provinces of origin leads to a 0.94-centimetre increase in offspring's height, and a 0.01 increase in parental genetic distance raises offspring's height by 0.38 centimetres.

In short, the above results obtained using a supplementary sample from CHNS suggest that the hypothesized heterosis effect is not limited to the domain of educational performance of offspring, but may exist at a more general level associated with biological fitness, such as height.

Table 5. Estimation results from HDFE regressions of parental genetic diversity on offspring height

	(1)	(2)	(3)
Hybrid Marriage	0.876** (0.378)		
Geographic Distance		0.937** (0.425)	
Genetic Distance			37.601** (16.760)
Age	-0.139 (0.089)	-0.139 (0.089)	-0.189 (0.191)
Age Squared	0.000 (0.002)	0.000 (0.002)	0.001 (0.002)
Male	11.249*** (0.122)	11.249*** (-0.122)	11.213*** (-0.125)
Ethnicity	-0.372 (0.229)	-0.379* (0.229)	-0.391* (0.232)
Average Parental Height	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)

Province of Residence FE	Yes	Yes	Yes
Community of Residence FE	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes
Birthplace × Birthyear FE	Yes	Yes	Yes
Father's Birthplace FE	Yes	Yes	Yes
Mother's Birthplace FE	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes
Total No. of Fixed Effects	814	814	814
N	10,412	10,412	10,412

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

4.3 Gender differences

Since there are potential structural differences between male and female offspring in the accumulation process of human capital attainment, we further analyse these two subsamples separately to detect potential gender difference in the hypothesized heterosis effect (results are reported in Appendix Tables A4 and A5, with first-stage estimates included). Similar patterns emerge for both genders in these tables, although the effect for male offspring is in general stronger than the effect for female offspring for both outcomes. For males, offspring of hybrid marriages completed 0.13 more years of education and are 1.15 centimetres taller than offspring of within-province marriages, while for female offspring the corresponding estimates are, respectively, 0.08 years and 0.35 centimetres, which are smaller in magnitude and less statistically significant. Likewise, an additional 1000-kilometre increase in the distance between two parents' birthplaces increases years of schooling by 0.30 years for male offspring, but only 0.19 years for female offspring. Such gender differences are consistent with existing biological findings that fitness indicators are normally greater for males than for females, likely due to sexual selection (Prokosch et al. 2005; Koziel et al. 2011; Gangestad and Thornhill 1999).⁸

⁸ Classical sexual selection theory provides a well-supported conceptual framework which predicts that, compared with females,

V. Conclusion

In summary, the results of the present study suggest that heterosis can be considered as a contributing factor to both the improved levels of educational attainment and increased body height in offspring from hybrid marriages. The estimated heterosis effect is robust across various measurements of parental genetic diversity and remains significantly positive after environmental influences and potential endogeneity of hybrid marriages have been controlled for. Moreover, the effect is stronger for male and higher educated offspring. To the best of our knowledge, these findings are the first to provide evidence of phenotypic heterosis among human populations from a socioeconomic perspective. They also offer new explanations for the outperformance of mixed-ethnic/race people found by social scientists and economists (Rhodes et al. 2005). The current study sheds light on the potentially beneficial effects of the intensified interprovincial migration on province-level human capital accumulation. As geographic boundaries become less important in the marriage market due to the modernization of transportation methods, and as “hybrid marriages” (in terms of state/province/race/ethnicity/country) become more commonly observed both in China and worldwide (Treiman 1970; Maas and Zijdeman 2010; Cai and Wang 2008), the implied heterosis effects on educational attainment and height could be profound for *Homo sapiens* as a species from an evolutionary point of view.

Although we have attempted to address potential endogeneity in hybrid marriage using a number of strategies, the problem may remain due to the lack of a source of entirely exogenous variation

males usually evolve to develop greater characters/traits (e.g. taller and stronger) due to their need for competition over mates (Andersson and Iwasa, 1996; Campbell 1972). Sexual selection would therefore lead to the evolution of several male-biased phenotypic characters such as height. The current results are consistent with such predictions in evolutionary biology.

in hybrid marriage that we can explore. However, we believe that our strategies have at least reduced the endogeneity problem. And despite the remaining problem, the paper serves among the first major steps toward understanding heterosis in humans beyond the molecular level. We hope that this paper can attract more studies of heterosis in humans and that future works can address endogeneity in hybrid marriage more thoroughly.

