I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17986

Does College Education Make Women Less Likely to Marry? Evidence from the Chinese Higher Education Expansion

Bin Huang
Massimiliano Tani
Lei Xu
Yu Zhu

JULY 2025

DISCUSSION PAPER SERIES

IZA DP No. 17986

Does College Education Make Women Less Likely to Marry? Evidence from the Chinese Higher Education Expansion

Bin Huang
Nanjing University

Lei Xu
Loughborough University and IZA

Massimiliano Tani
UNSW Canberra and IZA

Yu Zhu
University of Dundee and IZA

JULY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Does College Education Make Women Less Likely to Marry? Evidence from the Chinese Higher Education Expansion*

We study the impact of higher education (HE) on marriage incidence in China using the 2017 China Household Finance Survey. Taking advantage of the dramatic HE expansion starting in 1999, we explore the effect of education on marriage outcomes by instrumenting years of schooling using the interaction of childhood urban hukou status and a set of time dummy and trend variables capturing the exposure to the expansion. Contrary to conventional wisdom, the 2SLS results suggest that increased education induced by the HE expansion leads to higher marriage rates. These positive effects tend to be larger for women living in coastal areas or larger cities. The estimates are robust to alternative specifications, age range, the age cut-offs for childhood hukou status and controls for birth cohort-city specific sex ratios. Our findings imply that the strong negative relationship observed between college education and marriage outcomes for women is likely driven by educational assortative mating due to persistent gender norms in favour of status hypergamy, which prevents the Chinese marriage market from adjusting to the reversed gender gap in HE post-expansion.

JEL Classification: I23, J12

Keywords: marriage market outcomes, 2SLS, higher education expansion, educational assortative mating, China

Corresponding author:
Yu Zhu
University of Dundee
School of Business
Dundee
DD1 4HN
United Kingdom
E-mail: yuzhu@dundee.ac.uk

* We thank the Southwestern University of Finance and Economics (SWUFE) of China for data access. Yu Zhu thanks the UNSW Visiting Professorial Fellowship for financial support. All errors remain our own. The authors declare that they have no conflict of interest.

1. Introduction

Being married to someone with the same level of education is a common feature of contemporary society (Mare, 1991; Smith and Park, 2009; Ganguli, Hausmann and Viarengo 2014), as it is shared by about 50% of all married couples (OECD 2011). It is also a topic of high interest due to its potential role in maintaining social and economic inequality within and across generations (d'Addio, 2007; Black and Devereux, 2010; Eika, Mogstad and Zafar, 2019; Bingley, Cappellari and Tatsiramos, 2022). Educational assortative mating is the non-random matching of couples based on education which may reflect differential trends in educational achievement by gender as well as the gendered returns to education that underpin individual schooling investments. Research has highlighted that returns to education differ by gender (Psacharopoulos, 1994; Patrinos, 2008). As a result, when trends emerge or educational reforms are implemented, they tend to trigger different educational investment responses in males and females. These in turn influence types and numbers of people available for marriage, and marriage rates.

The variation in marriage rates of individuals across educational levels has been widely observed, but the extent to which this reflects a causal effect of education is far from being well established due to possible issues with aggregation (Gihleb and Katz 2016) and the challenge in disentangling education measures from unobserved determinants of marriage choices such as cultural and social norms which in turn may shift patterns of educational assortative mating. Existing research has therefore relied on exogenous shocks to educational settings, such as changes in compulsory schooling laws to trace subsequent marital choices (Rauscher 2005; Holmlund 2006; Hahn, Nuzhat and Yang 2018), or applied various econometric strategies to address the endogeneity of education as a determinant of marital outcomes.

We contribute to this literature by studying the rapid expansion of higher education (HE) in China since 1999, which we use as a shock to estimate its effect on women's marriage rates. In particular, we adopt the Instrumental Variable (IV) approach developed by Huang *et al.*

(2022) to measure education to investigate if the increased college attainment by women (in both absolute and relative terms to their male counterparts) **cause** lower marriage incidence.

China presents an interesting case study for this topic for a number of reasons. Over the past two decades China has transited rapidly from elitist to mass HE, with participation rates now exceeding 50%. During this period, women have also overtaken men in university enrolment, a phenomenon known as the *reversal of the gender gap in education* in the literature. Furthermore, Chinese women have historically preferred partners with at least equal educational qualifications due to strong social and cultural norms, known as *status hypergamy* (Hu 2016). These two trends coincide with the growing phenomenon of "*leftover women*" (literally translated from the Chinese term *shengnü*) which refers to highly educated women in their late twenties or older who are still single (Hong-Fincher 2014).[1]

Our analysis focuses on women in the nationally representative China Household Finance Survey (CHFS) 2017, instrumenting their education using the interaction between *hukou* (household registration) status in childhood and a set of time dummy and trend variables capturing the exposure to the massive HE expansion which started in 1999.[2] In other words, the treatment group only consists of people with an urban *hukou* in childhood and born from 1980 onwards, who received comparatively higher quality compulsory education and thus were better able to take advantage of an unanticipated expansion in higher education starting in 1999, compared to their rural origin peers or older cohorts (both in the control group). The results suggest that education has a positive effect on marriage rates and does not cause the phenomenon of "*leftover women*", effectively ruling out the possibility of a causal effect of increasing college education on non-marriage for educated women in China. At the same time,

[1] Hong-Fincher (2014) argues that a discourse around "*leftover women*" is propagated by the Chinese state media to encourage college educated women to marry and have children sooner, as the country faces an unmarriage crisis given the persisting imbalance in sex-ratio resulting in 30 million surplus men under the age of 20.
[2] The *hukou* (household registration) system, which connects access to education, labour market and social security programmes to one's status at birth, is the key institutional feature underpinning China's urban-rural divide. Section 4.1 is devoted to a detailed explanation of background of the education reforms in the context of the *hukou* system.

the results imply that the strong negative correlation between the level of education and marriage in OLS (or Probit) models is likely driven by persistent gender norms in favour of *status hypergamy*, which prevent the marriage market from adjusting to the reversed gender gap in higher education in China. Our findings are in line with recent studies which present strong evidence that HE has increased educational assortative mating among post-expansion couples in Shanghai and enhanced horizontal educational assortative mating among female graduates from elite universities (Hu and Qian, 2016; Feng 2002).

We contribute to the evidence base on the complex relationship between education and marriage incidence in a developing country context, by combining rigorous causal inference with descriptive analysis of the role of gender norms. Conceptually, changes in educational sorting of marriages can be decomposed into an exogenous structural component, driven by the HE expansion and the reversal of the gender gap in educational achievement in the Chinese context, and an endogenous component driven by changing patterns of educational assortative mating (Leesch, Katrňák and Skopek, 2024). While the *end of hypergamy hypothesis* (Esteve *et al.* 2016; De Hauw, Grow and Van Bavel 2017; Han 2022) suggests that drastic changes in the structural component should result in increases in hypogamy (i.e. women marrying down) which has been observed in most Western societies, we find very little empirical support of this in our data, especially for the birth cohorts immediately before and after the 1980 birth cutoff for the HE expansion. In contrast, we show that weaker gendered social norms of the older generation (aged 60+) at the city level is associated with higher marriage rates among younger graduates (aged below 40) who are non-migrants. Our heterogeneous effect analysis showing the positive causal effect of HE on marriage is more pronounced in bigger cities/coastal regions where traditional gender norms are expected to be weaker, lends further support to the notion that strong educational assortative mating due to persistent norms in favour of *status hypergamy* in education acts as a barrier to the social transition in China as predicted by *the end of hypergamy hypothesis*.

Moreover, using an alternative IV strategy which exploits potential heterogenous exposure to HE expansion by birth cohort and geographical area, we find the 2SLS estimates of education on marriage remain consistent and statistically significant, despite a notable reduction in sample size.

To the extent that our IV strategy exploits the differential exposure to the massive HE expansion by *hukou* status at childhood, the estimates should be interpreted as the local average treatment effect (LATE) of HE for urban natives after netting out any common trend between urban and rural origin college-educated women including socio-norms. Urban origin females who received college education as a result of the HE expansion might increase their likelihood of marriage due to improved economic status, compared to less-educated urban origin females, when the greater gender equality in the areas they live in facilitate more rapid transition to increases in hypogamy. However, this finding might not generalise to rural origin female graduates who were exposed to stronger gendered social norms, even if they have migrated to urban areas.

The rest of the paper is organised as follows. Section 2 reviews the literature. Section 3 presents the data. Section 4 discussed the identification strategy based on the 1999 HE expansion. Section 5 presents the results. Section 6 offers further discussions, with regard to the role of gender norms and assortative mating. Finally, Section 7concludes.

2. Literature

The interest of economists on marriage rates between people of similar or different levels of education and their realisation in the 'marriage market'[3] has surged in recent times as gender patterns in HE completion has become more divergent. Over the past decades women have overtaken men in university graduation rates in many OECD countries despite receiving lower

[3] The marriage market, as originally discussed by Becker (1973), highlights that individuals choosing rationally will marry if the utility of doing so is higher than that obtained by remaining single.

wages in the labour market and working more hours doing housework relative to men, which should discourage investing in HE (Goldin, Katz and Kuziemko 2006).

Marriage of people with similar levels of education arises if the schoolings of both partners complement each other in household and/or market production.[4] Moreover, educational assortative matching may lead to better marriage returns via higher household income and improved intra-marital spousal roles.[5] However, in Becker's classic model (Becker 1991), which focuses on division of labour in joint household production, the marriage of people with different levels of education might be optimal, in line with empirical evidence of husbands that are more highly educated and older than wives.

The literature has thus developed two main competing hypotheses to explain the gender gap in education and age within marriages. The *financial support (evolutionary) hypothesis* postulates that women have a relative preference for more educated and older partners who can better provide for them and their offspring (Li *et al*., 2023). In contrast, the *social equality hypothesis* suggests that both men and women prefer partners with the same age and education to enhance marital stability (Groot and van den Brink 2002). Research has also noted that women have an incentive to invest in education if higher levels of education becomes an effective channel to reduce their labour market discrimination (Chiappori, Iyigun and Weiss 2009).

Empirical work supports the gender-specific returns to education in labour and marriage markets, and the strong link between education and marriage within educational levels (Lefgren and McIntryre 2006; Bredemeier and Juessen 2013). For example, Attanasio and Kaufmann (2017) find that expected labour and marriage market outcomes significantly affect university enrollment decisions among high school graduates, but boys give a higher relative

[4] Some of the gains can also spill-over to the next generation as in Edwards and Roff (2016), or arise from common preferences for urban centers as in Mariotti, Mumford and Pena-Boquete (2017).

[5] Tampieri (2022) shows marital satisfaction increases with spousal educational qualifications under his proposed theoretical model which allows university attendance to increase chance of marrying an educated partner and accounts for the role of relative income in utility.

weight to labour market outcomes relative to those arising from future partnering, while the opposite applies for girls. British panel data evidence suggests a trade-off between labour supply and hours supplied for household duties, and a wage penalty for the latter, which rises when the couple has children (Bryan and Sevilla-Sanz 2011). Intra-household specialisation generates a wage premium for highly educated men (Bardasi and Taylor 2008), while higher returns to schooling lead females to supply more hours of work, especially when married to high-income husbands (Bredemeier and Juessen 2013).

Conceptually, changes in educational sorting of marriages over time can be decomposed into an exogenous structural component explained by changes in demographic factors including the relative higher educational attainment by gender and an endogenous component driven by changing patterns of educational assortative mating (Leesch *et al.,* 2024). With the reversal of the gender gap in education taking place in most countries in recent decades, the dominance of the traditional hypergamy in education is expected to be weakened among the younger cohorts, known as *end of hypergamy hypothesis* (Esteve *et al*. 2016; De Hauw *et al.*, 2017; Han 2022). Using vital statistics data on all marriages from 2000 to 2020 in Sweden, the Czech Republic and Italy, Leesch *et al.,* (2024) show that rising educational attainment and the reversal of the gender gap in education are driving forces of declining hypergamy in all three countries, with variations in assortative mating accounting for most of the cross-country differences. Moreover, research focusing on marriage rates also tends to find positive relationships in developed countries between HE and marriage rates, marital stability (Isen and Stevenson 2010; Geruso and Royer 2018), and delay in the search of a suitable partner (Gould and Paserman 2003). However, college educated women in the "East Asian tiger" economies have experienced decreased rates of marriages, leading to the phenomenon of "Gold Misses".

Using a dynamic model, Hwang (2016) argues that this can be explained by the interaction of Asia's rapid wage growth with intergeneration transmission of gender norms.[6]

To assign a causal interpretation to the relationship between education and marriage choices the literature typically uses an augmented regression model where the spouse's education is added to the determinants of his/her partner's labour market outcome. Various instrumental variables have been used to address the endogeneity of the spouse's education choice, as this is not independent of unobserved individual-specific characteristics underpinning marriage and the partner's labour market performance. These include the birth quarter (Lefgren and McIntyre 2006), twins' samples (Huang *et al*. 2009), siblings (Holmlund 2006), university admission scores (Kaufmann, Messner and Solis 2013), parents' occupational status (Hu 2016), and unique features of the education system such as within-cohort variation in the length of compulsory education induced by school exit rules (Anderberg and Zhu 2014). In some cases, researchers have been able to exploit exogenous changes in education and apply difference-in-differences or regression discontinuity designs to study ensuing changes in marriage decisions (Holmlund 2006; Kırdar, Dayıoğlu and Koç 2018).

