Gender Differences in College Migration

Zihan Hu

Hyejin Ku

Xinzheng Shi

Yizhi Wang

Ming-ang Zhang*

Version: March 26, 2026

Abstract

We examine gender differences in college migration—attending university outside one's hometown—using administrative data covering the universe of National College Entrance Examination takers in China. Conditional on exam performance, girls are 4.5 percentage points less likely than boys to enroll outside their home province. This gap is larger in areas with more traditional gender norms. A new survey we conducted shows that parents of high school girls are less supportive of college migration than parents of boys. This disparity is driven primarily by lower perceived returns to migration, greater safety concerns, and stronger preferences for daughters to find a local spouse, whereas risk preferences, family background, and parental old-age care contribute little. Using instrumental variables that exploit variation in university expansions and cross-province admission quotas, we estimate the causal migration premium and find it to be similar for boys and girls. These estimates imply that closing the gender gap in college migration would reduce the gender gap in elite university attendance by 20.9% and in expected wages by 11.4%.

Keywords: gender gap, geographic labor mobility, college choice

JEL classification: J16, J61, I23

*Zihan Hu: School of Economics, Singapore Management University. zihanhu@smu.edu.sg; Hyejin Ku: Department of Economics, University College London. h.ku@ucl.ac.uk; Xinzheng Shi: School of Economics, Peking University. shixzh@pku.edu.cn; Yizhi Wang: School of Economics, Singapore Management University. yizhi.wang.2025@phdecons.smu.edu.sg; Ming-ang Zhang: School of Public Policy and Management, Tsinghua University. zhangmingang@tsinghua.edu.cn. We thank Joseph Altonji, Pedro Carneiro, Raj Chetty, Adeline Delavande, Christian Dustmann, Hank Farber, John Friedman, Claudia Goldin, Christine Ho, Larry Katz, Attila Lindner, Lin Ma, Xin Meng, Claudia Olivetti, Jessica Pan, and conference and seminar participants for helpful discussions and comments. Ruijie Cai provided outstanding research assistance.

1. Introduction

Despite recent convergence, a persistent gender wage gap remains (Goldin, 2014; Blau and Kahn, 2017). Various factors have been identified in the literature, such as occupational segregation (Cortés and Pan, 2018), commuting preferences (Le Barbanchon, Rathelot and Roulet, 2021), demands for time flexibility (Bertrand, Goldin and Katz, 2010; Wiswall and Zafar, 2018), the career costs of children (Adda, Dustmann and Stevens, 2017), job search process (Cortés et al., 2023), and discrimination (Goldin and Rouse, 2000). This paper explores an understudied perspective relevant to the gender wage gap: the gender difference in migration for college.

The consequences of gender differences in college migration are multifaceted. First, if female students are less likely to migrate for college, they have a more restricted set of colleges to choose from, often leading to attendance at lower-quality local universities. This effect is likely exacerbated in disadvantaged regions where high-quality universities are sparse. Second, given the established link between college location and job migration (Dustmann and Glitz, 2011), lower college mobility among women may also imply a lower propensity to relocate for better job opportunities later in their careers. Taken together, attending lower-quality institutions and exhibiting reduced job mobility could contribute to the gender wage gap among the college-educated.

The ideal setting for estimating gender differences in college migration involves comparing boys and girls who are eligible for admission to the same set of universities but make different choices. This allows us to isolate individual migration decisions from other factors, such as university admission rules (that may differentially impact girls and boys). The college admission system in China offers an ideal testing ground for this analysis. All high school graduates who want to attend college must take the National College Entrance Exam (NCEE), and admission is purely based on exam scores (Li and Qiu, 2025). Critically for our identification, students taking the NCEE within the same province and academic track (humanity vs. science) and with the same NCEE score are eligible for the same set of universities. Therefore, conditional on the NCEE scores within a given province-track, the observed gender differences in college migration outcomes can be attributed to the choices made by students (or their families), rather than to university admission criteria.

We use administrative data from China's Ministry of Education (MOE) covering the universe of NCEE test takers from 1999 to 2003. The dataset contains 22 million test takers

with detailed information on their NCEE score, the university and major they were admitted to, high-school location, home county, and academic track. The dataset also includes demographic variables, including gender, ethnicity, hukou status (agricultural vs. non-agricultural), and birth year and month. The population-level coverage of this data offers several key advantages, as it ensures our analysis is nationally representative, permits a granular examination of migration patterns, and allows for the accurate derivation of key university-level characteristics such as province-specific enrollment quotas.

We begin our analysis by documenting key facts about the gender gap in college migration. While male and female students with the same NCEE rank are eligible for the same set of universities, their choices diverge significantly. Female students are 4.5 percentage points less likely than their male counterparts of the same ranking in the same county and academic track to enroll in an out-of-province university, representing 17.9% of the sample mean. Since high-achieving students are more likely to attend universities outside their hometown provinces, the documented gender gap in college migration is comparable to the effect of an 11 percentage point drop in ranking for female students. The results are robust to alternative measures of migration, such as the distance between university and home cities or the probability of attending college in top-tier cities. Moreover, the pattern persists even for shorter-distance migration: among students attending an in-province university, female students are still less likely than male students to choose an institution in the provincial capital.

The gender gap in college migration is most pronounced for students in the middle and upper quantiles of the academic performance distribution. These students typically face a broader and more varied set of university choices, magnifying the differences in their migration patterns. Conversely, the gap is minimal among students in the lower quantiles. For this group, the quality differential between local and distant institutions is often less pronounced, and attractive out-of-province options are more limited.

We then investigate the effect of traditional gender norms on the gender gap in college migration. It is theoretically ambiguous since restrictive gender norms could either limit female mobility ("constraint effect") or motivate girls' migration to areas with greater equality ("incentive effect"). Empirically, we find a significantly larger gender gap in college migration among students from rural areas and counties where traditional gender norms are more prevalent, such as the beliefs that men should prioritize their careers while women should focus on family, and

that marrying well is more important for a woman than her own professional achievements.[1] This evidence strongly suggests that local environments characterized by traditional gender norms are more powerful in limiting female students' mobility than in motivating them to leave.

To unpack the broad notion of "traditional gender norms" into concrete concerns and trade-offs, we conducted a new survey targeting parents of upper-level high school students in May 2025.[2] The survey elicits parents' preferences over their child's college location and relevant beliefs concerning the costs and benefits of migrating for college. Notably, despite surveying parents approximately 25 years after the period of our main administrative data, parents of daughters are still more likely to report a stronger preference for their children to stay in their home province for college.

To understand what drives this preference gap, we examine the beliefs parents hold about the benefits and costs of out-of-province enrollment. The most influential factor is the perceived returns to college migration, accounting for approximately half of the gender gap in parental migration preferences. Girls' parents are significantly more likely to believe that the returns to migration are higher for sons than for daughters. This difference may reflect either an underestimation of the pecuniary returns to college migration for girls or parents placing lower importance on higher income for daughters. Moreover, parents of daughters express a significantly stronger preference for their children to find a local spouse and also report greater safety concerns should their child migrate for college, accounting for 13.3% and 7.9% of the gender gap in migration preferences, respectively. Notably, while parental expectations for old-age care from children influence parental preferences regarding children's migration, we find no gender differences in these expectations, suggesting they do not drive the observed migration gap.

Beyond parental attitudes reflecting gender-related norms, we also test mechanisms rooted in gender differences in decision-making under uncertainty, particularly risk aversion and self-confidence (see Eckel and Grossman (2008) and Croson and Gneezy (2009) for comprehensive reviews). The hypothesis is twofold. First, if female students are more risk-averse, they may favor "safer" in-province universities, which often have larger admission quotas for local stu-

[1] These attitudes are measured using county-level average responses of the corresponding questions from the China General Social Survey (CGSS).

[2] The survey was conducted in May 2025, one month before the 2025 NCEE. Therefore, most of our respondents should have thought extensively about their college application preferences.

dents.[3] Second, if male students are more confident and optimistic when forecasting their eventual scores and ranking, they may be more willing to apply to more selective, distant universities.

To test this hypothesis, we leverage variation in the timing of university applications across provinces and over time. We classify admission systems into three tiers of risk: high-risk (applications submitted before the exam), medium-risk (submitted after the exam but before scores are known), and low-risk (submitted after scores are known). If risk aversion and self-confidence under uncertainty were key drivers, the gender gap should shrink in the low-risk regimes where uncertainty is largely resolved. By exploiting within-province changes of admission systems, we do not find a narrower gender gap in middle- and low-risk admission systems. Such results suggest that the gender differences in risk aversion and self-confidence are unlikely to be the primary drivers of the gender migration gap.

Another potential factor is family socioeconomic status (SES). If female students disproportionately come from lower-SES households, resource constraints could limit their ability to migrate for college. However, our analysis of the China College Student Survey (CCSS) data reveals the contrary: female college students tend to come from families with higher incomes and more educated parents than their male counterparts from the same home city.[4] Consequently, family SES cannot explain the gender gap in college migration.

In the final part, we aim to understand the consequences of the gender gap in college migration: how much of the gender gap in wages and elite university attendance would narrow if girls had the same migration propensity as boys? We begin by estimating the causal "migration premium" using an instrumental-variables design that exploits university expansion and differential cross-province admission quotas. The underlying identification source is that faster expansion in out-of-province universities that allocate more seats to a student's home province serves as an exogenous pull factor relative to expansion at hometown universities.

Our IV estimates indicate that migrating for college doubles the likelihood of attending a prestigious (Project 985) university and raises expected early-career wages by about 20%.[5] Notably, we find that the causal migration premia in terms of both expected wages and university quality are statistically indistinguishable between male and female students. Using

[3] See Chapter 2 in Jia, Li and Cousineau (2025) for the description of the regional quota system in China.

[4] This finding likely reflects positive selection of girls into higher education since low-SES households typically exhibit stronger son preference and are less likely to invest in daughters' education.

[5] Using CCSS, we compute the expected wage associated with each university by averaging the "best salary offer" received by final-year students, separately by local vs. non-local status and by gender.

these estimates in a counterfactual that assigns boys' migration propensity to girls, we find that gender differences in college migration account for 20.86% of the gender gap in elite university (Project 985) attendance and 11.44% of the gender gap in expected wages.

In addition, the estimated returns to college migration—higher expected wages and university prestige—are significantly larger in resource-poor regions, where the gender gap in migration is also larger. This reveals a concerning pattern of spatial-gender mismatch: the places where migration can deliver the most substantial benefits are precisely those where the gender gap in college migration is largest.

We contribute to several strands of the literature. First, we contribute to the literature on the gender wage gap (Bertrand, 2011; Goldin, 2014; Olivetti and Petrongolo, 2016; Blau and Kahn, 2017). To the best of our knowledge, we are the first to highlight that gender differences in college migration may partly explain the observed gender wage gap. This contributes to efforts to move beyond educational attainment when accounting for gender differences in earnings. For instance, Altonji et al. (2025) investigate the separate roles of gender differences in returns to fields of study and in field-of-study choices. Our results suggest that comparing college-attending boys and girls, girls tend to attend lower-ranked universities and remain in their less developed hometowns due to their lower likelihood of migrating for college education. It demonstrates how gender differences in college migration choices, long before entering the labor market, can have long-lasting impacts on the gender wage gap.

Second, the existing literature on the economic impacts of gender norms has examined how these norms affect women's labor supply, educational attainment, occupation choices, or preferences for job flexibility (Alesina, Giuliano and Nunn, 2013; Bertrand, Kamenica and Pan, 2015; Cortés and Pan, 2018; Booth et al., 2019; Gu et al., 2024; Huynh and Ku, 2025). We extend the literature by highlighting a distinct and crucial channel: migration for college education. We demonstrate how traditional gender norms can act as a powerful constraint, effectively tying highly educated women to their home regions. This spatial misallocation, driven by social norms rather than unequal migration premiums, results in these women attending lower-quality universities and facing diminished economic opportunities, even when they possess the same academic qualifications as their male peers.

Third, we add to the literature on the drivers of gender differences in college choices (Delavande and Zafar, 2019; Zafar, 2013; Wiswall and Zafar, 2018). It is widely accepted that, at least in developed countries, girls have higher achievement than boys in terms of educational

attainment (Goldin, Katz and Kuziemko, 2006). This paper highlights that, among boys and girls with the same education level, it is crucial to delve into the gender differences in university locations and the ensuing labor market opportunities attached to that to draw a more comprehensive picture of the role of education in the gender wage gap.

Lastly, we contribute to the literature on internal migration. While many studies focus directly on migration for jobs (Jia et al., 2023; Molloy, Smith and Wozniak, 2011), sometimes highlighting the differential migration propensity of college educated (vs. non-college) workers (Wozniak, 2010), within-country migration for college is an important topic heavily understudied. One exception is Hu and Ma (2025), who combine the uneven distribution of educational resources and migration for college to understand the spatial distribution of skilled workers in China. Our paper focuses specifically on the gender dimension of this mobility, examining the relevant drivers and consequences of the gender gap in college migration.

2. China's Higher Education and College Admissions

2.1. Higher Education in China

China's higher education system has expanded rapidly over the past few decades. The quality of these colleges varied immensely. At the very top of the pyramid are 39 universities belonging to Project 985 (called Project 985 Universities), a government initiative launched in May 1998 to develop world-class institutions. Below them are Project 211 Universities— another 68 institutions as of 2010—created under a 1995 initiative to strengthen about 100 key universities for the 21st century.[6] The central government concentrated resources heavily on these elite universities. For example, in 2008, Project 985 and 211 Universities together received roughly 70% of all higher education research funding in China.[7] Attending these high-quality colleges pays in the labor market. For example, Li et al. (2012) and Jia and Li (2021) find that students attending Project 211 Universities or national first-tier colleges earn higher wages in their first job than those attending regular colleges.

Educational resources in China, symbolized by elite universities, are highly unevenly dis-

[6]985 Project Universities automatically belong to the 211 Project. there were 107 universities belong to the 211 Project in year 2010.

[7]Data source: People's Daily Online (http://english.people.com.cn/90001/6381319.html). Although 985 and 211 Projects were replaced by a new program called Double First Class Program in 2016, which aimed to build up first-class universities and first-class majors, this new program covers the majority of colleges in 985 and 211 Projects.

tributed across provinces. Figure 1 shows the spatial distribution of Project 985 Universities across provinces.For example, Beijing and Shanghai alone host 8 and 4 Project 985 universities respectively, together accounting for 30% of all elite universities despite representing a tiny fraction of China's population and geographic area. In sharp contrast, 13 out of 31 provinces have zero Project 985 universities. As a result, migration out of one's home province is often necessary for aspiring students, especially those from resource-poor provinces who want to attend better universities.

2.2. *China's College Admission System*

College admissions in China are centrally structured around the National College Entrance Examination (NCEE, or Gaokao in Chinese), taken at the end of high school. Students typically choose between two tracks in the second year of high school: the science track, which emphasizes math and sciences and is the main route to STEM majors, and the humanities track, which emphasizes literature, history, and politics and channels students into social sciences and humanities majors. Admissions are organized within each province and track: students compete only with peers in the same province and track, and outcomes depend almost entirely on their NCEE scores (Li and Qiu, 2025).[8] For each college–major, quotas approved by the Ministry of Education are determined and publicly announced before students submit their applications.

College applications and admissions are conducted in several rounds, determined by the tier that a student's score falls into. Colleges are correspondingly divided into admission tiers,[9] with higher-tier colleges having priority in admitting students. Lower-tier colleges can only begin their admissions after higher-tier colleges have completed their admission rounds. Students are required to complete an application form ranking their preferred colleges within each admission tier (typically 6–8 colleges). For each college, they must also rank their preferred majors (usually 3–5 majors) (Yang, 2024). If their score does not meet the admission cutoff for the

[8]While the vast majority of students rely solely on their exam performance, a small fraction receive bonus points for specific characteristics, such as ethnic minority status or exceptional athletic talent. To account for this, we use students' total score (including bonus points) rather than the raw examination score in our empirical analysis, ensuring comparability within the same admission choice set, as detailed in Section 4.

[9]In our sample period, most provinces organized college admissions into several tiers. The first tier (Tier 1) refers to the most prestigious undergraduate programs, typically including Project 985 and 211 universities as well as a small number of other top provincial universities. The second tier (Tier 2) includes regular 4-year undergraduate institutions that are less selective than Tier 1 but still grant bachelor's degrees (Benke in Chinese). Below these are the specialist or vocational colleges (Zhuanke), which usually offer 3-year diploma programs.

majors or colleges they listed, they may be considered for reassignment. Accordingly, students must indicate in advance whether they are willing to be reassigned to other majors within the same college that are not among their listed preferences, and whether they accept reassignment to colleges outside their submitted list of preferences. The timing of this application differs across provinces: in some provinces, students apply before the NCEE; in others, they apply after the exam but before knowing their scores; and in still others, they apply after learning their scores (we discuss these admission mechanisms in detail in Section 6.1).

The most common admission procedure during our sample period is the Boston mechanism, which operates as a sequential matching system (Chen and Kesten, 2017; Li and Qiu, 2025). Admissions take place in multiple rounds, beginning with students' first-choice colleges. In the first round, each college only considers applicants who ranked it as their first choice (regardless of major preferences). Among these applicants, the college ranks students by their scores and admits them until its quota is fulfilled. Students with scores above the cutoff are admitted to the colleges, while those falling short are passed on to the next round, where their second-choice colleges evaluate them in the same way. Once a student is admitted to a college, the allocation of majors occurs internally. Within each college, students whose scores are above the college cutoff are ranked by their scores, and higher-scoring students are considered first for their preferred majors. The departments stop admitting students until the slots are filled. Students who cannot meet the cutoff for any of their listed majors and who do not agree to be reassigned are removed from the college and transferred to their next-choice institution. Because rejection at a given stage can mean falling to a much lower-ranked college, the costs of misjudging one's application strategy are high. This design naturally makes students prioritize college choice over major choice.[10]

Importantly, each student can receive only one admission offer, as the process terminates once they are admitted to either a college on their preference list or a reassigned college. This structure makes application strategy, risk preferences, and perceptions of admission chances

[10] In Appendix Table A2, we examine the relative importance of college and major in the sorting of students across the ability distribution. We regress rank percentiles on alternative sets of fixed effects and find that college-level fixed effects—specifically, Project 985, Project 211 (excluding Project 985), other four-year universities (excluding Project 211 and above), and non-four-year colleges—explain 24% of the variation in students' NCEE score rankings . By contrast, major-level fixed effects—based on the 12 broad disciplinary categories in the Chinese higher education system, such as Economics, Engineering, Science, and Medicine—explain only 4%. In the female sample, the corresponding figures are 21% and 3%, compared with 27% and 5% in the male sample. These results indicate that students—both male and female—place substantially greater weight on college choice than on major choice, and that females do not attach disproportionately higher priority to majors relative to colleges.

central to shaping educational and migration outcomes.