Before closing, a note on two caveats of the present study is in order. Another limitation of this study is that we did not measure genetic distance based on individual genomic information. Although the interprovincial migration rate of the grandparental generation is considerably low (1.51% for individuals born before 1950), which largely mitigates the concern about misclassification of parents' genetic origin based on their birthplace, our inability to directly link individual genome-wide heterozygosity to improved socioeconomic performance remains a problem. Also, if there is any unobserved cultural fusion advantage in the child-nurturing process of hybrid marriages, the estimated heterosis effect may be somewhat overestimated. Future research may use more direct measures of genetic distance or genome heterozygosity, preferably at the individual level, to gain a deeper understanding of phenotypic traits from heterosis in humans.

Appendix

Table A1. First-stage estimation results from HDFE-2SLS regressions of parental genetic diversity on offspring EA

	HDFE-2SLS, First Stage		
	(1) Instrumented Var: hybrid marriage	(2) Instrumented Var: geographic distance	(3) Instrumented Var: genetic distance
Exogenous parental migration	0.020*** (0.005)	0.014*** (0.004)	0.00005** (0.00003)
Average marital distance	0.033** (0.017)	0.429*** (0.090)	0.00317*** (0.00090)
Age	0.000 (0.000)	0.000** (0.000)	0.00000 (0.00000)
Male	0.001 (0.001)	0.001 (0.001)	0.00000 (0.00001)
Ethnicity	-0.005* (0.003)	-0.006** (0.003)	0.00001 (0.00002)
Urban	0.001 (0.002)	0.001 (0.002)	-0.00003 (0.00001)
Married	-0.003** (0.002)	-0.003* (0.001)	-0.00002 (0.00001)
Ever Migrated	0.000 (0.002)	0.004** (0.002)	0.00001 (0.00002)
No. of Siblings	-0.001* (0.001)	-0.002** (0.001)	0.00000 (0.00001)
No. of Migrated Parents	0.489*** (0.002)	0.319*** (0.002)	0.00142*** (0.00003)
Average Years of Schooling of Parents	-0.004*** (0.001)	-0.005*** (0.001)	-0.00002*** (0.00000)
Average Years of Schooling of Parents Squared	0.000*** (0.000)	0.001*** (0.000)	0.00000*** (0.00000)
Province of Residence FE	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes
Birthplace × Birthyear FE	Yes	Yes	Yes
Father's Birthplace FE	Yes	Yes	Yes
Mother's Birthplace FE	Yes	Yes	Yes
Total No. of Fixed Effects	1,537	1,537	1,537

First-stage F statistic	19.797***	16.284***	14.454***
p-value	(0.000)	(0.000)	(0.000)
Hansen-Sargan statistic	2.641	0.284	1.475
p-value	(0.104)	(0.594)	(0.225)
N	85,972	85,972	85,972

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. These are first-stage estimates corresponding to HDFE-2SLS results presented in Table 4, panel B.

Table A2. Key estimates from HDFE-2SLS regressions of subsamples by offspring EA

Measurements of genetic diversity	(A) Splitting by primary school		(B) Splitting by secondary school	
	(1) Finished primary school or below	(2) Finished secondary school or above	(3) Finished secondary school or below	(4) Finished high school or above
(a) Hybrid Marriage	0.044 (0.149)	0.131*** (0.037)	0.041 (0.055)	0.147*** (0.054)
(b) Geographic Distance	0.058 (0.181)	0.188*** (0.039)	0.044 (0.066)	0.132** (0.053)
(c) Genetic Distance	0.112 (23.602)	20.094*** (4.543)	1.453 (9.463)	20.382*** (6.018)
N	18,148	68,261	67,824	17,711

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Estimates of key independent variables are reported from 12 separate HDFE-2SLS regressions. Full estimation results are available from the authors upon request.