With reference to China, there is an emerging literature that provides evidence of higher marriage rates between partners with similar levels of education. Using census and survey data since 1982, Dong and Xie (2023) show a strong growing trend in educational assortative mating as measured by couple rank-rank correlations for people born after 1966. Using the 2002 Chinese Twins Survey data, Huang *et al*. (2009) find that more schooling by the husband significantly increases wife's earnings but not vice-versa. Qian and Qian (2014) note that higher levels of education lead to higher marriage rates, but only for men. In the case of highly educated women, it is more common to remain single after 30, implying some form of gender-biased trade-off between education and age in the marriage market. Hu and Qian (2016) focus

[6] Using the Asian Barometer Survey data, Chang (2018) also shows that strong Asian values constrain public support for social welfare spending, due to the traditional beliefs of self-reliance and filial duty prevalent in East and Southeast Asian societies.

on Shanghai's case after China's rapid expansion of HE, noting higher levels of couples with similar education by birth cohort as the number of highly educated individuals increased. Hu (2016) finds a strong association between the occupational status of an individual's father and of his or her spouse and the marriage choices of husband and wife. Parental *hukou* status is also found to play a pivotal role in marital outcomes, in that an individual's father and father-in-law tend to have the same rural or urban *hukou*. You, Yi and Chen (2021) highlight the role of personalities and persistent patrilocal social norms in explaining the *left-over women* phenomenon in China, while acknowledging the reduced-form regression should only be considered as suggestive evidence. Using the 2010 Chinese Family Panel Study (CFPS), Feng (2022) finds that female graduates of elite universities are increasingly more likely to marry male elite university graduates, after the HE expansion, relative to female graduates of lower-ranked universities. He interprets this as evidence consistent with the *social closure* (i.e. elites marrying within status group to maintain privilege or *culture matching* hypotheses.

While our paper is similar to Hu and Qian (2016) and Feng (2022) in exploring the HE expansion, there are substantive differences among several important dimensions. Firstly, the three studies differ in aims and objectives. Whereas Hu and Qian (2016) and Feng (2022) focus exclusively on the effect of HE expansion on marital assortative mating (i.e. conditional on marriage) from a quantitative sociological perspective, we provide direct <u>causal</u> evidence of the effect of HE induced by a college expansion on marital choices. The second difference is in the methodological approach. Both Hu and Qian (2016) and Feng (2022) use the log-multiplicative layer effect model to identify the changes in partner preference and availability in the marginal distribution of a contingency table (capturing partners' educational attainment) across cohorts. In contrast, we follow an Instrumental Variable (IV) approach which allows a causal interpretation of the estimated effect of education on outcomes, first proposed by Huang *et al.* (2022), by taking advantage of the exogenous variation in educational attainment induced by the 1999 HE expansion. Lastly, the findings of Hu and Qian (2016) may not be generalizable to the population at large, as they are based on a sample of 2357 individuals born in the 1980s in Shanghai and surveyed in 2013. In our case, the use of a nationally representative survey with nearly 20 thousand individuals in the analytical sample guarantees both external validity and statistical power.

3. Data

We use the 2017 China Household Finance Survey (CHFS), conducted by the Survey and Research Centre for China Household Finance at the Southwestern University of Finance and Economics (SWUFE) of China. CHFS is a nationally representative household survey of income, expenses, assets, liabilities, insurance and securities, of more than 40,000 households.

Using a stratified three-stage probability proportional to size (PPS) random sample design, CHFS covers the whole mainland China except Tibet, Xinjiang and Inner Mongolia. Counties, county-level cities or city districts (of prefectural-level cities) constitute the primary sampling units (PSU) in the first-stage. The second stage of sampling select villages in rural areas and residential committees in urban areas within each selected PSU. Lastly, in the third-stage, households are chosen from the selected villages and residential committees, respectively. Every stage of samplings is carried out with PPS method and weighted by its population size.

Our main sample consists of all women in the 2017 CHFS who are aged 23 and above, and born in 1970 or later. Age 23 is chosen as the minimum age, not only to allow time to complete university education which normally takes 4 years from age 18 or 19, but also to reflect China's high legal minimum age for marriage of 20 for women (and 22 for men).[7] Our final sample after excluding observations with missing values on key variables consists of 19,581 women, all of whom born between 1970 and 1994, and aged 23 to 47 at the time of the 2017 survey.

[7] Huang *et al.* (2022) show that pre-1970 birth cohorts suffered from educational disruptions due to the "Cultural Revolution (1966-1976)" and were not exposed to the 9-year compulsory education introduced in 1986.

Table 1 shows the self-reported marital status by education levels. Overall, 13% and 84% of women in our sample are single or married (for the first time) respectively. It also appears that more educated women are less likely to be currently married, although this relationship is non-monotonic. However, this is just a correlation rather than causation. Compared to developed countries, the share of cohabiting individuals is surprisingly low, reflecting the fact that China is still a traditional society with strong cultural and social norms. Divorcees account for roughly 2% of both genders, and separation is even rarer.

Using this sample, we can study the effect of education on two key marriage outcomes, namely:

1) being currently married, which includes being remarried;

2) have ever been married, which includes all those who report being married, separated, divorced, widowed and remarried.

Table 2 presents summary statistics by 5-year birth cohort bands first before formally testing the equality of sample means between pre and post HE expansion cohorts, defined by whether born before 1980. While there appears to be a declining trend in marriage rates (including remarriage) and ever married, this is to be expected given that both are likely an increasing function in age. Importantly, there is very little difference between the two adjacent cohorts at either side of the HE expansion cut-off. In contrast, there is a clear surge in years of schooling between pre- and post-expansion cohorts. It is also worth mentioning that *hukou* status at either birth or age 12, does not vary much over time, including around the cut-off.

Conditional on currently being married, we are also interested in the effect of education on key indicators of quality of the match. Earlier literature (e.g. Groot and van den Brink 2002) have highlighted spousal gaps in education and age. For this purpose, we will use a subsample of the main sample, consisting of couples only.

For the 12,398 couples we identified from the main sample, Appendix Table A1 presents evidence on educational assortative mating, by cross-tabulating the wife's education level against that of the husband. Note that we restrict husband's age to the range 20-60, to mitigate potential censoring at either end of wives' age distribution. Due to concerns for small cell sizes, we also grouped academic and vocational high schools together into a single High School category, and all postgraduates into Masters+.

While the traditional gender norms in China are consistent with educational hypergamy and homogamy (i.e. wives have lower or same level of education as their husbands respectively), this is not entirely borne out by the data. The strongest evidence of educational homogamy as highlighted by the diagonal cells is observed at the levels of Middle (lower secondary) School, Bachelors and Masters+ Degrees for the wife, with 63%, 61% and 59% reporting husbands having the same level of education, respectively. For women with no more than lower secondary education, which has become compulsory in principle for all sample members (most of the non-compliance occurs at the primary school level possibly due to early dropouts), there is strong evidence of educational hypergamy. For wives holding a High School qualification or Vocational College, there is a 1.2 odds ratio that the husband has lower education (hypogamy). The odds ratio of hypogamy to hypergamy increases to almost 4 for wives with a Bachelor's degree. For wives with postgraduate qualifications (i.e. Masters or doctoral degrees), more than 40% report educational hypogamy, most of which are College/Bachelor's Degrees.

For the matched couples subsample, Appendix Table A2 shows key marriage characteristics in terms of the differences in years of schooling and age, by the wife's birth year bands. Note that the massive HE expansion started in 1999, which corresponds to being born in 1980. As we move across the columns, from older to younger birth cohorts, there is a clear downward trend in the spousal educational gap in years of schooling, from 0.93 years to 0.04 years, whereas women's average years of schooling increased by around 2.4 years. In the meantime, the proportion of educational homogamy couples had remained almost constant, at around 50%, which on its own would suggest little change in educational homogamy over time. However, this finding masked a significant 11 percentage points decrease in the proportion of hypergamy couples, which is almost fully offset by a corresponding increase in the share of hypogamy couples, for cohorts born between 1990-94 compared to their counterparts born two decades earlier. This pattern would be consistent with the *end of hypergamy hypothesis* induced by the reversal of the gender gap in education (Esteve *et al.* 2016; De Hauw *et al*. 2017; Han 2022). Nevertheless, the pace of transition is very slow, with hypergamy reducing by merely 1.9 percentage points for cohorts born between 1980-1984 who were exposed to the HE expansion starting in 1999, relative to cohorts born between 1975-1979 who narrowly missed out on the opportunity.

In contrast, there is a significant increase in the spousal age gap, from 1.8 years for the oldest cohorts to 3.0 years for the youngest cohorts. However, this pattern might be spurious due to censoring on both ends of the age distribution. Focusing on the cohorts groups immediately before and after the HE expansion, the change in the spousal age gap becomes much less drastic. While there is a 0.2 year increase in the age gap between the 1975-79 and the 1980-84 cohorts, the proportions of couples where the husband is of the same age as, older than, or younger than the wife remain virtually unchanged.

4. Identification Strategy based on the 1999 HE Expansion

4.1. Background information on the Chinese education system and the hukou system

Following the founding of the People's Republic of China in 1949, Soviet style central planning was adopted in the development of a new state-run education system, geared towards the needs of rapid industrialization. However, the education system was severely disrupted during the Cultural Revolution (1966-1976), when ideology was emphasized at the expense of professional competence. It was not until 1978 that the university entrance exam was reintroduced, when Deng Xiaoping initiated China's reform and opening-up.

After the introduction of the Law on Nine-Year Compulsory Education in 1986, the Middle (aka. Junior High) School Exit Examinations, known as *zhongkao*, are used to stream students into either the academically or vocationally oriented Senior High schools, both lasting 3 years

(OECD 2016). After obtaining Senior High School qualifications, one can apply for entry into vocational colleges or universities, which would normally take 2-3 and 4+ years to complete, respectively. However, admissions into HE was highly competitive, at least before the massive HE expansion starting in 1999. Moreover, HE in China was free of tuition fees up to the early 1990s, when modest tuition fees were introduced.

Before the massive expansion, the HE sector was tightly controlled by the Ministry of Education, which sets provincial, university and subject quotas annually (OECD 2016). HE enrolment has always been administered by a centralised admissions system which proceeds sequentially in tiers on the basis of one's performance in standardized National College Entrance Examinations (*gaokao*), with little regard for gender, *hukou* status and family background. For instance, college enrolment only increased by an average of 4.7% per annum between 1995 and 1998 (Che and Zhang 2018).

An important institutional feature of China is the *hukou* (household registration) system, which determines at birth one's status as either rural or urban, usually according to the mother's *hukou* status. The *hukou* system originated from the 1950s, as an instrument for social control and to prevent rural-to-urban migration. Education resources at primary and secondary level are unequally distributed in China, with a strong bias in favour of urban residents. For instance, despite the significant improvements in recent years, the senior high school public expenditure per student in rural areas remains 25.2% lower below the national average in 2017 (National Bureau of Statistics 2017).

As a result, urban *hukou* holders, especially those living in the major cities, enjoy much better access to HE in general. For instance, Qin and Buchanan (2019) show that while 5.6% of *gaokao* entrants from the predominantly urban Shanghai in 2016 entered the prestigious "Project 985 Universities", a group of 39 comprehensive universities selected by the central government for research intensity and excellence, only 1.2% of their counterparts made it in the same year from Henan province, a province with less than half the urbanization rate as Shanghai.[8]

In this paper, we derive the *hukou* status at age 12 for all sample members, who are all subject to the 9-year compulsory education regime by construction. Intuitively, childhood *hukou* status determines access to urban or rural secondary schools, which vary systematically in quality.[9]

4.2 The 1999 HE Expansion

The year 1999 marked the beginning of a decade of massive HE expansion in China, totally unprecedented among major economies in terms of both the scale and the speed. Between 1998 and 2008, annual HE enrolment in China grew from 1.08 million to 6.08 million.

The HE expansion came as a totally unanticipated policy shock, with no public consultation, and HE institutions around the country were only given a few months to prepare for the surge in intake (Wan 2006; Wu and Zhao 2010; Li *et al*. 2017). In response to the rising youth unemployment in the aftermath of the 1997 Asian Financial Crisis, the Ministry of Education suddenly announced in spring 1999 a 47% increase in college and university intake for the September entry. This was followed by increases of 38% and 22% for the following two academic years, and subsequent more modest double-digit growth year on year on average for the rest of the decade (Che and Zhang 2018). The expansion was made possible by significant supply-side growth, in the hiring of new staff and not least in the construction of over 60 new "university towns", i.e. suburban districts of several adjacent university campuses, in China's HE conglomerates by 2006 (Rouppila and Zhao 2017). This phenomenal growth was only eased off after the Global Financial Crisis in 2008, as the graduate labour market became increasingly challenging.