3. Data

Our primary data is the individual-level administrative records for the universe of Chinese students (excluding those in Tibet) who took the NCEE between 1999 and 2003. The dataset is administered by the Ministry of Education (MOE) of China. It contains detailed information on around 23.35 million test takers, such as their gender, birth year, hukou (agricultural vs. non-agricultural), exam-taking year, academic track (humanity vs. science), NCEE raw test score, bonus point eligibility, and their hometown county.[11] Since universities admit students based on the sum of the NCEE raw score and bonus points, we also use this sum to characterize students' ranking and university eligibility.

Critically for our analysis, the dataset also includes final college and major admission outcomes for each student. Unlike the university application and admission process in the U.S., China's centralized admission system admits each student to only one university, which will be the institution the student attends unless they forgo enrollment in any university that year. Although this dataset does not contain students' complete preference lists of colleges and majors, we do observe whether a student is admitted to their first-choice university. In Appendix Table A3, we show that our conclusions based on the finally admitted college are robust to restricting to those students admitted by their first-choice university.

Since our analysis focuses on migration for college, we exclude the 38.7% of students who were not admitted to any college and therefore lack a well-defined migration outcome.[12] We further exclude students with missing information on crucial variables such as NCEE scores, academic track, and home origin county, resulting in the exclusion of an additional 3.46% of observations. After these refinements, our final analytical sample consists of around 13.5 million students.

Our main outcome variable—college migration—is defined as a dummy variable equal to one if the admitted college is located in a different province from the student's home province. We use provinces as the relevant geographic boundary for several reasons. First, provinces rep-

[11] While our data specifies a student's hukou type, it does not contain their hukou registration county. We overcome this data limitation by using the county where each student took the exam as their home county, since students are required to take the NCEE in their hukou registration location.

[12] We later verify the robustness of our findings by defining non-admitted students as non-migrants in a separate analysis.

resent significant administrative, economic, and cultural units in China, with provincial boundaries often demarcating distinct labor markets, policy regimes, and social networks.[13] Second, the NCEE is administered provincially with scores comparable only within each province-track. Third, related to the second point, universities set province-specific admission quotas, allocating a predetermined number of seats to students from each province. This quota system means that students compete only against other applicants from their home province. Provincial boundaries are therefore not merely geographic divisions but are structurally embedded in the admission process itself, making them the natural unit for analyzing college migration decisions.

In addition to this extensive margin of migration, we also compute the straight-line geographic distance between the city where the university is located and the city of the student's hometown.[14] This intensive margin of migration measurement captures migration intensity, includes within-province moves, and reveals both whether and how far students travel for college education.

Table A1 provides summary statistics for key variables, separately for boys and girls. Several patterns are worth noting. There are significantly more boys than girls in the analytical sample, with girls accounting for only 44% of the sample, yielding a male-female sex ratio of 1.27 (= 56/44). This gap is not an artifact of our sample restrictions but rather reflects deep-seated demographic and social patterns.[15] First, there is a skewed sex ratio at birth of approximately 1.08 for the relevant birth cohorts (1980-1985), due to the widespread son preference in China (Almond, Li and Zhang, 2019). Second, girls in China, especially those from rural households, have lower probabilities of matriculating to or completing senior high school and are thus less likely to become eligible to take the NCEE (Zeng et al., 2014; Connelly and Zheng, 2003).

As shown in Table A1, boys are 6 percentage points more likely to choose a university outside their home province and migrate 19% farther on average. However, this gender gap in college migration cannot simply be attributed to differences in migration preferences or environmental constraints. Boys are simultaneously more likely to be in the science track and have slightly higher NCEE scores and rankings. Comparing students with identical scores from

[13] As one concrete example of the salience of provincial boundaries, cross-provincial phone calls incurred substantial extra charges during the data period, making communication across provincial lines more costly and reinforcing the economic significance of these boundaries.

[14] The distance is coded as zero if the university is located in the student's hometown city.

[15] Girls account for 42% of the sample in the raw data without any sample restrictions.

the same province-track is therefore essential. Notably, boys are also more likely to come from rural households, likely reflecting stronger sex selection at birth in rural areas and positive selection among girls who successfully reach the stage of taking the NCEE, given that rural girls are less likely to complete high school and appear in our sample.

Additionally, we draw on multiple complementary data sources to examine different dimensions of the analysis. We use the Population Census/Survey and the China General Social Survey (CGSS) to measure local social norms associated with gender differences in college migration; we conduct a self-designed survey to investigate parental attitudes towards their children's migration for college; and we use the China College Student Survey (CCSS), which provides detailed family background and labor market outcome information on college graduates, to examine gender differences in family background and estimate the wage migration premium. The specific details and use of each dataset are described in the corresponding sections of the paper.

4. Gender Differences in College Migration

4.1. Main Results

4.1.1. Graphical Evidence

We begin our analysis by presenting the raw gender differences in college migration for each rank percentile. To do so, we first rank students based on their raw NCEE scores within each year–province–academic track and divide them into 100 groups according to their percentile ranks, with higher percentiles indicating higher rankings. For each percentile, we then compute the average migration indicators separately for boys and girls. Figure 2 Panel (a) shows the results for migration out of the home province for college education, our main outcome of interest. Circles represent girls, and triangles represent boys. For both boys and girls, students with higher percentiles are substantially more likely to migrate out of province, from around 15% for the bottom-ranked students to over 70% for the top-ranked students. This pattern implies that broader access to high-quality universities increases the propensity to leave the home province.

Furthermore, boys are more likely to go out of province for college education compared

to girls at the same CEE percentile.[16] The gap is persistent across most ability levels (except the bottom 20%). As shown in panel (b) of Figure 2, which calculates raw gender differences across rank percentiles, the gender gap increases for higher-ranking students, indicating a larger gap when students face a less constrained set of eligible universities. The gap still exists but becomes somewhat smaller for the very top students. This could reflect the spatial concentration of elite universities in a few major cities, such as Beijing and Shanghai, which are usually different from their home provinces.[17] Therefore, for top-ranking students who are very likely to be admitted by the elite universities, staying local would mean a substantial downgrade in university quality, making the migration probability very high for both boys and girls. In contrast, for non-top students, the quality penalty from staying local is smaller, allowing other factors to play a larger role in their decision-making.

In panels (c) and (d) of Figure 2, we show similar patterns of the results using the alternative migration measurement, the distance between the university city and the hometown city. There is a significant gender gap in migration distance, especially for students in the middle- and high-percentile ranks. Taking students at the 40th, 60th, 80th, and 100th percentiles as examples, boys, on average, migrate approximately 42, 53 , 72, and 16 kilometers further from home than girls,respectively-—corresponding to about 2.2% to 17.9% greater distances for boys.

4.1.2. Regression Analysis

One potential concern in Figure 2 is that boys and girls at the same CEE percentile may come from different cities with varying hometown university quality or levels of economic development. If so, the observed gender difference in Figure 2 might be due to different "necessities" to migrate for boys and girls from different hometowns. For example, in the sample of CEE test takers, girls are more likely than boys to come from provincial capitals, the top-tier developed cities in each province.[18]

To ease the concern, we compare the same CEE ranking boys and girls from the same

[16] In Appendix Figure A2, we examine the gender gap in college migration separately for students in the STEM and humanities tracks. We observe a similar pattern across both tracks. This suggests that the aggregate gender gap is not driven by the disproportionate representation of male students in the STEM track.

[17] For example, among the 39 most elite universities (Project-985 universities) designated by the Ministry of Education in China, 12 are located in Beijing and Shanghai, and 35 are in provincial capitals or provincial-level cities (Beijing, Shanghai, Tianjin, and Chongqing). Of the top 10 universities in China (QS ranking 2004), 5 are in Beijing and Shanghai.

[18] This result is likely due to the positive selection of girls entering the sample of students admitted to universities. For example, parents of girls in provincial capitals may be more likely to invest in their daughters' education compared to parents in more remote areas, whereas this difference might be minimal for boys.

hometown who are eligible to be admitted by the exact same set of universities. We estimate the following equation controlling for hometown and CEE rank percentile:

$$Migration_i = \beta \cdot Female_i + X_i'\gamma + \phi_{R_i} + \psi_{cst(i)} + \varepsilon_i \tag{1}$$

where $Migration_i$ is a measure of college migration for a student i. $Female_i$ indicates whether student i is female. We also include a set of demographic controls, X_i, including hukou status, ethnic minority indicator, and age at the time of taking the NCEE. R_i is the percentile rank (grouped into 100 percentiles) of the individual i within the corresponding province-track-year, the level at which students are ranked in the college admission process based on their CEE scores. We flexibly control for the student's percentile rank through a full set of percentile rank fixed effects,ϕ_{R_i}, which non-parametrically control for the CEE performance.

Once we observe student i, we know his or her home county c, track s, and the year t when they took the NCEE. Therefore, we also include county-track-year fixed effects, $\psi_{cst(i)}$, to account for common shocks affecting college choices and migration decisions of individuals in the same home county-track-year. Our coefficient of interest is β, which shows the gender gap in college migration conditional on the same home county-track-year and the same performance on the college entrance exam. Standard errors are clustered by home county.

Table 1 presents the main results regarding the gender gap in migration for college education. The dependent variable is an indicator of attending university outside the hometown province, with the female dummy being the primary variable of interest. In column (1), without any controls, the results indicate that girls are 6.2 percentage points less likely to migrate out of the province for college education compared to boys. This gender gap is substantial, representing 25% of the dependent variable mean (0.248).

However, as argued above, this result might reflect that boys and girls are likely to come from different types of hometowns, which may have varying university quality and levels of economic development. Furthermore, students in China are required to choose between science and humanities tracks during high school, and there is a significant difference in the proportion of girls between these two tracks. According to our data, the proportion of females in the humanities is 60.11%, while in the sciences it is 37.42%. Consequently, the gender gap observed might be influenced by differing migration necessities or premiums associated with these tracks. To address this concern, column (2) includes county-year-track fixed effects, ensuring that the comparison is made between boys and girls who take the CEE in the same

county, year and academic track. The estimated coefficient of the female indicator decreases to -0.051, which remains statistically significant at the 1% level. The reduction in the gender gap estimate suggests that the different migration premiums between the science and humanities tracks or across counties account for some of the observed gender gap in column (1). However, a substantial portion of the gap remains unexplained by track or county differences.

In column (3), we further include the percentile rank of each student's CEE score alongside the county-year-track fixed effects. Consistent with the findings in Figure 2, panel (a), the coefficient for the percentile rank (0.004) is positive and statistically significant, indicating that improving a student's ranking by 1 percentage point increases the probability of migrating out of the province by 0.4 percentage point. The inclusion of the percentile rank results in a further reduction of the female coefficient to -0.045. This suggests that after controlling for both track selection and academic performance, the gender gap in migration remains substantial. The magnitude of this gap is comparable to the effect of an 11 percentage point drop in ranking for female students (-0.045/0.004).

Column (3) assumes a linear relationship between percentile rank and migration, which might oversimplify the relationship between academic performance and migration decisions. To address this, in column (4), we show that the results are robust to including percentile rank fixed effects by grouping students into 100 percentile categories based on their rank within each province-year-track, allowing for a more flexible control of the students' academic ranking.[19].

In column (5), our preferred specification, we further add student demographic characteristics, including dummies for agricultural hukou, ethnic minority, and age of taking the CEE exam. The corresponding result indicates that girls are, on average, 4.5 percentage points less likely to migrate out of their province for college education compared to boys from the same county-year-track who are eligible for a comparable set of university admissions. This remaining gender gap, after including the full set of controls, is statistically significant at the 1% level and represents 17.9% of the dependent variable mean.

To go beyond the average results presented in Table 1 and explore the heterogeneity of the gender gap across different levels of academic ability, we provide a more granular analysis by splitting the full sample into 100 percentile groups. Separate regressions are run for each percentile subsample based on equation (1). Figure A4 displays the regression coefficients of the female indicator for each percentile subsample, using out-of-province migration as the de-

[19]The results are highly robust if we group the percentile rank into 200 or 500 groups and include the corresponding fixed effects, as shown in Appendix Table A3.

pendent variable. The figure reveals a pattern consistent with the descriptive analysis in Figure 2, where the gender gap widens for students with higher CEE scores, except for a reverse trend among those in the top 20% of the distribution. This aligns with the reason provided to explain Figure 2. Higher CEE scores give students more options, allowing more differing choices between boys and girls. However, due to the geographical concentration of the most elite institutions, staying local means a significant sacrifice in university ranking for top students, making other factors less significant.

4.2. Robustness of Main Results

In this section, we show that the gender gap in college migration is highly robust to alternative migration measures, specifications, and sample constructions.

We begin by testing robustness to alternative definitions of migration. The baseline analysis focuses on out-of-province moves. Column (1) of Table 2 instead examines within-province migration, measured by moves to the provincial capital among students who enroll within their home province and are not originally from the provincial capital. Girls are 3.5 percentage points less likely than academically comparable boys from the same county–year–track to migrate to the provincial capital—7.3% of the dependent-variable mean—confirming the gap exists even for shorter-distance moves, though its magnitude is smaller than the primary out-of-province gap. Column (2) replaces the binary outcome with the university–hometown distance, capturing the intensive margin that pools within- and out-of-province moves. We find that boys, on average, enroll at institutions located about 12.2% further from home.

Columns (3) and (4) use indicator variables for migrating to high-opportunity destinations: (i) any top-30 GDP city and (ii) any top-10 GDP city. For each column, students originally from the corresponding set of cities are excluded. The estimates show that female students remain less likely than male students to migrate to economically developed cities. Specifically, females are 1.9 percentage points less likely to move to top-30 GDP cities and 0.8 percentage points less likely to move to top-10 GDP cities, corresponding to 12.9% and 8.6% of the respective means.

In comparison, the gender gap in out-of-province migration represents 17.3% of the dependent mean. The narrower gender gap for migration to economically developed cities likely reflects larger benefits from studying there (elite universities, stronger labor markets, broader networks), which attenuate gender-specific constraints. As the benefits of migration become

more substantial, gender-specific constraints that influence migration decisions will play a relatively smaller role, resulting in a smaller gender gap in these contexts. This pattern parallels our finding of a smaller gender gap among the very top students, for whom the expected returns to migration are especially large.

To assess the robustness of our findings, we conduct a series of additional checks addressing potential concerns related to sample construction, score measurement, and institutional features of the NCEE admission system and report the results in Appendix Table A3. First, because our baseline sample focuses on students admitted to college, the estimated migration gap could be affected if admission probabilities differ by gender (as shown in Appendix Figure A3). We therefore re-estimate the analysis using the full population of high school students, coding those not admitted to college as non-migrants. Second, to ensure that comparisons across students with similar academic ability are not sensitive to the discretization of score rankings, we adopt finer divisions of NCEE score percentiles (200 and 500 groups instead of the baseline 100 groups). Third, since our data only record final admission outcomes but not the full preference lists submitted by students, gender differences in application strategies could potentially affect migration outcomes. To address this concern, we restrict the sample to students admitted to their first-choice university, whose admissions more directly reflect intended choices rather than fallback placements. In addition, we perform several checks to account for institutional features of the NCEE system that may affect migration patterns. We exclude students admitted through special examination tracks (such as sports, arts, or ethnic minority admissions) whose admission rules differ from the standard academic track. We also exclude students whose high school province differs from their NCEE exam province to rule out potential bias from pre-college migration. Furthermore, because males and females may differ in their likelihood of receiving bonus points that affect total NCEE scores, we re-estimate the results after excluding students who received bonus points. Finally, to allow for more flexible comparisons across students with similar academic performance, we include province–track–year–percentile fixed effects to account for differences in the meaning of a given percentile rank across provinces, tracks, and cohorts. Across all these exercises, the estimated gender gap in college migration remains large and statistically significant.

4.3. Heterogeneity by Family and Hometown Characteristics

Having documented that female students are, on average, less likely to migrate for college than their male counterparts, we now examine how this gender gap varies with family and hometown characteristics. In the main text, we focus on the heterogeneity driven by rural–urban status and prevailing hometown gender norms.

We begin by investigating heterogeneity according to family rural–urban *hukou* status. Because educational attainment and income are generally lower in rural areas, students with a rural *hukou* are more likely to come from families with lower socioeconomic status (SES). In column (1) of Table 3, we present the interaction between the female indicator and an agricultural (rural) *hukou* indicator. the coefficient on the interaction term is similar in magnitude to the main female effect, implying that the gender migration gap is nearly twice as large for rural students as it is for urban students.

It is important to note that the heterogeneity by *hukou* status likely captures more than just differences in parental SES; it also reflects variation in hometown gender norms, as these norms tend to be more conservative in rural China than in urban areas.To isolate the cultural channel, we now turn to a direct investigation of how the gender gap in college migration responds to more explicit measures of hometown gender norms.

The effect of an environment with more traditional gender norms on this gender gap is theoretically ambiguous. On one hand, restrictive gender norms could limit female mobility (the "constraint effect") by decreasing girls' aspirations, lowering the willingness to invest in their human capital, or imposing additional psychological costs on moving away from parents. On the other hand, a local environment with a stronger bias against females may incentivize girls' migration to areas with greater equality (the "incentive effect").

To more explicitly measure local gender norms, we construct three different county-level gender norm measurements. The first one is the sex ratio at birth. Based on the *Tabulation on the 2000 China Population Census (By County)*, we calculate the male-to-female sex ratio at birth for children aged 0–4.[20] A higher sex ratio indicates a stronger preference for sons and more conservative local gender norms (Edlund, 1999; Choi and Hwang, 2015).

The second and third gender norm measurements are derived from the China General Social Survey (CGSS). Similar to the General Social Survey in the US, CGSS collects information on

[20]For sex ratios at birth, ideally, we would use data for 0-year-olds in the 2000 Census. However, the county-level tabulation only reports the number of boys and girls aged 0-4 instead of a single-year age 0.

social behaviors, beliefs, ideology, and lifestyles. Using the 2010 CGSS data, which covers 12,000 households in 122 counties, we calculate the county-level average agreement with the following two statements separately: (1) "Men should focus on career, while women should focus on family," and (2) "Marrying well is more important for women than working well." Responses are rated on a scale from 1 (strongly disagree) to 5 (strongly agree), where higher averages indicate stronger traditional gender norms in the local county. For comparability and ease of interpretation, the two county-level measures are then standardized to a mean of zero and a unit standard deviation.

We interact the female indicator with each gender norm measurement as key independent variables, with county-year-track fixed effects absorbing the single terms of county-level gender norm measurements. The results are summarized in Columns (2)-(4) of Table 3.[21] The coefficients of the interaction terms are significantly negative across specifications, suggesting that the gender migration gap is wider in counties with stronger traditional norms. In columns (1)-(3), a one-standard-deviation increase in these social norm measures expands the gender gap in college migration by 7.0% ($= \frac{0.003}{0.043}$), 17.9% ($= \frac{0.007}{0.039}$), and 15.8% ($= \frac{0.006}{0.038}$) of the baseline migration gap, respectively.

We interpret the interaction coefficients in Table 3 as the net effects of two opposing mechanisms: the incentive for females to migrate toward more gender-egalitarian regions versus the constraints imposed by local conservatism. The consistently negative coefficients imply that the binding constraints of traditional gender norms outweigh the "escape" motive.

