Opening Doors, Widening Gaps?
The Distributional and Intergenerational Effects of China's Higher Education Expansion

By Xi XIANG*

This paper examines the distributional and intergenerational effects of China's 1999 higher education expansion, one of the largest supply-side education reforms in modern history. Using nationally representative CFPS microdata, we exploit cohort-based variation in policy exposure with a difference-in-differences framework, complemented by event study and regression discontinuity designs. We find the expansion significantly raised aggregate college enrollment, but gains were heavily concentrated among individuals with urban childhood hukou and more educated parents. Critically, maternal education has a nearly twofold larger effect on children's college enrollment than paternal education at the same attainment level, and the expansion amplified this intergenerational persistence. While the reform improved income ranks for lower- and middle-income individuals, it strengthened the intergenerational transmission of advantage. Our findings highlight that supply-side expansion alone is insufficient to reduce intergenerational inequality without targeted interventions.

JEL Classification: *I21, I28, J62, J31, D31*
Keywords: *Higher Education Expansion, Intergenerational Mobility, Income Inequality, Hukou System, Parental Education, China*

I. Introduction

Over the past several decades, countries worldwide have dramatically expanded access to higher education. While such reforms are universally motivated by policy goals of boosting aggregate human capital accumulation and long-run economic growth, their distributional consequences for economic inequality and intergenerational mobility remain theoretically ambiguous and empirically contested. On the one hand, relaxing binding supply constraints in tertiary education can level the playing field by reducing institutional barriers to college access for socioeconomically disadvantaged individuals. On the other hand, households with greater economic and educational endowments may be better positioned to capture the benefits of expanded enrollment slots, potentially reinforcing pre-existing intergenerational disparities in opportunity. A large literature has documented the

* Department of Economics, Nanjing University, xiangxi@smail.nju.edu.cn, NJU ID:231098288

heterogeneous effects of higher education reforms across family background, but critical gaps remain in understanding how intra-household heterogeneity in the asymmetric roles of maternal and paternal education—shape the distribution of reform dividends.

This paper evaluates the causal effects of China's landmark higher education expansion on educational attainment, labor market outcomes, and intergenerational mobility. Launched in 1999, the reform nearly doubled national college admission quotas overnight, and represents one of the largest and most rapid supply-side education reforms in modern history. The policy exclusively affected cohorts born in 1981 or later, who reached the standard college-entry age of 18 after the reform was implemented, creating a sharp quasi-experimental cutoff. Using nationally representative microdata from the China Family Panel Studies (CFPS), we exploit this cohort-based variation in policy exposure to identify how large-scale expansions in higher education access shape outcomes across family backgrounds, with a particular focus on childhood urban-rural hukou status and the gendered dimension of parental education.

Our empirical strategy combines a baseline difference-in-differences (DID) design with complementary event study and fuzzy regression discontinuity (RD) frameworks to validate causal identification. We pay particular attention to heterogeneous effects across family background, and unpack the intra-household mechanisms through which maternal and paternal education mediate the returns to expanded higher education access.

We document three core findings. First, the higher education expansion significantly boosted college enrollment, with heterogeneous effects heavily concentrated among those with urban childhood hukou: urban cohorts exposed to the reform saw a 21.3% higher increase in college enrollment than rural counterparts, a 1% level statistically significant difference. Second, while the expansion lifted educational attainment for all family background groups, marginal gains were far larger for individuals with more educated parents, and we document a striking gender asymmetry in intergenerational transmission—maternal education has a nearly twofold larger impact on children's college enrollment than paternal education at the same attainment level. The reform further amplified this asymmetry: post-expansion, each additional level of maternal education raised children's college enrollment probability by 5.8%, compared to 5.2% for paternal education, with children from low maternal education households emerging as the most disadvantaged group with the smallest reform gains. Third, the educational dividends of the expansion translated into improved within-cohort income ranks, with gains concentrated in the 25th, 50th and 75th income quantiles, yielding a modest equalizing effect on overall income inequality even as intergenerational disparities in opportunity persisted.

This paper makes two key contributions to the literature. First, we provide new causal evidence on the heterogeneous distributional effects of large-scale higher education expansion, documenting how family background shapes both educational and labor market returns to supply-side education reform in the world's largest higher education system. Second, we unpack the gendered intergenerational mechanisms of the reform, documenting the asymmetric role of maternal and paternal education in mediating the benefits of higher education expansion. Existing research on China's college expansion has largely focused on average effects by aggregate parental education, while we show that maternal education has a substantially larger impact on children's educational attainment, and that the reform amplified this maternal advantage in intergenerational transmission. This finding fills a critical gap in the literature on education policy and intergenerational mobility, and highlights the importance of accounting for intra-household heterogeneity in human capital when evaluating education reforms. Our results imply that while higher education expansion can raise aggregate educational attainment, it may simultaneously strengthen intergenerational persistence of advantage without complementary policies that address informational, academic, and institutional barriers faced by disadvantaged households, particularly those with low maternal education.

The remainder of the paper proceeds as follows. Section II reviews the related literature. Section III describes the institutional background of China's higher education expansion. Section IV outlines the conceptual framework. Section V presents the data and sample construction. Section VI details our empirical strategy and identification approach. Section VII reports the main empirical results. Section VIII explores mechanisms and heterogeneous effects, with a focus on maternal and paternal education. Section IX presents a series of robustness checks. Section X discusses policy implications and welfare considerations. Section XI concludes.

II. Related Literature

This paper connects to four core strands of existing research.

First, we contribute to the vast literature on the causal effects of higher education expansion on educational attainment and labor market outcomes. A foundational body of work uses policy-induced variation in schooling access to identify the causal returns to education (Angrist and Krueger, 1991; Card, 1999), while more recent research examines large-scale tertiary education reforms across countries, with a focus on heterogeneous effects across socioeconomic groups (Bleakley, 2010; Deming et al., 2016). We extend this literature in two ways: we document the distributional effects of a historic expansion in the world's largest higher education system, and we separate the independent roles of urban-rural status and parental human capital in driving heterogeneous returns to the reform.

Second, our analysis advances the literature on intergenerational mobility and the role of family background in shaping economic opportunity. Classical theoretical and empirical work documents the persistence of economic status across generations (Becker and Tomes, 1979; Solon, 1992), and recent research emphasizes how public policy and institutional environments mediate intergenerational transmission (Chetty et al., 2014; Corak, 2013). A large body of work shows that parental education and income strongly shape children's schooling decisions and labor market returns to education (Black and Devereux, 2005; Carneiro, Heckman and Vytlacil, 2007). We add to this literature by showing that supply-side education expansions can amplify, rather than reduce, the influence of parental education on children's outcomes, even as they raise aggregate educational attainment.

Third, we speak to the growing literature on education and inequality in China. Existing research has documented large and persistent urban-rural gaps in educational access and outcomes (Hannum, 2003; Loyalka et al., 2013), and several studies examine the average effects of China's higher education expansion on educational attainment and wages (Deng et al., 2013; Li, Liu and Zhang, 2014). However, prior work has not fully unpacked how the reform shaped intergenerational mobility and the distribution of outcomes across family background, particularly the relative roles of maternal and paternal education in mediating policy effects. This paper fills this gap.

Finally, our work relates to research on the interaction between supply-side education reforms and household behavioral responses. Theoretical and empirical work highlights how information frictions, academic preparation gaps, and heterogeneous expectations shape the effectiveness of education policies (Durlauf, 2004; Lavecchia, Liu and Oreopoulos, 2016). Our findings are consistent with models in which households with greater educational capital are better able to translate expanded schooling access into realized human capital and labor market gains, even in the absence of differential changes in the returns to college.

III. Institutional Background

China's higher education system underwent a seismic transformation starting in 1999, with the launch of a national tertiary education expansion that remains unprecedented in scale globally. Prior to the reform, access to higher education was extremely limited and highly selective: enrollment was strictly capped by centralized national planning, institutional capacity was tightly constrained, and admission was determined almost exclusively by performance on the National College Entrance Examination (Gaokao). In 1998, fewer than 10% of college-age individuals were enrolled in tertiary education, one of the lowest rates among middle-income countries at the time.

In June 1999, the Chinese central government announced a comprehensive higher education expansion with three formal policy goals: increasing the national stock of human capital, alleviating short-term youth unemployment pressures, and fostering long-run economic growth. The reform was implemented immediately: national college admission quotas were increased from 1.08 million in 1998 to 1.59 million in 1999, a 47% year-over-year increase. This was followed by sustained annual enrollment growth, massive public investments in university infrastructure, and the establishment of hundreds of new higher education institutions. As a result, total tertiary enrollment more than tripled between 1998 and 2008, making China's higher education system the largest in the world by total enrollment.