[8] Using the China Family Panel Studies, Kang, Peng and Zhu (2021) show that annual returns to more selective key universities are significantly higher than ordinary universities or vocational colleges regardless of subjects studied.

[9] Current *hukou* status suffers from endogeneity due to possible *hukou* status change through marriage, acquirement of properties in cities, and obtaining HE qualifications (Xing 2013).

Figure 1 shows the years of schooling by gender and birth cohort using the CHFS. Similar to many countries, China experienced a reversal of the gender education gap in recent decades. While men born in 1970 have one more year of education than their female counterparts, the gender gap turns into minus 0.6 years for people born in 1994. The turning point is around 1986, which corresponds to the 2004-5 university entry cohort. This implies that women benefited disproportionately from the HE expansion. Whereas male graduates outnumber their female counterparts before the HE expansion, this was reversed approximately 5 years after the expansion began. Interestingly, the trends for both genders appear to be quite smooth, with no apparent jumps at the threshold of the HE expansion.[10]

4.3 Instrumental Variable Strategy

It is well known that measures of education including years of schooling suffer from endogeneity problems, due to self-selection, ability bias and measurement errors. All three sources of endogeneity apply here. First, HE is certainly a matter of individual choice. Second, in CHFS we do not have any ability measure such as scores in standardized exams or test, or good measures of the quality of the education qualifications. Third, the years of schooling variable is imputed from people's self-reported highest qualification obtained or attempted.

To overcome endogeneity in years of schooling, we explore a large and unanticipated HE expansion which started in 1999 as a source of exogenous variation in the educational attainment. Moreover, given the systematic difference in school resources and quality between urban and rural areas in China, it is important to allow the impact of HE expansion to vary according to people's original *hukou* status.

Figure 2 shows the average years of schooling by *hukou* status at age 12 and birth cohort for our sample members. It is striking that for pre-1980 birth cohorts rural and urban female

[10] Appendix Figure A1 showing the HE attainment by gender and birth cohort displays similar patterns, with the reversal of the gender gap in Degree+ attainment taking place about two years sooner than for vocational colleges.

students share the same time trend in years of schooling, despite a staggering nearly 4-year gap in favour of urban students. Moreover, there is only a visible jump for urban students in years of schooling around the HE expansion cut-off, of about 0.5 years. For rural students, there appears to be no discontinuity in either the intercept or slope.[11]

These patterns motivate our choice of identification strategy. As Figure 2 suggests that the impact of the HE expansion differs for rural and urban *hukou* holders, we need to interact an urban *hukou* at age 12 dummy with a set of time dummy and trend variables capturing the exposure to the HE expansion. Specifically, in order to parameterize the potential change in both the intercept and the slope induced by the HE expansion, one needs to include 3 HE expansion main effect variables in the regressions for individual *i* in city *j*:

1) *Birth year trend (T_{ij})*: a linear time trend before the expansion as the baseline;
2) *Post-1980 birth (D_{ij})*: a dummy for being born in 1980 or later to capture the instantaneous effect of the HE expansion (on the intercept);
3) *Post-1980 birth year trend (TD_{ij})*: an interaction term between the linear *birth year trend* and the *post-1980 birth* dummy to capture the change in the time trend from the pre-expansion baseline.

In our Two Stage Least Squares (2SLS) setting, we include an urban *hukou* at age 12 dummy, U_{ij}, and the main effects of the 3 variables in the second-stage equations. Therefore, the identification of the causal effect of education on marriage market outcomes relies only on the interactions of the urban *hukou* at age 12 dummy, with these main effects. In other words, we assume that the interaction effects of the *hukou* system and the HE expansion have no direct effect on marriage market outcomes over and above their impact through the education attainment. Compared to an instrumental variable strategy which relies ***only*** on the main effects

[11] Appendix Figure A2 focuses on HE attainment by childhood *hukou* status and birth cohort. Consistent with Figure 2, it indicates that there is only a significant jump for urban women of around 8 percentage points, in Degree+ attainment around the HE expansion cut-off.

of HE expansion variables and the urban *hukou* dummy, our identification strategy is much less restrictive. Because our model is over-identified, we will also be able to formally test the exogeneity of our instruments using the over-identification tests. Moreover, if any of the main HE expansion effects above is found to be statistically significant, that will lend further support to our identification strategy which relies only on the interaction terms.

Formally, the 2SLS is a two-equation system defined as follows. The first-stage involves estimating an OLS equation of years of schooling for individual i in city j on exogenous controls X_{ij}, $U_{ij}, Tij, D_{ij,}$ and TD_{ij}, as well as the instrumental variables comprising **only** the interaction terms $U_{ij}T_{ij,}$ $U_{ij}D_{ij,}$ and $U_{ij}TD_{ij}$, which are excluded from the second-stage equation:

$$S_{ij} = \alpha_1 + \gamma_1 X_{ij} + \delta_{10}U_{ij} + \delta_{11}T_{ij} + \delta_{12}D_{ij} + \delta_{13}(TD_{ij}) + \left[\pi_1\left(U_{ij}T_{ij}\right) + \pi_2\left(U_{ij}D_{ij}\right) + \pi_3\left(U_{ij}TD_{ij}\right)\right] + \theta_1^j C_j + \lambda_{11}^p H_{ip} + \lambda_{12}^p U_{ip}^b + \lambda_{13}^p (H_{iP}U_{ip}^b) + \varepsilon_{ij} \qquad (1)$$

Note that we have allowed for fixed-effects of the current (prefectural) city of residence C_j, as well as full interaction of birth province H_{ip} and the *hukou* status at birth U_{ip}^b.

The second-stage involves regressing marriage market outcomes M_{ij} for the same individual i (in city j) on one's own years of schooling and the same set of control variables. The difference between 2SLS and OLS is that in the former we simply replace the observed S_{ij} with the fitted value estimated from equation (1):

$$M_{ij} = \alpha_2 + \beta\,\widehat{S_{ij}} + \gamma_2 X_{ij} + \delta_{20}U_{ij} + \delta_{21}T_{ij} + \delta_{22}D_{ij} + \delta_{23}(TD_{ij}) + \theta_2^j C_j + \lambda_{21}^p H_{iP} + \lambda_{22}^p U_{ip}^b + \lambda_{23}^p (H_{iP}U_{ip}^b) + e_{ij} \qquad (2)$$

For both equations, we control for city-level fixed-effects, C_j, which capture any time-invariant unobservable features of the local marriage market. This is important, given the huge variation in demographic structure such as sex-ratios, and the level of economic development across different geographic areas in China.

Technically, our IV estimator is analogous to a difference-in-differences (DID) estimator that estimates the effect of own education net of an HE expansion on marriage market effect (common for both rural and urban *hukou* holders). The fact that rural *hukou* holders have the same pre-treatment time trend in years of schooling with their urban counterparts as indicated in Figure 2, ensures that they serve as an ideal control group for the urban *hukou* holders.

When the marriage market outcome is binary, e.g. an indicator for being currently married, we will estimate an IV Probit model instead of 2SLS. However, both specifications use the same linear reduced form for the endogenous explanatory variable. Therefore, the diagnostic tests for IV relevance and exogeneity could be based on the 2SLS specification (Wooldridge 2010, Chap 15.7.5).

5. Results

5.1 Effect of education on marriage and ever married

Table 3 presents the Linear Probability Model (LPM) and 2SLS estimates for being currently married (including remarried), for women. Note that we control for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province and *hukou* status at birth in all specifications. The former might capture any time-invariant city characteristics which might affect marriage market outcomes, such as the tendency for later marriages in big cities. The latter will pick up the family planning policy regimes which varies by province and *hukou* status in China. It is well documented that in the 1980s each province enacted its own family-planning regulations which also tended to vary by *hukou* status, thus resulting in substantial variations in fertility rates across the country (e.g. Short and Zhai 1998; Attane 2002).

The LPM marginal effects of women's education on both being married and ever married are virtually identical, implying that each year's increase in education reduces the probability of being married or ever married by 1.1 percentage points.

However, this is overturned once we account for endogeneity of education using 2SLS. Across all 2SLS specifications, the interactions of childhood *hukou* status with the post-1980 birth and with the post-1980 birth year trends are both significant at the 1% level individually. Importantly, the joint significance tests show F-statistics of over 40, well above the threshold of 10 for IV relevance. This suggests that we do not have issues with weak instruments. The endogeneity tests overwhelmingly reject the null of exogeneity of years of schoolings, thus justify the use of instrumental variables to correct for endogeneity. Lastly, for currently being married or remarried, Sargan's over-identification tests indicate that we cannot reject at even the 45% significance level the null of exogeneity of our all instruments provided as least one of instruments is exogenous. However, the over-identification test is more borderline for the outcome of ever been married, with significance at 0.051.

It is worth noting that the instruments are highly significant in the first stage. Consistent with Figure 2, we find that urban women increased their years of schooling by almost 2 years as a result of the HE expansion, although there is evidence of a slow catching up by rural women after the expansion. This was despite a 1.9 years education gap in favour of urban women in the first place.

Appendix Table A3 shows the corresponding Probit and IV-Probit estimates for being currently married/remarried or ever been married, which are qualitatively similar to the corresponding LPM and 2SLS IV-Probit estimates in Tables 3. Probit results suggest that education is negatively associated with both currently married and ever married for women. In both cases, the coefficients for years of schooling are statistically significant at the 1% level. However, once we instrument years of schooling, education turns out to have a positive and statistically significant effect on marriage for women. Note that the negative correlation coefficients between the residuals of the first-stage (years of schooling) equation with those of the marriage outcome equations are in excess of 0.6 in absolute value in both equations, indicating women who tend to have higher education attainments are less prone to be married or ever married, due to unobservable personal attributes. Indeed, the Wald tests of exogeneity of years of schooling are overwhelmingly rejected. Therefore, the negative correlation between education and marriage for women is spurious and there is no evidence to support the notion that HE *causes* the “left-over ladies” problem. In contrast, the evidence points to a positive causal effect of education on marriage for women, holding constant all factors including individual attributes which might often be (partially) unobservable to researchers.

5.2 Heterogeneous effect of education by geographical region and type of residence area

In Table 4, we explore the heterogeneous effect of education on the probability of being currently married, by geographical region and type of residence area. The positive effect of years of schooling is larger for more economically developed coastal region which also includes the Northeast, than for the inland provinces. Similarly, the positive effect of years of schooling is also larger in major cities, comprising the 4 metropolises and all provincial capitals, relative to small cities, towns and rural areas. Note also that for each subgroup, all the diagnostic tests for the 2SLS are satisfied.

It is well established that in developing countries the gender gap in urban areas, in a wide range of socio-economic indicators including education and labour outcomes, are in general smaller than in rural areas (e.g. Chant 2013). One could reasonably expect that the traditional gender norms become weaker with urbanization, especially in metropolises with over-concentration of highly educated young migrants.

5.3 Robustness w.r.t. different model specifications and age cut-offs

Next, we check the robustness of the main specification with respect to different model specifications and age cut-offs.

The first 3 columns of Table 5 show that controlling for city fixed effects, and the full interaction of birth *hukou* province and birth *hukou* status actually makes little difference to the 2SLS estimates of education on currently married. All three specifications also pass the diagnostic tests for 2SLS. However, our main specification in Table 3 which corresponds to columns 3 is still preferred due to the improved precision of estimates and the more favourable F-statistic for IV relevance.

One might be concerned that the critical birth cohort cut-off for HE expansion might coincides with the implementation of the one-child policy starting from 1978-1980, which leads to large increases in the sex ratio over time. There is ample literature to show that the sex ratio likely influences the behaviour of both men and women on the marriage market (e.g. Ong, Yang and Zhang 2020). Column 4 shows the robustness of the preferred specification with additional controls for the sex ratio, by the relevant 5-year birth city-cohort derived from the 2000 Census tables. Column 5 only includes the sex ratio as an additional IV. While sex ratio appears to be significant for women's years of schooling suggesting that skewed sex ratio in favour of boys reflecting strong son preferences is associated with more education for girls all else being equal, it makes virtually no difference to women's incidence of marriage.

One might also be concerned with the sensitivity of the results with respect to the age cut-off. So, we re-estimate the LPM and 2SLS models using only women aged 28 and above, or women aged 20 and above. The higher age cutoff at 28 implies dropping the youngest birth cohorts who were born in 1990 or later, resulting in a more balanced sample with 10 years both before and after the HE expansion cut-off. On the other hand, the lower age cutoff at 20 corresponds to the minimum legal age of marriage for women in China. The marginal effect of (one extra year of) education in Appendix Table A4 is reduced from 0.129 in Table 3 to 0.077 for age 28 cutoff or to 0.072 for age 20 cutoff, but remain statistically significant at the 1% level in both cases. Moreover, the 2SLS models still pass all diagnostic tests. The lower magnitudes of the estimated effects might be partly explained by the postponement of marriage for university graduates, who would not complete full-time education until age 22 at least.