Table A3. Additional HDFE-2SLS results of parental genetic diversity on offspring EA

	(1)	(2)	(3)
Hybrid Marriage	0.109** (0.046)		
Geographic Distance		0.172*** (0.048)	
Genetic Distance			20.644*** (5.998)
Age	-0.001	-0.002	0.001

	(0.003)	(0.003)	(0.003)
Male	0.014 (0.018)	0.025 (0.019)	0.034* (0.021)
Ethnicity	-0.138*** (0.042)	-0.114** (0.045)	-0.200*** (0.046)
Urban	2.032*** (0.026)	2.015*** (0.028)	2.053*** (0.031)
Married	-0.009 (0.024)	-0.005 (0.025)	0.005 (0.028)
Ever Migrated	0.526*** (0.029)	0.482*** (0.034)	0.540*** (0.035)
No. of Siblings	-0.126*** (0.011)	-0.120*** (0.011)	-0.130*** (0.012)
No. of Migrated Parents	-0.494 (1.560)	-0.972 (0.875)	-0.832 (0.779)
Father_edu_primary	0.303*** (0.032)	0.282*** (0.034)	0.296*** (0.037)
Father_edu_secondary	0.740*** (0.035)	0.706*** (0.038)	0.734*** (0.040)
Father_edu_high	1.181*** (0.049)	1.186*** (0.047)	1.156*** (0.052)
Father_edu_college	2.084*** (0.133)	1.872*** (0.153)	2.069*** (0.134)
Mother_edu_primary	0.280*** (0.022)	0.273*** (0.024)	0.279*** (0.026)
Mother_edu_secondary	0.630*** (0.030)	0.610*** (0.033)	0.626*** (0.034)
Mother_edu_high	1.151*** (0.077)	1.060*** (0.082)	1.194*** (0.055)
Mother_edu_college	2.016*** (0.090)	1.657*** (0.157)	1.970*** (0.118)
Province of Residence FE	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes
Birthplace × Birthyear FE	Yes	Yes	Yes
Father's Birthplace FE	Yes	Yes	Yes
Mother's Birthplace FE	Yes	Yes	Yes
Total No. of Fixed Effects	1,537	1,537	1,537
<i>First-stage:</i>			
Exogenous parental migration	0.020*** (0.005)	0.015*** (0.004)	0.00007* (0.00004)
Average marital distance	0.036* (0.020)	0.435*** (0.090)	0.00316*** (0.00090)
Age	0.000* (0.000)	0.000*** (0.000)	0.00000 (0.00000)

Male	0.001 (0.001)	-0.001 (0.001)	0.00000 (0.00001)
Ethnicity	-0.005* (0.003)	-0.006** (0.003)	0.0000106 (0.00002)
Urban	0.001 (0.002)	0.003 (0.002)	-0.00002 (0.00001)
Married	-0.003** (0.002)	-0.003* (0.001)	-0.00002 (0.00001)
Ever Migrated	-0.001 (0.002)	0.004** (0.002)	0.00001 (0.00002)
No. of Siblings	-0.001* (0.001)	-0.002** (0.001)	0.00000 (0.00001)
No. of Migrated Parents	0.489*** (0.002)	0.319*** (0.002)	0.00141*** (0.00003)
Father_edu_primary	-0.001 (0.002)	0.001 (0.002)	-0.00001 (0.00002)
Father_edu_secondary	-0.001 (0.002)	0.003 (0.002)	-0.00001 (0.00002)
Father_edu_high	0.008 (0.003)	0.005** (0.003)	0.00003 (0.00002)
Father_edu_college	0.037*** (0.004)	0.051*** (0.004)	0.00021*** (0.00004)
Mother_edu_primary	-0.001 (0.002)	0.000 (0.002)	0.00000 (0.00001)
Mother_edu_secondary	0.002 (0.002)	0.004* (0.002)	0.00001 (0.00002)
Mother_edu_high	0.019*** (0.003)	0.024*** (0.003)	0.00002 (0.00003)
Mother_edu_college	0.015*** (0.005)	0.049*** (0.005)	0.00012** (0.00005)
First-stage F statistic	11.805	17.015	14.542
p-value	(0.000)	(0.000)	(0.000)
Hansen-Sargan statistic	2.561	0.010	1.089
p-value	(0.110)	(0.922)	(0.297)
Observations	85,972	85,972	85,972

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Father_edu_primary – Mother_edu_college are defined as: (1) father completed primary school; (2) father completed secondary school; (3) father completed high school; (4) father completed college or above; (5) mother completed primary school; (6) mother completed secondary school; (7) mother completed high school; and (8) mother completed college or above.