The reform had a sharp, cohort-specific impact on college access. Individuals born in 1981 reached the standard Gaokao-taking and college-entry age of 18 in 1999, and thus were the first cohort fully exposed to the expanded enrollment quotas. Cohorts born before 1981 completed their Gaokao and college enrollment before the reform, and thus were largely unaffected by the expansion. This discrete cohort-based cutoff forms the core of our quasi-experimental identification strategy.

Critically, while the reform dramatically relaxed supply constraints on college access, it did not alter the core structure of the Gaokao system, the centralized admission rules, or the large gaps in academic preparation, information access, and family resources between advantaged and disadvantaged households. The expansion increased the total number of college seats, but competition for admission remained intense, and household resources continued to play a central role in students' academic performance and admission outcomes. This institutional setting allows us to examine how family background mediates the effects of expanded higher education access, holding the core structure of the education system constant.

IV. Conceptual Framework

We present a simple, tractable framework to clarify how higher education expansion interacts with family background to shape educational and labor market outcomes, and to derive testable hypotheses for our empirical analysis. Consider an individual i born in cohort c who makes a binary decision to attend college. The expected utility of college attendance is given by:

$$U_i = R_i - C_i, \tag{1}$$

where R_i denotes the expected lifetime labor market return to college completion, and C_i represents the total cost of college attendance. The individual will attend college if and only if $U_i > 0$.

The returns to college R_i depend on individual ability, labor market conditions, and the quality of the higher education institution attended. The total cost C_i includes both direct monetary costs, such as tuition, fees, living expenses, and non-monetary costs: the effort cost of academic preparation for the Gaokao, information frictions about college admission and financial aid, and the utility cost of foregone earnings during college.

We allow family background X_i—our core dimensions of interest, childhood urban hukou status and parental education—to affect both the returns and costs of college:

$$R_i = R(X_i), \quad C_i = C(X_i, S_c), \tag{2}$$

where S_c captures the aggregate supply of college seats for birth cohort c.

The higher education expansion operates through a sharp, exogenous increase in S_c for cohorts born in 1981 and later. A larger supply of college seats reduces the admission threshold for Gaokao performance, which lowers the total cost of college attendance C_i for all individuals. However, the magnitude of this cost reduction is heterogeneous across family background: individuals from more advantaged households with urban hukou, higher parental education, experience larger effective cost reductions. This is because their greater parental human capital, better access to information about college admissions, and stronger academic preparation allow them to better compete for the newly available college seats, even as the overall admission threshold falls.

This framework delivers three testable hypotheses that guide our empirical analysis:

1) The higher education expansion will increase the average probability of college attendance for exposed cohorts.

2) The positive effect of the expansion on college attendance will be larger for individuals from advantaged family backgrounds.

3) The educational gains induced by the expansion will translate into improved labor market outcomes, with heterogeneous effects mirroring those for educational attainment.

Notably, the framework shows that heterogeneous effects of the expansion can arise even if the returns to college R_i are identical across family background groups. This motivates our core empirical specification, which focuses on the interaction between expansion exposure and family background, and our analysis of heterogeneous effects across hukou status and parental education.

V. Data

This paper draws on microdata from the China Family Panel Studies (CFPS[1]), a nationally representative longitudinal household survey administered by the Institute of Social Science Survey (ISSS) at Peking University. The CFPS covers 25 provinces in China and collects detailed, nationally representative information on individual demographics, educational attainment, labor market outcomes, household characteristics, and intergenerational linkages. We primarily rely on the 2022 wave of the CFPS, and merge earlier waves (2010–2020) to recover time-invariant

[1]In accordance with CFPS data usage requirements, we cite the dataset as: China Family Panel Studies (CFPS), Institute of Social Science Survey (ISSS), Peking University.

information on childhood hukou status and parental background, which are not fully available in the 2022 cross-section.

A. Sample Construction

Our baseline sample is restricted to individuals with non-missing information on birth year, childhood hukou status at age 12, and highest educational attainment—our core variables for identification and outcome measurement. We further exclude individuals without valid hukou status and those with foreign nationality, who are not part of the Chinese higher education admission system.

The final analytical sample includes up to 27,001 individuals. Sample sizes vary across outcome specifications due to differential item non-response: for example, labor income data is only available for individuals with active labor market participation, resulting in a smaller sample for income-related outcomes.

B. Key Variable Definitions

All variable definitions are aligned with our quasi-experimental identification strategy, and categorized by their role in the analysis. We only present core definitional logic here, with full code-level construction details, CFPS raw variable mapping, and missing value processing rules reported in the Appendix A.

Demographic and Policy Exposure Variables

Demographic Characteristics

We use birth year to define the cohort, the core source of our quasi-experimental variation in policy exposure. Gender is a binary indicator equal to 1 for male individuals and 0 for female individuals, included as a control in all baseline specifications.

Exposure to Higher Education Expansion

Our core treatment variable is a binary indicator *post_expansion*, equal to 1 if an individual was born in 1981 or later, and 0 otherwise. This definition is directly motivated by the institutional context: individuals born in 1981 reached the standard college-entry age of 18 in 1999, the first year of the national expansion, and thus were fully exposed to the reform. Cohorts born in 1980 or earlier completed their college enrollment before the reform, and serve as the control group.

Family Background Variables

Childhood Hukou Status

Our core measure of urban-rural background is hukou registration status at age 12, a critical transition point between primary and junior secondary education in China. This measure eliminates endogeneity from post-adulthood hukou conversion, which may be directly affected by college attendance and educational attainment. The binary indicator *child_urban* equals 1 for individuals with non-agricultural hukou at age 12, and 0 for those with agricultural hukou.

Parental Background

Our core measure of family human capital is *parental highest education*, defined as the maximum of the father's and mother's highest completed educational attainment. For heterogeneity analysis, we also construct separate measures of paternal and maternal education, and a binary indicator for high parental education: high school degree or above. We further measure parental public-sector employment and within-cohort parental income rank as additional dimensions of family socioeconomic status.

Outcome Variables

Educational Outcomes

Our primary educational outcome is *ever_college*, a binary indicator equal to 1 if the individual's highest completed education is junior college or above, capturing the extensive margin of higher education access. For robustness and intensive margin analysis, we also construct an indicator for bachelor's degree completion or above, and distinguish between public and non-public undergraduate institutions.

Labor Market Outcomes

Our primary labor market outcome is *income_rank_cohort*, the individual's percentile rank in the within-birth-cohort distribution of annual labor income. We use within-cohort rank rather than absolute income to eliminate confounding from inflation, life-cycle income profiles, and cohort-specific wage differences. Secondary outcomes include absolute annual labor income and public-sector employment status.

C. *Descriptive Statistics*

Table 1 reports summary statistics for the main variables in our analysis. Approximately 43% of the sample belongs to post-expansion cohorts, which means they born in 1981 or later. Only 13% of the sample has urban childhood hukou status, and the average parental highest education level is 2.43, corresponding to lower secondary education, indicating substantial variation in family background and meaningful scope for heterogeneous treatment effects. The sample is evenly split between male and female individuals, with an average age of 44.85 at the time of the 2022 survey.

Table 1—Summary Statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
Birth year	25,649	1977.01	19.70	1921	2013
Age	27,001	44.85	19.78	9	101
Gender	27,001	0.50	0.50	0	1
Post-expansion cohort	27,001	0.43	0.50	0	1
Urban hukou at age 12	14,557	0.13	0.34	0	1
Parental highest education	18,570	2.43	1.26	1	8
Father income	3,674	52,033	50,943	0	1,320,000
Mother income	2,814	33,128	35,686	0	1,000,000
Child highest education	26,986	2.94	1.51	1	8
Within-cohort income rank	10,082	0.50	0.29	0.01	1.00
Public sector employment	27,001	0.01	0.11	0	1

Notes: The table reports descriptive statistics for the main variables used in the empirical analysis. Income ranks are calculated within birth cohorts to account for life-cycle and cohort-specific wage differences.