5.4 Robustness w.r.t. critical age for childhood hukou status in IV

Finally, we test the robustness of the IV results with regard to childhood urban *hukou* status at ages other than 12, which roughly corresponds the age at the end of primary education. Table 6 present IV Probit estimates using the similar identification strategies but based on the interaction of HE expansion timing with *hukou* status at birth, age 6 and age 15 respectively. The last two age cut-offs correspond to the normal age starting formal schooling and the age completing the compulsory 9-year education in China.

Compared to Table 3, Table 6 shows that the strong positive causal effects of education on either being married or ever married are insensitive to alternative critical age at which the childhood *hukou* status in measured.

5.5 Alternative IV strategy allowing for heterogenous exposure to higher education expansion by birth cohort

The validity of our 2SLS strategy relies on the assumption that the interaction effects of childhood *hukou* status and the HE expansion have no direct effect on marriage market outcomes over and above their impact through the education attainment. One might be concerned that a lot is happening in China around 1999 and essentially labelling all Chinese millennials as "treated" by the HE expansion might be too strong an assumption, even after we control for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province and rural birth (by *hukou* status).

In Tables A5 in the Appendix, we further relax the identifying assumption by allowing for heterogenous exposure to higher education expansion by region and birth cohort, in the spirit of the identification strategy used in Huang, Pan & Zhou (2023). Following Li *et al.* (2017) and Dai *et al*. (2022), we construct an IV for HE expansion based on the year and province of residence when the individuals are in high school to measure the temporal and geographical variations in the scale their exposure to HE expansion.[12] In order to compare with our main results, we interact the newly constructed instruments with urban *hukou* at age 12, which

[12] $ln(Expansion_{p,t}) = ln(TotalExpansion_t \times EnrolmentShare_{p,base})$, where $TotalExpansion_t$ represents the national HE enrolment for year t, exceeding the average national enrolment during the baseline period prior to the expansion, and $EnrolmentShare_{p,base}$ denotes the average provincial share of enrolment during this baseline period. In the context of central planning, both the provincial enrolment shares before expansion and the annual national enrolment expansion rates are considered exogenous. The natural logarithm of the sum of potential college enrolment expansion in province p in year t, $ln(Expansion_{p,t})$, can also be considered exogenous. The numbers of HE expansion are from the Education Statistical Yearbooks.

captures the birth-cohort specific exposure to the expansion between urban and rural students by province in a single cross-sectional data like ours. Specifically, the negative interaction terms for the most recent year dummies suggest diminishing advantages for urban *hukou* students compared to their rural peers due to a catching-up effect for the latter group. While slightly smaller in magnitude, the 2SLS estimates in the new Tables A5 are still statistically significant and consistent with the preferred specifications reported in Tables 3 and 4. These lend strong support to the credibility of our main 2SLS results.[13]

Table A6 presents the corresponding 2SLS estimates with alternative critical age for childhood *huko*u status. It is reassuring that all estimates of education on marriage or ever married remain positive and statistically significant at the 1% level.

6. Further discussions with regard to the role of gender norms and assortative mating

Our causal estimates in Section 5 indicate that HE *per se* has positive effects on marriage or ever married, and hence does not contribute to observed negative correlation between HE and marriage. Conceptually, changes in marriage rates can be decomposed into exogenous changes in structural opportunities which shift the educational composition of potential partners and changes in assortative mating by education. While *the end of hypergamy hypothesis* suggests that changes in the structural component should result in increases in educational hypogamy (women marrying down), there is only very weak support for this in our data (see Table A2), especially for the 10 cohorts immediately before and after the 1980 birth cutoff for the HE expansion.

The fact that there is little decrease in hypergamy/homogamy despite the very substantial structural change implies that increased educational assortative mating (due to persistent norms) acts as a barrier to the social transition predicted by *the end of hypergamy hypothesis*. This

[13] We acknowledge that the first stage results are a bit weak, due to the requirement of availability of childhood *hukou* status resulting in the HE enrolment cutoff at 2008 (and consequently birth cohort cutoff at 1990 instead of 1994). This reduces the effective sample size by almost 20%, from 19,581 to 15,718.

interpretation is in line with the evidence by Hu & Qian (2016) showing HE promote educational homogamy in Shanghai among post-HE expansion couples, and by Feng (2022) showing increased horizontal educational assortative mating using the 2010 CFPS, in that female graduates from elite institutions are increasingly marrying elite male graduates, although the converse is not true.

Moreover, our own heterogeneous effect analysis showing the positive causal effect of HE on marriage is more pronounced in bigger cities and in coastal regions where traditional gender norms are expected to be weaker, lends further support the notion that China (and developed East Asia) still has a long way to go to reach the new marriage market equilibrium as predicted by *the end of hypergamy hypothesis*. In societies with strong traditional gender roles, highly educated women are not valued in the marriage market (Qian & Qian, 2014). Moreover, many women would rather remain single than marry down (Edin and Kefalas 2011; Lichter *et al.* 1995).

Finally, we present further descriptive analysis using the rich information on gender norms in the China General Social Survey (CGSS) for the waves 2010-2017 to show the relationship between gender inequality and marriage rates at the province level. Gender inequality is measured by the principal component derived from five questions on gender norms, with higher values indicating stronger agreements with on traditional gender norms in a province. To mitigate concerns for endogeneity (reverse causation), we measure agreements with traditional gendered social norms among the people aged 60 and above, and marriage outcomes using the subsample of people aged below 40. We also distinguish between natives and immigrants, and those with and without college degrees.

Figure A3 in the Appendix shows clear evidence that regions with lower gender inequality have moderately higher marriage rates for graduates, but significantly lower marriage rates for non-graduates, among natives. On the other hand, regions with lower gender inequality tend to have lower marriage rates for immigrants regardless of education level.

Taken together, this suggests that gender norms play a significant role in explaining the regional variation in marriage rates between graduates and non-graduates. In conservative regions with stronger gender inequality, female university graduates have much lower marriage rates compared to female non-university graduates.

7. Concluding remarks

Using the 2017 China Household Finance Survey, we study the effect of HE on marriage outcomes, in particular the probability of being married or have been ever married. To overcome the endogeneity of education, we exploit China's dramatic HE expansion over the decade from 1999. Specifically, we instrument years of schooling using the interaction of childhood urban *hukou* status and a set of time dummy and trend variables capturing the exposure to the expansion. This identification strategy is analogous to a difference-in-differences estimator using rural students as a control for any common time trend.

The apparent strong negative relationship between education and marriage outcomes is completely overturned once we instrument education attainment using the policy-induced HE expansion. This implies the existence of strong unobserved cultural and social norms or individual preferences driving the spurious relationship. For instance, strong preference for education hypergamy is likely to result in college-educated women having increasing difficulties of matching with better or at least equally educated men, especially when females account for an ever-growing share of graduates. Moreover, career-oriented women are not only more likely to have higher educational attainment, but also have strong incentives to delay (rather than forego) marriage in a highly competitive labour market with very weak protection and support for women (Hwang 2016; Wang and Klugman 2020).

Our causal estimates show that HE *per se* has positive effects on marriage or ever married, and hence does not contribute to the phenomenon of "*leftover women*" that worries some segments of China's society. To the extent that the negative relationship between education and marriage outcomes is driven by cultural and social norms or individual preferences, there is no easy quick fix to the problem, at least in the short run. However, a promising policy approach could start by improving employment protection, maternity leave and childcare subsidies.

References

Anderberg, D., and Y. Zhu. 2014. "What a difference a term Makes: The effect of educational attainment on marital outcomes in the UK." *Journal of Population Economics 27*, 387-419.

Attanasio, O., and K.M. Kaufmann. 2017. "Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations." *Journal of Economic Behavior and Organization* 140, 35-55.

Attane, I. 2002. "China's Family Planning Policy: An overview of its past and future." *Studies in Family Planning* 33(1), 103-113.

Bardasi, E., and M. Taylor. 2008. "Marriage and wages: a test of the specialization hypothesis." *Economica* 75, 569–591.

Becker, G.S. 1973. "A theory of marriage: Part 1." *Journal of Political Economy* 81(4), 813-846.

Becker, G.S. 1991. *A Treatise on the Family*. 2nd ed. Cambridge, MA: Harvard University Press.

Bingley, P., Cappellari, L. and K. Tatsiramos. 2022. "Parental assortative mating and the intergenerational transmission of human capital", *Labour Economics* 77, Article 102047.

Black, S. E., and P. J. Devereux. 2010. "Recent developments in intergenerational mobility." *NBER Working Paper* 15889.

Bredemeier, C., and F. Juessen. 2013. "Assortative mating and female labor supply." *Journal of Labor Economics* 31(3), 603-631.

Bryan, M. L., and A. Sevilla-Sanz. 2011. "Does housework lower wages? Evidence for Britain." *Oxford Economic Papers* 63(1), 187-210.

Chang, A. C-h. 2018. "How do Asian values constrain public support for redistribution?", *Journal of Behavioral and Experimental Economics* 77, 139-150.

Chant, S. 2013. "Cities through a 'gender lens': A golden 'urban age' for women in the global South." *Environment & Urbanization* 25 (1), 9-29.

Che, Y., and L. Zhang. 2018. "Human capital, technology adoption and firm performance: Impacts of China's higher education expansion in the late 1990s." *Economic Journal* 128, 2282-2320.

Chiappori, P. A., M. Iyigun, and Y. Weiss. 2009. “Investment in schooling and the marriage market.” *American Economic Review 99*(5), 1689-1713.

d'Addio, A.C. 2007. “Intergenerational transmission of disadvantage: Mobility or Immobility across Generations?” Paris: *OECD Social, Employment and Migration Working Papers* No. 52.

Dai, F., Xu, L. and Zhu, Y., 2022. Higher education expansion and supply of teachers in China. *China Economic Review*, *71*, p.101732.

De Hauw, Y., Grow, A. and J. Van Bavel. 2017. “The reversed gender gap in education and assortative mating in Europe”, *European Journal of Population* 33(4), 445-474.

Dong, H. and Y. Xie. 2023. “Trends in Educational Assortative Marriage in China Over the Past Century.” *Demography* 60 (1): 123–145.

Edwards, R.D., and J. Roff. 2016. “What mom and dad’s match means for junior: Marital sorting and child outcomes.” *Labour Economics* 40, 43-56.

Eika, L., M. Mogstad, and B. Zafar. 2019. “Educational assortative mating and household income inequality.” *Journal of Political Economy 127*(6), 2795-2835.

Esteve, A., Schwartz, C. R., Van Bavel, J., Permanyer, I., Klesment, M. and J. Garcia, 2016. “The end of hypergamy: global trends and implications”, *Population and Development Review* 42(4): 615–625.

Feng, A.J. 2022. “Revisiting horizontal stratification in Higher Education: College prestige hierarchy and educational assortative mating in China”. *Demography* 59(1), 349–369.

Ganguli, I., R. Hausmann, and M. Viarengo. 2014. “Marriage, education and assortative mating in Latin America.” *Applied Economics Letters*,*21*(12), 806-811.

Geruso, M., and H. Royer. 2018. “The impact of education on family formation: Quasi-experimental evidence from the UK.” *NBER Working Paper* No. 24332.

Gihleb, R., and K. Lang. 2016. “Educational homogamy and assortative mating have not increased.” *NBER Working Paper* No. 22927.

Goldin, C., L. Katz, and I. Kuziemko. 2006. “The homecoming of American college women: A reversal of the college gender gap.” *NBER Working Paper* No. 12139.

Gould, E. D., and M. D. Paserman. 2003. “Waiting for Mr. Right: Rising inequality and declining marriage rates.” *Journal of Urban Economics 53*(2), 257-281.

Groot, W. and H.M. van den Brink. 2002. “Age and education differences in marriages and their effects on life satisfaction.” *Journal of Happiness Studies* 3, 153-165.

Hahn, Y., K. Nuzhat, and H.S. Yang. 2018. “The effect of female education on marital matches and child health in Bangladesh.” *Journal of Population Economics 31*(3), 915-936.

Han, S. W. 2022. “Is it only a numbers game? A macro-level study of educational hypogamy’, *Demography* 59(4), 1571–1593.

Holmlund, H. 2016. “Education and equality of opportunity: what have we learned from educational reforms?” *IFAU-Institute for Evaluation of Labour Market and Education Policy Working Paper* No. 2016: 5

Hong-Fincher, L. 2014. *Leftover Women: The Resurgence of Gender Inequality in China.* Zed Books.

Hu, A. and Z. Qian. 2016. “Does higher education expansion promote educational homogamy? Evidence from married couples of the post-80s generation in Shanghai, China.” *Social Science Research* 60, 148-162.

Hu, Y. 2016. “Marriage of matching doors: Marital sorting on parental background in China.” *Demographic Research* 35, 557-580.

Huang, B., Tani, M., Wei, Y. and Y. Zhu. 2022. “Returns to education in China: Evidence from the Great Higher Education Expansion”, *China Economic Review* 74, 101804.

Huang, C., H. Li, P.W. Liu, and J. Zhang. 2009. “Why does spousal education matter for earnings? Assortative Mating and Cross-Productivity.” *Journal of Labor Economics* 27(4), 633-652.