Table A4. HDFE-2SLS estimation results of offspring EA by gender

	(A) Male			(B) Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Hybrid Marriage	0.127** (0.055)			0.075* (0.040)		
Geographic Distance		0.295*** (0.089)			0.192*** (0.067)	
Genetic Distance			22.820** (11.456)			19.231* (9.942)
Age	-0.013*** (0.003)	-0.012*** (0.003)	-0.011*** (0.003)	0.044*** (0.006)	0.046*** (0.006)	0.044*** (0.006)
Ethnicity	-0.139*** (0.052)	-0.136*** (0.045)	-0.174*** (0.041)	-0.160** (0.068)	-0.157** (0.074)	-0.182** (0.071)
Urban	1.895*** (0.033)	1.897*** (0.028)	1.902*** (0.028)	2.333*** (0.043)	2.351*** (0.050)	2.282*** (0.044)
Married	0.110*** (0.029)	0.089*** (0.023)	0.090*** (0.022)	-0.473*** (0.047)	-0.471*** (0.051)	-0.502*** (0.051)
Ever Migrated	0.540*** (0.038)	0.521*** (0.035)	0.573*** (0.032)	0.572*** (0.046)	0.559*** (0.050)	0.606*** (0.050)
No. of Siblings	-0.093*** (0.014)	-0.100*** (0.012)	-0.105*** (0.011)	-0.153*** (0.015)	-0.161*** (0.018)	-0.158*** (0.016)
No. of Migrated Parents	-1.326 (1.482)	-0.563 (0.572)	-0.430 (0.314)	2.693** (1.151)	2.766** (1.336)	1.355** (0.560)
Average Years of Schooling of Parents	0.009 (0.015)	0.007 (0.014)	-0.014 (0.010)	0.001 (0.017)	-0.018 (0.024)	0.001 (0.017)
Average Years of Schooling of Parents Squared	0.015*** (0.002)	0.015*** (0.001)	0.018*** (0.001)	0.017*** (0.001)	0.020*** (0.002)	0.017*** (0.001)
Province of Residence FE	Yes	Yes	Yes	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes
Birthplace × Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes
Father's Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes
Mother's Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes
Total No. of Fixed Effects	1,451	1,451	1,451	1,060	1,060	1,060
<i>First-stage:</i>						
Exogenous parental migration	0.020*** (0.006)	0.013** (0.005)	0.00008* (0.00005)	0.017** (0.007)	0.015** (0.007)	0.00002* (0.00001)
Average marital distance	0.423*** (0.117)	0.771*** (0.107)	0.00782*** (0.00113)	0.296** (0.146)	0.619*** (0.188)	0.00736*** (0.00142)
Age	0.000 (0.000)	0.000 (0.000)	0.00000 (0.00000)	0.000 (0.000)	0.000 (0.000)	0.00000 (0.00000)
Ethnicity	-0.005 (0.003)	-0.008** (0.003)	-0.00002 (0.00003)	-0.003 (0.005)	-0.002 (0.005)	0.00001 (0.00003)
Urban	0.000 (0.002)	0.000 (0.002)	-0.00003* (0.00002)	0.004 (0.003)	0.005* (0.003)	0.00000 (0.00002)