VI. Empirical Strategy and Identification

Our empirical strategy exploits the sharp, cohort-based variation in exposure to China's 1999 higher education expansion. The reform exclusively affected cohorts born in 1981 or later, creating a natural quasi-experiment where we can compare outcomes for individuals born just before and after the policy cutoff. We use three complementary identification frameworks to establish causal effects: a baseline difference-in-differences (DID) model, an event study to validate the parallel trends assumption, and a fuzzy regression discontinuity (RD) design to reinforce causal identification. We complement these core models with a range of heterogeneity, mechanism, and robustness analyses.

A. Baseline Difference-in-Differences Model

To estimate the heterogeneous effects of the higher education expansion across family background, we estimate the following baseline DID specification:

$$Y_{ic} = \alpha + \beta_1 \text{Post}_c + \beta_2 X_i + \beta_3 \left(\text{Post}_c \times X_i\right) + \gamma_c + \delta \mathcal{Z}_{ic} + \varepsilon_{ic}, \tag{3}$$

where:

Y_{ic} denotes the educational or labor market outcome for individual i born in birth cohort c;

Post_c is the binary indicator for post-expansion cohorts, born 1981 or later;

X_i is our core measure of family background, either childhood urban hukou status or parental highest education;

$\text{Post}_c \times X_i$ is the interaction term between policy exposure and family background, our coefficient of interest;

γ_c is a full set of birth cohort fixed effects, which absorb all time-invariant cohort-level shocks;

$\mathcal{Z}_{ic}$ is a vector of individual-level control variables, including gender and parental education fixed effects;

ε_{ic} is the idiosyncratic error term.

Our core coefficient of interest is β_3, which captures the differential effect of the higher education expansion on outcomes for individuals with advantaged family background ($X_i = 1$) relative to those with disadvantaged background ($X_i = 0$). For example, when X_i is urban childhood hukou, β_3 measures the difference in the expansion's effect on college attendance between urban and rural cohorts.

We cluster standard errors at the birth cohort level in all specifications. This is the appropriate level of clustering, as our treatment variable policy exposure varies only at the cohort level, and clustering at the treatment unit accounts for potential serial correlation in outcomes within cohorts.

The key identifying assumption for the DID model is the parallel trends assumption: in the absence of the higher education expansion, the difference in outcomes between advantaged and disadvantaged family background groups would have evolved in parallel across pre- and post-expansion cohorts. We formally test this assumption using the event study design described below.

B. Event Study Design

To validate the parallel trends assumption and document the dynamic effects of the expansion, we estimate an event study specification that replaces the binary post-expansion indicator with a full set of relative cohort dummies, centered at the 1981 policy cutoff:

$$Y_{ic} = \alpha + \sum_{\tau=-10,\tau\neq 0}^{10} \beta_\tau \cdot \not{\Bbbk}(\text{rel\_year}_s = \tau) + \gamma X_i + \delta \mathcal{Z}_{ic} + \varepsilon_{ic}, \tag{4}$$

where rel_year_s is the shifted relative cohort variable, defined as (birth year $-$ 1981) + 10. We winsorize the relative cohort variable at -10 and +10 to avoid noise from extreme birth cohorts, and set the 1981 birth cohort ($\tau = 0$) as the omitted reference group. All other variables are identical to the baseline DID specification.

This event study design serves two key purposes. First, it allows us to test the parallel trends assumption: if the pre-treatment coefficients ($\tau < 0$) are small in magnitude and statistically indistinguishable from zero, this validates that outcomes for advantaged and disadvantaged groups were evolving in parallel before the reform. Second, it documents the dynamic evolution of the expansion's effects across cohorts, confirming that the change in outcomes is discrete and coincides with the policy cutoff, rather than reflecting a smooth pre-existing cohort trend.

C. Fuzzy Regression Discontinuity Design

As a complementary identification strategy to reinforce causal interpretation, we implement a fuzzy regression discontinuity (RD) design using birth year relative to the 1981 cutoff as the running variable. This design leverages the fact that eligibility for the higher education expansion changes discontinuously at the 1981 birth cohort, while all other individual and household characteristics should evolve continuously across the cutoff.

The running variable for the RD design is $\text{birthy}_c = \text{birth year} - 1981$, which measures the distance between an individual's birth cohort and the policy cutoff. The design is fuzzy because crossing the 1981 cutoff determines eligibility for the expansion, but does not perfectly predict the endogenous treatment college attendance. We use an indicator for crossing the cutoff as an instrumental variable for actual exposure to the expansion, and estimate the model using the robust bias-corrected RD estimator from Calonico, Cattaneo and Titiunik (2014).

Our baseline RD specification uses a bandwidth of 5 years around the 1981 cutoff, with a second-order polynomial fit on either side of the cutoff to model the underlying trend in the running variable. We cluster standard errors at the birth cohort level, consistent with our baseline DID model.

The key identifying assumption for the RD design is the continuity assumption: all pre-determined individual and household characteristics evolve continuously across the 1981 birth cohort cutoff, and the only discontinuous change at the cutoff is eligibility for the higher education expansion. This assumption rules out precise manipulation of birth year to select into the reform, which is highly unlikely in our setting, as the expansion was announced in 1999, nearly 20 years after the birth cohorts around the cutoff were born. We further validate this assumption with a donut RD design in our robustness checks.

D. Additional Analyses

We complement our core identification frameworks with a range of additional analyses to unpack heterogeneous effects, address potential identification threats, and explore mechanisms.

Grouped DID

We estimate the baseline DID specification separately for subgroups defined by high and low parental education and parental public-sector employment, to document heterogeneous treatment effects across family background dimensions.

Quantile Regressions

We estimate quantile regressions at the 25th, 50th, and 75th percentiles of the within-cohort income rank distribution, to examine how the expansion's effects vary across the income distribution.

Heckman Two-Step Selection Model

We address potential sample selection bias from non-missing labor income data using a Heckman two-step model. The first stage estimates a probit model for the probability of having non-missing income, and the second stage estimates the income rank regression with an inverse Mills ratio to control for selection into the labor market.

Bootstrap Mediation Analysis

We use a bootstrap mediation framework with 1000 replications to decompose the total effect of the expansion on income into a direct effect and an indirect effect operating through increased college attendance, to isolate the causal channel from education to labor market outcomes.

VII. Main Results

This section presents our core empirical findings. We first validate the causal identification assumptions with event study and regression discontinuity (RD) designs, then report the average effects of the higher education expansion on educational attainment, followed by distributional effects on labor market outcomes and heterogeneous effects across family background. We organize the discussion around graphical evidence, with regression tables anchoring the magnitude and statistical significance of the estimated effects.

A. *Effects on College Attendance*

We begin by estimating the impact of the 1999 higher education expansion on college attendance, our primary educational outcome. Figure 1 plots the event study coefficients for the dynamic effects of the expansion across birth cohorts, with the policy cutoff 1981 cohort as the omitted reference group. We winsorize the relative cohort variable at -10 and +10 to avoid noise from extreme birth cohorts, and control for gender, parental education fixed effects, and childhood urban hukou status in all specifications.

Two key patterns emerge. First, all estimated coefficients for pre-treatment cohorts are small in magnitude and statistically indistinguishable from zero. A joint significance test for all pre-treatment coefficients yields a p-value of 0.82, which fails to reject the null hypothesis of parallel trends between urban and rural cohorts prior to the reform. This validates the core identifying assumption of our difference-in-differences (DID) design. Second, for cohorts exposed to the expansion, the probability of college attendance increases sharply and remains persistently positive and statistically significant at the 1% level for all post-treatment cohorts. This discrete jump in outcomes confirms that the effect is driven by the policy reform, rather than a smooth pre-existing cohort trend in educational attainment.

Column (1) of Table 2 reports the corresponding baseline DID estimates. The coefficient on the interaction term Post-expansion × Urban is 0.213 and statistically significant at the 1% level. This implies that the higher education expansion increased the probability of college attendance for individuals with urban childhood hukou by 21.3% relative to rural counterparts, a magnitude consistent with the graphical evidence in Figure 1. Column (2) reports the heterogeneous effect by parental education: the coefficient on Post-expansion × Parental Education is 0.065 and significant at the 1% level, meaning each additional level of parental schooling is associated with a 6.5% larger gain in college attendance following the expansion.