Huang, W., Pan, Y. and Zhou, Y. 2023. “One-Child Policy, Marriage Distortion, and Welfare Loss.” *The Review of Economics and Statistics*. doi: https://doi.org/10.1162/rest_a_01332.

Hwang, J. 2016. “Housewife, ‘Gold Miss’, and equal: The evolution of educated women’s role in Asia and the US”, *Journal of Population Economics* 29, 529-570.

Isen, A., and B. Stevenson. 2010. “Women's education and family behavior: Trends in marriage, divorce and fertility.” *NBER Working Paper* No. 15725.

Kang, L., F. Peng, and Y. Zhu. 2021. “Returns to university subjects and tiers in China - Evidence from the China Family Panel Studies.” *Studies in Higher Education* 46, 1682-1695.

Kaufmann, K.M., M. Messner, and A. Solis. 2013. “Returns to elite Higher Education in the marriage market: Evidence from Chile.” *mimeo*.

Kırdar, M. G., M. Dayıoğlu, and İ. Koç, 2018. “The effects of compulsory-schooling laws on teenage marriage and births in turkey.” *Journal of Human Capital 12*(4), 640-668.

Leesch, J. Katrňák, T. and Skopek, J. 2024. “Structural opportunities or assortative mating? – Decomposing trends and country differences in educational sorting outcomes in marriages”, *European Societies* 26(4), 1137-1169.

Lefgren, L., and F. McIntyre. 2006. “The relationship between women’s education and marriage outcomes.” *Journal of Labor Economics* 24(4), 787–830.

Li, H., Ma, Y., Meng, L., Qiao, X. and X. Shi. 2017. “Skill complementarities and returns to higher education: Evidence from college enrollment expansion in China.” *China Economic Review* 46, 10-26.

Li, W., Lou, X., Yi, J. and J. Zhang. 2023. Unobserved endowments and gender differences in marriage matching. *Journal of Labor Economics*, *41*(4), 859-886.

Lichter, D.T., Anderson, R.N. and M.D. Hayward. 1995. “Marriage markets and marital choice”. *Journal of Family Issues* 16(4), 412-431.

Mare, R. D. 1991. “Five decades of educational assortative mating.” *American Sociological Review* 56(1), 15-32.

Mariotti, F., Mumford, K.A. and Y. Pena-Boquete. 2017. “Power-couples and the Colocation Hypothesis revisited.” *IZA Journal of Migration* 6 (1), 1-18.

National Bureau of Statistics. 2017. *China Educational Finance Statistical Yearbook 2016*.

OECD. 2011. *Doing Better for Families*. Paris: Organisation for Economic Cooperation and Development. http://dx.doi.org/10.1787/888932394015. Accessed 3 January 2020.

Ong, D., Yang, Y. and J. Zhang. 2020. “Hard to get: The scarcity of women and the competition for high-income men in urban China.” *Journal of Development Economics* 144, 1–19.

Patrinos, H. 2008. Returns to education: The gender perspective. *Girls' Education in the 21st Century: Gender Equality, Empowerment and Economic Growth. Washington, DC: The World Bank*, 53-661.

Psacharopoulos, G. 1994. "Returns to investment in education: A global update." *World Development 22*(9), 1325-1343.

Qian, Y., and Z. Qian. 2014. "The gender divide in urban China: Singlehood and assortative mating by age and education." *Demographic Research* 31, 1337-1364.

Qin, X., and R. Buchanan. 2019. "Policy and public preferences regarding the university enrollment quotas system across Chinese provinces." *Higher Education Policy* 34, 881-901.

Rauscher, E. 2015. "Effects of early US compulsory schooling laws on educational assortative mating: the importance of context." *Demography 52*(4), 1219-1242.

Rouppila, S. and F. Zhao. 2017. "The role of universities in developing China's university towns: The case of Songjiang university town in Shanghai." *Cities* 69, 56-63.

Short, S.E., and F. Zhai. 1998. "Looking locally at China's one-child policy." *Studies in Family Planning* 29(4), 373-387.

Smith, J., and H. Park. 2009. "Five decades of educational assortative mating in 10 East Asian societies." *Social Forces 88*(1), 227-255.

Tampieri, A. 2022. "The effects of educational assortative matching on job and marital satisfaction", *Journal of Behavioral and Experimental Economics* 98, Article 101883.

Wang, L. and J. Klugman. 2020. "How women have fared in the labour market with China's rise as a global economic power." *Asia & The Pacific Policy Studies* 7(1), 43-64.

Wooldridge, J. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press. Cambridge, Massachusetts.

Xing, C. 2013. "Education expansion, migration and rural-urban education gap: A case study on the effect of university expansion." *China Economic Quarterly* 13(1), 207-232 (*in Chinese*).

You, J., Yi, X., and M. Chen. 2021. "Love, life and 'leftover ladies' in urban China: Staying modernly single in patriarchal traditions." *China Economic Review* 68, Article 101626.

Tables

Table 1: Marital Status by education, percentages

	Single	Married	Cohabit	Separated	Divorce	Widowed	Re-married	Obs	Share (%)
No Schooling	8.50	88.10	0.26	0.13	1.18	1.57	0.26	765	3.91
Primary	2.34	94.71	0.36	0.13	1.20	1.07	0.19	3,083	15.74
Mid School	5.04	92.23	0.16	0.19	1.54	0.60	0.24	6,345	32.40
Voc High School	11.76	84.63	0.13	0.20	2.87	0.40	0.00	1,496	7.64
Aca. High School	8.12	88.15	0.25	0.10	2.84	0.40	0.15	2,008	10.25
Voc. College	22.07	75.33	0.21	0.04	1.81	0.46	0.08	2,379	12.15
Bachelor’s	34.83	63.70	0.33	0.10	0.98	0.03	0.03	3,069	15.67
Masters	41.92	56.57	0.25	0.00	1.26	0.00	0.00	396	2.02
Doctoral	25.00	75.00	0.00	0.00	0.00	0.00	0.00	40	0.20
Total %	13.10	84.18	0.23	0.13	1.64	0.56	0.15	19,581	100.00

Table 2: Summary Statistics by birth cohort bands

	Pre-expansion		Post-expansion			Pre-expansion	Post-expansion	Diff	p-value
Birth cohorts	**1970-74**	**1975-79**	**1980-84**	**1985-89**	**1990-94**				
Age	45.0	40.1	35.1	30.2	25.7	42.9	30.1	12.8	0.000
Year of birth	1971.9	1976.9	1982.0	1987.0	1991.9	1974.1	1986.9	-12.8	0.000
Married (incl. remarried)	0.945	0.949	0.933	0.857	0.479	0.947	0.762	0.185	0.000
Ever married	0.986	0.981	0.952	0.870	0.485	0.984	0.774	0.209	0.000
Years of schooling	8.77	9.77	10.82	11.53	12.48	9.20	11.60	-2.40	0.000
Urban *hukou* at birth	0.269	0.277	0.293	0.259	0.261	0.273	0.270	0.002	0.703
Urban *hukou* at age 12	0.324	0.345	0.355	0.320	0.317	0.333	0.330	0.003	0.665
Observations	4,838	3,733	3,579	3,914	3,477	8,611	10,970	-	-
%	24.7	19.3	18.3	20.0	17.8	44.0	56.0	-	-

Table 3: LPM and 2SLS estimates of being married or ever married

	Married		Ever married	
	LPM	2SLS	LPM	2SLS
	(1)	(2)	(3)	(4)
Second Stage:				
Years of schooling	-0.011***	0.129***	-0.011***	0.150***
	(0.001)	(0.016)	(0.001)	(0.017)
Age	0.393***	0.490***	0.402***	0.513***
	(0.015)	(0.027)	(0.015)	(0.028)
Age sq	-0.005***	-0.005***	-0.005***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.000)
Post-1980 birth	-0.675***	-0.885***	-0.709***	-0.951***
	(0.049)	(0.087)	(0.045)	(0.093)
Post-1980 birth year trend	0.069***	0.089***	0.073***	0.095***
	(0.004)	(0.008)	(0.004)	(0.008)
Urban *hukou* at age 12	-0.007	-0.217***	-0.003	-0.245***
	(0.009)	(0.029)	(0.008)	(0.031)
Rural birth	0.108***	0.190***	0.085**	0.179***
	(0.039)	(0.057)	(0.036)	(0.060)
Constant	-7.335***	-11.565***	-7.523***	-12.385***
	(0.337)	(0.710)	(0.319)	(0.758)
First stage: Years of schooling				
Age		-0.711***		-0.711***
		(0.140)		(0.140)
Age sq.		0.006***		0.006***
		(0.002)		(0.002)
Post-1980 birth		0.915*		0.915*
		(0.498)		(0.498)
Post-1980 birth year trend		-0.105***		-0.105***
		(0.044)		(0.044)
Urban *hukou* at age 12		1.919***		1.919***
		(0.166)		(0.166)
Rural birth		-0.559*		-0.559*
		(0.324)		(0.324)
Instruments:				
Post-1980 birth X Urban *hukou*		1.842***		1.842***
		(0.303)		(0.303)
Birth year linear trend X Urban *hukou*		-0.027		-0.027
		(0.024)		(0.024)
Post-1980 birth year trend X Urban *hukou*		-0.112***		-0.112***
		(0.028)		(0.028)
F-stat of IV relevance		40.148		40.148
(p-value)		(0.0000)		(0.0000)
Endogeneity Test χ^2		233.271		388.68
(p-value)		(0.0000)		(0.0000)
Sargan's overidentification test χ^2		1.512		5.940
(p-value)		(0.470)		(0.051)
Observations	19,581	19,581	19,581	19,581

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province and *hukou* status at birth.

Table 4: 2SLS estimates of being married, by geographical region and area type

	By geographical Region		By area type	
	Coastal	Inland	Major cities	Small cities, towns and rural areas
	(1)	(2)	(3)	(4)
Second Stage:				
Years of schooling	0.156***	0.093***	0.157***	0.136***
	(0.027)	(0.017)	(0.026)	(0.029)
Age	0.528***	0.444***	0.385***	0.563***
	(0.040)	(0.033)	(0.041)	(0.046)
Age sq	-0.006***	-0.005***	-0.004***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.001)
Post-1980 birth	-1.024***	-0.745***	-0.640***	-1.046***
	(0.132)	(0.110)	(0.142)	(0.133)
Post-1980 birth year trend	0.101***	0.078***	0.064***	0.106***
	(0.012)	(0.010)	(0.013)	(0.012)
Urban *hukou* at age 12	-0.238***	-0.173***	-0.249***	-0.241***
	(0.042)	(0.038)	(0.045)	(0.052)
Rural birth	0.223***	-0.149	0.215***	0.194
	(0.066)	(0.356)	(0.066)	(0.550)
First Stage: Years of schooling				
Age	-0.705***	-0.707***	-0.021	-1.143***
	(0.184)	(0.218)	(0.211)	(0.189)
Age sq.	0.006***	0.006***	-0.002	0.011***
	(0.002)	(0.003)	(0.002)	(0.002)
Post-1980 birth	1.361**	0.369	-0.471	1.934***
	(0.647)	(0.773)	(0.741)	(0.669)
Post-1980 birth year trend	-0.142***	-0.064	0.040	-0.207
	(0.058)	(0.069)	(0.067)	(0.059)
Urban *hukou* at age 12	1.780***	2.249***	1.981***	1.949***
	(0.144)	(0.275)	(0.235)	(0.241)
Rural birth	-0.752***	4.153*	-0.643**	1.079
	(0.325)	(2.427)	(0.325)	(3.113)
Instruments:				
Post-1980 birth X Urban *hukou*	1.286***	2.469***	1.311***	1.617***
	(0.379)	(0.508)	(0.428)	(0.452)
Birth year linear trend X Urban *hukou*	-0.040	-0.030	-0.036	-0.031
	(0.030)	(0.040)	(0.035)	(0.035)
Post-1980 birth year trend X Urban *hukou*	-0.071**	-0.146***	-0.084**	-0.088**
	(0.035)	(0.047)	(0.040)	(0.041)
F-stat of IV relevance	17.789	22.768	19.337	12.396
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Endogeneity Test χ^2	128.852	87.031	132.859	91.654
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Sargan's overidentification test χ^2	1.616	0.504	0.225	0.765
(p-value)	(0.446)	(0.777)	(0.894)	(0.682)
Observations	10,745	8,836	8,145	11,436

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province and rural birth (by *hukou* status).