Married	-0.004 (0.002)	-0.003* (0.002)	-0.00001 (0.00001)	-0.001 (0.003)	-0.001 (0.003)	-0.00003 (0.00002)
Ever Migrated	0.000 (0.003)	0.006** (0.002)	0.00001 (0.00002)	0.000 (0.004)	-0.003 (0.004)	-0.00002 (0.00002)
No. of Siblings	-0.002 (0.001)	-0.001* (0.001)	0.00000 (0.00001)	-0.001 (0.001)	-0.002 (0.001)	-0.00001 (0.00001)
No. of Migrated Parents	0.478*** (0.003)	0.306*** (0.003)	0.00140*** (0.00003)	0.533*** (0.004)	0.355*** (0.004)	0.00135*** (0.00004)
Average Years of Schooling of Parents	-0.003*** (0.001)	-0.005*** (0.001)	-0.00002*** (0.00001)	-0.003*** (0.001)	-0.005*** (0.001)	-0.00002** (0.00001)
Average Years of Schooling of Parents Squared	0.000*** (0.000)	0.001*** (0.000)	0.00000*** (0.00000)	0.000*** (0.000)	0.001*** (0.000)	0.00000*** (0.00000)
First-stage F statistic	11.941***	18.425***	15.174***	19.776***	16.425***	13.494***
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Hansen-Sargan statistic	0.169	1.671	2.095	0.049	0.358	0.337
p-value	(0.684)	(0.196)	(0.148)	(0.824)	(0.550)	(0.562)
N	57,612	57,612	57,612	28,360	28,360	28,360

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A5. HDFE estimation results of offspring height by gender

	(A) Male			(B) Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Hybrid Marriage	1.152** (0.483)			0.352* (0.185)		
Geographic Distance		1.550*** (0.545)			0.478** (0.020)	
Genetic Distance			17.991 (221.862)			394.359 (301.184)
Age	-0.155 (0.106)	-0.152 (0.106)	-0.184 (0.168)	-0.147 (0.176)	-0.146 (0.176)	-0.221 (0.181)
Age Squared	0.000 (0.002)	0.000 (0.002)	0.001 (0.002)	0.001 (0.003)	0.001 (0.003)	0.003 (0.003)
Ethnicity	-0.487 (0.306)	-0.503 (0.306)	-0.503 (0.311)	0.120 (0.491)	0.117 (0.491)	0.094 (0.507)
Average Parental Height	0.004*** (0.001)	0.004*** (0.001)	0.004** (0.001)	0.005** (0.002)	0.005** (0.002)	0.006** (0.002)
Province of Residence FE	Yes	Yes	Yes	Yes	Yes	Yes
Community of Residence FE	Yes	Yes	Yes	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes
Birthplace × Birthyear FE	Yes	Yes	Yes	Yes	Yes	Yes

Father's Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes
Mother's Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes	Yes	Yes
Total No. of Fixed Effects	766	766	766	722	722	722
N	7,095	7,095	7,095	3,317	3,317	3,317

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Acknowledgements

We thank Timothy Beatty, Junfei Bai, Yaohui Zhao, Xiaoyuan Dong, Fangwen Lu, Jianfeng Xu, Jonathan Beauchamp, Yong Zhu, Yaping Cheng, as well as participants of the 2017 Chinese Economists Society annual meeting in Nanjing, Chinese Women Economists International Workshop in Beijing, and Behavior Genetics Association annual meeting in Oslo for their helpful comments. The study is financially supported by the National Social Science Fund of China (14CJY018), the National Science Foundation of China (71603261), the Ministry of Education of Humanities and Social Science Project (16YJC880107), and the Chinese Universities Scientific Fund #2017RW005. This paper uses data from the China Health and Nutrition Survey (CHNS). We thank the National Bureau of Statistics of China, National Institute of Nutrition and Food Safety, China Center for Disease Control and Prevention; the Carolina Population Center, University of North Carolina at Chapel Hill; the National Institutes of Health (NIH; R01-HD30880, DK056350, and R01-HD38700); and the Fogarty International Center, NIH, for financial support for the CHNS data collection and analysis files since 1989.