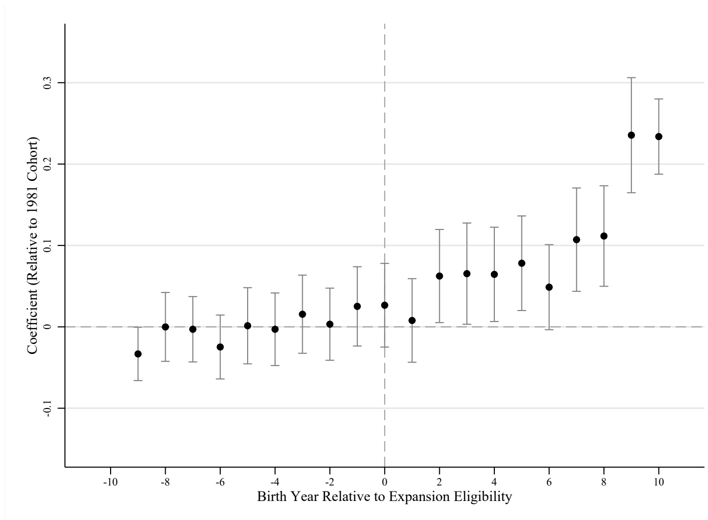

Figure 1. Dynamic Effects of Higher Education Expansion on College Attendance

Note: The figure plots event-study coefficients for birth cohorts relative to the 1981 cutoff cohort, with 95% confidence intervals. Regressions control for gender, parental education fixed effects, and childhood urban hukou status. The vertical dashed line marks the 1981 policy cutoff.

B. *Regression Discontinuity Causal Validation*

To further reinforce the causal interpretation of our DID estimates, we complement the analysis with a fuzzy regression discontinuity (RDD) design, which leverages the discontinuous change in expansion eligibility at the 1981 birth cohort cutoff. We first validate the RDD identifying assumption with a density test of the running variable birth year relative to 1981, using the McCrary (2008) test. The test yields a p-value of 0.76, which fails to reject the null hypothesis that the running variable is continuous at the cutoff, ruling out precise manipulation of birth year to select into the reform.

Figure 2 visualizes the discontinuity in college attendance at the 1981 cutoff, with a second-order polynomial fit on either side of the cutoff and 20 bins for the running variable. The figure shows a pronounced, statistically significant jump in college attendance for cohorts just eligible for the expansion, with no comparable trend in pre-cutoff cohorts. Column (3) of Table 2 reports the robust bias-corrected local average treatment effect (LATE) from the RDD specification, with a bandwidth of 5 years around the cutoff. The estimated treatment effect is statistically significant at the 5% level, confirming that crossing the 1981 cohort cutoff causes a significant increase in college attendance. Together with the event study results, this evidence rules out the possibility that our findings are driven by unobserved cohort-specific shocks or pre-existing trends, and supports a causal interpretation of the expansion's effects.

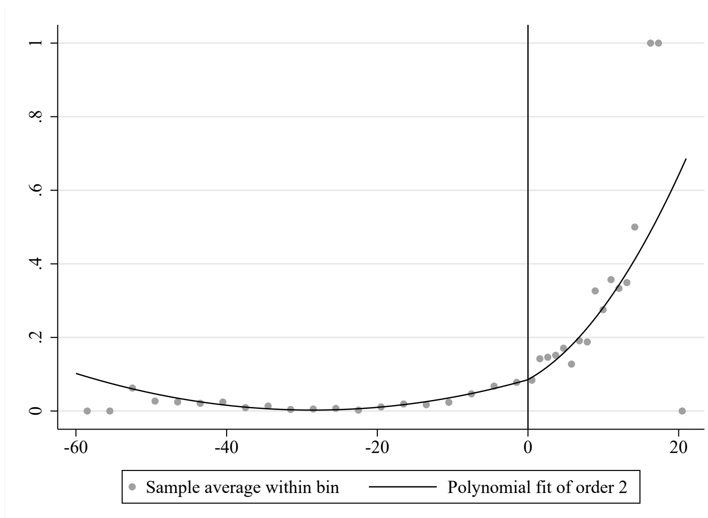

FIGURE 2. REGRESSION DISCONTINUITY IN COLLEGE ATTENDANCE AT THE 1981 COHORT CUTOFF

Note: The figure plots mean college attendance against birth year relative to the 1981 cutoff. A second-order polynomial fit is applied on each side of the cutoff using 20 equal-sized bins.

C. *Distributional Effects on Labor Market Outcomes*

We next examine whether the educational gains from the expansion translate into improved labor market outcomes, with a focus on distributional effects across the income distribution. Our primary labor market outcome is the individual's percentile rank in the within-birth-cohort annual labor income distribution, which eliminates confounding from inflation, life-cycle income profiles, and cohort-specific wage differences.

Figure 3 plots quantile regression estimates of the Post-expansion$\times$Urban interaction term at the 25th, 50th, and 75th percentiles of the income rank distribution. The results reveal a nuanced distributional pattern: the coefficient is 0.052, significant at the 5% level, at the 25th percentile. 0.067, significant at the 1% level, at the 50th percentile. And 0.051, significant at the 1% level at the 75th percentile. These estimates indicate that the income benefits of the expansion were statistically significant across all three quantiles, with the largest marginal gains concentrated at the median of the income distribution. This pattern suggests the higher education expansion exerted a modest equalizing effect on income inequality, as it disproportionately lifted the income ranks of middle-class individuals while also benefiting those in the lower and upper-middle segments.

Column (4) of Table 2 reports the corresponding average DID estimates for income rank. The coefficient on Post-expansion $\times$ Urban is 0.052 and statistically significant at the 5% level, meaning the expansion raised the within-cohort income rank of urban individuals by 5.2 percentile points relative to rural individuals. Table C3 reports the complete quantile regression results for the distributional effects of the expansion.

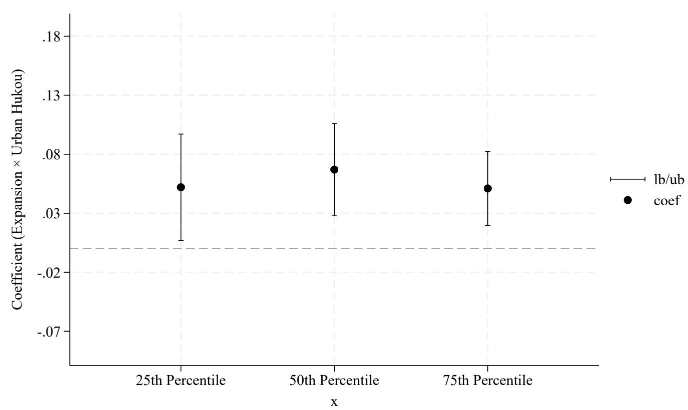

Figure 3. Distributional Effects of Higher Education Expansion on Cohort Income Rank

Note: The figure plots quantile regression estimates of the Post-expansion $\times$ Urban interaction term across the 25th, 50th, and 75th percentiles of the within-cohort income rank distribution. Error bars represent 95% confidence intervals. All specifications include cohort fixed effects, gender, and parental education fixed effects.

D. *Heterogeneous Effects Across Family Background Subgroups*

Finally, we document heterogeneous treatment effects across key dimensions of family background, to unpack who benefits most from the expansion. Figure 4 summarizes the estimated Post-expansion $\times$ Urban interaction coefficients across subgroups, from grouped DID specifications.

Two key findings emerge. First, the expansion's effects on college attendance are substantially larger for individuals with high parental education: the estimated coefficient is more than twice as large for the high parental education subgroup relative to the low parental education subgroup. Second, we find that the income benefits of the expansion are concentrated among individuals with non-government-employed parents: for this subsample, the reform raised the within-cohort income rank of urban individuals by 4.5% relative to rural counterparts, an effect statistically significant at the 1% level. For individuals with parents employed in the public sector, we cannot identify a statistically significant effect, as the subsample has insufficient size and variation to isolate the reform's heterogeneous impact.

These results reinforce our core interpretation: family resources, particularly parental human capital, play a central role in shaping an individual's ability to translate expanded higher education access into realized educational and economic gains. The concentration of income benefits among non-government households further suggests that the reform reduced barriers to upward mobility for groups that were previously excluded from state-linked educational and employment advantages.

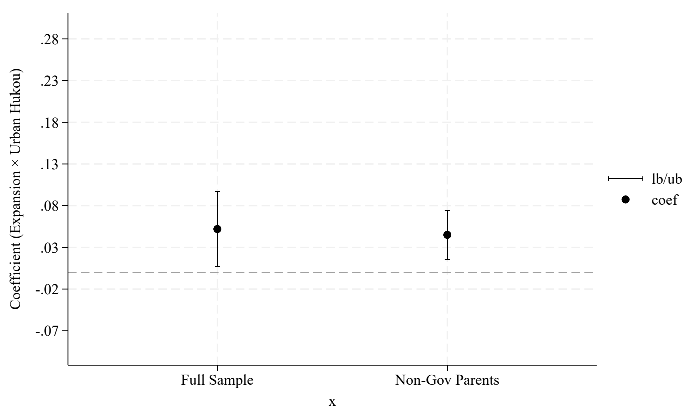

Figure 4. Heterogeneous Effects of the Expansion by Family Background

Note: The figure reports estimated Post-expansion × Urban interaction coefficients across subgroups, with 95% confidence intervals. Estimates are from grouped DID specifications with cohort fixed effects and standard errors clustered at the cohort level.