Table 5: Robustness of 2SLS of being married, to city fixed effects, interaction of birth *hukou* province and birth *hukou* status, and birth city-cohort sex ratio

	No sex-ratio control			**Sex ratio as control**	**Sex ratio as IV**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Second Stage:					
Years of schooling	0.136***	0.133***	0.129***	0.132***	0.129***
	(0.019)	(0.017)	(0.016)	(0.017)	(0.016)
Age	0.495***	0.502***	0.490***	0.496***	0.496***
	(0.029)	(0.028)	(0.027)	(0.028)	(0.028)
Age sq	-0.006***	-0.006***	-0.005***	-0.006***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Post-1980 birth	-0.934***	-0.921***	-0.885***	-0.943***	-0.940***
	(0.098)	(0.092)	(0.087)	(0.093)	(0.093)
Post-1980 birth year trend	0.092***	0.092***	0.089***	0.093***	0.093***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Urban *hukou* at age 12	-0.604***	-0.505***	-0.217***	-0.227***	-0.223***
	(0.070)	(0.053)	(0.029)	(0.032)	(0.030)
Rural birth			0.190***	0.191***	0.188***
			(0.057)	(0.057)	(0.056)
Birth city-cohort sex ratio				-0.032	
				(0.061)	
City FE	No	Yes	Yes	Yes	Yes
Birth province X Rural birth	No	No	Yes	Yes	Yes
First stage: Years of schooling					
Age	-0.677***	-0.776***	-0.711***	-0.606***	-0.606***
	(0.155)	(0.144)	(0.140)	(0.148)	(0.148)
Age sq.	0.006***	0.007***	0.006***	0.005***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Post-1980 birth	1.104**	1.166**	0.915*	0.966*	0.966*
	(0.546)	(0.508)	(0.498)	(0.518)	(0.518)
Post-1980 birth year trend	-0.114***	-0.130***	-0.105***	-0.102**	-0.102**
	(0.049)	(0.045)	(0.044)	(0.046)	(0.046)
Urban *hukou* at age 12	4.128***	3.562***	1.919***	1.977***	1.977***
	(0.156)	(0.146)	(0.166)	(0.166)	(0.166)
Rural birth			-0.559*	-0.591*	-0.591*
			(0.324)	(0.318)	(0.318)
Birth city-cohort sex ratio				0.924***	0.924***
				(0.332)	(0.332)
City FE	No	Yes	Yes	Yes	Yes
Birth province X Rural birth (full interaction)	No	No	Yes	Yes	Yes
Instruments:					
Post-1980 birth X Urban *hukou*	1.762***	1.683***	1.842***	1.581***	1.581***
	(0.309)	(0.309)	(0.303)	(0.312)	(0.312)
Birth year linear trend X Urban *hukou*	-0.035	-0.038	-0.027	-0.034	-0.034
	(0.027)	(0.025)	(0.024)	(0.025)	(0.025)
Post-1980 birth year trend X Urban *hukou*	-0.097***	-0.094***	-0.112***	-0.093***	-0.093***
	(0.031)	(0.029)	(0.028)	(0.029)	(0.029)
F-stat of IV relevance	28.392	36.142	40.148	33.831	28.232
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Endogeneity Test χ^2	210.6	233.123	233.271	198.855	212.113
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Sargan's overidentification test χ^2	3.869	2.343	1.512	0.862	1.167
(p-value)	(0.145)	(0.310)	(0.470)	(0.650)	(0.761)
Observations	19,581	19,581	19,581	17,456	17,456

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province and *hukou* status at birth.

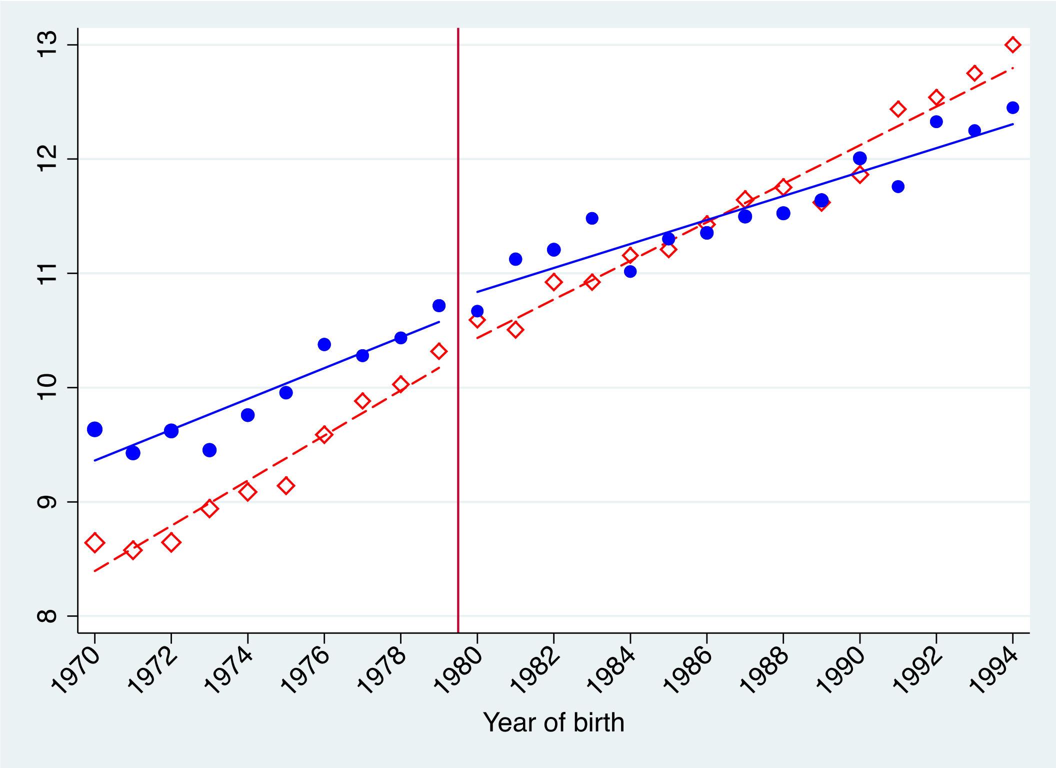
Table 6: 2SLS estimates with alternative critical age for childhood *hukou* status in IV

	Married			Ever Married		
	At birth (1)	**Age 6 (2)**	**Age 15 (3)**	**At birth (4)**	**Age 6 (5)**	**Age 15 (6)**
Second Stage:						
Years of schooling	0.130***	0.126***	0.136***	0.152***	0.144***	0.157***
	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)	(0.017)
Age	0.489***	0.488***	0.493***	0.513***	0.510***	0.516***
	(0.027)	(0.026)	(0.027)	(0.029)	(0.028)	(0.029)
Age sq	-0.005***	-0.005***	-0.005***	-0.006***	-0.006***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Post-1980 birth	-0.884***	-0.883***	-0.887***	-0.952***	-0.947***	-0.952***
	(0.088)	(0.086)	(0.090)	(0.094)	(0.091)	(0.096)
Post-1980 birth year trend	0.089***	0.089***	0.089***	0.095***	0.095***	0.095***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
Urban *hukou* in childhood	-0.369***	-0.172***	-0.259***	-0.384***	-0.201***	-0.287***
	(0.061)	(0.026)	(0.031)	(0.066)	(0.028)	(0.033)
Rural birth		0.212***	0.185***		0.196***	0.176***
		(0.056)	(0.058)		(0.059)	(0.062)
First Stage: Years of schooling						
Age	-0.706***	-0.718***	-0.692***	-0.706***	-0.717***	-0.690***
	(0.141)	(0.140)	(0.140)	(0.141)	(0.140)	(0.140)
Age sq.	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Post-1980 birth	1.022**	0.946*	0.827*	1.036**	0.959*	0.828*
	(0.494)	(0.494)	(0.493)	(0.495)	(0.494)	(0.494)
Post-1980 birth year trend	-0.112**	-0.108**	-0.096**	-0.113**	-0.108**	-0.095**
	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)
Urban *hukou* in childhood	2.325***	1.758***	2.007***	2.325***	1.758***	2.007***
	(0.350)	(0.169)	(0.162)	(0.350)	(0.169)	(0.162)
Rural birth		-0.659**	-0.545*		-0.659**	-0.545*
		(0.327)	(0.321)		(0.327)	(0.321)
Instruments:						
Post-1980 birth X Urban *hukou*	1.823***	1.877***	1.852***	1.823***	1.877***	1.852***
	(0.321)	(0.305)	(0.300)	(0.321)	(0.305)	(0.300)
Birth year linear trend X Urban	-0.0319	-0.0266	-0.0167	-0.0319	-0.0266	-0.0167
hukou	(0.0256)	(0.0242)	(0.0239)	(0.0256)	(0.0242)	(0.0239)
Post-1980 birth year trend X Urban	-	-	-	-	-	-
hukou	0.114***	0.116***	0.121***	0.114***	0.116***	0.121***
	(0.0300)	(0.0284)	(0.0280)	(0.0300)	(0.0284)	(0.0280)
F-stat of IV relevance	43.202	42.295	40.981	43.202	42.295	40.983
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Endogeneity Test χ^2	258.4	233.926	261.432	437.00	382.655	430.609
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Sargan's overidentification test χ^2	1.617	1.314	1.039	4.887	5.713	4.833
(p-value)	(0.445)	(0.518)	(0.594)	(0.086)	(0.057)	(0.089)
Observations	19,581	19,581	19,581	19,581	19,581	19,581

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province and rural birth (by *hukou* status).

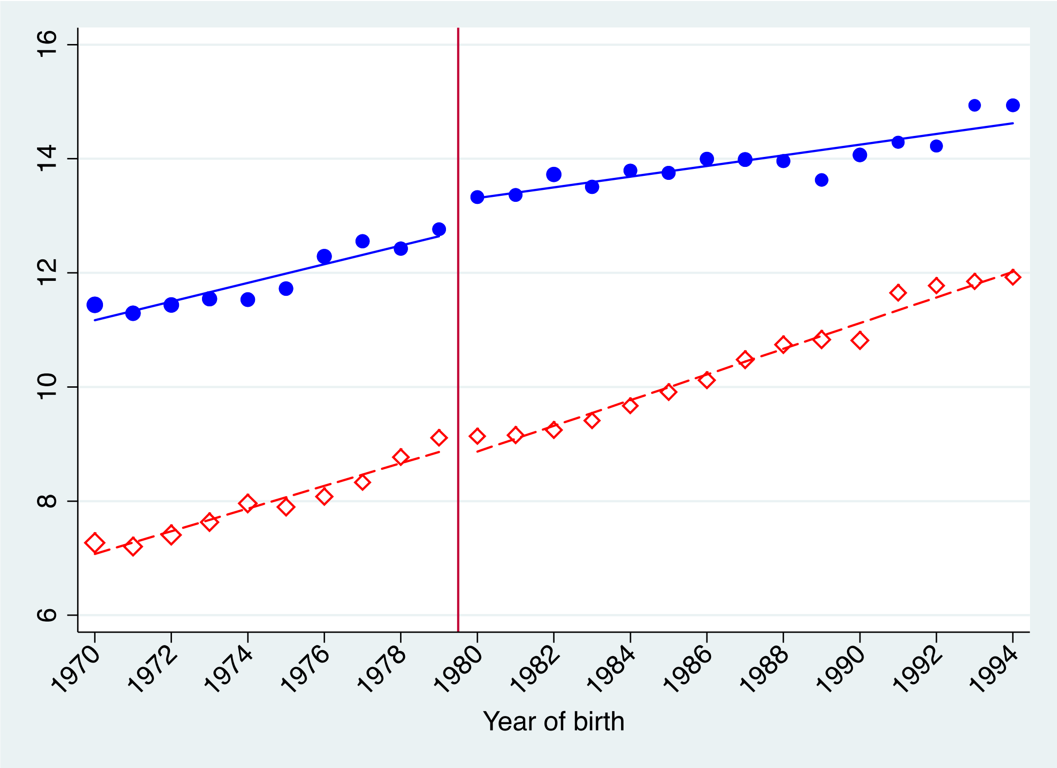
Figures

Figure 1: Years of schooling by gender and birth cohort



Note: Blue solid circles and red hollow diamonds indicate males and females respectively. Bubbles proportional to cell sizes.

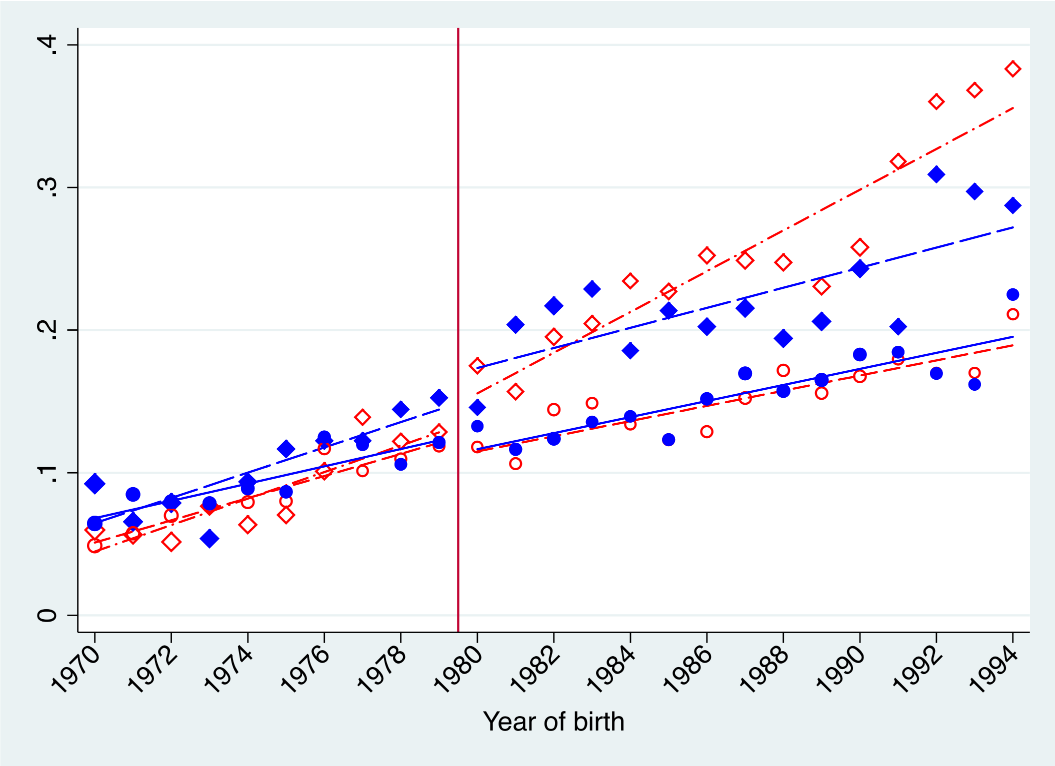
Figure 2: Years of schooling by birth cohort and *hukou* status at age 12, women only



Note: Blue solid circles for non-agricultural *hukou*. Red hollow diamonds for agricultural *hukou*. Size of bubble proportional to cell sizes.