References

- [1] Aidoo, M., Terlouw, D. J., Kolczak, M. S., McElroy, P. D., ter Kuile, F. O., Kariuki, S., ... & Udhayakumar, V. (2002). Protective effects of the sickle cell gene against malaria morbidity and mortality. *The Lancet*, 359(9314), 1311-1312.
- [2] Andersson, M. and Iwasa, Y. (1996). Sexual selection. *Trends in ecology & evolution*, 11(2), pp.53-58.
- [3] Baten, J. and Murray, J.E. (2000). Heights of men and women in 19th-century Bavaria: economic, nutritional, and disease influences. *Explorations in Economic History*, 37(4), pp.351-369.
- [4] Baten, J. (2009). Protein supply and nutritional status in nineteenth century Bavaria, Prussia and France. *Economics & Human Biology*, 7(2), pp.165-180.
- [5] Beauchamp, J.P. (2016). Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences*, 113(28), pp.7774-7779.
- [6] Bikard, D., Patel, D., Le Metté, C., Giorgi, V., Camilleri, C., Bennett, M.J. and Loudet, O. (2009). Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science*, 323(5914), pp.623-626.
- [7] Bittles, A. H. (2002). Endogamy, consanguinity and community genetics. *Journal of genetics*, 81(3), 91-98.
- [8] Cai, F., Wang, D. (2008). Impacts of internal migration on economic growth and urban development in China. In: DeWind, Josh, Holdaway, Jennifer (Eds.), *Migration and Development Within and Across Borders: Research and Policy Perspectives on Internal and International Migration*. International Organization for Migration and the Social Science Research Council, New York, pp. 245–272.

- [9] Campbell, B.G. ed. (1972). Sexual selection and the descent of man, 1871-1971 (pp. 87-104). Heinemann.
- [10] Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling (No. w4483). *National Bureau of Economic Research*.
- [11] Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton university press.
- [12] Chen, Z. J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nature Reviews Genetics*, 14(7), 471-482.
- [13] Chen, J., Zheng, H., Bei, J. X., Sun, L., Jia, W. H., Li, T.,... & Liu, J. (2009). Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *The American Journal of Human Genetics*, 85(6), 775-785.
- [14] Chu, J. Y., Huang, W., Kuang, S. Q., Wang, J. M., Xu, J. J., Chu, Z. T.,... & Geng, Z. C. (1998). Genetic relationship of populations in China. *Proceedings of the National Academy of Sciences*, 95(20), 11763-11768.
- [15] Comings, D. E., & MacMurray, J. P. (2000). Molecular heterosis: a review. *Molecular genetics and metabolism*, 71(1), 19-31.
- [16] Connelly, R., Zheng, Z. (2003). Determinants of school enrollment and completion of 10 to 18 year olds in China. *Economics of Education Review*. 22, 379–388.
- [17] Currie, J., & Cole, N. (1993). Welfare and child health: The link between AFDC participation and birth weight. *The American Economic Review*, 83(4), 971-985.
- [18] Deary, I.J., Strand, S., Smith, P. and Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1), pp.13-21.
- [19] Du, R., Xiao, C., & Cavalli-Sforza, L. L. (1997). Genetic distances between Chinese

populations calculated on gene frequencies of 38 loci. *Science in China Series C: Life Sciences*, 40(6), 613-621.

[20] Emmrich, P.M.F., Roberts, H.E. and Pancaldi, V. (2015). A Boolean gene regulatory model of heterosis and speciation. *BMC evolutionary biology*, 15(1), p.24.

[21] Gangestad, S. W., & Thornhill, R. (1999). Individual differences in developmental precision and fluctuating asymmetry: a model and its implications. *Journal of Evolutionary Biology*, 12(2), 402-416.

[22] Guimaraes, P. and Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal*, 10(4), p.628.

[23] Guo, G., Wang, L., Liu, H., & Randall, T. (2014). Genomic assortative mating in marriages in the United States. *PloS One*, 9(11), e112322.

[24] Harwell, L.H., Goldberg, M.L., Fischer, J.A., Hood, L., Aquadro, C. F. (2014). *Genetics from genes to genomes* (5th edition). McGraw-Hill.

[25] Joniak-Lüthi, A. (2013). Han Migration to Xinjiang Uyghur Autonomous Region: Between State Schemes and Migrants' Strategies. *Zeitschrift für Ethnologie*, pp.155-174.

[26] Jorde, L. B., & Wooding, S. P. (2004). Genetic variation, classification and 'race'. *Nature Genetics*, 36, S28-S33.

[27] Knight, J. and Li, S. (1996). Educational Attainment and the Rural--Urban Divide in China. *Oxford Bulletin of Economics and Statistics* 58(1), 83-117.

[28] Koziel, S., Danel, D. P., & Zaręba, M. (2011). Isolation by distance between spouses and its effect on children's growth in height. *American Journal of Physical Anthropology*, 146(1), 14-19.