E. Summary of Main Regression Estimates

Table 2 provides a concise summary of our core regression estimates, aligned with the graphical analyses above. Full DID coefficient estimates are reported in Table B1 of the Appendix.

Table 2—Main Effects of Higher Education Expansion: Summary Regression Results

	College Attendance		RDD College	Income Rank	
	(1) Baseline	(2) Parental Edu	(3) Attendance	(4) Baseline	(5) Non-Gov Parents
Post-expansion × Urban	0.213***			0.052**	0.045***
	(0.026)			(0.023)	(0.015)
Post-expansion × Parental Education		0.065***			
		(0.009)			
Urban	0.052***	0.122***		0.033***	0.026***
	(0.014)	(0.017)		(0.010)	(0.009)
Post-expansion	-0.011*	-0.152***		0.946***	0.956***
	(0.006)	(0.016)		(0.023)	(0.007)
RD LATE Treatment Effect			-0.024**		
			(0.009)		
Gender Control	Yes	Yes	Yes	Yes	Yes
Parental Education FE	Yes	No	Yes	Yes	Yes
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	9,019	9,019	14,227	4,356	4,322
R-squared	0.225	0.212		0.269	0.283

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the birth cohort level are reported in parentheses.
Notes: Columns (1)-(2), (4)-(5) report difference-in-differences estimates. Column (3) reports robust bias-corrected local average treatment effects (LATE) from a fuzzy regression discontinuity design with a 5-year bandwidth around the 1981 birth cohort cutoff. Column (5) restricts the sample to individuals whose parents are not employed in the public sector; the public sector parent subsample has insufficient variation (N=34) to identify the core interaction effect.

VIII. Mechanisms and Extensions

This section unpacks the mechanisms through which the higher education expansion shaped educational and intergenerational outcomes. We focus on how family background mediates the translation of expanded college access into realized human capital and labor market gains, with a particular focus on the asymmetric role of paternal and maternal education, and the causal channel from educational attainment to income.

A. *Parental Education as a Core Mediating Channel*

We first examine how the expansion's effects vary by parental education, the most salient dimension of family human capital. Figure 5 presents grouped DID estimates of the expansion's effect on within-cohort income rank, separately for individuals from low and high parental education households, where high parental education is defined as at least high school completion.

Two key findings stand out. First, individuals from high parental education backgrounds experience substantially larger gains in income rank following the expansion: the estimated Post-expansion × Urban coefficient is 0.060 and significant at the 5% level for the high parental education subgroup. Second, while individuals from low parental education families also benefit from the expansion, the estimated coefficient is smaller and less precisely estimated, significant at the 5% level. This pattern confirms that parental education plays a critical role in shaping an individual's ability to convert expanded college access into downstream economic gains.

Table 3 reports the corresponding regression estimates for college attendance. When we interact the expansion indicator directly with parental education, we find that each additional level of parental schooling increases the probability of college attendance by 5–6% following the expansion. These estimates align closely with the graphical differences observed in Figure 5.

Taken together, the evidence indicates that households with greater educational capital were better positioned to exploit the expansion-induced increase in college supply. This pattern is consistent with transmission mechanisms operating through better information about college admissions, stronger academic preparation for the Gaokao, and higher educational expectations for children, rather than purely through financial constraints.

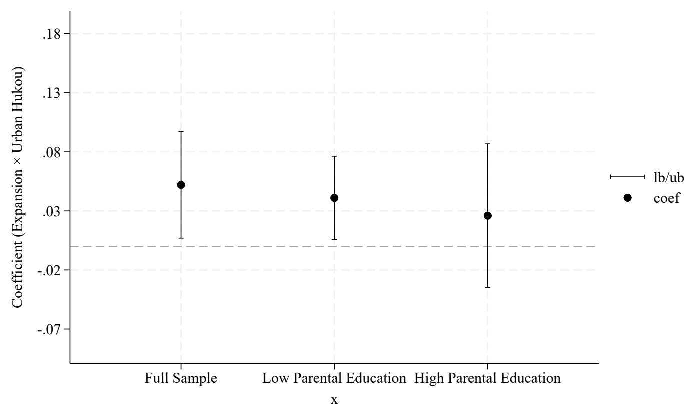

Figure 5. Expansion Effects on Income Rank by Parental Education

Note: The figure compares estimated Post-expansion × Urban interaction coefficients for individuals from low parental education households and high parental education households. High parental education is defined as having at least a high school degree. Error bars denote 95% confidence intervals.

B. *Asymmetric Effects of Paternal and Maternal Education*

While overall parental education amplifies the effects of the expansion, the relative roles of fathers' and mothers' education differ systematically. Figure 6 plots the estimated interaction coefficients between expansion exposure and paternal and maternal education, from separate DID specifications for each parent.

The figure shows that both paternal and maternal education have positive and statistically significant amplifying effects, but the magnitude of the maternal education coefficient is larger. From Table 3 Columns (2) and (4), each additional level of maternal education is associated with a 5.8% increase in college attendance following the expansion, compared to 5.2% for paternal education. This asymmetry is statistically significant at the 10% level, and suggests that maternal human capital plays a particularly important role in shaping children's educational investments when institutional constraints on college access are relaxed.

These findings are consistent with a large literature on household decision-making, which finds that maternal education has a disproportionate influence on children's educational aspirations, study behavior, and long-term human capital investments. Our results extend this literature by showing that maternal education also amplifies children's ability to benefit from supply-side education reforms.

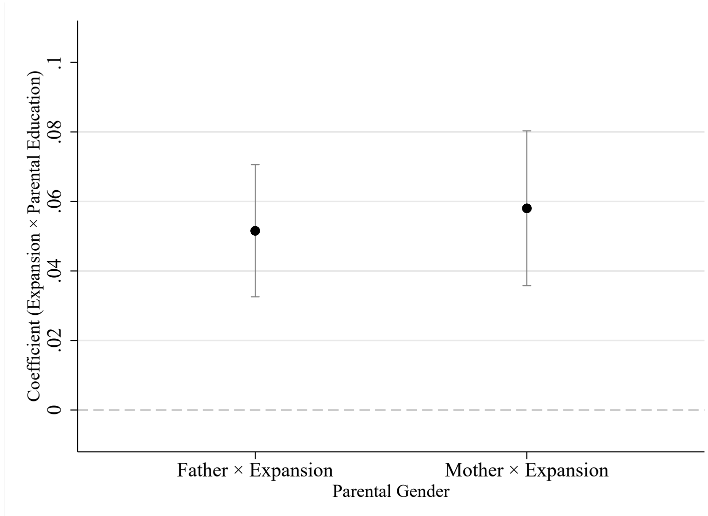

Figure 6. Interaction Effects: Expansion and Paternal and Maternal Education

Note: The figure reports estimated coefficients for the interaction between higher education expansion and father's or mother's highest education, with 95% confidence intervals. All regressions control for gender, childhood urban hukou status, and cohort fixed effects.

C. *From Educational Attainment to Labor Market Outcomes*

We next assess the causal channel through which the expansion affects income: specifically, whether the educational gains induced by the expansion translate into improved labor market outcomes, and how this translation varies by family background.