However, it is important to note that a high sex ratio reflects not only son preference (culture) but also a scarcity of women. This scarcity could increase the bargaining power of women in the local marriage market, potentially creating an incentive to stay that is distinct from cultural restrictions. While using direct measurements of local gender attitudes in columns (2) and (3) eases this concern to some degree, it does not completely eliminate it, as sex ratios are often correlated with local gender attitudes.

To disentangle the heterogeneity driven by gender norms from that driven by sex-ratio-induced marriage market conditions, we estimate a specification in Table A4 that simultaneously includes the interactions of the female indicator with both the sex ratio and the CGSS gender attitude measurements. The coefficients on the interaction between the female indicator

[21] It's worth noting that, when matched with the CGSS data, the number of observations in columns (3)-(4) decreases, as the matched dataset includes only 122 counties. However, the estimated coefficients for the Female variable remain of similar magnitude across these different samples, indicating the robustness of the results.

and gender attitudes remain comparable to those in columns (2) and (3) of Table 3. This robustness suggests that the observed gender gap is not merely a rational economic response to potential favorable local marriage markets for girls, but is indeed influenced by the constraints of traditional gender norms.

5. Causes of Gender Gap in College Migration

In this section, we explore potential reasons that could explain the observed gender gap in migration for college. Specifically, we examine the following candidates that may explain the gender gap in migration we documented in Section 4: 1) Gender differences in parental attitudes such as elderly care expectations, perceived returns, and safety concerns regarding migration; 2) gender differences in decision-making under incomplete information and uncertainty ; 3) gender differences in family background.

5.1. Gender Differences in Parental Beliefs and Concerns

5.1.1. Survey Design

The robust gender gap in migration documented in Section 4 is based on the cohorts taking the NCEE between 1999 and 2003. Given China's rapid economic development and shifting cultural attitudes over the past two decades, a natural question is whether this gender gap in migration preferences persists today. Furthermore, guided by our previous finding that the gap is wider in counties with stronger traditional gender norms, we aim to unpack the broad concept of gender norms to identify the specific constraints, concerns, and trade-offs that drive gender differences in migration decisions.

To address both questions, we designed and conducted an original online survey in May 2025 in collaboration with a professional survey company. The survey specifically targets parents with children currently in the final two years of senior high school in China (grade 11 or 12), whose children will take the NCEE approximately one year later or one month after the survey. This timing is strategically chosen to capture families at a critical decision-making period. By May of their final or penultimate high school year, families are deeply engaged in NCEE preparation, with university selection often among the most discussed household topics. Respondents are therefore more likely to offer well-considered views on university selections.

Due to budget constraints, we only survey parents in Guangdong, Hubei, and Shaanxi provinces with targeting sample sizes proportional to their respective populations of high school graduates.[22] We distributed 3,995 questionnaires and collected 2,992 valid responses, achieving an average response rate of 74.9%.[23] This high response rate is likely attributable to the high salience and relevance of the topic for those parents. After excluding observations with invalid high school city information, we obtained a final sample of 2,940 parents, with 1,447 respondents from Guangdong, 980 from Hubei, and 513 from Shaanxi.

The survey questionnaire collects demographic and academic information about the child, including the child's gender, number of sisters and brothers, high school city, academic track (science or humanities), and parental expectations regarding university attendance (e.g., Project-985 university, Tier 1 university, vocational college). We also collect information on parental background, including the gender of the survey respondent (whether the father or mother was surveyed), hukou type (agricultural or non-agricultural), education level, and monthly income (in bracketed categories). Appendix Table B2 summarizes the descriptive statistics of the basic information for children and parents in our sample. Notably, we observe a reasonable distribution of expected university levels: 16% of parents expect their child to be admitted to highly prestigious universities (Project-985 institutions), while the majority (58%) expect their children to attend normal Tier 1 or Tier 2 universities. This distribution suggests that our sample provides relatively representative coverage of students across various academic performance levels.

We construct a measure of college migration preference based on the following survey question: "If your child could enter a university in the hometown province, a university in another province (excluding Beijing and Shanghai), or a university in Beijing or Shanghai, assuming the universities have the same quality and same level, how would you like your child to choose? Please rank the following three options from most preferred to least preferred." The dummy variable migration preference takes the value of 1 if the parent ranks the local-province university below both alternatives. The summary statistics in Panel A of Appendix Table B3 suggest that girls' parents are 5 percentage points less likely to prefer outside home province

[22] We selected Guangdong, Hubei, and Shaanxi provinces for two main reasons. First, they represent the eastern, central, and western regions of China with distinct levels of GDP per capita, thereby capturing geographic and socioeconomic diversity. Second, all three provinces host several good universities, including Project-985 and Project-211 institutions, which ensures that parents face a meaningful choice between encouraging children to remain in the local province versus migrating out of the province for higher education.

[23] Only respondents who took more than 120 seconds to complete the questionnaire are considered valid responses.

universities compared to boys' parents.

We inquire about specific concerns and trade-offs regarding college migration only after parents have indicated their general location preference, in order to prevent their university location preference rankings from being biased by the salience of specific concerns raised in the subsequent questions. In this part, we ask parents to rate the degree of their concern regarding safety, the ability to provide timely assistance, financial costs in attending a university in another province, and finding a spouse if their child were to migrate. We also ask whether they have a strong preference for their child finding a spouse from the local province, and solicit their expectations regarding future elderly care, specifically distinguishing between financial support and time spent accompanying them. The summary statistics of corresponding variables are summarized in panel B of Appendix Table B3, and the exact wording of each question is attached in Appendix Table B1.

One important factor in migration decisions is the perceived benefit of migration. We ask parents whether they believe attending university outside the province offers greater benefits for boys, for girls, or similar benefits for both. This question directly elicits parents' subjective assessments of whether out-of-province education generates gender-specific returns. Based on the gender of the respondent's child, we transform this categorical response into a binary indicator of whether parents believe migration benefits are larger for their child's own gender than for the opposite gender. As shown in Panel B of Appendix Table B3, one quarter of boys' parents believe migration benefits are larger for boys, while only 7.4% of girls' parents believe migration benefits are larger for girls.

5.1.2. Gender Differences in Parental Preference for Colleges Migration

We begin our analysis by examining whether contemporary parents exhibit gender differentiated preferences for out-of-province college attendance, mirroring the patterns documented in our administrative NCEE data from approximately 25 years ago. In Table 4, we regress the migration preference indicator on an indicator for child being a girl. Column (1) reports the regression without any controls, while column (2) controls for background characteristics. Column (3) includes a complete set of fixed effects for the highest university level parents realistically expect their child to be able to attend.[24] This set of controls ensures the estimated female coefficient represents a comparison between boys and girls who are expected to attend

[24] University levels are categorized as the following mutually exclusive levels: Project 985; Project 211 (non-985); Tier 1 (below 211); Tier 2; vocational; no college.

similar levels of universities, mirroring the percentile rank fixed effects in our main specification using administrative data.[25] Also similar to our main specification using administrative data, column (4) adds fixed effects for the interaction between home city and academic track.

The regression results in Table 4 show that parents with higher education have stronger migration preferences. Additionally, parents exhibit weaker college migration preferences when they have only one child. These intuitively sensible patterns in the control variable coefficients support the construct validity for our migration preference measure.

More importantly, across all four specifications in Table 4, the coefficient on the female child dummy is consistently negative and statistically significant, indicating that parents are less likely to prefer a out-of-province university for daughters than for sons. Focusing on our preferred specification in column (4), the estimated gender gap in migration preference is 5.4 percentage points. This magnitude is reasonably similar to the gender gap of 4.3 percentage points documented in Table 1 using administrative data approximately 25 years ago.[26] Such robust results suggest a persistent gender gap in university location preferences in China, despite the significant economic development and rising female educational attainment.

5.1.3. Potential Factors in Migration Preference

We next examine how parental attitudes toward migration concerns and benefits correlate with the preference for their child to attend college out of the home province. We classify parental attitudes into four key dimensions: marriage market considerations, expectations for elderly care, perceived benefits of out-of-province study, and concerns regarding safety, cost, and timely assistance. We regress the migration preference indicator on these factors jointly, while controlling for demographic controls, income-level fixed effects, parental expectations of university tier, and home–city–by–academi–track fixed effects.

The results based on both genders are summarized in column (1) of Appendix Table B4 and highlight several patterns. First, marriage market considerations play a prominent role: parents who express concerns about their child finding a spouse if they migrate are 3.8 p.p.

[25] One concern is that parental expectations regarding children's eligible university level may be systematically associated with their stated preferences for out-of-province colleges. For instance, parents of high-achieving students might more readily think about distant elite universities. These expectations could differ systematically by child gender due to actual performance differences or biased parental beliefs. As shown in Appendix Table B3, parents of boys hold higher expectations for their children to attend a Project-985 university compared to parents of girls.

[26] When considered relative to the mean, the gap in Table 1 represents 17.1% of the mean migration rate. In this survey analysis, the corresponding ratio is 13.7%($= \frac{0.054}{0.394}$). Both calculations indicate that the gender gap represents approximately 14-17% of the relevant baseline rate.

less likely to prefer out-of-province universities, and parents who specifically prefer their child to find a spouse from the local province are 10.6 p.p. less likely to prefer out-of-province universities. This pattern suggests that marriage market concerns, particularly preferences for geographically proximate marriage partners, constitute an important driver of local college preferences.

Second, while parents expecting financial support from their children show no significant association with migration preferences, parents expecting more time from their children to accompany them exhibit significantly weaker migration preferences (8 percentage points). This distinction makes intuitive sense. Studying outside the province is often associated with long-term migration away from the hometown, while companionship requires physical proximity. In contrast, geographic distance does not pose a significant barrier to providing financial support to parents.

Third, parents who believe out-of-province study offers greater benefits for their child's gender are significantly more likely to prefer out-of-province universities (9.7 percentage points higher), indicating that perceived returns to migration are a key determinant of location preferences. Finally, practical concerns about safety and parents' ability to provide timely assistance are both negatively and significantly associated with migration preferences (6.6 and 5.9 percentage points, respectively), while cost concerns show no significant relationship.

To investigate whether the determinants of migration preferences differ systematically by child gender, in columns (2) and (3) of Appendix Table B4, we conduct the same analysis separately for parents of boys and girls. Overall, the results reveal that the factors affecting migration preferences operate in broadly similar ways for parents of both boys and girls. However, one notable disparity worth mentioning is that cost concerns are statistically significant predictors of migration preferences for parents of boys but show no significant effect for parents of girls. This asymmetry is particularly informative because it contradicts a potential explanation for gender gaps in migration: namely, that parents of girls exhibit weaker migration preferences because they are more cost-sensitive or less willing to invest in their daughters' education. If this explanation were correct, we would expect cost concerns to be more strongly associated with migration preferences for parents of girls. The opposite pattern suggests gender gaps reflect non-financial factors rather than a stronger willingness to invest in sons.

5.1.4. Gender Differences in Perceived Migration Concerns and Benefits

Now we examine whether there are systematic gender differences in these attitudes. If parents hold different concerns, perceived migration benefits, or preferences when thinking about daughters versus sons, these gender-differentiated attitudes could serve as mechanisms explaining the observed gender gap in college location preferences. Table 5 reports regressions of each parental attitude on a female child indicator. We include the same comprehensive set of controls employed in our previous analysis: demographics, income-level fixed effects, expected university-level fixed effects, and home city-by-academic track fixed effects.

Focusing on columns (1) and (2) regarding marriage market concerns, a striking result emerges. Parents of girls are significantly more likely to desire a spouse from the home province, yet they report lower levels of concern regarding their child's general marriageability in distant markets. This could reflect the "marrying up" tradition in China (Rossi and Xiao, 2025). Since most students who migrate attend universities in more economically developed cities, this directional pattern of migration creates asymmetric marriage market opportunities by gender.[27] In the Chinese marriage market, it is generally easier for women from less developed regions to marry men from more prosperous cities than for men to "marry up" geographically. For sons, the marriage market is often more rigid; parents of boys may anticipate that their sons will struggle to compete with local men in destination provinces—especially if local girls' parents strongly prefer "home-grown" sons-in-law, as shown in column (2).

The fact that parents of boys express greater concerns about spouse-finding in destination provinces indicates that this particular mechanism cannot explain—and in fact works against—the observed gender gap in migration preferences. In contrast, the stronger desire for daughters to marry someone from the home province operates in the expected direction and contributes positively to explaining why parents exhibit weaker migration preferences for daughters, regardless of the daughter's perceived marriageability elsewhere.

Turning to columns (3) and (4) of Table 5 regarding elderly care, the results indicate that parents of daughters have lower expectations for post-retirement financial support and report similar levels of expectation for time and companionship support compared to parents of sons. Given that financial concerns show no significant association with migration preference and that expectations for companionship are statistically similar across genders, it appears that elderly

[27] In our CEE administrative data, 73.9% of students of students who migrate for college attend universities in cities with higher GDP per capita compared to their hometown.

care considerations play only a limited role in explaining the gender gap in college migration.

Columns (5)-(7) of Table 5 focus on other parental concerns if their child migrates. Understandably, parents of daughters are significantly more likely to express safety concerns than parents of sons, a factor that likely contributes to the gender gap in college migration. This difference in safety concern could reflect both objective differences in vulnerability–particularly in unfamiliar environments where they lack established social networks and local knowledge—and potentially gendered cultural norms about appropriate protection of daughters versus sons. Meanwhile, there are no clear gender differences in concerns about the financial cost of migration or the ability to provide timely assistance.

To quantify the relative importance of these mechanisms, we combine the estimated coefficients from Table 5 and Appendix Table B4. Assuming these migration concerns shift migration preferences according to the full-sample estimates in Column (1) of Appendix Table B4, eliminating the gender differences in those concerns would reduce the overall gender gap in migration preference by: (1) 13.3% ($=\frac{0.068\times0.106}{0.054}$) due to local spouse preferences; (2) -7.4% due to concerns about finding a spouse; (3) 7.9% due to safety concerns; and (4) 2.9% due to elderly care companionship expectations. If we instead use the girls-only sample coefficients from Column (3) of Appendix Table B4 to calibrate these shifts, the reduction in the gender gap is even more pronounced: 17.0% for local spouse preferences, -9.3% for general spouse-finding concerns, 10.2% for safety concerns, and 2% for elderly care companionship expectations.

It is important to recognize the limitations of this decomposition exercise, which is fundamentally based on the associational relationships estimated in Appendix Table B4 rather than causal effects. Therefore, the absolute magnitudes should be interpreted with considerable caution. However, cross-concern comparisons remain informative for understanding the relative importance of various concerns. The fact that local spouse preferences and safety concerns generate substantially larger numbers than elderly care companionship expectations, and that spouse-finding concerns exhibit negative explanatory power, provides useful qualitative information about which concerns are likely more important for the gender gap in migration preferences.

Column (8) of Table 5 examines gender differences in the perceived benefits of migration. Parents of daughters are significantly less likely to believe that out-of-province study yields greater advantages for their child's gender. This suggests that parents systematically evaluate the potential advantages of migration as less applicable or less valuable when the migrating

child is a daughter rather than a son.

Following the same decomposition approach as for the concern variables, we find that eliminating gender differences in perceived migration benefits could reduce the gender gap in migration preference by 30.4% using the full sample estimation results or 51.0% using the girls' sample estimation results. This represents a substantially larger contribution than any of the individual concern variables examined earlier. Such significant figures qualitatively show the perceived migration benefits as a critical mechanism driving gender gaps in college location preferences.

The discrepancy in perceived benefits may be driven by three distinct mechanisms. First, it could accurately reflect real gender differences in migration returns due to labor market discrimination. Under this interpretation, the gender gap in perceived benefits could be a rational response to discriminatory labor market structures. As detailed in Section 6, our causal estimates—leveraging university quotas as a quasi-experimental instrument—reveal that actual economic returns to migration are remarkably consistent across genders regarding university quality and initial wages. It suggests that the gender gap in perceived benefits is unlikely to be a rational response to discriminatory labor market structure. Admittedly, parents may still be acting rationally if they anticipate long-term gender disparities beyond the scope of our data, such as slower promotion trajectories, earlier retirement, or higher chances of career interruptions for women (Cortés and Pan, 2023). It is still possible that parental beliefs might reflect accurate expectations of lifetime returns rather than pure misperception, even if initial returns appear similar.[28]

Second, given the similar short-term return to boys and girls, the difference in perceived migration benefits could also reflect misperceptions or misinformation that parents underestimate the migration return for girls. Under this interpretation, the gender gap in perceived benefits represents an information failure that could potentially be corrected through role model effects or interventions that update parental beliefs. Third, the perception gap may reflect differential subjective valuations: parents of girls may weigh material benefits less heavily than non-pecuniary factors like job security or work-life balance, even while believing material returns are similar. This represents preference differences rather than information gaps. Disentangling and evaluating the relative importance of these three explanations more cleanly—through experimental information interventions, long-term career tracking, or detailed preference elici-

[28] Fully assessing this possibility would require longitudinal data on lifetime earnings and labor force participation patterns, which exceeds our current data scope.

tation—represents an important direction for future research that could yield actionable insights for policy interventions aimed at reducing gender gaps in college migration.

5.1.5. Limitations of the Survey Study

Two limitations of the survey study should be noted. First, our survey measures parents' preferences without measuring children's own preferences.[29] While we acknowledge that parents cannot fully dictate their children's college choices, they nevertheless exert considerable influence, especially in the Chinese context, where strong respect for parental authority remain culturally salient. In the China College Student Survey (CCSS) conducted by the China Data Center of Tsinghua University, college students are asked to rank the relative importance of different individuals (themselves, parents, teachers, peers) in their college application decision-making process.[30] The response patterns reveal both high overall parental influence and striking gender differences in decision-making dynamics.

Among CCSS respondents, 45.7% of girls rank parents as the most important decision-maker, while only 35.7% of girls rank themselves as the most important decision-maker. This pattern—girls identify their parents rather than themselves as the primary driver of their college choices—corroborates that parents play a dominant or co-dominant role in daughters' application decisions in China. and suggests that parental preferences likely translate quite directly into daughters' realized college choices. Interestingly, 43.1% of boys rank themselves as the primary decision-maker compared to 38.7% who rank their parents first. Although parents clearly still play a significant role in college application decisions for boys, parents' opinions appear to carry relatively less weight in sons' decisions compared to daughters' decisions.