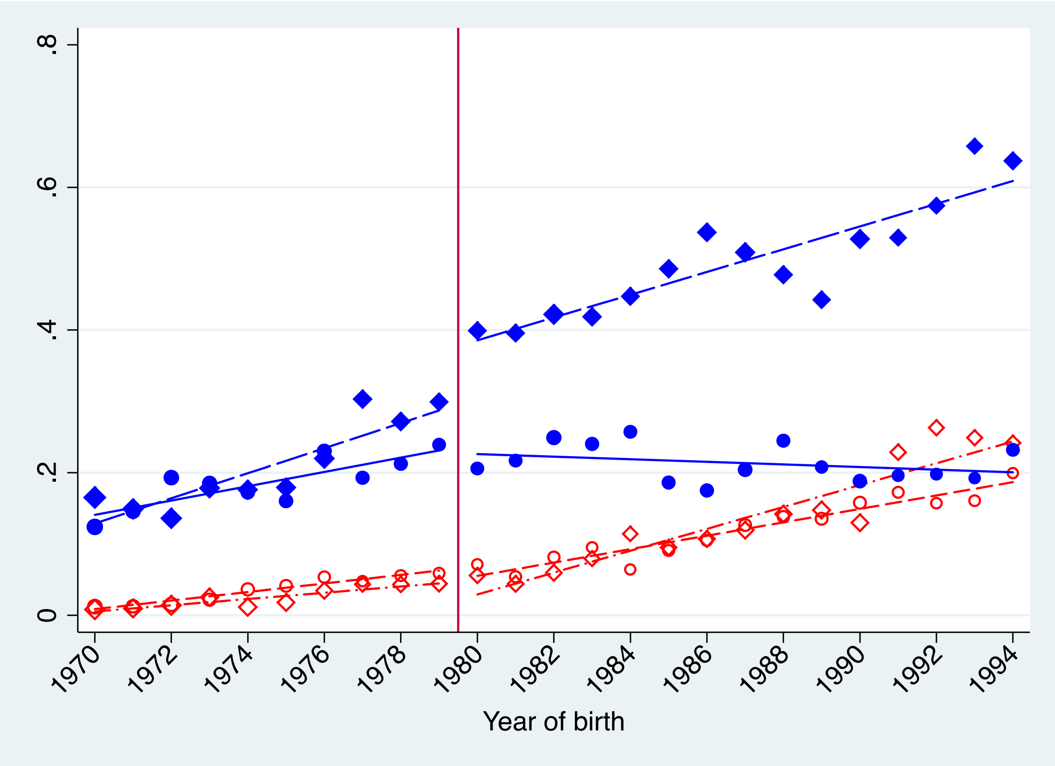
Appendix

Figure A1: Attainment of Vocational College and Degree+ qualifications by gender and birth cohort



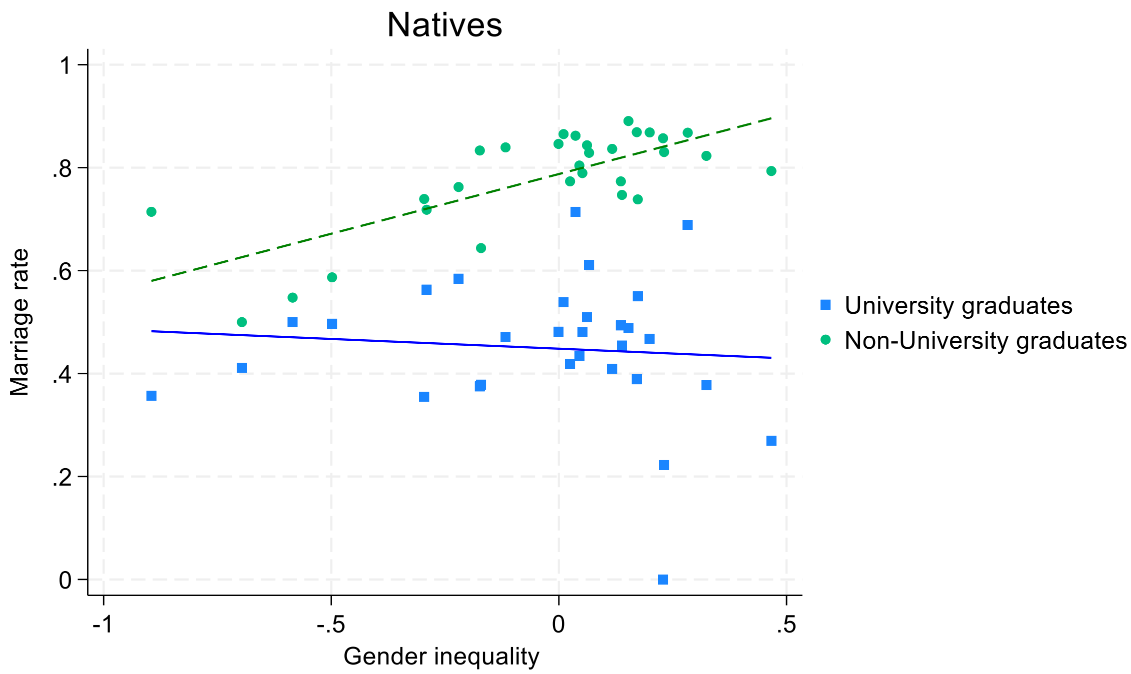
Note: Circle and diamonds indicate vocational college and degree+ qualifications respectively. Red dash (-dot) and blue solid lines indicate fitted lines for females and males respectively. Bubble proportional to cell sizes.

Figure A2: Attainment of Vocational College and Degree+ qualifications by gender and *hukou* status at 12, women only



Note: Circle and diamonds indicate vocational college and degree+ qualifications respectively. Red and blue lines indicate fitted lines for rural and urban *hukou* at age 12 respectively. Bubble proportional to cell sizes.

Figure A3: Marriage rates and agreement with traditional gender norms by province, *hukou* status and college degree



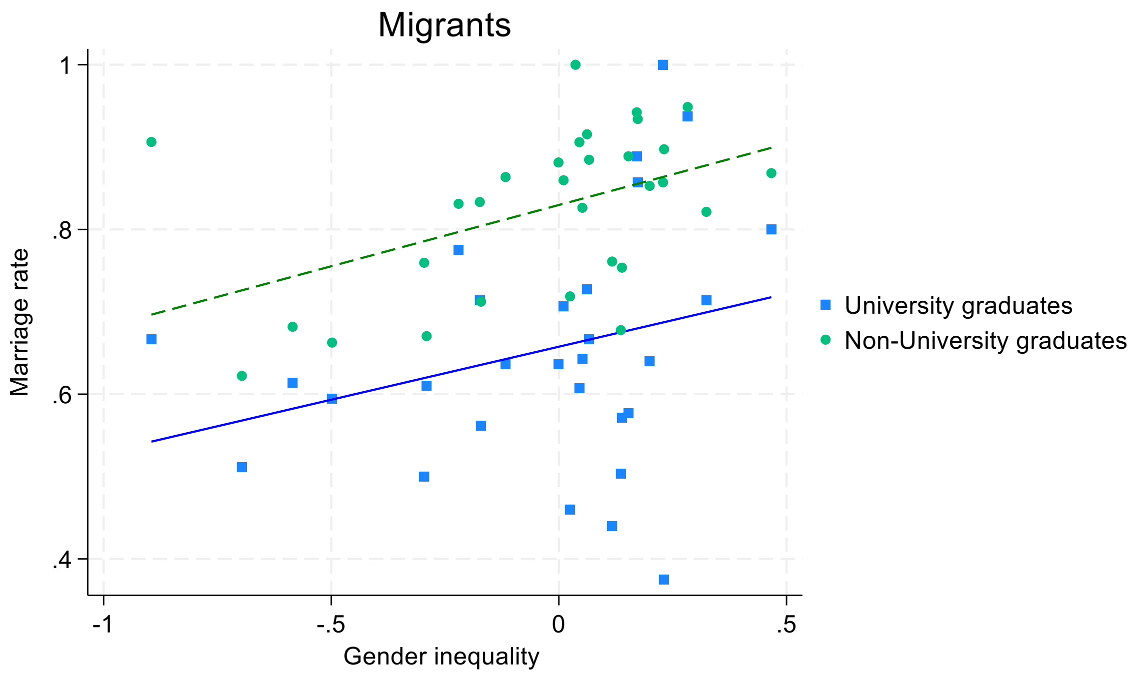


Table A1: Educational assortative mating, percentages

Husbands' qualification	No schooling	Primary	Middle School	High School	Voc College	Bachelors	Masters+	Observa-tions	Row %
Wives' education:									
No Schooling	**22.02**	43.25	27.78	5.75	0.60	0.60	0.00	504	4.07
Primary	1.75	**44.02**	43.70	8.56	1.30	0.67	0.00	2,231	17.99
Middle School	0.48	11.85	**63.34**	18.31	4.55	1.47	0.00	4,453	35.11
High School	0.31	3.18	26.45	**44.54**	17.29	7.96	0.27	2,261	18.24
Voc College	0.14	0.64	8.73	22.69	**40.87**	24.84	2.08	1,397	11.27
Bachelors	0.07	0.14	2.75	10.26	18.11	**60.61**	8.06	1,452	11.71
Masters+	0.00	0.50	1.50	0.50	2.50	36.50	**58.50**	200	1.61
Total observations	181	1,800	4,635	2,491	1,460	1,562	269	12,398	100.00
Column %	1.46	14.52	37.39	20.09	11.78	12.60	2.17	100.00	-

Note: Diagonal cells in bold.

Table A2: Key marriage characteristics by wife's birth cohort bands

Birth cohorts	Pre-expansion		Post-expansion			Overall
	1970-74	1975-79	1980-84	1985-89	1990-94	
Age	45.0	40.1	35.1	30.2	25.7	38.0
Years of schooling (own)	8.84	9.80	10.85	11.45	11.19	10.10
(husband-wife) schooling gap	0.93	0.55	0.37	0.21	0.04	0.54
% wife less schooling (*hypergamy*)	36.1	31.6	29.7	28.1	24.9	31.3
% with same schooling (*homogamy*)	48.1	48.6	51.9	49.8	49.6	49.4
% wife more schooling (*hypogamy*)	15.9	19.7	19.5	22.1	25.5	19.2
Age gap	1.78	2.27	2.51	2.65	3.04	2.28
% with same age	18.2	16.8	16.8	17.0	16.4	17.2
% wife younger	64.9	69.6	71.1	72.2	76.5	69.4
% wife older	16.9	13.5	12.0	10.8	7.1	13.4
Observations	3,723	3,133	2,757	2,082	703	12,398
%	30.0	25.3	22.2	16.8	5.7	100.0

Table A3: Probit & IV Probit Estimates of Being Married and Ever Married

	Married		Ever Married	
	Probit	IV-Probit	Probit	IV-Probit
	(1)	(2)	(3)	(4)
Second Stage: Marriage outcomes				
Years of Schooling	-0.059***	0.159***	-0.083***	0.135***
	(0.005)	(0.031)	(0.006)	(0.038)
Age	1.039***	1.000***	1.087***	1.055***
	(0.079)	(0.083)	(0.088)	(0.093)
Age sq	-0.012***	-0.012***	-0.013***	-0.012***
	(0.001)	(0.001)	(0.001)	(0.001)
Post-1980 birth	-1.159***	-1.372***	-1.525***	-1.708***
	(0.297)	(0.259)	(0.340)	(0.295)
Post-1980 birth year trend	0.124***	0.139***	0.149***	0.162***
	(0.026)	(0.023)	(0.029)	(0.026)
Urban *hukou* at age 12	-0.069	-0.315***	-0.067	-0.279***
	(0.058)	(0.060)	(0.068)	(0.067)
Rural birth	0.349**	0.429***	0.286	0.395**
	(0.162)	(0.149)	(0.179)	(0.161)
Constant	-19.810***	-22.671***	-20.887***	-23.793***
	(1.742)	(1.591)	(1.956)	(1.762)
First Stage: Years of schooling				
Age		-0.712***		-0.711***
		(0.140)		(0.140)
Age sq		0.006***		0.006***
		(0.002)		(0.002)
Post-1980 birth		0.904*		0.912*
		(0.493)		(0.494)
Post-1980 birth year trend		-0.106**		-0.105**
		(0.044)		(0.044)
Urban *hukou* at age 12		1.923***		1.918***
		(0.157)		(0.162)
Rural birth		-0.557*		-0.559***
		(0.322)		(0.322)
Post-1980 birth X Urban *hukou*		1.873***		1.850***
		(0.285)		(0.294)
Birth year linear trend X Urban *hukou*		-0.030		-0.027
		(0.022)		(0.023)
Post-1980 birth year trend X Urban *hukou*		-0.110***		-0.112***
		(0.026)		(0.027)
Corr (ε_{ij}, e_{ij})		-0.619***		-0.605***
(sd)		(0.081)		(0.096)
Wald test of exogeneity (χ^2)		30.22		21.56
(p-value)		(0.0000)		(0.0000)
Observations	19,581	19,581	19,581	19,581
Pseudo R^2	0.309		0.422	

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province and *hukou* status at birth.