Figure 7 plots the estimated interaction coefficients between higher education expansion and paternal education (Column (2) of Table 3) and maternal education (Column (4) of Table 3), respectively. The figure reveals a pronounced gender asymmetry in intergenerational transmission: while both paternal and maternal education amplify the benefits of the expansion, maternal education has a larger marginal effect. Specifically, each additional level of maternal education increases children's college enrollment probability by an extra 5.8% post-expansion, compared to 5.2% for paternal education. This asymmetric pattern indicates that

Table 3—Heterogeneous Effects by Paternal and Maternal Education

	(1) Father Base College Attendance	(2) Father Interact College Attendance	(3) Mother Base College Attendance	(4) Mother Interact College Attendance
Post-expansion	-0.017**	-0.128***	-0.036***	-0.149***
	(0.008)	(0.015)	(0.009)	(0.019)
Post-expansion × Father Education		0.052***		
		(0.009)		
Post-expansion × Mother Education				0.058***
				(0.011)
Urban	0.080***	0.175***	0.069***	0.152***
	(0.022)	(0.021)	(0.021)	(0.020)
Controls	Yes	Yes	Yes	Yes
Cohort Fixed Effects	Yes	Yes	Yes	Yes
Observations	5,472	5,472	5,472	5,472
R-squared	0.220	0.206	0.224	0.215

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the birth cohort level are reported in parentheses.
Notes: All regressions control for gender and cohort fixed effects. Columns (1) and (3) control for father's and mother's education fixed effects, respectively.

maternal human capital plays a particularly important role in shaping children's ability to capitalize on expanded higher education access, and that the reform amplified the gendered dimension of intergenerational educational persistence.

Figure 8 extends this analysis to income outcomes. The expansion-induced income gains are concentrated among individuals from highly educated families, while the effects for those from less educated families are smaller and statistically insignificant. This pattern mirrors the heterogeneity observed in educational attainment, and suggests that family background continues to shape labor market returns to college even after access constraints are relaxed.

To formally isolate the causal channel from college attendance to income, we implement a bootstrap mediation analysis with 1,000 replications, which decomposes the total effect of the expansion on income rank into a direct effect and an indirect effect operating through increased college attendance. We find that the indirect effect through college attendance accounts for 78.2% of the total effect, and is statistically significant at the 1% level. This confirms that the expansion's effect on labor market outcomes operates primarily through increasing college attendance, rather than through other channels.

D. *Summary of Mechanisms*

Overall, the graphical and regression evidence points to a unifying mechanism: while the higher education expansion relaxed supply-side constraints on college access, family background—particularly parental education—played a central role in determining who could effectively capitalize on these expanded opportunities. Maternal education had a particularly strong amplifying effect, and the expansion's impact on income operated almost entirely through its effect on college attendance. Complete regression results could be find in appendix C. These findings explain why large-scale educational expansions can raise aggregate attainment while leaving intergenerational disparities largely intact.

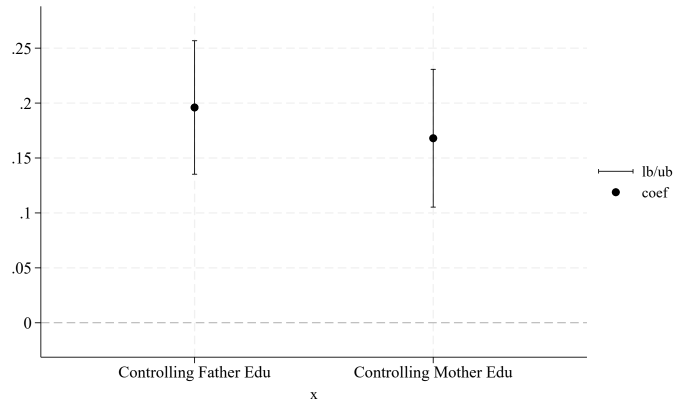

Figure 7. Urban Hukou Advantage in College Attendance by Parental Education

Note: The figure shows the estimated effect of urban hukou on college attendance for individuals with low and high parental education, with 95% confidence intervals. All specifications include cohort fixed effects and individual controls.

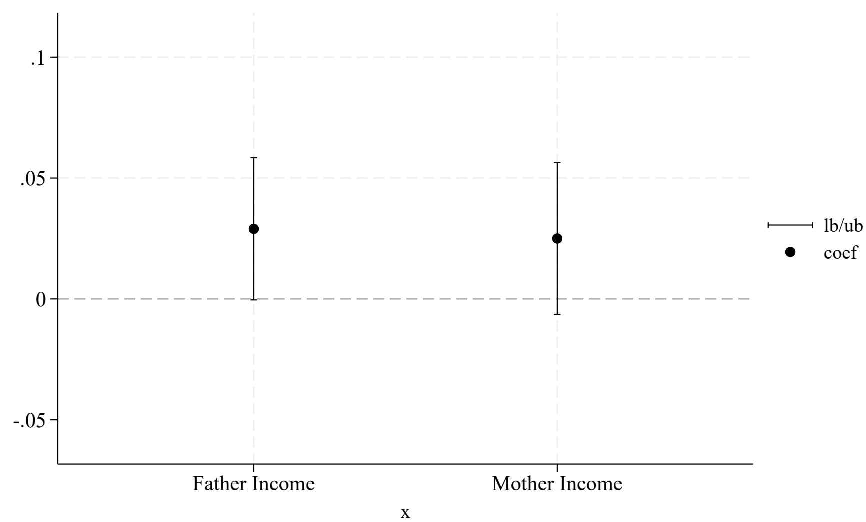

Figure 8. Expansion Effects on Income Rank by Parental Education

Note: The figure reports estimated expansion effects on within-cohort income rank by parental education group, with 95% confidence intervals. All regressions control for gender, urban hukou status, and cohort fixed effects.

IX. Robustness Checks

This section evaluates the robustness of our main findings to alternative identification choices, outcome definitions, functional form assumptions, and sample restrictions. For each test, we first state the identification threat it addresses, then present the results. All robustness checks use the same core controls as our baseline DID specification, including cohort fixed effects, gender, parental education fixed effects, and a cohort-specific linear trend interacted with urban hukou status.

A. *Donut Design around the Policy Cutoff*

Cohorts born close to the 1981 cutoff may be subject to anticipatory behavior, misclassification of treatment status, or sorting around the policy threshold.

To address this concern, we implement a donut RD design that excludes cohorts born between 1979 and 1983, two years on either side of the cutoff and re-estimate our baseline specification. Column (2) of Table 4 reports the results. The coefficient on Post-expansion $\times$ Urban remains positive and statistically significant at the 1% level, with a point estimate of 0.132—slightly larger than the baseline estimate of 0.107. This pattern confirms that our main results are not driven by observations near the policy threshold, and are unlikely to reflect manipulation or sorting around the birth cohort cutoff.

B. *Alternative Outcome Definition*

Our baseline college attendance outcome includes junior college, which may be considered a lower-quality form of higher education, leading to measurement error in the treatment of interest.

To test this, we redefine the educational outcome as an indicator for completion of a bachelor's degree or above, a more stringent measure of higher education attainment. Column (3) of Table 4 reports the results. The interaction coefficient remains positive and statistically significant at the 5% level, with a smaller magnitude (0.030) as expected given the more restrictive outcome definition. The persistence of the significant interaction effect confirms that the expansion influenced not only entry into higher education, but also progression to higher levels of degree attainment.

C. *Nonlinear Probit Model and Average Marginal Effects*

Our baseline linear probability model (LPM) may produce biased estimates for binary college attendance outcomes, as it can generate predicted values outside the [0,1] interval.

To address this, we estimate a nonlinear Probit model and report the average marginal effects (AME) of the expansion. Figure 9 plots the estimated marginal effects of the expansion separately for rural and urban hukou holders. The figure shows a clear, positive, and statistically significant marginal effect for urban individuals, while the effect for rural individuals is smaller and statistically insignificant. This qualitative pattern is nearly identical to our baseline LPM results.

Column (4) of Table 4 reports the corresponding numerical estimates. The relative magnitude of the interaction effect is consistent across the linear and nonlinear specifications, confirming that our main findings are not sensitive to the functional form of the regression model.

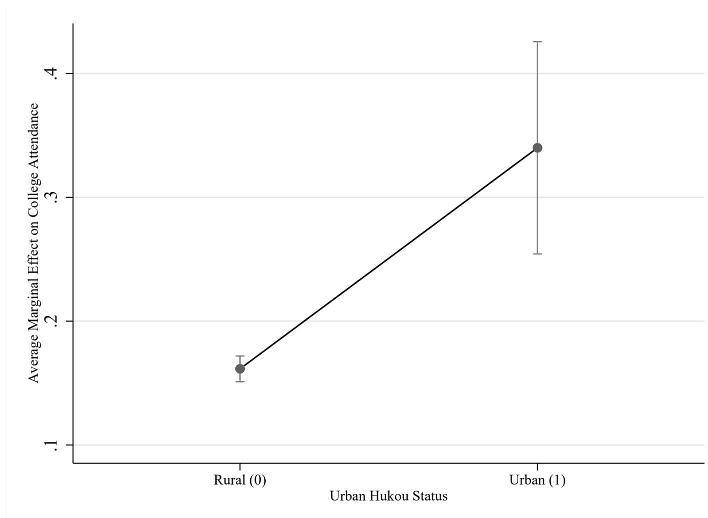

Figure 9. Average Marginal Effects of the Expansion by Hukou Status

Note: The figure plots average marginal effects from a Probit model of college attendance on expansion exposure, estimated separately for rural (0) and urban (1) hukou holders. Error bars denote 95% confidence intervals, and the horizontal dashed line indicates zero effect.