This gender difference in the relative importance of parental influence introduces another

[29] Ideally, we would survey both parents and students to understand the preferences of both parties toward university migration and to examine potential intergenerational differences in attitudes toward studying outside one's home province. However, we face a practical constraint. Since students in our target population are not yet adults, the survey company informed us that surveying minors requires approval from both the student's parents and their schools. Given the heavy administrative burden associated with such procedures, we choose to survey exclusively among parents

[30] The CCSS was conducted in May and June of each year (from 2010 to 2015) to collect information on colleges and graduates. The CCSS first randomly selected colleges using the method of stratified random sampling, with locations (Beijing, Shanghai, Tianjin, Northeastern China, Eastern China, Central China, and Western China) and type of colleges (1-7 Tier colleges) as the stratifying variables. In 2010, a pilot study included 19 randomly selected colleges from a total of 2,305 nationwide. This number expanded to 50 colleges in 2011 and 2012, and then to 65 in 2013. In 2014 and 2015, an unexpected cut in financial support resulted in less than 20 colleges participating in the survey. For each selected college, about 300 students from the graduating class were randomly chosen. The survey questionnaire, designed collaboratively by experts in economics, sociology, and education, collects not only the basic information on students' characteristics and their family backgrounds but also their CEE scores and college activities.

potential mechanism of the observed gender gap in college migration. Under a plausible assumption that parents on average hold stronger local university preferences than their children, gender differences in the relative importance of parental preferences would also generate gender gaps in children's migration outcomes even with identical parental preferences across children's genders. Unfortunately, without directly measuring students' local preferences in our survey, we are unable to quantify the relative importance of "who decides" mechanism from the "what parents prefer" mechanism. [31]

The second limitation is the temporal gap between our primary analysis sample (1999–2003 cohorts) and the survey (May 2025). While the overall magnitude of the gender gap in migration preferences has remained stable over two decades, the underlying mechanisms may have evolved significantly. For example, the safety concern serves as a prime example of this evolution as it likely plays a much smaller role today than it did twenty years ago, following a fundamental improvement in China's public security.[32] Consequently, while contemporary parents still express differential safety concerns for daughters, these concerns were likely far more acute and a far more potent driver of the migration gap two decades ago than they are today.

5.2. Gender Differences in Decision-making under Uncertainty

Beyond differential parental attitudes between boys and girls, another potential explanation for the gender gap in migration for college is gender differences in decision-making under uncertainty, given the substantial amount of uncertainty and risk embedded in the college application process in China. For example, during our data period, students in most provinces were required to submit their rank-ordered lists of universities before knowing the score or even before taking the exam. Moreover, as introduced in detail in Section 2, the sequential processing rules of the admission system mean that failing to gain admission to one's first-choice university can result in a sharp drop in university quality rather than a smooth transition to a next-best alternative.

[31] Both channels may operate simultaneously and potentially reinforce each other: parents may hold weaker migration preferences for daughters than sons (as our survey documents) and have their preferences weighted more heavily by daughters than by sons in the ultimate decision-making process (as the CCSS data suggest).

[32] China's homicide rate plummeted from 2.04 per 100,000 people in 2002 to 0.50 in 2020—a 75% reduction that ranks the country among the safest globally (https://data.worldbank.org/indicator/VC.IHR.PSRC.P5). Data from the Institute for Economics and Peace indicates that residents in China now feel safer from violent crime than those in most other countries (https://www.economist.com/china/2023/11/23/china-says-it-has-achieved-a-miraculously-low-crime-society) .

Such high-stake environment with incomplete information may interact with gendered behavioral traits in two ways. First, it is well established that, on average, girls tend to be more risk-averse than boys (see Croson and Gneezy (2009) for a comprehensive review). As a result, they may favor "safer" in-province universities, which often have larger admission quotas for local students.[33] Second, if boys are more prone to overestimating their performance and relative rank, as documented by Niederle and Vesterlund (2007) and Adamecz-Völgyi and Shure (2022), such overconfidence may facilitate more aggressive application strategies by applying to more selective universities typically located in top-tier cities.

Being more risk-taking or overconfident can lead to a higher probability of listing universities beyond one's reach as the first choice, resulting in a lower likelihood of first-choice admission. Consistent with this, Appendix Figure C1, Panel (a), shows that girls have a higher success rate with their first-choice applications, suggesting a more conservative or more "realistic" selection process.

However, the gender gaps in first-choice admission (Appendix Figure C1 panel (b)) and the gender gap in college migration (Figure 2) show quite different patterns across percentile ranks. While the migration gap is negligible for low-ranking students and most salient around the 80th percentile, the gap in first-choice admission is most pronounced for those in the bottom 20% and the middle (40%–60%) of the distribution, fluctuating around zero for high-ranking students. This lack of alignment suggests that while gendered behavioral traits clearly shape application tactics and first-choice success rates, they are unlikely to largely account for the geographic migration patterns observed among top-tier students.

To test this mechanism more formally, we examine heterogeneity across varying admission policies that allow different levels of information when submitting the university preference. Provinces adopt one of three admission systems: (1) low-information: students submit their application forms prior to taking the CEE, (2) medium-information: after taking the CEE but before knowing their scores, or (3) high-information: after knowing their scores. In provinces where students submit their applications after knowing their scores, they also have a good sense of their ranking within the province-track-year, as the score distribution is typically an-

[33] For example, in 2008, Tsinghua University planned to admit 296 students from Beijing, accounting for 15.7% of its overall admission quotas (https://edu.sina.com.cn/gaokao/2008-04-15/1659134342.shtml). However, CEE takers in Beijing only accounted for 1.1% of the total CEE takers across the country. Similarly, in 2008, Shenzhen University allocated 86.7% of its quotas to the local province, Guangdong, while Guangdong CEE takers represented only 5.8% of all CEE takers nationwide (https://yz.szu.edu.cn/info/1041/10357.htm).

nounced.[34]

Moving from low- to high-information systems, there is a substantial increase in the amount of information available at the time of application submission, significantly reducing the inherent risk of the application process.[35] Furthermore, since students in high-information systems already know their scores and relative rankings, there is little remaining scope for gender differences in subjective confidence or self-assessment accuracy to influence application strategies. In summary, if overconfidence or differential risk aversion are important drivers of the migration gender gap, we should observe significantly larger gaps in lower-information systems.

Directly comparing provinces with different admission policies raises a concern that other regional characteristics may also be related to the gender gap in migration. To address this, we examine changes in admission policies within the same provinces during our study period. Between 1999 and 2003, 11 provinces altered their admission policies, allowing us to compare the gender gap within the same province under different policies at different times. The corresponding results are summarized in Appendix Table C1.

In column (1) of Appendix Table C1, we replicate our main specification on the sample of 11 provinces that changed their admission policies. The results are consistent with those from the main sample, showing a coefficient of -0.042. In column (2), the main coefficients of interest are $Female \times High\ Information$ and $Female \times Low\ Information$, with the gender gap under medium-information admission policies as the omitted group. We account for time-invariant regional characteristics associated with gender gap heterogeneity by including a full set of $Female \times Province$ dummies, which flexibly allow each province to have its own baseline gender gap level that can differ arbitrarily across provinces for any time-invariant reasons. If overconfidence or differential risk aversion are important drivers of the migration gender gap, we should expect positive and significant coefficients for $Female \times High\ Information$, indicating a less negative and therefore smaller gender gap, and a negative and significant coefficient for $Female \times Low\ Information$. However, we observe neither pattern in column (2). In column (3), we further include a complete set of $Female \times Year$ dummies to control for time trends associated with gender gap heterogeneity, and the results remain robust compared to those in column (2).

Taken together, two distinct and complementary pieces of empirical evidence point to the

[34] Normally, the number of students above each round score (e.g., 600, 610, 620) will be announced.

[35] Risk is not entirely eliminated even under the high-information system, as admission cutoffs for specific universities remain unknown beforehand.

same conclusion. First, there is a clear distributional misalignment between the gender gap in migration and the gender gap in first-choice admission. Second, providing students with more information before they submit their applications—thereby significantly reducing risk and uncertainty—fails to narrow the gender gap in migration. These findings consistently show that the observed disparity is unlikely to be driven by girls' higher risk aversion or boys' overconfidence during the application process.

5.3. Gender Differences in Parental Social-economic Status

Migrating for college education requires additional resources due to transportation costs and potentially higher living expenses in larger cities. Students from wealthier households are more likely to afford this migration. In this subsection, we examine whether gender differences in family background could drive the gender gap in college migration.

The relationship between gender and family socioeconomic status among CEE test takers in China remains theoretically ambiguous. On one hand, given the well-documented son preference in China, wealthier households might be more likely to afford sex selection interventions at birth, making boys more likely to come from affluent families. Additionally, having a son may alter parental behavior after birth. For example, parents with a son may tend to increase their savings rate (Wei and Zhang, 2011) or earnings (Li and Yi, 2015) in response to a competitive marriage market with a skewed sex ratio. On the other hand, if traditional son-preference values are more prevalent in less developed areas, boys might be disproportionately represented in lower-SES households. In addition, the positive selection of girls who successfully reach the CEE stage may mean that, in the CEE sample, girls could come from more affluent families than boys. These contrasting dynamics suggest that we need to investigate this empirically.

To explore the gender differences in family background, we resort to the China College Student Survey (CCSS) data, which collects information on college students' household income, parental SES background (such as education and hukou status).[36]. Appendix Table C2 reports the difference in family background between female and male college students.[37] Female students are more likely to come from families with higher family income, better-educated parents, and are less likely to be from rural areas.

[36]The details of this dataset are introduced in footnote 30

[37]Here, we include NCEE score percentile rank FE and city-year-track FE in the regression analysis to align with the main specification in Table 1. We do not exactly follow the main specification using county-year-track fixed effects due to the limited sample size in each cell. Percentile rank of each student is calculated using the CEE score reported in the survey and ranked among CCSS sample in that province-year.

Such findings suggest that girls are more likely to come from higher SES households, consistent with the positive selection of girls entering the entrance exam stage and the lower proportion of girls in the CEE sample documented in Section 3. It indicates that differences in family background cannot be the primary factor driving the lower likelihood of girls migrating for college education.

Another potential concern is that, in the Chinese context, girls are more likely to have siblings due to traditional son preference.[38] As shown in column (1) of Appendix Table C3, female students in our sample have, on average, 0.06 more siblings than their male counterparts. This raises the possibility of a 'resource dilution' effect: even if a girl's family has higher total resources, the per-child allocation might be lower. In column (2) of Table C3, we examine the gender difference in family income per child. While the coefficient of 0.064 is significantly lower than the corresponding coefficient in column (1) of Table C2, it remains positive and statistically significant. This suggests that while family size does somewhat dilute resources, female students are still not more likely to come from disadvantaged households even when measured by per-capita family income.

6. Consequences and Counterfactual Exercise

Having established a robust pattern of gender gaps in college migration and discussed potential drivers, this section examines the corresponding consequences. Geographic immobility mechanically restricts the university choice set, likely leading students to attend lower-quality universities. This effect can be exacerbated in disadvantaged regions where high-quality institutions are sparse. Beyond university quality, students who remain local, especially those from less developed provinces, are less likely to access superior economic opportunities and labor market networks in more developed cities. Data from the CCSS sample confirm this geographic path dependency: 76.28% of students who attend university in their home province take their first job there, whereas only 27.42% of those who migrate for college return home for their first job. Both university quality and job market access channels can lead to lower expected future income.

Our goal is to evaluate the extent to which the gender gap in college migration accounts for

[38] Households with a strong preference for sons were historically more likely to have a second birth if the first child was a daughter—a practice often permitted under the 'One-and-a-Half Child Policy' implemented in rural China.

the gender gap in university quality and expected earnings. We first use instrumental variables to obtain causal estimates of the college migration premium both in terms of university quality and expected income. Combining these estimated premiums with the gender gap in migration documented earlier, we quantify the contribution of the gender migration gap to the gender wage gap through counterfactual analyses simulating scenarios in which female students exhibit the same geographic mobility as their male counterparts.

6.1. Instrumental Variable Construction and First Stage Results

To address potential endogenous selection into out-of-province colleges and estimate the causal return to out-of-province college migration, we need exogenous shocks that push some students toward out-of-province universities for reasons unrelated to those students' underlying ability, motivation, labor market prospects, or perceived migration benefit. To this end, we construct a Bartik-style shift-share instrument exploiting the supply-side variation of university admission quotas. The exogenous shocks arise from the interaction of two key components: (1) shift: the large-scale policy-driven nationwide expansion of aggregate admission quotas initiated in 1999, and (2) share: universities' heterogeneous province-specific and institutionally persistent quota allocation patterns. For instance, the aggregate quota expansion of a Beijing-based university that historically allocates a high proportion of its admission quota to Sichuan province will disproportionately increase migration opportunities for Sichuan students to Beijing.

A primary concern with this identification strategy is whether the shift component is endogenous—specifically, whether the national aggregate quota expansion of each university could be a response to rising student demand in specific provinces (Sichuan in the above-mentioned example). This is unlikely, as individual universities generally do not adjust aggregate quotas based on province-specific demand shocks. Moreover, in 1999, the Chinese State Council initiated a nationwide policy to expand college enrollment across all universities (Che and Zhang, 2018). This expansion was centrally mandated and applied broadly rather than targeting specific provinces, rendering the overall expansion of admission quotas plausibly exogenous.

The second concern is that universities might adjust their cross-province quota distributions in response to differential changes in student demand. Following the classical shift-share approach, we address this by fixing the cross-province quota distributions at their 1999 levels—the first year of our sample—to measure each university's baseline propensity to admit students from specific provinces. Correspondingly, we will exclude the year 1999 from all

the following instrumental analyses. Furthermore, we argue that these historical shares are 'sticky'. As documented by Li et al. (2012), a university's propensity to admit students from various provinces is largely dictated by historical patterns rather than immediate local demand. Supporting this assumption, Panel (a) of Appendix Figure D1 confirms a strong positive correlation between the 1999 admission quota share for each university-track-province (as a fraction of the total admission quota for each university-track) against the average 2000–2003 quota share.

Formalizing the idea above, we construct the projected admission quota under track s for university j allocated to province p in year t as follows:

$$\frac{Quota^{js}_{p,1999}}{Quota^{js}_{1999}} \times Quota^{js}_{t}, \quad t \geq 2000 \tag{2}$$

Here, we calculate the ratio of university-track js's quota for province p ($Quota^{js}_{p,1999}$) to its total admission quota ($Quota^{js}_{1999}$) in the base year (1999). This ratio serves as the "exposure weight," capturing the baseline propensity of university-track js to admit students from province p.[39] The second term, $Quota^{js}_{t}$, represents total admission quota of university-track js in year t, capturing the aggregate expansion of university j in response to the nationwide expansion policy. We calculate the projected admission quota only for years after 2000. Panel (b) of Appendix Figure D1 demonstrates a high correlation between projected and actual quotas at the university-track-year level, with the data points clustering along the 45-degree line.

Since migration decisions are fundamentally determined by both push and pull factors, we construct two separate but complementary instrumental variables to capture these distinct channels of influence on out-of-province migration decisions. The first instrument is based on the out-of-province projected quota, capturing the pull factor that draws students away from their home province by expanding opportunities in destination provinces. The second instrument is based on the in-province projected quota, capturing the retention force that affects students' propensity to leave.

Specifically, for a student under track s taking the CEE in province p at year t, we define

[39] Since we observe the universe of admitted students by college and year, we proxy quotas using actual admissions for each university-track-province-year, given the excess demand for college seats in China.

the out-province instrument as follows:

$$Out_{pst} = \frac{\sum_{j \notin p} \frac{Quota^{js}_{p,1999}}{Quota^{js}_{1999}} \times Quota^{js}_{t}}{N_{pst}} \tag{3}$$

Here, the numerator aggregates projected quotas for students under track s from all colleges located outside province p, Crucially, we normalize this aggregate shock by N_{pst}, the total number of CEE candidates in the province-track-year. This normalization ensures that the instrument reflects the true per-capita probability of migration. Symmetrically, we construct the in-province instrument to measure the localized retention capacity by aggregating the projected quotas from all colleges located within the student's home province p:

$$In_{pst} = \frac{\sum_{j \in p} \frac{Quota^{js}_{p,1999}}{Quota^{js}_{1999}} \times Quota^{js}_{t}}{N_{pst}} \tag{4}$$

Leveraging these two instruments, we estimate the following first-stage equation:

$$Migration_i = \alpha + \beta_1 In_{pst(i)} + \beta_2 Out_{pst(i)} + X_i'\gamma + \delta_{t(i)} + \theta_{cs(i)} + \phi_{R_i} + \varepsilon_i \tag{5}$$

where the dependent variable $Migration_i$ is an indicator for whether student i migrates out of the province for college. Once we observe student i, we know his or her track s, home county c within province p, and the year t when they took the NCEE. Because our instrument variables, $In_{pst(i)}$ and $Out_{pst(i)}$, varies at the province-track-year level, we cannot follow the main specification to include county-track-year fixed effects, as doing so would absorb all the variation required for identification. Instead, our specification includes year fixed effects (δ_t) and county-track fixed effects (θ_{cs}). We further control for individual demographic characteristics (X_i) and percentile rank fixed effects, $\phi_{R(i)}$.

Table 6 presents the first-stage results, confirming that our instruments possess strong predictive power in the anticipated directions. In the full sample (Column (1)), a one-unit increase in the per-capita out-of-province quota significantly increases the probability of out-migration by 0.332 percentage points, acting as a robust pull factor. Conversely, a one-unit increase in the in-province quota decreases the out-migration probability by 0.224 percentage points, highlighting a strong domestic retention effect. Both estimates are statistically significant at the 1% level. Columns (2) and (3) reveal that this push-pull dynamic holds across genders,

though male students in Column (3) exhibit a slightly higher elasticity to both quota expansions compared to their female counterparts in column (2). The Kleibergen-Paap F-statistics are 107.97 for the full sample, 55.47 for females, and 121.66 for males. All values substantially exceed the conventional threshold, ruling out weak instrument concerns.

It is important to highlight the Local Average Treatment Effect (LATE) nature of our IV estimations. For example, very low-ranking students with close-to-zero chances of being admitted by an out-of-province university will not be affected by our instrument. Only students at the margin of migration decision will be affected by the relative increase in admission opportunities outside the province. Crucially for our counterfactual exercise, this LATE interpretation fits our analytical purpose. By asking how outcomes would change if female migration rates increased to match males, we are simulating a shift in behavior for marginal students who could potentially be induced to migrate through policy interventions, reduced barriers, or changed preferences. These are likely to be the students our instrument isolates, making the LATE the relevant policy parameter for explaining the gender gap.

Another concern relevant to the LATE nature of our IV estimations is whether the compliers come from segments of the ability distribution where the gender gap in migration is negligible. If this were the case, the resulting LATE would be uninformative for understanding the consequences of the gender gap in migration, limiting the relevance of the counterfactual analysis.

Appendix Figure D2 addresses this by plotting first-stage estimates across percentiles for both genders. Both pull (out-of-province quota) and retention (in-province quota) effects show strong predictive power across ability levels. Crucially, instruments exhibit the most significant influence at upper-middle and top percentiles (60th-90th), precisely where gender migration gaps are largest. Since quota expansions induce compliance among high-achieving students driving gender differences in migration, our LATE estimates are highly relevant for counterfactual analyses.

6.2. Outcome Construction and Second Stage Results

Our analysis focuses on two dimensions of student outcomes: university quality and subsequent labor market outcomes. To measure university quality, we require a metric that is both objective and widely accepted. Due to the lack of reliable, comprehensive university rankings during our sample period (early 2000s), we employ "Project 985" status as our primary proxy for institutional quality. Initiated by the central government to build world-class research institutions,

Project 985 is the most widely recognized indicator of university prestige in China, carrying a reputation and labor market signal analogous to the Ivy League in the United States.