[29] Lewis, M. B. (2010). Why are mixed-race people perceived as more attractive?. *Perception*, 39(1), 136-138.

- [30] Li, H., Zhang, J., Zhu, Y. (2008). The quantity-quality trade-off of children in a developing country: identification using Chinese twins. *Demography* 45 (1), 223–243.
- [31] Liang, Z. (2001). The age of migration in China. *Population and development review*, 27(3), pp.499-524.
- [32] Lippman, Z. B., & Zamir, D. (2007). Heterosis: revisiting the magic. *Trends in genetics*, 23(2), 60-66.
- [33] Liu, X. and Zhu, C. (2014). Will knowing diabetes affect labor income? Evidence from a natural experiment. *Economics Letters*, 124(1), pp.74-78.
- [34] Maas, I., & Zijdeman, R. L. (2010). Beyond the local marriage market: The influence of modernization on geographical heterogamy. *Demographic Research*, 23, 933.
- [35] Manica, A., Prugnolle, F., & Balloux, F. (2005). Geography is a better determinant of human genetic differentiation than ethnicity. *Human genetics*, 118(3-4), 366-371.
- [36] Mascie-Taylor, C.N. (1991). Biosocial influences on stature: a review. *Journal of biosocial science*, 23(01), pp.113-128.
- [37] Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, 114(3), 806.
- [38] Moll, R.H., Lonnquist, J.H., Fortuno, J.V. and Johnson, E.C. (1965). The relationship of heterosis and genetic divergence in maize. *Genetics*, 52(1), p.139.
- [39] Morris, S. (2006). Body mass index and occupational attainment. *Journal of Health Economics*, 25(2), 347-364.
- [40] Ngok, K. (2008). The changes of Chinese labor policy and labor legislation in the context of market transition. *International Labor and Working-Class History*, 73(1), 45-64.
- [41] Plug, E. and Vijverberg, W. (2003). Schooling, family background, and adoption: Is it nature

or is it nurture? *Journal of political Economy*, 111(3), pp.611-641.

[42] Prokosch, M. D., Yeo, R. A., & Miller, G. F. (2005). Intelligence tests with higher g-loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence*, 33(2), 203-213.

[43] Polasek, O. (2009). Investigating the role of human genomewide heterozygosity as a health risk factor.

[44] Popkin, B.M., Du, S., Zhai, F. and Zhang, B. (2009). Cohort Profile: The China Health and Nutrition Survey—monitoring and understanding socio-economic and health change in China, 1989–2011. *International Journal of Epidemiology*, 39(6), pp.1435-1440.

[45] Robinson, M. R., Kleinman, A., Graff, M., Vinkhuyzen, A. A., Couper, D., Miller, M. B., ... & van Vliet-Ostaptchouk, J. V. (2017). Genetic evidence of assortative mating in humans. *Nature Human Behaviour*, 1, 0016.

[46] Rhodes, G., Lee, K., Palermo, R., Weiss, M., Yoshikawa, S., Clissa, P.,... & Jeffery, L. (2005). Attractiveness of own-race, other-race, and mixed-race faces. *Perception*, 34(3), 319-340.

[47] Treiman, D. J. (1970). Industrialization and social stratification. *Sociological Inquiry*, 40(2), 207-234.

[48] Unger, J. (1982). *Education under Mao: Class and competition in Canton schools, 1960-1980*. Columbia University Press.

[49] Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y.,... & Qian, J. (2004). Genetic evidence supports demic diffusion of Han culture. *Nature*, 431(7006), 302-305.

[50] Wright, S. (1949). The genetical structure of populations. *Annals of Human Genetics*, 15(1), 323-354.

[51] Zhang, F., Su, B., Zhang, Y. P., & Jin, L. (2007). Genetic studies of human diversity in East

Asia. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1482), 987-996.

[52] Zhang, Y., Zhang, F., Lin, H., Shi, L., Wang, P., Shi, L.,... & Chu, J. (2010). Nucleotide polymorphism of the TNF gene cluster in six Chinese populations. *Journal of Human Genetics*, 55(6), 350-357.