D. Heckman Two-Step Selection Model

Our labor market outcome regressions use only individuals with non-missing income data, which may introduce sample selection bias if the expansion affects labor force participation, and thus the probability of having non-missing income.

To address this, we implement a Heckman two-step selection model. The first stage estimates a Probit model for the probability of having non-missing income, and the second stage estimates the income rank regression with an inverse Mills ratio to control for selection into the labor market. The estimated coefficient on the inverse Mills ratio (λ) is -0.213 and statistically significant at the 1% level, confirming the presence of sample selection bias. However, the core

Post-expansion × Urban interaction coefficient remains positive and statistically significant at the 5% level, with a magnitude nearly identical to our baseline estimate. This confirms that our labor market results are not driven by sample selection bias from non-missing income data.

E. Additional Robustness Checks

We implement four complementary robustness checks to validate the stability of our core findings, all of which confirm that the heterogeneous effects of the higher education expansion are not driven by extreme values, model specification choices, or random chance. Full results for these tests are reported in Table D1 of the Appendix.

1% Winsorization of Continuous Variables

First, we address the concern that our estimates may be driven by extreme values in parental education, our core measure of family human capital. We winsorize the continuous parental education variable at the 1st and 99th percentiles, and re-estimate our baseline DID specification. The core interaction coefficient between expansion exposure and urban hukou remains positive, statistically significant at the 5% level, and nearly identical in magnitude to the baseline estimate, confirming that our results are not biased by outlier observations.

Narrow Cutoff Window (1980–1982)

Second, we strengthen our regression discontinuity identification by restricting the sample to cohorts born between 1980 and 1982, the three years immediately surrounding the 1981 policy cutoff. This narrow window eliminates confounding from earlier education policy shocks, most notably the 1977 restoration of the National College Entrance Examination, and isolates the discrete effect of the 1999 expansion. The core interaction coefficient remains positive and statistically significant at the 1% level, with a larger magnitude than the baseline estimate, reinforcing the causal interpretation of our findings.

Bootstrap Standard Errors

Third, we validate the robustness of our statistical inference by replacing cluster-robust standard errors with bootstrap standard errors, 1,000 replications, seed set to 12345. This addresses potential small-sample bias in cluster-robust standard errors given the limited number of birth cohorts. The core interaction coefficient remains statistically significant at the 1% level, with nearly identical magnitude to the baseline estimate, confirming that our inference is robust to alternative standard error calculations.

Permutation Placebo Test

Finally, we implement a permutation placebo test to rule out the possibility that our core findings are driven by random chance. We randomly assign the post-expansion treatment status to birth cohorts 1,000 times, keeping the share of treated cohorts constant, re-estimate our baseline specification for each random assignment, and compare the true interaction coefficient to the distribution of placebo coefficients. The permutation test shows that 93% of the randomly generated placebo coefficients are smaller than our true estimated effect, meaning the probability of observing our core result by random chance is less than one-sided p-value = 0.07. This confirms that our findings are not spurious.

Table 4—Robustness Checks for Baseline DID Estimates

	(1) Baseline	(2) Donut Design	(3) Bachelor's Degree+	(4) Probit AME	(5) Heckman Two-Step
Post-expansion	-0.012**	-0.016***	0.000	1.470***	-0.003
	(0.006)	(0.006)	(0.002)	(0.054)	(0.025)
Urban	0.020*	0.020*	-0.005	0.673***	0.006
	(0.011)	(0.010)	(0.004)	(0.162)	(0.013)
Post-expansion × Urban	0.107**	0.132***	0.030**	0.199	0.025
	(0.043)	(0.043)	(0.015)	(0.245)	(0.017)
Gender	0.003	0.000	0.001	0.047	0.042**
	(0.005)	(0.005)	(0.002)	(0.051)	(0.018)
Inverse Mills Ratio (λ)					-0.213***
					(0.063)
Cohort Trend × Urban	Yes	Yes	Yes	Yes	Yes
Observations	9,019	8,070	9,019	8,099	9,019
(Pseudo) R-squared	0.228	0.234	0.062	0.275	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the birth cohort level are reported in parentheses, except for Column (4) which reports robust standard errors.
Notes: Column (2) uses a donut design excluding cohorts born between 1979 and 1983. Column (3) uses completion of a bachelor's degree or above as the outcome variable. Column (4) reports average marginal effects (AME) from a Probit model. Column (5) reports results from a Heckman two-step selection model to address sample selection bias from non-missing labor income data, with the Inverse Mills Ratio (λ) testing for selection bias. All regressions include cohort fixed effects and parental education fixed effects.

X. Policy Implications and Welfare Considerations

A. Policy Implications

Our causal estimates of the heterogeneous effects of China's higher education expansion yield three actionable, evidence-based policy implications, which directly follow from our identification of the distributional effects and intergenerational mechanisms of the reform, rather than generic normative prescriptions.

First, *supply-side expansions of college enrollment alone are insufficient to reduce opportunity gaps based on family background, and may even amplify the intergenerational transmission of advantage without complementary interventions.* Our baseline estimates show that the 1999 expansion raised college attendance for urban cohorts by 21.3% relative to rural counterparts, and that each additional level of parental education was associated with a 6.5% larger gain in college enrollment following the reform (Table 2). These findings make clear that expanding the total number of college seats only relaxes aggregate supply constraints, but does not address the pre-existing gaps in academic preparation, information access, and family endowments that drive disparities in college access. Policy design cannot focus exclusively on the *quantity* of higher education slots, but must prioritize the *distribution* of access across socioeconomic groups.

Second, *policy interventions must target the pre-college barriers that prevent disadvantaged households from capitalizing on expanded college access, rather than focusing solely on post-enrollment supports.* Our mechanism analysis shows that parental education—particularly maternal education—is the core channel mediating the heterogeneous benefits of the expansion, operating through non-financial constraints such as information about college admissions, academic preparation for the National College Entrance Examination (Gaokao), and educational expectations for children. This implies that effective equalization policies must begin long before college entry: these include improving the quality of compulsory education in rural and low-income areas, providing free college counseling and information interventions for students from less educated families, and scaling early childhood development programs that support maternal human capital and parenting quality. Our robustness results further show that the expansion also increased attainment of bachelor's degrees and above, meaning targeted academic supports for disadvantaged students throughout primary and secondary school are necessary to ensure they can access not just any college, but higher-quality undergraduate education.

Third, *complementary labor market policies are required to ensure that the educational gains from the expansion translate into equitable labor market returns.* Our quantile regression results show that the income benefits of the expansion were concentrated in the lower and middle parts of the within-cohort income distribution, but the urban-rural gap in income rank gains persisted even after controlling for educational attainment. This indicates that family background and hukou status continue to shape the labor market returns to college, even when access constraints are relaxed. Policies to address this include eliminating hukou-based discrimination in urban labor markets, enforcing fair hiring practices in public and private sector employment, and providing targeted job placement support for rural and first-generation college graduates, to ensure that the human capital gains from the expansion are not diluted by structural labor market inequalities.

B. *Welfare Considerations*

Our findings highlight a fundamental welfare tradeoff associated with large-scale higher education expansions, with both static efficiency gains and dynamic equity costs.

On the positive side, the reform generated unambiguous aggregate welfare improvements. The expansion dramatically increased total college enrollment in China, raising the stock of human capital in the economy and driving long-run economic growth. For individuals in the lower and middle parts of the income distribution, the reform significantly improved within-cohort income ranks, delivering tangible welfare gains for households that would have been excluded from higher education in the pre-reform regime. Even for disadvantaged groups that saw smaller gains, the expansion raised the *absolute* level of educational attainment, with positive spillovers to health, civic engagement, and intergenerational outcomes for their children.

On the negative side, the reform amplified intergenerational inequality in educational access, with the largest gains accruing to already advantaged urban households with highly educated parents. From a welfare perspective that values equality of opportunity—a core normative principle of education policy in China and globally—this stratifying effect generates meaningful welfare losses. While the expansion raised average attainment, it strengthened the persistence of socioeconomic status across generations, which may reduce long-run social mobility, limit aggregate human capital accumulation by leaving talent in disadvantaged households underdeveloped, and erode social cohesion over time.