Measuring labor market outcomes is more challenging since we cannot track students after their graduation from universities. To address this, we proxy for labor market performance using expected income derived from the CCSS, which records the highest salary offer received prior to graduation for each final-year student. We calculate the average monthly starting salary for each university, further stratified by gender and student origin (local vs. non-local to the university's province).[40] We then map these university-gender-origin specific averages to the NCEE admission records, restricting the sample to the universities covered by the CCSS. This allows us to impute expected incomes for over 1.2 million students. Essentially, our labor market outcome measurement represents a gender- and origin-specific university value, capturing the returns to both institutional quality and university location.

Our migration premium analysis focuses on the first-job wage after graduation rather than tracking long-term outcomes due to data limitations. Addressing concerns regarding the persistence of these initial impacts, Oreopoulos, Von Wachter and Heisz (2012) and Carr and Wiemers (2016) have documented a strong correlation between starting wages and long-term labor market success in the U.S. Specifically for Chinese college graduates, Jia and Li (2021) show that for college graduates, a one s.d. increase in the first-job wage rank is associated with a 0.73 s.d.increase in the wage rank over the subsequent five years, remaining highly persistent at 0.68 over a 6-to-10-year horizon. This evidence suggests that the migration premiums we estimate for initial wages likely translate into substantial, persistent earnings advantages that compound over a worker's career, rather than temporary entry effects that quickly dissipate.

Table 7 reports the second stage results of our 2SLS estimation for the 2000–2003 sample. Regarding university quality, migration significantly improves the probability of admission to elite institutions. Female students who migrate are 10.2 percentage points more likely to attend a Project 985 university, while male students see an 12.3 percentage point increase. To contextualize these magnitudes, it is important to note that only 2.5% of girls and 3.5% of boys who remain in their home province attend Project 985 universities. Against these low baseline rates, the estimated migration effects represent dramatic improvements in Project 985 access for complier students, who are more likely to be upper-middle- or high-ranking students.

[40] Since local students may have labor market advantages associated with their local networks or hukou status (Schmutte, 2015), we calculate expected income separately for local and non-local students. Ideally, we would further differentiate by each home province within each college–gender cell, but this level of disaggregation is infeasible given the limited sample size at the college level.

These estimates underscore the critical role of geographic mobility in accessing prestigious education, particularly because elite educational resources are heavily unevenly distributed across provinces in China.

To further illustrate the importance of geographic mobility in accessing prestigious education, Table D4 separates provinces into those with and without Project 985 universities. The presence or absence of Project 985 universities is not only an indicator of the availability of elite university seats, but also serves as a broad predictor of general educational resources and economic development opportunities.[41] The results show that migration premiums for accessing prestigious universities are concentrated among students from provinces without any Project 985 universities. In stark contrast, students from provinces that do host Project 985 universities experience close-to-zero migration premiums for elite university access.

In terms of labor market outcomes, migrating out of province yields a substantial return: expected monthly income increases by 701.8 yuan for females and 599.8 yuan for males, corresponding to a 24.9% and 19.4% increase compared to the respective sample means. As further demonstrated in Table D4, the migration premium in expected income is also substantially and significantly larger for students from provinces without any Project 985 universities. For these individuals, migration is necessary for them to access prestigious education and superior labor markets. In contrast, the migration premium is much smaller for students in resource-rich provinces, as the marginal benefit of leaving a resource-rich home province is naturally lower. It is worth mentioning that the gender gap in college migration is also larger in provinces without Project 985 universities (see Appendix Table D5). This reveals a concerning pattern of spatial-gender mismatch: the places where migration can deliver the most substantial economic and educational benefit are precisely the locations where female students face greater geographic mobility constraints relative to their male counterparts.

Comparing the 2SLS results between boys and girls in Table 7 show that the returns to migration are not statistically distinguishable between genders. If anything, the point estimates suggest that girls may exhibit slightly higher migration premiums in terms of expected income. This finding suggests that differential migration behavior by gender does not reflect rational responses to differential objective returns to migration. Instead, as our earlier survey analysis documented, the gender gap in migration must arise from other mechanisms, such as differ-

[41] Provinces lacking Project 985 universities, such as Ningxia and Guizhou, tend to be less developed, and offer fewer high-quality educational and employment options. In contrast, Beijing and Shanghai together host 12 Project 985 universities, accounting for 30% of all Project 985 universities in China.

ential safety concerns, marriage market considerations, lower subjective weight on material benefits for girls, or misaligned beliefs about migration returns.

6.3. Counterfactual Exercise

Given the substantial gains in university quality and labor market outcomes associated with out-of-province college migration, girls' lower probability of migrating may contribute to gender disparities in university quality and subsequent labor market performance. In this subsection, we conduct the counterfactual analysis to examine how much the gender gap in university quality and initial wage after graduation can be attenuated, when girls increase their migration probability to be the same as boys. Specifically, we define the share of the total gender gap in an outcome explained by the gender gap in college migration as:

$$\text{Explained Ratio} = \frac{\textit{Gender gap in migration} \times \textit{MigPremium}^{G}}{\textit{Gender gap in Outcome}} \tag{6}$$

where $\textit{MigPremium}^{G}$ is the female migration premium from the 2SLS estimates, and the denominator is the observed gender gap in the corresponding outcome variable such as expected income or Project-985 attendance. Intuitively, the numerator captures the counterfactual improvement in female outcomes if girls migrated at the same rate as boys, holding their migration premium fixed. The ratio thus measures the fraction of the gender gap that can be attributed to differential migration rates.

Here, we use the female coefficient (0.045) from Table 1 column (5) as the estimated gender gap in migration. As shown by the dependent mean in Table 7, boys are 2.1 p.p. more likely to attend Project 985 universities and have expected initial incomes 274.2 yuan higher than girls, which serves as the denominator for Equation 7. Applying these estimated parameters, we find that equalizing the female migration probability to that of males would narrow the gender gap in Project 985 attendance by 20.86% and the gap in expected initial income by 11.44%.

7. Conclusion

This paper documents a substantial gender gap in college migration in China and establishes its role as a contributing factor to the gender wage gap among the college-educated. Using administrative data covering the universe of NCEE test takers from 1999 to 2003, we show

that female students are significantly less likely than observationally equivalent male students to enroll in out-of-province universities, despite being eligible for the same set of institutions. This gap is most pronounced among high-achieving students who face the broadest choice sets, and it is amplified in environments characterized by traditional gender norms.

Our analysis of potential mechanisms points to parental attitudes rooted in gender norms as the primary drivers, particularly differential perceptions of the benefits of migration, preferences for local spouses, and safety concerns. We also rule out gender differences in risk aversion, self-confidence under uncertainty, expectations regarding elderly care, and family socioeconomic status as main explanations for our findings

The consequences of this spatial constraint are economically meaningful. Our instrumental variable estimates indicate that migrating for college doubles the probability of attending an elite (Project 985) university and raises expected early-career income by approximately 16 percent, with these premiums statistically indistinguishable across genders. A counterfactual exercise assigning male migration propensities to female students suggests that gender differences in college migration account for over one-fifth of the gender gap in elite university attendance and roughly one-tenth of the gender gap in early-career income. These findings underscore how decisions made years before labor market entry—shaped by social norms rather than by differential returns to migration—can have lasting effects on gender inequality.

References

Adamecz-Völgyi, Anna, and Nikki Shure. 2022. "The gender gap in top jobs–the role of overconfidence." *Labour Economics*, 79: 102283.

Adda, Jérôme, Christian Dustmann, and Katrien Stevens. 2017. "The career costs of children." *Journal of Political Economy*, 125(2): 293–337.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the origins of gender roles: Women and the plough." *Quarterly Journal of Economics*, 128(2): 469–530.

Almond, Douglas, Hongbin Li, and Shuang Zhang. 2019. "Land reform and sex selection in China." *Journal of Political Economy*, 127(2): 560–585.

Altonji, Joseph G, John Eric Humphries, Yagmur Yuksel, and Ling Zhong. 2025. "Decomposing Trends in the Gender Gap for Highly Educated Workers." National Bureau of Economic Research.

Bertrand, Marianne. 2011. "New perspectives on gender." In *Handbook of Labor Economics*. Vol. 4, 1543–1590. Elsevier.

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz. 2010. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic Journal: Applied Economics*, 2(3): 228–255.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender identity and relative income within households." *Quarterly Journal of Economics*, 130(2): 571–614.

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." *Journal of Economic Literature*, 55(3): 789–865.

Booth, Alison, Elliott Fan, Xin Meng, and Dandan Zhang. 2019. "Gender differences in willingness to compete: The role of culture and institutions." *The Economic Journal*, 129(618): 734–764.

Carr, Michael D, and Emily E Wiemers. 2016. "The decline in lifetime earnings mobility in the US: Evidence from survey-linked administrative data." *Washington Center for Equitable Growth*.

Chen, Y, and Onur Kesten. 2017. "Chinese college admissions and school choice reforms: A theoretical analysis." *Journal of Political Economy*, 125(1): 99–139.

Che, Yi, and Lei Zhang. 2018. "Human Capital, Technology Adoption and Firm Performance: Impacts of China's Higher Education Expansion in the Late 1990s." *The Economic Journal*, 128(614): 2282–2320.

Choi, Eleanor Jawon, and Jisoo Hwang. 2015. "Child gender and parental inputs: No more son preference in Korea?" *American Economic Review*, 105(5): 638–643.

Connelly, Rachel, and Zhenzhen Zheng. 2003. "Determinants of school enrollment and completion of 10 to 18 year olds in China." *Economics of Education Review*, 22(4): 379–388.

Cortés, Patricia, and Jessica Pan. 2018. "Occupation and gender." *The Oxford Handbook of Women and the Economy*, 425–452.

Cortés, Patricia, and Jessica Pan. 2023. "Children and the remaining gender gaps in the labor market." *Journal of Economic Literature*, 61(4): 1359–1409.

Cortés, Patricia, Jessica Pan, Laura Pilossoph, Ernesto Reuben, and Basit Zafar. 2023. "Gender differences in job search and the earnings gap: Evidence from the field and lab." *The Quarterly Journal of Economics*, 138(4): 2069–2126.

Croson, Rachel, and Uri Gneezy. 2009. "Gender differences in preferences." *Journal of Economic Literature*, 47(2): 448–474.

Delavande, Adeline, and Basit Zafar. 2019. "University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints." *Journal of Political Economy*, 127(5): 2343–2393.

Dustmann, Christian, and Albrecht Glitz. 2011. "Migration and education." In *Handbook of the Economics of Education*. Vol. 4, 327–439. Elsevier.

Eckel, Catherine C, and Philip J Grossman. 2008. "Men, women and risk aversion: Experimental evidence." *Handbook of Experimental Economics Results*, 1: 1061–1073.

Edlund, Lena. 1999. "Son preference, sex ratios, and marriage patterns." *Journal of Political Economy*, 107(6): 1275–1304.

Goldin, Claudia. 2014. "A grand gender convergence: Its last chapter." *American Economic Review*, 104(4): 1091–1119.

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating impartiality: The impact of "blind" auditions on female musicians." *American Economic Review*, 90(4): 715–741.

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko. 2006. "The homecoming of American college women: The reversal of the college gender gap." *Journal of Economic Perspectives*, 20(4): 133–156.

Gu, Yizhen, Naijia Guo, Jing Wu, and Ben Zou. 2024. "Home location choices and the gender commute gap." *Journal of Human Resources*, 59(2): 545–575.

Hu, Naiyuan, and Lin Ma. 2025. "Educational Migration in China." Working paper.

Huynh, Quynh, and Hyejin Ku. 2025. "Gender norms and female labor supply: Evidence from export shocks in Vietnam." *Review of Economics and Statistics*, Forthcoming.

Jia, Ning, Raven Molloy, Christopher Smith, and Abigail Wozniak. 2023. "The economics of internal migration: Advances and policy questions." *Journal of Economic Literature*, 61(1): 144–180.

Jia, Ruixue, and Hongbin Li. 2021. "Just above the exam cutoff score: Elite college admission and wages in China." *Journal of Public Economics*, 196: 104371.

Jia, Ruixue, Hongbin Li, and Claire Cousineau. 2025. *The highest exam: How the Gaokao shapes China.* Harvard University Press.

Le Barbanchon, Thomas, Roland Rathelot, and Alexandra Roulet. 2021. "Gender differences in job search: Trading off commute against wage." *Quarterly Journal of Economics*, 136(1): 381–426.

Li, Hongbin, and Xinyao Qiu. 2025. "Heuristics in self-evaluation: Evidence from the centralized college admission system in China." *Review of Economics and Statistics*, 1–10.

Li, Hongbin, Lingsheng Meng, Xinzheng Shi, et al. 2012. "Does attending elite colleges pay in China?" *Journal of Comparative Economics*, 40(1): 78–88.

Li, Wenchao, and Junjian Yi. 2015. "The competitive earning incentive for sons: Evidence from migration in China." IZA Discussion Papers.

Molloy, Raven, Christopher L Smith, and Abigail Wozniak. 2011. "Internal migration in the United States." *Journal of Economic Perspectives*, 25(3): 173–196.

Niederle, Muriel, and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, 122(3): 1067–1101.

Olivetti, Claudia, and Barbara Petrongolo. 2016. "The evolution of gender gaps in industrialized countries." *Annual Review of Economics*, 8(1): 405–434.

Oreopoulos, Philip, Till Von Wachter, and Andrew Heisz. 2012. "The short-and long-term career effects of graduating in a recession." *American Economic Journal: Applied Economics*, 4(1): 1–29.

Rossi, Pauline, and Yun Xiao. 2025. "Left over or opting out? Squeeze, mismatch and surplus in Chinese marriage markets." *Journal of Development Economics*, 103629.

Schmutte, Ian M. 2015. "Job referral networks and the determination of earnings in local labor markets." *Journal of Labor Economics*, 33(1): 1–32.

Wei, Shang-Jin, and Xiaobo Zhang. 2011. "The competitive saving motive: Evidence from rising sex ratios and savings rates in China." *Journal of Political Economy*, 119(3): 511–564.

Wiswall, Matthew, and Basit Zafar. 2018. "Preference for the workplace, investment in human capital, and gender." *Quarterly Journal of Economics*, 133(1): 457–507.

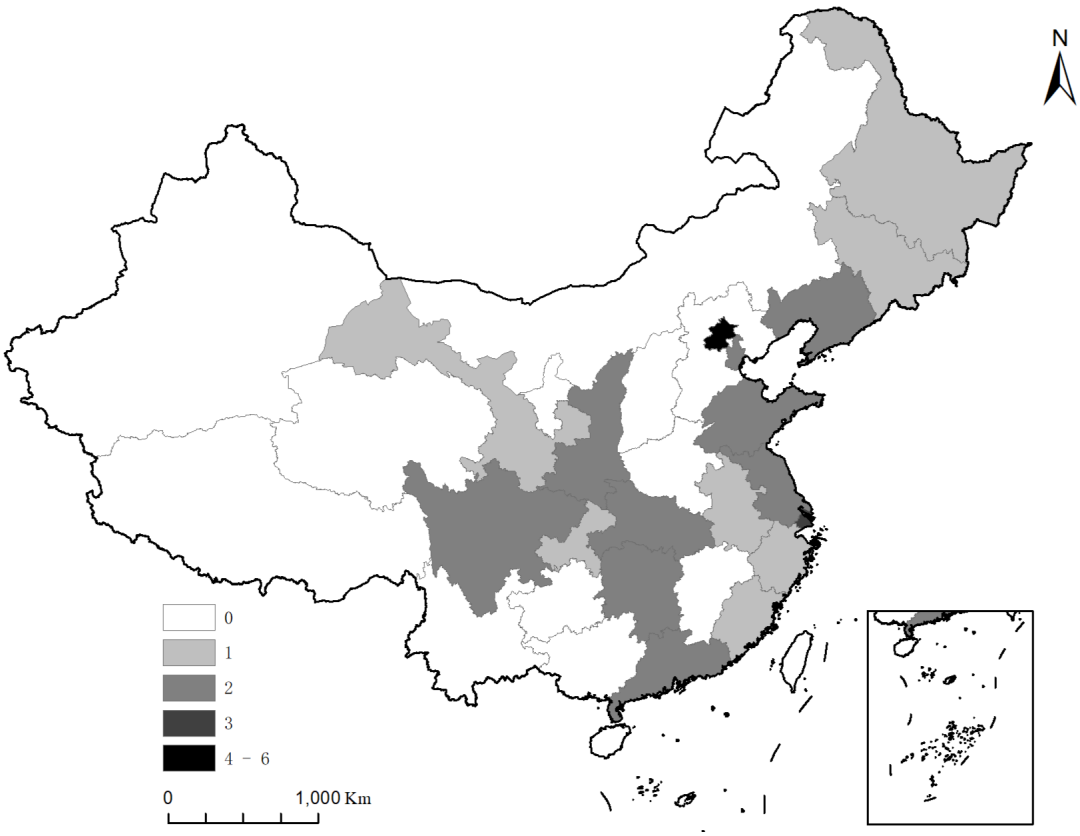
Wozniak, Abigail. 2010. "Are college graduates more responsive to distant labor market opportunities?" *Journal of Human Resources*, 45(4): 944–970.

Yang, Y. Angela. 2024. "On the Spatial Allocation of College Seats: Human Capital Production and the Distribution of Skilled Labor." Working paper.

Zafar, Basit. 2013. "College major choice and the gender gap." *Journal of Human Resources*, 48(3): 545–595.

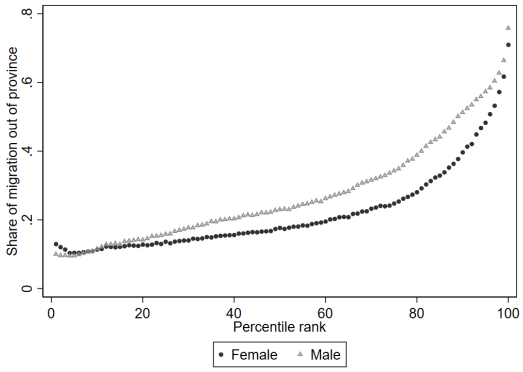
Zeng, Junxia, Xiaopeng Pang, Linxiu Zhang, Alexis Medina, and Scott Rozelle. 2014. "Gender inequality in education in China: a meta-regression analysis." *Contemporary Economic Policy*, 32(2): 474–491.

Figure 1: Spatial Distribution of Elite Universities

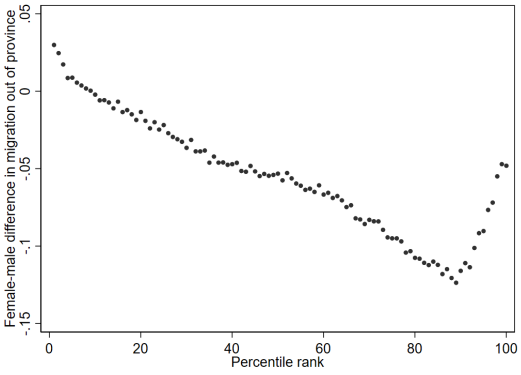


Notes: This figure illustrates the spatial distribution of Project 985 universities (elite universities). The map plots the provincial count of Project 985 universities as of 2003. Darker shades of grey indicate a higher count. Provinces shaded in solid white indicate zero Project 985 universities.