Table A4: LPM and 2SLS estimates of being married, sample aged 28+ or 20+

	Sample aged 28+		Sample aged 20+	
	LPM	2SLS	LPM	2SLS
	(1)	(2)	(3)	(4)
Second Stage: Marriage Outcomes				
Years of schooling	-0.002**	0.077***	-0.012***	0.072***
	(0.001)	(0.020)	(0.001)	(0.008)
Age	0.080***	0.137***	0.332***	0.361***
	(0.024)	(0.036)	(0.009)	(0.013)
Age sq	-0.001***	-0.001***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Post-1980 birth	0.003	-0.120	-0.493***	-0.535***
	(0.058)	(0.090)	(0.037)	(0.050)
Post-1980 birth year trend	0.002	0.013	0.053***	0.057***
	(0.006)	(0.009)	(0.003)	(0.004)
Rural birth	-0.011	-0.133***	-0.008	-0.130***
	(0.009)	(0.033)	(0.009)	(0.017)
Urban *hukou* at age 12	0.119***	0.167***	0.094**	0.139***
	(0.040)	(0.043)	(0.037)	(0.041)
Constant	-0.806	-3.258***	-6.008***	-7.929***
	(0.515)	(0.927)	(0.215)	(0.337)
First stage: Years of schooling				
Age		-0.693**		-0.386***
		(0.280)		(0.0940)
Age sq.		0.006*		0.00222**
		(0.003)		(0.00109)
Post-1980 birth		0.828		-0.0793
		(0.743)		(0.391)
Post-1980 birth year trend		-0.098		-0.0146
		(0.070)		(0.0342)
Urban *hukou* at age 12		1.750***		2.032***
		(0.169)		(0.163)
Rural birth		-0.603*		-0.458
		(0.351)		(0.313)
Instruments:				
Post-1980 birth X Urban *hukou*		1.968***		2.000***
		(0.431)		(0.263)
Birth year linear trend X Urban *hukou*		-0.023		-0.0284
		(0.024)		(0.0241)
Post-1980 birth year trend X Urban *hukou*		-0.122***		-0.121***
		(0.035)		(0.0265)
F-stat of IV relevance		11.765		82.521
(p-value)		(0.0000)		(0.0000)
Endogeneity Test χ^2		29.510		175.55
(p-value)		(0.0000)		(0.0000)
Sargan's overidentification test χ^2		3.188		5.525
(p-value)		(0.203)		(0.063)
Observations	16,104	16,104	21,486	21,486

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province and *hukou* status at birth.

Table A5: LPM and 2SLS estimates of being married or ever married allowing for heterogenous exposure to higher education expansion by birth cohort

	Married		Ever married	
	LPM	2SLS	LPM	2SLS
	(1)	(2)	(3)	(4)
Second Stage:				
Years of schooling	-0.002**	0.073***	-0.003***	0.099***
	(0.001)	(0.018)	(0.001)	(0.020)
Age	0.007	0.083*	0.005	0.109**
	(0.028)	(0.047)	(0.017)	(0.051)
Age sq	-0.000	-0.001	-0.000	-0.001*
	(0.000)	(0.001)	(0.000)	(0.001)
Post-1980 birth	0.521*	0.712**	0.339	0.601*
	(0.274)	(0.334)	(0.249)	(0.360)
Post-1980 birth year trend	-0.036*	-0.039*	-0.023	-0.028
	(0.019)	(0.022)	(0.017)	(0.024)
Urban *hukou* at age 12	-0.009	-0.124***	-0.007	-0.163***
	(0.009)	(0.031)	(0.007)	(0.033)
Rural birth	0.120***	0.166***	0.091**	0.154***
	(0.040)	(0.042)	(0.036)	(0.045)
Constant	0.753	-2.024*	0.783**	-3.009**
	(0.596)	(1.154)	(0.370)	(1.246)
First stage: Years of schooling				
Age		-0.179***		-0.179***
		(0.0269)		(0.0269)
Age sq.		-0.00226***		-0.00226***
		(0.000135)		(0.000135)
Post-1980 birth		-4.653*		-4.653*
		(2.762)		(2.762)
Post-1980 birth year trend		0.258		0.258
		(0.173)		(0.173)
Urban *hukou* at age 12		1.653***		1.653***
		(0.115)		(0.115)
Rural birth		-0.559*		-0.559*
		(0.324)		(0.324)
Instruments:				
Urban *hukou* at age 12 X HEE 1999		0.0186		0.0186
		(0.0261)		(0.0261)
Urban *hukou* at age 12 X HEE 2000		0.0287		0.0287
		(0.0237)		(0.0237)
Urban *hukou* at age 12 X HEE 2001		0.0316		0.0316
		(0.0220)		(0.0220)
Urban *hukou* at age 12 X HEE 2002		-0.0121		-0.0121
		(0.0235)		(0.0235)
Urban *hukou* at age 12 X HEE 2003		-0.00199		-0.00199
		(0.0244)		(0.0244)
Urban *hukou* at age 12 X HEE 2004		-0.0121		-0.0121
		(0.0234)		(0.0234)
Urban *hukou* at age 12 X HEE 2005		-0.0204		-0.0204
		(0.0230)		(0.0230)
Urban *hukou* at age 12 X HEE 2006		-0.0499**		-0.0499**
		(0.0222)		(0.0222)
Urban *hukou* at age 12 X HEE 2007		-0.0779***		-0.0779***
		(0.0231)		(0.0231)
Urban *hukou* at age 12 X HEE 2008		-0.0951***		-0.0951***
		(0.0236)		(0.0236)
F-stat of IV relevance		4.433		4.433
(p-value)		(0.0000)		(0.0000)
Endogeneity Test χ^2		20.189		69.809
(p-value)		(0.0000)		(0.0000)

Sargan's overidentification test χ^2		7.947		16.623
(p-value)		(0.539)		(0.055)
Observations	15,718	15,718	15,718	15,718

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province, *hukou* status at birth, and Li's instruments between 1999 and 2008.

Table A6: 2SLS estimates with alternative critical age for childhood *hukou* status allowing for heterogenous exposure to higher education expansion by birth cohort

	Married			Ever Married		
	At birth	**Age 6**	**Age 15**	**At birth**	**Age 6**	**Age 15**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Second Stage:						
Years of schooling	0.124***	0.075***	0.066***	0.124***	0.097***	0.094***
	(0.025)	(0.018)	(0.016)	(0.025)	(0.019)	(0.017)
Age	0.172***	0.134***	0.128***	0.172***	0.156***	0.153***
	(0.045)	(0.036)	(0.033)	(0.045)	(0.037)	(0.036)
Age sq	-0.002***	-0.001***	-0.001***	-0.002***	-0.002***	-0.002***
	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Post-1980 birth	-0.216*	-0.119	-0.100	-0.216*	-0.183*	-0.170*
	(0.114)	(0.091)	(0.085)	(0.114)	(0.095)	(0.092)
Post-1980 birth year trend	0.022**	0.013	0.011	0.022**	0.019**	0.018**
	(0.011)	(0.009)	(0.008)	(0.011)	(0.009)	(0.009)
Urban *hukou* in childhood	-0.339***	-0.102***	-0.132***	-0.339***	-0.139***	-0.172***
	(0.068)	(0.028)	(0.029)	(0.068)	(0.030)	(0.031)
Rural birth		0.187***	0.155***		0.167***	0.144***
		(0.043)	(0.039)		(0.045)	(0.042)
First Stage: Years of schooling						
Age	-0.833***	-0.924***	-0.907***	-0.833***	-0.924***	-0.907***
	(0.305)	(0.308)	(0.307)	(0.305)	(0.308)	(0.307)
Age sq.	0.00755**	0.00864**	0.00847**	0.00755**	0.00864**	0.00847**
	(0.00357)	(0.00361)	(0.00359)	(0.00357)	(0.00361)	(0.00359)
Post-1980 birth	1.466*	1.476*	1.385*	1.466*	1.476*	1.385*
	(0.801)	(0.809)	(0.806)	(0.801)	(0.809)	(0.806)
Post-1980 birth year trend	-0.144*	-0.151**	-0.141*	-0.144*	-0.151**	-0.141*
	(0.0754)	(0.0761)	(0.0758)	(0.0754)	(0.0761)	(0.0758)
Urban *hukou* in childhood	2.064***	1.515***	1.800***	2.064***	1.515***	1.800***
	(0.350)	(0.120)	(0.106)	(0.350)	(0.120)	(0.106)
Rural birth		-0.687*	-0.607*		-0.687*	-0.607*
		(0.355)	(0.351)		(0.355)	(0.351)
Instruments:						
Urban *hukou* at age 12 X HEE 1999	0.0269	0.0316	0.0267	0.0269	0.0316	-0.607*
	(0.0251)	(0.0237)	(0.0235)	(0.0251)	(0.0237)	(0.351)
Urban *hukou* at age 12 X HEE 2000	-0.00915	0.0187	0.0128	-0.00915	0.0187	0.0267
	(0.0230)	(0.0212)	(0.0211)	(0.0230)	(0.0212)	(0.0235)
Urban *hukou* at age 12 X HEE 2001	0.0152	0.0292	0.0234	0.0152	0.0292	0.0128
	(0.0203)	(0.0188)	(0.0187)	(0.0203)	(0.0188)	(0.0211)
Urban *hukou* at age 12 X HEE 2002	-0.0269	-0.0126	-0.0186	-0.0269	-0.0126	0.0234
	(0.0213)	(0.0198)	(0.0196)	(0.0213)	(0.0198)	(0.0187)
Urban *hukou* at age 12 X HEE 2003	0.00361	0.00618	-0.000587	0.00361	0.00618	-0.0186
	(0.0226)	(0.0208)	(0.0206)	(0.0226)	(0.0208)	(0.0196)
Urban *hukou* at age 12 X HEE 2004	-0.0196	-0.0119	-0.0193	-0.0196	-0.0119	-0.000587
	(0.0221)	(0.0201)	(0.0199)	(0.0221)	(0.0201)	(0.0206)
Urban *hukou* at age 12 X HEE 2005	-0.0311	-0.0326*	-0.0403**	-0.0311	-0.0326*	-0.0193
	(0.0216)	(0.0197)	(0.0196)	(0.0216)	(0.0197)	(0.0199)
Urban *hukou* at age 12 X HEE 2006	-0.043**	-0.0438**	-0.052***	-0.043**	-0.0438**	-0.0403**
	(0.0211)	(0.0194)	(0.0193)	(0.0211)	(0.0194)	(0.0196)
Urban *hukou* at age 12 X HEE 2007	-0.081***	-0.067***	-0.075***	-0.081***	-0.067***	-0.052***
	(0.0222)	(0.0206)	(0.0204)	(0.0222)	(0.0206)	(0.0193)
Urban *hukou* at age 12 X HEE 2008	-0.104***	-0.108***	-0.118***	-0.104***	-0.108***	-0.075***
	(0.0244)	(0.0221)	(0.0219)	(0.0244)	(0.0221)	(0.0204)
F-stat of IV relevance	3.406	4.081	4.690	3.407	4.071	4.690
(p-value)	(0.0002)	(0.0000)	(0.0000)	(0.0002)	(0.0000)	(0.0000)
Endogeneity Test χ^2	42.48	32.774	29.531	117.345	85.782	90.631
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Sargan's overidentification test χ^2	5.890	5.605	3.271	13.046	17.656	11.730
(p-value)	(0.751)	(0.779)	(0.952)	(0.161)	(0.039)	(0.229)
Observations	15718	15718	15718	15718	15718	15718

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Controls for fixed-effects of the current (prefectural) city of residence, and the full interaction of birth province and rural birth (by *hukou* status).