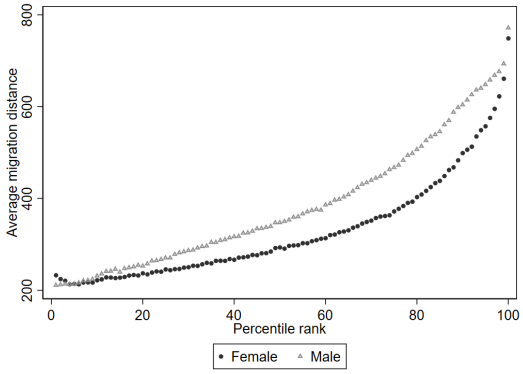
Figure 2: Gender Differences in College Migration by Exam Percentile Rank



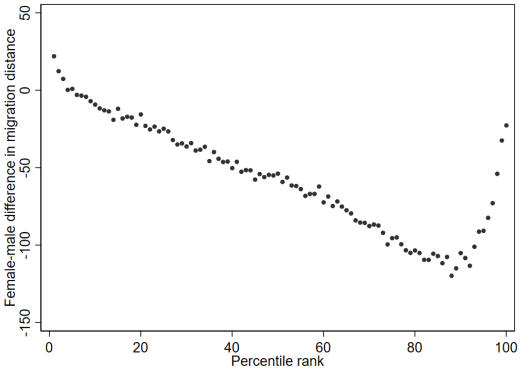
(a) Share of choosing college out of province



(b) Female-male difference



(c) Average distance between home and college cities



(d) Female-male difference

Notes: This figure plots college migration outcomes and gender differences across percentile ranks. Panels A and C report the average outcomes by gender. Panel A plots the share of students migrating out of their home province. Panel C plots the average distance (in kilometers) between the home city and the college city. Panels B and D report the gender difference (calculated as the female mean minus the male mean) for the outcomes in Panels A and C, respectively. In all panels, the x-axis represents the student's percentile rank, divided into 100 integer bins. In Panels A and C, grey hollow triangles denote male students, and black solid circles denote female students. In Panels B and D, black solid circles represent the difference in means between female and male students within each bin. The sample consists of all students admitted to college.

Table 1: Gender Gap in Out-of-Province College Migration: Baseline Results

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Out of province				
Female	-0.062***	-0.051***	-0.045***	-0.043***	-0.045***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Percentile rank			0.004***		
			(0.000)		
County-year-track FE		Yes	Yes	Yes	Yes
Percentile rank FE				Yes	Yes
Demographic characteristics					Yes
Dependent mean	0.251	0.251	0.251	0.251	0.251
Observations	13,524,632	13,524,632	13,524,632	13,524,632	13,524,632

Notes: This table reports gender differences in cross-province college migration for the full sample of students admitted to college. The dependent variable is an indicator equal to 1 if the student migrates out of their home province, and 0 otherwise. The independent variable of interest is an indicator for being female. Column (1) reports baseline estimates without fixed effects. Column (2) controls for county-by-year-by-track fixed effects. Column (3) adds a continuous control for the student's percentile rank (calculated within each province-year-track cohort). Column (4) replaces the continuous rank control with fixed effects for each of the 100 integer percentile bins. Column (5) adds student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table 2: Gender Gap Using Alternative Migration Measures

Dependent variable:	(1) To provincial capital	(2) Migration distance	(3) To top 30 GDP cities	(4) To top 10 GDP cities
Female	-0.042***	-45.155***	-0.022***	-0.011***
	(0.001)	(0.669)	(0.000)	(0.000)
County-year-track FE	Yes	Yes	Yes	Yes
Percentile rank FE	Yes	Yes	Yes	Yes
Demographic characteristics	Yes	Yes	Yes	Yes
Dependent mean	0.479	360.424	0.147	0.093
Observations	7,755,958	13,524,632	10,027,874	11,843,492
Sample	Students within province and hometown not in provincial capital	Full sample	Exclude students from top 30 GDP cities	Exclude students from top 10 GDP cities

Notes: This table reports gender differences in college location choices. The independent variable of interest in all columns is an indicator for being female. Column (1) restricts the sample to students attending university within their home province, excluding those originating from provincial capitals or direct-controlled municipalities; the dependent variable is an indicator equal to 1 if the student attends a university in the provincial capital. Column (2) uses the full sample, and the dependent variable is the migration distance (in kilometers) between the home city and the college city. Columns (3) and (4) exclude students originating from the top 30 and top 10 cities (ranked by total GDP in 2000), respectively. The dependent variables in these columns are indicators for attending a university in a top 30 city (Column 3) or a top 10 city (Column 4). Demographic characteristics include agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table 3: Heterogeneity by Hukou Status and Local Gender Norms

	(1)	(2)	(3)	(4)
Dependent variable:	Out of province			
Female × Agricultural hukou	-0.028***			
	(0.001)			
Female × Gender norm		-0.003***	-0.007***	-0.006***
		(0.001)	(0.002)	(0.002)
Female	-0.031***	-0.045***	-0.039***	-0.039***
	(0.001)	(0.001)	(0.002)	(0.002)
Gender norm measurements		Sex ratio at birth	Gender role	Marriage importance
County-year-track FE	Yes	Yes	Yes	Yes
Percentile rank FE	Yes	Yes	Yes	Yes
Demographic characteristics	Yes	Yes	Yes	Yes
Dependent mean	0.251	0.251	0.182	0.182
Observations	13,524,632	13,524,632	1,137,525	1,137,525
Sample	Full	Full	CGSS merged	CGSS merged

Notes: This table reports estimates of the heterogeneous effects of hukou and local gender norms on cross-province college migration. The dependent variable is an indicator equal to 1 if the student migrates out of their home province, and 0 otherwise. Column (2) measures gender norms using the male-to-female sex ratio for the population aged 0–4 based on the 2000 NBS survey. Columns (3) and (4) measure norms using the county-level average agreement with the statement "Men should focus on career, while women should focus on family" (Column 3) and "Marrying well is more important for women than doing well" (Column 4), based on data from the 2010 China General Social Survey (CGSS). Responses in the CGSS are rated on a scale from 1 (strongly disagree) to 5 (strongly agree). All gender norm measures are standardized to have a mean of zero and a standard deviation of one across counties. Demographic characteristics include agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table 4: Gender Differences in Parental Preferences for Out-of-Province College

	(1)	(2)	(3)	(4)
Dependent variable		Migration preference		
Female	-0.050**	-0.053**	-0.054**	-0.054**
	(0.023)	(0.021)	(0.022)	(0.023)
Mother		0.018	0.021	0.010
		(0.022)	(0.022)	(0.021)
Agricultural hukou		-0.018	-0.017	-0.026
		(0.016)	(0.016)	(0.017)
Years of education		0.008**	0.007*	0.007*
		(0.003)	(0.004)	(0.004)
Only child		-0.019	-0.020	-0.035*
		(0.021)	(0.021)	(0.020)
Expected university level FE		Yes	Yes	Yes
Income level FE			Yes	Yes
City-academic track FE				Yes
Observations	2,940	2,940	2,940	2,940
Dependent mean	0.394	0.394	0.394	0.394

Notes: This table reports estimates on gender differences in parental preference for out-of-province college attendance. The dependent variable, migration preference, equals one if the parent ranks an in-province university above both an out-of-province university (excluding Beijing and Shanghai) and a university in Beijing or Shanghai, assuming equal quality across options. The Female indicator equals one if the child is female. Mother is a dummy equal to one if the respondent is the child's mother. Agricultural hukou equals one if the respondent holds an agricultural hukou. Years of education refers to the respondent's years of schooling. Only child equals one if the respondent has only one child. Expected university level FE includes fixed effects for parents' expectations regarding the highest university level their child can attain. Income level FE denotes fixed effects for household income categories. City–academic track FE includes fixed effects for the child's home city interacted with academic track (science vs. humanities). Robust standard errors clustered by city are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Gender Differences in Parental Concerns and Perceived Benefits of Migration

Dependent variable	(1) Spouse concern	(2) Spouse in province	(3) Financial support	(4) Time support	(5) Safety concern	(6) Cost concern	(7) Assistance concern	(8) Larger benefit for child gender
Female	-0.111***	0.068**	-0.028*	0.021	0.064***	0.002	0.013	-0.169***
	(0.020)	(0.030)	(0.016)	(0.025)	(0.010)	(0.020)	(0.008)	(0.017)
Demographic characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Expected university level FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Income level FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-academic track FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,940	2,940	2,940	2,940	2,940	2,940	2,940	2,940
Dependent mean	0.482	0.328	0.098	0.540	0.856	0.567	0.910	0.177

Notes: This table reports estimates on gender differences in parental perceived migration concerns and benefits. Spouse concern equals one if the respondent reports being at least "slightly concerned" about whether the child would be able to find a suitable spouse if studying outside the home province. Spouse in province equals one if the respondent agrees (to some extent or strongly) that the child's future spouse should be from the home province. Financial support equals one if the respondent indicates that their most important expectation for the child's elderly-care responsibility is providing financial support rather than spending time accompanying parents or having no expectations. Time support equals one if the respondent's most important expectation is that the child spend time accompanying parents. Safety concern, cost concern, and assistance concern equal one if the respondent reports being at least "slightly concerned" about city safety, the financial costs of out-of-province study, or the family's ability to provide timely assistance or networking support, respectively. Larger benefit for child gender equals one if the respondent believes that studying outside the province yields greater benefits for the child's own gender than for the opposite gender. Robust standard errors clustered by city are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: First-Stage Estimates: Projected Admission Quotas and College Migration

	(1)	(2)	(3)
Dependent variable:	Out of province		
In-province quota	-0.224***	-0.161***	-0.260***
	(0.018)	(0.019)	(0.018)
Out-of-province quota	0.332***	0.267***	0.333***
	(0.038)	(0.043)	(0.037)
Year FE	Yes	Yes	Yes
County-track FE	Yes	Yes	Yes
Percentile rank FE	Yes	Yes	Yes
Demographic characteristics	Yes	Yes	Yes
Dependent mean	0.255	0.222	0.282
Observations	11,987,370	5,275,842	6,711,450
Kleibergen-Paap F-stat	107.967	55.467	121.664
Sample	Full sample	Female	Male

Notes: This table reports the first-stage estimates for the instrumental variable regressions. The sample is restricted to students admitted to college between 2000 and 2003. The dependent variable is an indicator equal to 1 if the student migrates out of their home province, and 0 otherwise. The excluded instruments are the predicted in-province and out-of-province admission quotas. These quotas are constructed using a shift-share approach, interacting the 1999 college-track admission shares in each home province with the 2000–2003 admission shifts. The predicted admissions are aggregated to the home province level for in-province and out-of-province colleges, respectively, and then divided by the total number of NCEE takers in the home province in that year. Column (1) reports estimates for the pooled sample. Columns (2) and (3) report estimates for the female and male subsamples, respectively. All columns control for year fixed effects, county-by-track fixed effects, percentile rank fixed effects, and student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. The Kleibergen-Paap rk Wald F-statistic for weak instruments is reported for each specification. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table 7: 2SLS Estimates of the Migration Premium: Elite University and Expected Income

	(1)	(2)	(3)	(4)
Dependent variable:	985 university	985 university	Expected income	Expected income
Out of province	0.102***	0.123***	701.847***	599.760***
	(0.029)	(0.021)	(223.802)	(75.989)
Year FE	Yes	Yes	Yes	Yes
County-track FE	Yes	Yes	Yes	Yes
Percentile rank FE	Yes	Yes	Yes	Yes
Demographic characteristics	Yes	Yes	Yes	Yes
Dependent mean	0.041	0.062	2813.735	3087.909
Observations	5,275,842	6,711,450	432,055	696,935
Sample	Female	Male	Female	Male

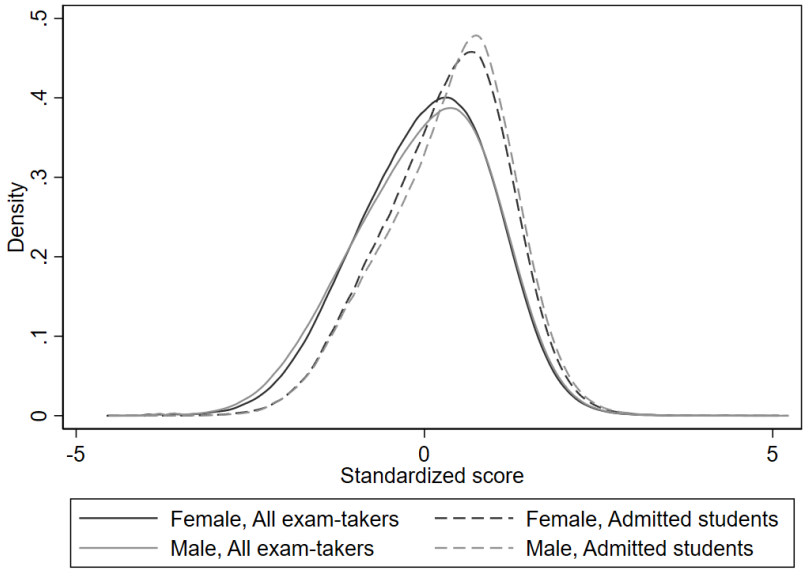
Notes: This table reports the second-stage estimates of the instrumental variable regressions. The sample is restricted to students admitted to college between 2000 and 2003. The endogenous variable is an indicator for migrating out of the home province, which is instrumented by the shift-share predicted in-province and out-of-province admission quotas. In Columns (1) and (2), the dependent variable is an indicator equal to 1 if the student is admitted to a Project 985 university, and 0 otherwise. In Columns (3) and (4), the dependent variable is the expected monthly income (in RMB). This expected income is constructed using the CCSS data by calculating the average monthly wage of the best job offer for graduates from each university, specific to the student's gender and college migration status. These averages are then matched to the NCEE administrative data by university name. Columns (1) and (2) use the full sample of students admitted between 2000 and 2003, while Columns (3) and (4) use the NCEE-CCSS matched sample. Across these respective samples, Columns (1) and (3) report estimates for the female subsample, while Columns (2) and (4) report estimates for the male subsample. All columns control for year fixed effects, county-by-track fixed effects, percentile rank fixed effects (calculated within each province-year-track cohort), and student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Online Appendix for

Gender Differences in College Migration

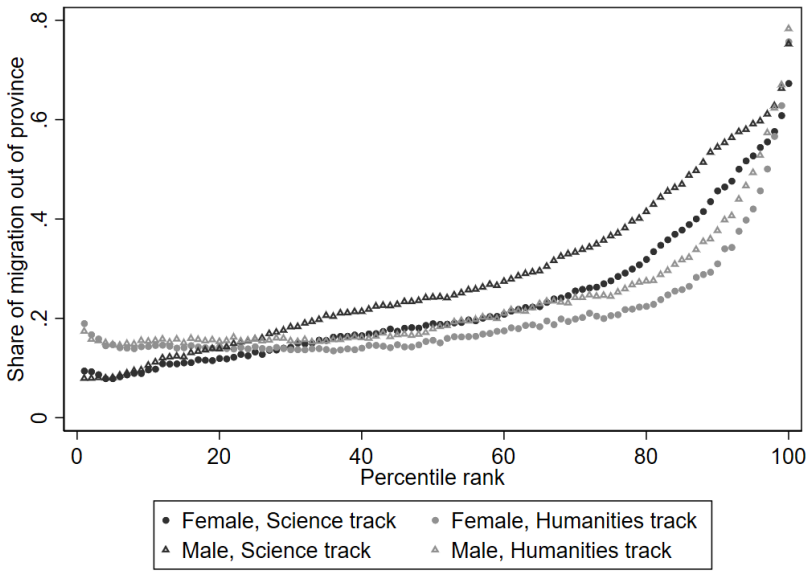
A. Summary and Robustness

Figure A1: Distribution of Standardized NCEE Scores by Gender and Admission Status



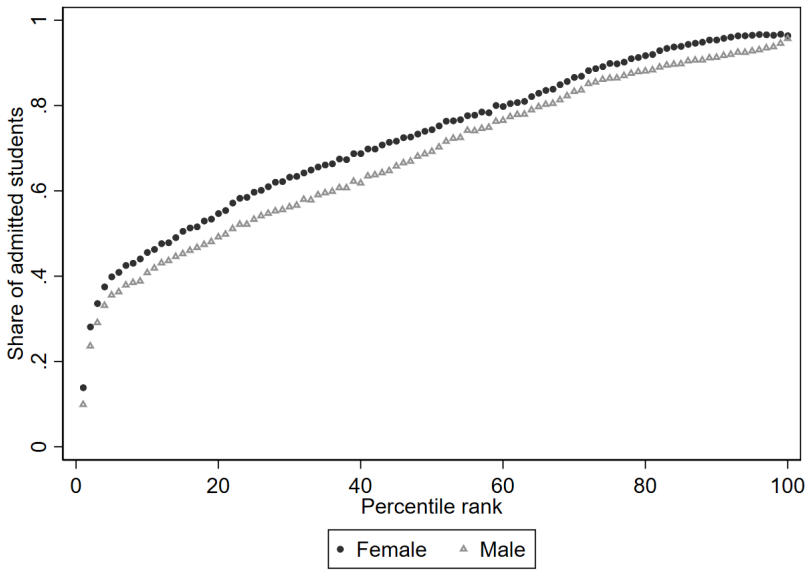
Notes: This figure plots the probability density functions of standardized NCEE scores by gender and admission status. The solid lines represent the full sample of all students who took the NCEE, while the dashed lines represent the subsample of students admitted to college. Black lines denote female students, and grey lines denote male students. Scores are standardized to have a mean of zero and a standard deviation of one within each province-year-track cohort, based on the sample of all exam takers.

Figure A2: Gender Gap in College Migration by Academic Track



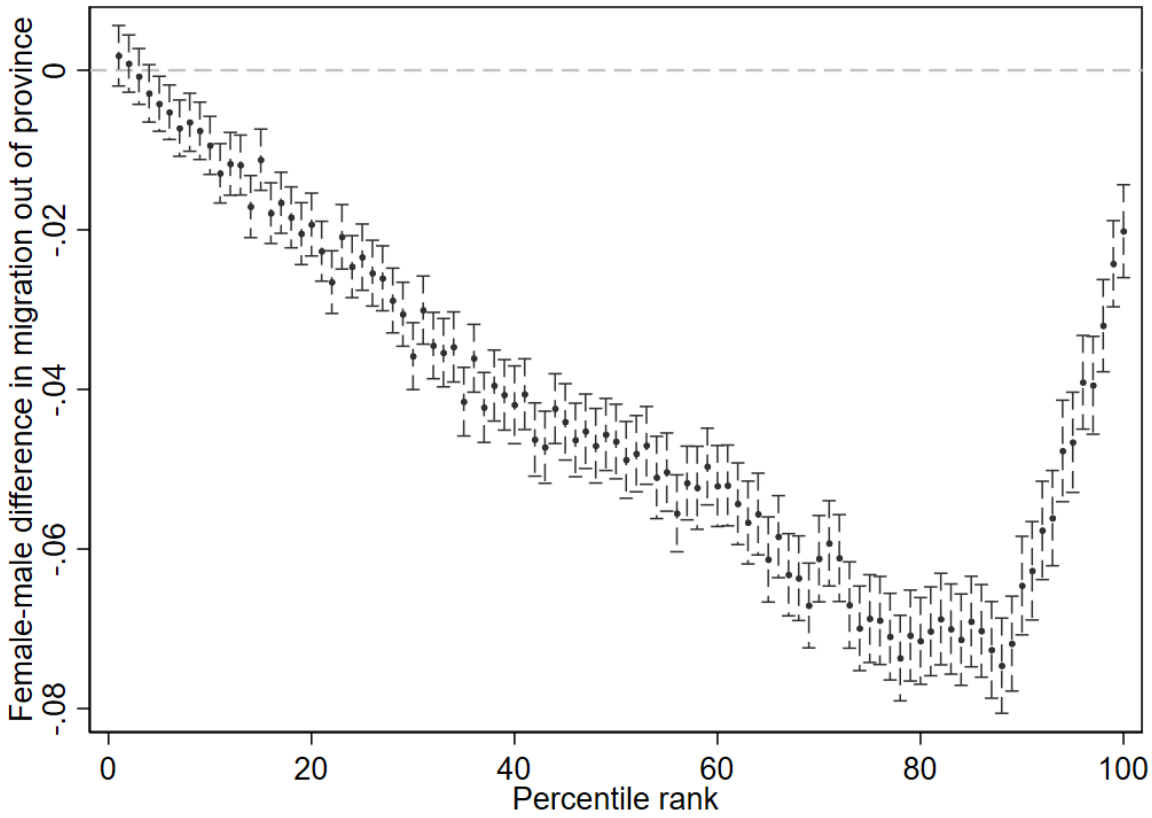
Notes: This figure plots the share of students migrating out of their home province by gender and academic track across percentile ranks. The sample consists of all students admitted to college. The x-axis represents the student's percentile rank. Each marker represents the mean share of migration within that specific bin. Black markers denote science track students, and grey markers denote humanity track students. Solid circles represent female students, and hollow triangles represent male students.

Figure A3: College Admission Rates by Gender and Percentile Rank



Notes: This figure plots the share of students admitted to college by gender and percentile rank. The sample consists of all NCEE takers, including those not admitted. The x-axis represents the percentile rank, divided into 100 integer bins. The percentile ranks are constructed based on the score distribution of the subsample of admitted students. Each marker represents the share of admitted students within that specific bin. Grey hollow triangles denote male students, and black solid circles denote female students.

Figure A4: Estimated Gender Gap in Out-of-Province Migration across Percentile Ranks



Notes: This figure plots the estimated gender gaps in out-of-province migration probability across percentile ranks. Each marker represents the point estimate of the coefficient on the female indicator from a separate regression estimated using the subsample of students within that specific percentile bin. The dependent variable is an indicator equal to 1 if the student migrates out of their home province, and 0 otherwise. All regressions control for county-by-year-by-track fixed effects and demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. The x-axis represents the student's percentile rank, calculated within each province-year-track cohort. Vertical dashed lines indicate 95 percent confidence intervals, based on robust standard errors clustered at the home county level.

Table A1: Summary statistics

	Full sample		Female		Male	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.44	0.50	1.00	0.00	0.00	0.00
Out of province	0.25	0.43	0.22	0.41	0.28	0.45
Migration distance	360.42	497.38	325.18	472.66	387.87	514.14
Science track	0.72	0.45	0.61	0.49	0.80	0.40
Percentile rank	50.00	28.87	49.33	28.55	50.53	29.10
Agricultural hukou	0.49	0.50	0.44	0.50	0.54	0.50
Age	18.71	1.14	18.58	1.08	18.81	1.17
Minority	0.07	0.25	0.07	0.26	0.07	0.25
985 university	0.05	0.23	0.04	0.20	0.06	0.24
Observations	13,524,632		5,922,069		7,602,563	

Notes: This table reports summary statistics for the individual-level variables in the baseline sample. Columns (1) and (2) report the means and standard deviations for the full sample, while Columns (3)–(4) and Columns (5)–(6) report statistics for the female and male subsample, respectively. *Female* is an indicator equal to 1 if the student is female. *Out of province* is an indicator equal to 1 if the student's university is located in a different province from their hometown. *Migration distance* represents the straight-line distance (in kilometers) between the hometown city and the university city. *Science track* is an indicator for students in the science track. *Percentile rank* denotes the student's percentile ranking within their province-year-track cohort based on the NCEE score. *Agricultural hukou* is an indicator for holding an agricultural hukou. *985 university* is an indicator equal to 1 if the university is part of Project 985 (a group of elite research universities). The number of observations is reported in the last row.

Table A2: Relative Importance of College vs. Major in Ranking-based Sorting

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Percentile rank					
Adjusted R-squared	0.238	0.039	0.204	0.033	0.263	0.050
University level FE	Yes		Yes		Yes	
Major FE		Yes		Yes		Yes
Observations	13,524,632	13,308,579	5,922,069	5,819,924	7,602,563	7,488,655
Sample	Full sample	Full sample	Female	Female	Male	Male

Notes: This table reports the adjusted R-squared values from regressions of student percentile ranks on university-level or major fixed effects. The dependent variable is the student's percentile rank in the CEE, calculated within each province-year-track cohort. Columns (1) and (2) use the full sample, Columns (3) and (4) restrict the sample to females, and Columns (5) and (6) restrict the sample to males. Odd-numbered columns report the adjusted R-squared from regressions including only university-level fixed effects. University-level fixed effects are categorized into four mutually exclusive tiers: Project 985, Project 211 (excluding Project 985), other 4-year universities (excluding Project 211), and non-4-year colleges. Even-numbered columns report the adjusted R-squared from regressions including only major fixed effects. Major fixed effects are defined using the 12 broad disciplinary categories (e.g., Economics, Engineering, Science, and Medicine) in the Chinese higher education system.

Table A3: Robustness of the Gender Gap in College Migration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:					Out of province			
Female	-0.045***	-0.045***	-0.045***	-0.050***	-0.044***	-0.045***	-0.045***	-0.041***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
County-year-track FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Percentile rank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Demographic characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province-year-track-percentile rank FE								Yes
Number of percentile rank FE groups	100	200	500	100	100	100	100	100
Dependent mean	0.251	0.251	0.251	0.264	0.259	0.250	0.247	0.247
Observations	13,524,632	13,524,632	13,524,632	8,753,911	13,030,153	13,474,805	12,239,724	13,524,421
Sample	Full sample	Full sample	Full sample	First-choice admission	Normal exam type	Exclude NCEE migrants	Exclude students with bonus scores	Full sample

Notes: This table reports robustness checks for the baseline results. The dependent variable is an indicator equal to 1 if the student migrates out of their home province for college, and 0 otherwise. Column (1) reports the baseline estimates. Columns (2) and (3) test the sensitivity to the granularity of ability controls by replacing the baseline 100-bin percentile rank fixed effects with 200-bin and 500-bin percentile rank fixed effects, respectively. Column (4) restricts the sample to students admitted to their first-choice university. Column (5) restricts the sample to students classified as the ”normal” examination type, excluding those in special categories (e.g., ”normal spring,” ”small language,” ”double degree,” ”intern,” ”teacher,” ”vocational,” ”athletic,” and ”other”). Column (6) excludes NCEE migrants, defined as students whose NCEE province differs from their high school province. Column (7) excludes students with bonus scores. Column (8) replaces the baseline fixed effects with province-by-year-by-track-by-percentile rank fixed effects, comparing students within the exact same province-year-track-score cell. Unless otherwise noted, all columns control for county-by-year-by-track fixed effects, percentile rank fixed effects, and student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table A4: Gender Migration Gap: Disentangling Gender Norms from Sex-Ratio Effects

	(1)	(2)
Dependent variable:	Out of province	
Female × Gender norm	-0.006***	-0.005**
	(0.002)	(0.002)
Female × Sex ratio at birth	-0.004*	-0.006**
	(0.002)	(0.002)
Female	-0.041***	-0.041***
	(0.002)	(0.002)
Gender norm measurements	Gender role	Importance of marriage
County-year-track FE	Yes	Yes
Percentile rank FE	Yes	Yes
Demographic characteristics	Yes	Yes
Dependent mean	0.182	0.182
Observations	1,137,525	1,137,525

Notes: This table reports estimates of the effects of traditional gender norms on gender gap in cross-province college migration, conditional on the local sex ratio. The dependent variable is an indicator equal to 1 if the student migrates out of their home province, and 0 otherwise. The sex ratio is defined as the male-to-female sex ratio for the population aged 0–4 based on the 2000 NBS survey. Column (1) measures gender norms using the county-level average agreement with the statement "Men should focus on career, while women should focus on family" (Gender Role), and Column (2) measures norms using the agreement with the statement "Marrying well is more important for women than doing well" (Importance of Marriage), with data from the 2010 CGSS. All gender norm and sex ratio measures are standardized to have a mean of zero and a standard deviation of one across counties. Demographic characteristics include agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

B. Self-designed Survey

Table B1: Survey Questions and Variables Definition

Variable	Survey Question	Answer Choices
Panel A: Child Characteristics		
Child gender	What is the gender of this child?	(1) Male (2) Female
Siblings	How many brothers and sisters does this child have?	Open-ended
High school location	Where does this child attend high school?	Open-ended (province and city)
Academic track	Did this child (plan to) choose the Humanities or Science track for the NCEE?	(1) Humanities (2) Science
Expected university level	What level of university do you think this child will be able to attend?	(1) 985, (2) 211 (excl. 985), (3) Tier 1, (4) Tier 2, (5) Vocational, (6) None
Panel B: Parent Characteristics		
Parent gender	What is your gender?	(1) Male (2) Female
Hukou type	What is your current hukou status?	(1) Agricultural (2) Non-agricultural
Education level	What is the highest level of education you have attained?	(1) None, (2) Elementary, (3) Middle school, (4) High school, (5) Vocational, (6) Associate, (7) Bachelor's, (8) Master's or above
Income level	How much do you earn per month?	(1) <2,000, (2) 2,000–5,000, (3) 5,000–10,000, (4) 10,000–20,000, (5) >20,000 RMB
Panel C: Parental Attitudes		
Migration preference for child college	If your child could enter a university in the hometown province, in another province (excl. Beijing & Shanghai), or in Beijing or Shanghai — all of equal quality — how would you rank these options?	Rank three options: (1) Local province, (2) Other province, (3) Beijing/Shanghai. *Binary: = 1 if local-province university is ranked below both alternatives.*

(Continued on next page)

(Continued from previous page)

Variable	Survey Question	Answer Choices
Concern: spouse	If your child were to attend university in another province, would you be concerned about the following?	*Scale:* (1) Very, (2) Somewhat, (3) Slightly, (4) Not concerned. *Binary: = 1 if (1)–(3). Factor: finding a spouse.*
Concern: safety		*Binary: = 1 if (1)–(3). Factor: safety of the city.*
Concern: cost		*Binary: = 1 if (1)–(3). Factor: financial costs (e.g., travel expenses).*
Concern: assistance		*Binary: = 1 if (1)–(3). Factor: ability to provide timely support for the child.*
Spouse in province	Do you hope this child's spouse will be from the same province?	(1) Strongly disagree $\cdots$ (5) Strongly agree. *Binary: = 1 if (4) or (5).*
Financial support	What is your most important expectation for your child's role in your elderly care?	(1) Financial support, (2) Time/companionship, (3) No expectation. *Binary: = 1 if (1).*
Time support		*Binary: = 1 if (2).*
Larger benefit for child's gender	Do you believe studying outside the province offers greater benefits for boys, girls, or both equally?	Categorical: boys $>$ girls / girls $>$ boys / similar. *Binary: = 1 if benefits perceived larger for child's own gender.*

Table B2: Descriptive Statistics (Survey): Child and Parent Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All		Boys		Girls		Mean diff.
Variables	Mean	SD	Mean	SD	Mean	SD	
Panel A. Child							
Expected university level							
985 university	0.163	(0.370)	0.174	(0.379)	0.149	(0.356)	0.025*
211 university	0.206	(0.404)	0.199	(0.399)	0.214	(0.411)	-0.015
Tier 1 university	0.429	(0.495)	0.425	(0.495)	0.434	(0.496)	-0.009
Tier 2 university	0.154	(0.361)	0.155	(0.362)	0.152	(0.360)	0.003
Vocational college	0.046	(0.209)	0.044	(0.205)	0.048	(0.214)	-0.004
No college	0.002	(0.045)	0.002	(0.049)	0.002	(0.040)	0.001
Science track	0.662	(0.473)	0.786	(0.410)	0.499	(0.500)	0.288***
Only child	0.467	(0.499)	0.456	(0.498)	0.482	(0.500)	-0.026
Panel B. Parent							
Mother	0.589	(0.492)	0.400	(0.490)	0.838	(0.368)	-0.439***
Agricultural hukou	0.475	(0.499)	0.471	(0.499)	0.480	(0.500)	-0.009
Years of education	14.917	(2.697)	14.955	(2.621)	14.867	(2.793)	0.088
Monthly income level (RMB)							
Less than 2000	0.011	(0.105)	0.010	(0.098)	0.013	(0.115)	-0.004
2000-5000	0.132	(0.338)	0.113	(0.316)	0.156	(0.363)	-0.044***
5000-10000	0.482	(0.500)	0.478	(0.500)	0.489	(0.500)	-0.011
10000-20000	0.303	(0.460)	0.328	(0.470)	0.271	(0.445)	0.057***
More than 20000	0.071	(0.258)	0.072	(0.259)	0.071	(0.256)	0.001
Observations	2940		1667		1273		-

Notes: Expected university level is based on parental expectations regarding the university attendance of their child. Science track dummy refers to the student in science track. Only child dummy takes the value of 1 if the respondent has only one child. Mother is a dummy representing the respondent is mother of the child. Agricultural hukou takes value of 1 if the respondent has an agricultural hukou. Years of education refers to the years of education of the respondent. Monthly income level refers to the monthly income level of the respondent.

Table B3: Descriptive Statistics (Survey): Subjective Attitudes and Migration Concerns

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All		Boys		Girls		Mean diff.
Variables	Mean	SD	Mean	SD	Mean	SD	
Panel A. College migration preference							
Migration preference	0.394	(0.489)	0.416	(0.493)	0.365	(0.482)	0.050***
Panel B. Subject attitude to migration							
Concerns about studying out of home province							
Spouse	0.482	(0.500)	0.541	(0.498)	0.404	(0.491)	0.137***
Safety	0.856	(0.351)	0.818	(0.386)	0.907	(0.290)	-0.090***
Costs	0.567	(0.496)	0.569	(0.495)	0.564	(0.496)	0.005
Assitance	0.910	(0.286)	0.900	(0.300)	0.922	(0.268)	-0.022**
Spouse in province	0.328	(0.470)	0.303	(0.460)	0.361	(0.480)	-0.058***
Financial support	0.098	(0.297)	0.103	(0.303)	0.092	(0.289)	0.011
Time support	0.540	(0.498)	0.554	(0.497)	0.523	(0.500)	0.031
Larger benefit for child gender	0.177	(0.382)	0.256	(0.436)	0.074	(0.262)	0.182***
Observations	2940		1667		1273		-

Notes: Please refer to Appendix Table B1 for the definition of the variables.

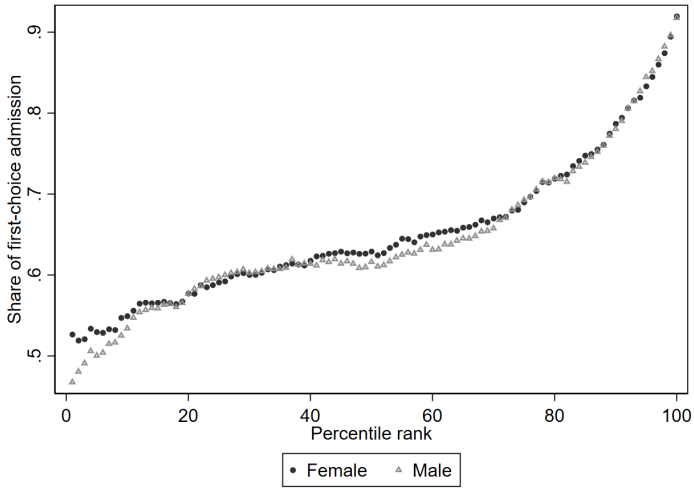
Table B4: Association between Parental Concerns and Migration Preferences

	(1)	(2)	(3)
Sample	All	Boys	Girls
Dependent variable	Migration preference		
Spouse concern	-0.038**	-0.034	-0.043
	(0.016)	(0.028)	(0.045)
Spouse in province	-0.106***	-0.086**	-0.135***
	(0.027)	(0.032)	(0.033)
Financial support	-0.028	-0.043	0.004
	(0.033)	(0.042)	(0.051)
Time support	-0.080***	-0.090***	-0.063***
	(0.022)	(0.032)	(0.020)
Safety concern	-0.066***	-0.064**	-0.086**
	(0.021)	(0.030)	(0.034)
Cost concern	-0.025	-0.054*	0.021
	(0.019)	(0.028)	(0.026)
Assistance concern	-0.059**	-0.054	-0.046
	(0.028)	(0.032)	(0.054)
Larger benefit for child gender	0.097***	0.079**	0.166***
	(0.027)	(0.033)	(0.050)
Demographic characteristics	Yes	Yes	Yes
Expected university level FE	Yes	Yes	Yes
Income level FE	Yes	Yes	Yes
City-academic track FE	Yes	Yes	Yes
Observations	2,940	1,666	1,270

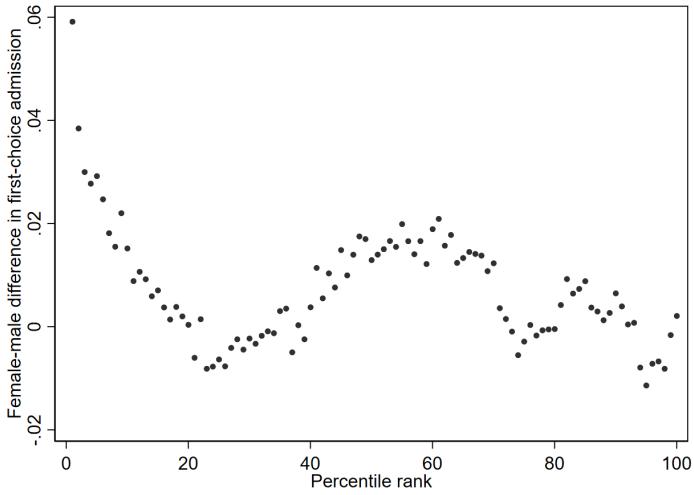
Notes: This table reports estimates on how parental attitudes are associated with preference for out-of-province college. Please refer to Appendix Table B1 for the definition of the variables. Demographic characteristics include whether the respondent is the child's mother, the respondent's hukou status (agricultural vs. non-agricultural), years of education, and whether the child has any siblings. Robust standard errors clustered by city are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

C. Gender Differences in Decision Making under Uncertainty and Parental Social-economic Status

Figure C1: Gender Differences in First-Choice University Admission



(a) Raw pattern by gender



(b) Female-male difference

Notes: This figure plots first-choice admission outcomes and gender differences across percentile ranks. Panel A reports the average outcome by gender, plotting the share of students admitted to their first-choice university. Panel B reports the gender difference (calculated as the female mean minus the male mean) for the outcome in Panel A. In all panels, the x-axis represents the student's percentile rank, divided into 100 integer bins. In Panel A, grey hollow triangles denote male students, and black solid circles denote female students. In Panel B, black solid circles represent the difference in means between female and male students within each bin. The sample consists of all students admitted to college.

Table C1: Gender Gap in Migration across Admission Information Regimes

	(1)	(2)	(3)
Dependent variable:		Out of province	
Female	-0.045***		
	(0.001)		
Female × Low information		0.004***	0.003*
		(0.001)	(0.001)
Female × High information		0.001	0.002
		(0.002)	(0.002)
Female-province FE		Yes	Yes
Female-year FE			Yes
County-year-track FE	Yes	Yes	Yes
Percentile rank FE	Yes	Yes	Yes
Demographic characteristics	Yes	Yes	Yes
Dependent mean	0.251	0.251	0.251
Observations	13,524,632	13,524,632	13,524,632
Sample	Provinces with policy changes	Provinces with policy changes	Provinces with policy changes

Notes: This table reports estimates of the heterogeneity in gender gaps across different college admission regimes, with sample restricted to the 13 provinces that experienced changes in admission policies between 1999 and 2003. The dependent variable is an indicator equal to 1 if the student migrates out of their home province, and 0 otherwise. The admission regimes are defined as follows: "Low information" (preference submission before the exam), "Medium information" (submission after the exam but before knowing scores), and "High information" (submission after knowing scores). Column (1) reports the average gender gap for the pooled sample. Columns (2) and (3) include interactions between the female indicator and the regime dummies (Low Information and High Information); the reference category is the Medium Information regime. Demographic characteristics include agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table C2: Gender Differences in Family Socioeconomic Status (CCSS Data)

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Ln(family income)	Father education	Mother education	Father agricultural hukou	Mother agricultural hukou
Female	0.105***	0.480***	0.682***	-0.053***	-0.050***
	(0.021)	(0.093)	(0.112)	(0.013)	(0.014)
Hometown city FE	Yes	Yes	Yes	Yes	Yes
Percentile rank FE	Yes	Yes	Yes	Yes	Yes
Dependent mean	10.587	10.988	9.961	0.551	0.569
Observations	17,311	16,341	16,343	17,311	17,311

Notes: This table reports gender differences in parents' socioeconomic status using survey data from the China College Student Survey (2011–2013). The independent variable of interest in all specifications is an indicator for being female. The dependent variables are indicated in the column headers: Column (1) is the natural logarithm of total household income in the previous year. Columns (2) and (3) report the father's and mother's years of schooling, respectively. Columns (4) and (5) are indicators for the father and mother holding an agricultural hukou, respectively. Panel A reports estimates controlling only for hometown city fixed effects. Panel B additionally controls for hometown city-by-year-by-track fixed effects and fixed effects for each of the 20 quantiles of the standardized college entrance exam score. Robust standard errors, clustered at the hometown city level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

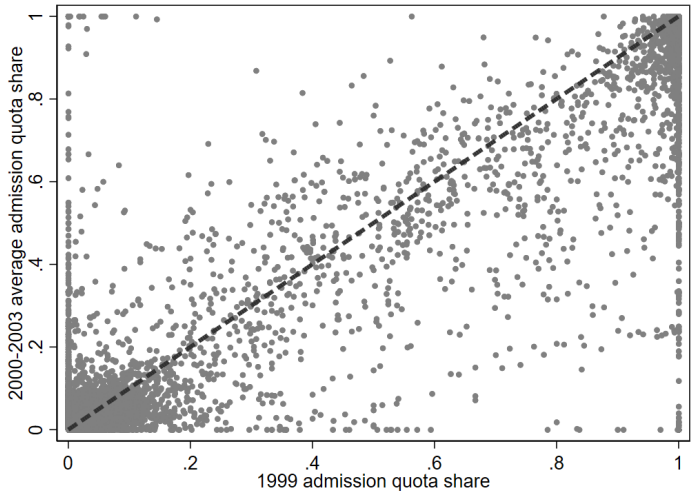
Table C3: Gender Differences in Number of Siblings and Per-Child Family Resources (CCSS Data)

	(1)	(2)
Dependent variable:	Number of siblings	Ln(family income per child)
Female	0.064***	0.069***
	(0.015)	(0.024)
Hometown city FE	Yes	Yes
Percentile rank FE	Yes	Yes
Dependent mean	1.757	10.145
Observations	17,293	17,293

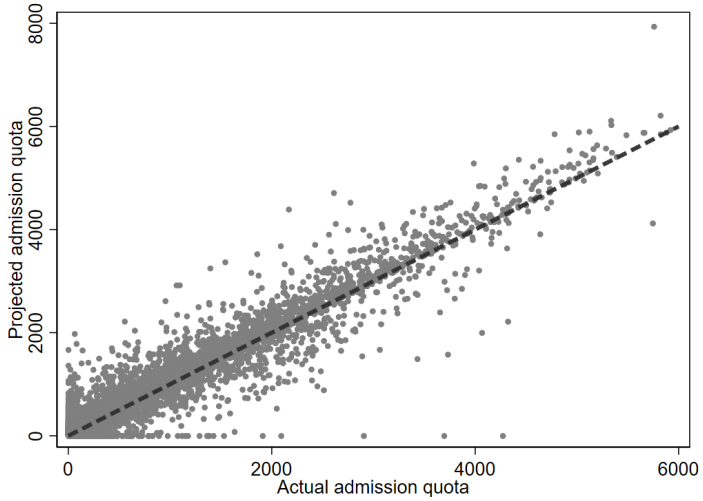
Notes: This table reports gender differences in the number of siblings and per-child family income using survey data from the China College Student Survey (2011–2012). The independent variable of interest is an indicator for being female. Column (1) reports the total number of siblings the student has. Column (2) reports the natural logarithm of family income per child, calculated as the total household income divided by the total number of children in the family (the number of siblings plus one). All columns control for hometown city fixed effects and fixed effects for each of the 20 quantiles of the standardized college entrance exam score. Robust standard errors, clustered at the hometown city level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

D. More Results on Consequences and Counterfactual Analysis

Figure D1: Validity of the Shift-Share Instrumental Variable Design



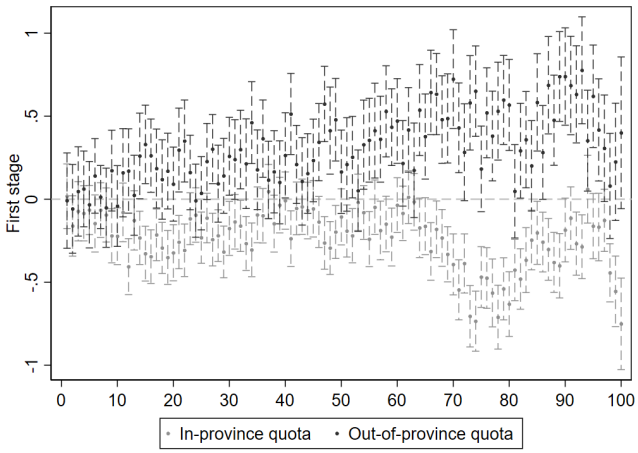
(a) Corr between 2000–2003 average quota share and the baseline 1999 quota share



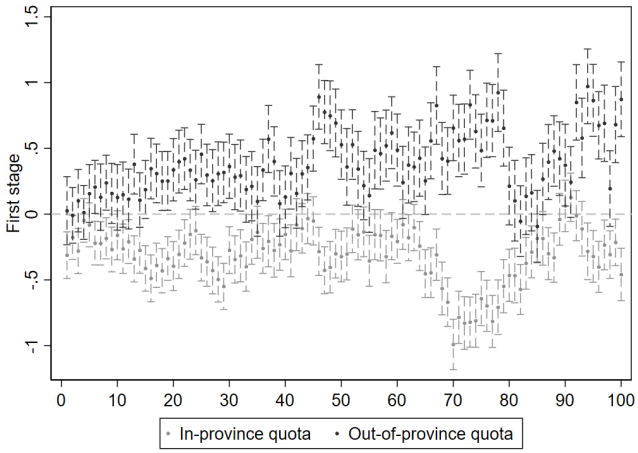
(b) Corr between shift-share projected quota and the actual quota

Notes: This figure illustrates the validity of the shift-share instrument design. Panel A plots the 2000–2003 average admission quota share against the baseline 1999 admission quota share. Each dot represents an observation at the university-province-track level. The admission quota share is defined as the number of students admitted from a specific home province divided by the total number of students admitted by that college-track. Panel B plots the shift-share projected admission quota against the actual admission quota. Each dot represents an observation at the university-year-province-track level for the years 2000 to 2003. In both panels, the dashed line represents the 45-degree line.

Figure D2: First-Stage Estimates by Percentile Rank and Gender



(a) First stage estimates by percentile for females



(b) First stage estimates by percentile for males

Notes: This figure plots the first-stage estimates of the instrumental variables across percentile ranks by gender. The sample is restricted to students admitted to college between 2000 and 2003. Panel A reports the estimates for the female subsample, and Panel B reports the estimates for the male subsample. Each marker represents the point estimate of the coefficient on the respective instrumental variable from a separate regression estimated using the subsample of students within that specific percentile bin. The dependent variable is an indicator equal to 1 if the student migrates out of their home province, and 0 otherwise. The y-axis displays the estimated coefficients for the two excluded instruments: the shift-share predicted in-province and out-of-province admission quotas. Grey dots denote the coefficients for the in-province quota, and black dots denote the coefficients for the out-of-province quota. All regressions control for year fixed effects, county-by-track fixed effects, and student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. The x-axis represents the student's percentile rank, divided into 100 integer bins. Vertical dashed lines indicate 95 percent confidence intervals, based on robust standard errors clustered at the home county level.

Table D1: First-Stage Estimates for the NCEE–CCSS Matched Sample

	(1)	(2)	(3)
Dependent variable:	Out of province		
In-province quota	-0.322***	-0.211***	-0.388***
	(0.024)	(0.029)	(0.026)
Out-of-province quota	0.316***	0.250***	0.333***
	(0.053)	(0.068)	(0.062)
Year FE	Yes	Yes	Yes
County-track FE	Yes	Yes	Yes
Percentile rank FE	Yes	Yes	Yes
Demographic characteristics	Yes	Yes	Yes
Dependent mean	0.358	0.315	0.384
Observations	1,129,238	432,055	696,935
Kleibergen-Paap F-stat	91.833	28.043	112.716
Sample	NCEE-CCSS matched	Female	Male

Notes: This table reports the first-stage estimates for the instrumental variable regressions. The sample is restricted to students admitted to college between 2000 and 2003 who can be matched to the CCSS data, corresponding to the sample used for the expected income analysis in the second stage. The dependent variable is an indicator equal to 1 if the student migrates out of their home province, and 0 otherwise. The excluded instruments are the predicted in-province and out-of-province admission quotas. These quotas are constructed using a shift-share approach, interacting the 1999 college-track admission shares in each home province with the 2000–2003 admission shifts. The predicted admissions are aggregated to the home province level for in-province and out-of-province colleges, respectively, and then divided by the total number of NCEE takers in the home province in that year. Column (1) reports estimates for the NCEE-CCSS matched pooled sample. Columns (2) and (3) report estimates for the female and male subsamples, respectively. All columns control for year fixed effects, county-by-track fixed effects, percentile rank fixed effects, and student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. The Kleibergen-Paap rk Wald F-statistic for weak instruments is reported for each specification. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table D2: 2SLS Estimates of the College Migration Premium: Pooled Sample

	(1)	(2)
Dependent variable:	985 university	Expected income
Out of province	0.128***	736.772***
	(0.021)	(102.002)
Year FE	Yes	Yes
County-track FE	Yes	Yes
Percentile rank FE	Yes	Yes
Demographic characteristics	Yes	Yes
Dependent mean	0.052	2982.972
Observations	11,987,370	1,129,238
Sample	Full sample	NCEE-CCSS matched

Notes: This table reports the second-stage estimates of the instrumental variable regressions. The sample is restricted to students admitted to college between 2000 and 2003. The endogenous variable is an indicator for migrating out of the home province, which is instrumented by the shift-share predicted in-province and out-of-province admission quotas. In Column (1), the dependent variable is an indicator equal to 1 if the student is admitted to a Project 985 university, and 0 otherwise. In Column (2), the dependent variable is the expected monthly income (in RMB). This expected income is constructed using the CCSS data by calculating the average monthly wage of the best job offer for graduates from each university, specific to the student's gender and college migration status (local or migrant). These averages are then matched to the NCEE administrative data by university name. Column (1) reports estimates for the pooled sample, while Column (2) reports estimates for the NCEE-CCSS matched sample. All columns control for year fixed effects, county-by-track fixed effects, percentile rank fixed effects (calculated within each province-year-track cohort), and student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table D3: First-Stage Estimates by Local Educational Resources

	(1)	(2)	(3)	(4)
Dependent variable:		Out of province		
Panel A		Sample: Full sample		
In-province quota	-0.040**	-0.144***	-0.304***	-0.474***
	(0.018)	(0.020)	(0.049)	(0.039)
Out-of-province quota	0.249***	0.278***	0.454***	0.602***
	(0.044)	(0.045)	(0.105)	(0.053)
Dependent mean	0.190	0.250	0.294	0.356
Observations	3,670,737	4,690,719	1,605,105	2,020,731
Kleibergen-Paap F-stat	15.831	31.694	52.783	141.190
Panel B		Sample: NCEE-CCSS matched sample		
In-province quota	-0.144***	-0.384***	-0.328***	-0.485***
	(0.033)	(0.034)	(0.056)	(0.043)
Out-of-province quota	0.222***	0.371***	0.474***	0.511***
	(0.084)	(0.086)	(0.113)	(0.086)
Dependent mean	0.267	0.350	0.433	0.460
Observations	306,700	482,090	125,355	214,845
Kleibergen-Paap F-stat	10.318	63.478	27.174	90.276
Year FE	Yes	Yes	Yes	Yes
County-track FE	Yes	Yes	Yes	Yes
Percentile rank FE	Yes	Yes	Yes	Yes
Demographic characteristics	Yes	Yes	Yes	Yes
Sample	Female	Male	Female	Male
	Home prov. with 985	Home prov. with 985	Home prov. without 985	Home prov. without 985

Notes: This table reports the first-stage estimates for the instrumental variable regressions by home province educational resources. The sample is restricted to students admitted to college between 2000 and 2003. The dependent variable is an indicator equal to 1 if the student migrates out of their home province, and 0 otherwise. The excluded instruments are the predicted in-province and out-of-province admission quotas. These quotas are constructed using a shift-share approach, interacting the 1999 college-track admission shares in each home province with the 2000–2003 admission shifts. The predicted admissions are aggregated to the home province level for in-province and out-of-province colleges, respectively, and then divided by the total number of NCEE takers in the home province in that year. Panel A uses the full sample, while Panel B uses the NCEE-CCSS matched sample. Across both panels, the sample is divided based on whether the student's home province has at least one Project 985 university. Columns (1) and (2) report estimates for students from home provinces with Project 985 universities, while Columns (3) and (4) report estimates for students from home provinces without Project 985 universities. Within these categories, Columns (1) and (3) report estimates for the female subsample, while Columns (2) and (4) report estimates for the male subsample. All columns control for year fixed effects, county-by-track fixed effects, percentile rank fixed effects (calculated within each province-year-track cohort), and student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. The Kleibergen-Paap rk Wald F-statistic for weak instruments is reported for each specification. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table D4: 2SLS Estimates by Local Educational Resources

	(1)	(2)	(3)	(4)
Panel A		Dependent variable: 985 university		
Out of province	-0.040	-0.013	0.283***	0.213***
	(0.047)	(0.045)	(0.028)	(0.016)
Dependent mean	0.048	0.071	0.025	0.041
Observations	3,670,737	4,690,719	1,605,105	2,020,731
Panel B		Dependent variable: Expected income		
Out of province	337.262	424.404***	1,166.889***	733.691***
	(430.698)	(101.741)	(170.922)	(76.842)
Dependent mean	2866.999	3139.477	2683.497	2972.225
Observations	306,700	482,090	125,355	214,845
Year FE	Yes	Yes	Yes	Yes
County-track FE	Yes	Yes	Yes	Yes
Percentile rank FE	Yes	Yes	Yes	Yes
Demographic characteristics	Yes	Yes	Yes	Yes
Sample	Female	Male	Female	Male
	Home prov. with 985	Home prov. with 985	Home prov. without 985	Home prov. without 985

Notes: This table reports the second-stage estimates of the instrumental variable regressions by home province educational resources. The sample is restricted to students admitted to college between 2000 and 2003. The endogenous variable is an indicator for migrating out of the home province, which is instrumented by the shift-share predicted in-province and out-of-province admission quotas. In Panel A, the dependent variable is an indicator equal to 1 if the student is admitted to a Project 985 university, and 0 otherwise. In Panel B, the dependent variable is the expected monthly income (in RMB). This expected income is constructed using the CCSS data by calculating the average monthly wage of the best job offer for graduates from each university, specific to the student's gender and college migration status. These averages are then matched to the NCEE administrative data by university name. Panel A uses the full sample of students admitted between 2000 and 2003, while Panel B uses the NCEE-CCSS matched sample. Across both panels, the sample is divided based on whether the student's home province has at least one Project 985 university. Columns (1) and (2) report estimates for students from home provinces with Project 985 universities, while Columns (3) and (4) report estimates for students from home provinces without Project 985 universities. Within these categories, Columns (1) and (3) report estimates for the female subsample, while Columns (2) and (4) report estimates for the male subsample. All columns control for year fixed effects, county-by-track fixed effects, percentile rank fixed effects, and student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.

Table D5: Gender Migration Gap by Local Availability of Elite Universities

	(1)	(2)
Dependent variable:	Out of province	
Female	-0.045***	-0.042***
	(0.001)	(0.001)
Female × Home prov. without 985		-0.010***
		(0.001)
County-year-track FE	Yes	Yes
Percentile rank FE	Yes	Yes
Demographic characteristics	Yes	Yes
Dependent mean	0.251	0.251
Observations	13,524,632	13,524,632
Sample	Full sample	Full sample

Notes: This table reports the baseline gender gap in college migration and its heterogeneity by home province Project 985 university availability. The dependent variable is an indicator equal to 1 if the student migrates out of their home province for college, and 0 otherwise. Column (1) reports the baseline estimates, where the independent variable of interest is an indicator for being female. Column (2) adds an interaction term between the female indicator and a dummy variable indicating whether the student's home province lacks a Project 985 university. All columns control for county-by-year-by-track fixed effects, percentile rank fixed effects, and student demographic characteristics, including agricultural hukou dummy, ethnic minority dummy, and age dummies. Robust standard errors, clustered at the home county level, are reported in parentheses. *** Significant at the 1 percent level; ** significant at the 5 percent level; * significant at the 10 percent level.