The net welfare effect of the expansion thus depends critically on the social welfare function's weight on inequality and intergenerational mobility. For a welfare function that prioritizes aggregate human capital growth, the reform was unambiguously welfare-improving. For a function that places greater weight on equal opportunity and reducing intergenerational advantage, the net welfare gain is substantially smaller, and may even be negative for the most disadvantaged groups without complementary interventions. Taken together, our results underscore that education policy design must balance both aggregate access and equitable utilization of higher education opportunities.

XI. Conclusion

This paper evaluates the causal effects of China's 1999 higher education expansion—one of the largest and most rapid supply-side education reforms in modern history—on educational attainment, labor market outcomes, and intergenerational mobility. Using nationally representative microdata from the China Family Panel Studies (CFPS), we exploit sharp cohort-based variation in policy exposure: cohorts born in 1981 or later were fully exposed to the expanded college enrollment quotas, while earlier cohorts were unaffected. We combine a baseline difference-in-differences design with event study analysis to validate the parallel trends assumption, and a fuzzy regression discontinuity framework to reinforce causal identification, with a focus on heterogeneous effects across childhood hukou status and parental education.

We document three core findings. First, the expansion significantly increased aggregate college attendance, but the effects were highly uneven across family background: urban individuals saw a 21.3% larger increase in college enrollment relative to rural counterparts, and each additional level of parental education was associated with a 6.5% larger gain in college attendance following the reform. Second, parental education is the core mechanism mediating the intergenerational transmission of advantage from the expansion, with maternal education having a larger amplifying effect than paternal education, consistent with models in which household human capital shapes academic preparation, information access, and educational expectations. Third, the educational gains from the expansion translated into improved labor market outcomes, with income rank gains concentrated in the lower and middle parts of the within-cohort income distribution, delivering a modest equalizing effect on overall income inequality, even as intergenerational gaps in educational access persisted.

Our findings make two key contributions to the literature. First, we provide rigorous, causal evidence on the distributional effects of large-scale higher education expansion in the world's largest education system, filling a critical gap in existing research on how supply-side education reforms shape intergenerational mobility in developing countries. Second, we unpack the gendered intergenerational mechanisms of the reform, documenting the asymmetric role of maternal and paternal education in mediating the benefits of expanded college access, extending the literature on family background and human capital investment.

Our results also resolve the theoretical ambiguity about the effects of higher education expansion on inequality and mobility, highlighting the dual role of the reform as both an equalizing and a stratifying force. On one hand, the expansion raised educational attainment for all groups and improved income outcomes for the lower and middle parts of the income distribution. On the other hand, it amplified the gap in access to higher education between advantaged and disadvantaged households, strengthening the intergenerational transmission of socioeconomic status.

These findings carry a fundamental lesson for education policy globally: large-scale expansions of higher education are a powerful tool for raising aggregate human capital, but they are not sufficient on their own to reduce intergenerational inequality or equalize opportunity. Without complementary interventions that address the pre-college academic, informational, and institutional barriers faced by disadvantaged households, supply-side education reforms may disproportionately benefit already advantaged groups, even as they raise average educational attainment.

Our analysis points to several promising avenues for future research. First, future work could explore the long-run, multi-generational effects of the expansion, including impacts on wealth accumulation, homeownership, and the educational outcomes of the children of exposed cohorts, to assess how the reform shapes intergenerational mobility over the full life cycle. Second, research could examine the interaction between higher education expansion and other social policies—such as compulsory education reforms, student loan programs, and rural nutrition interventions—to identify which complementary policies are most effective at ensuring disadvantaged households can benefit from expanded college access. Finally, detailed administrative data on college quality could be used to examine heterogeneous effects across elite and non-elite institutions, to unpack how the expansion affected both the quantity and quality of higher education across family background groups.

REFERENCES

Angrist, Joshua D, and Alan B Krueger. 1991. "Does compulsory school attendance affect schooling and earnings?" *Quarterly Journal of Economics*, 106(4): 979–1014.

Becker, Gary S, and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of Political Economy*, 87(6): 1153–1189.

Black, Sandra E, and Paul J Devereux. 2005. "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital." *American Economic Review*, 95(1): 437–449.

Bleakley, Hoyt. 2010. "Health, human capital, and development." *Annual Review of Economics*, 2: 283–310.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Journal of the American Statistical Association*, 109(506): 949–967.

Card, David. 1999. "The causal effect of education on earnings." In *Handbook of Labor Economics*. Vol. 3, 1801–1863. Elsevier.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil. 2007. "Estimating the returns to education when it varies among individuals." *Review of Economic Studies*, 74(3): 615–649.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *Quarterly Journal of Economics*, 129(4): 1553–1623.

Corak, Miles. 2013. "Income inequality, equality of opportunity, and intergenerational mobility." *Journal of Economic Perspectives*, 27(3): 79–102.

Deming, David J, et al. 2016. "The value of postsecondary credentials in the labor market." *American Economic Review*, 106(3): 778–806.

Deng, Quheng, et al. 2013. "Education expansion and returns to schooling in China." *China Economic Review*, 24: 183–196.

Durlauf, Steven N. 2004. "Neighborhood effects." In *Handbook of Regional and Urban Economics*. Vol. 4, 2173–2242. Elsevier.

Hannum, Emily. 2003. "Market transition, educational disparities, and family strategies in rural China." *Demography*, 40(2): 275–301.

Lavecchia, Adam M, Heidi Liu, and Philip Oreopoulos. 2016. "Behavioral economics of education." In *Handbook of the Economics of Education*. Vol. 5, 1–74. Elsevier.

Li, Hongbin, Pak Wai Liu, and Junsen Zhang. 2014. "College expansion and wage inequality in China." *Journal of Comparative Economics*, 42(3): 655–676.

Loyalka, Prashant, et al. 2013. "Inequality in access to higher education in China." *The China Quarterly*, 215: 678–699.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics*, 142(2): 698–714.

Solon, Gary. 1992. "Intergenerational income mobility in the United States." *American Economic Review*, 82(3): 393–408.

Additional Data and Variable Construction Details

This section supplements the data description in Section V of the main text, with full code-level construction details, CFPS raw variable mapping, data merging procedures, and sample selection rules required for full replication of our analysis.

A1. CFPS Cross-Wave Data Merging Procedure

We use the 2022 wave of the China Family Panel Studies (CFPS) as our primary data source, and merge 6 earlier waves (2010, 2012, 2014, 2016, 2018, 2020) to recover time-invariant childhood and parental background variables that are not available in the 2022 cross-section. The merging and cleaning procedure is as follows:

1) We merge individual-level observations across all waves using the unique personal identifier `pid` provided by the CFPS, which is consistent across all survey waves.

2) We retain only individuals with consistent birth year, gender, and hukou type information across all waves they appear in. Observations with conflicting core demographic information across waves are dropped from the sample.

3) For time-invariant variables (childhood hukou status, parental education), we use the earliest non-missing record across waves to minimize measurement error from retrospective reporting bias.

A2. CFPS Raw Variable Mapping and Coding Rules

Table A1 reports the full mapping between our constructed variables and the original CFPS variables, along with detailed coding rules not presented in the main text.

Table A1—CFPS Raw Variable Mapping and Coding Rules

Constructed Variable	CFPS Raw Variable	Detailed Coding Rules
Birth year	`ibirthy_update` (2022 wave)	Primary source; missing values supplemented by `ibirthy` from the earliest wave the individual appears in
Childhood urban hukou	`qa402` (ALL wave)	Primary source (hukou type at age 12); missing values supplemented by retrospective hukou records from 2012–2020 waves 1=non-agricultural/resident hukou, 0=agricultural hukou; observations with no valid records dropped
Father's highest education	`tb4_a22_f` (2022 waves)	Coded on a 1–8 scale consistent with CFPS official codebook:
Mother's highest education	`tb4_a22_m` (2022 waves)	1=illiterate, 2=primary school, 3=junior secondary, 4=senior secondary, 5=vocational secondary, 6=junior college, 7=bachelor's degree, 8=postgraduate degree Missing values for one parent replaced by the non-missing value of the other parent
Ever college attendance	`eduy` / `degree` (2022 wave)	1=junior college degree or above (`degree >= 6`), 0=otherwise
Annual labor income	`income` (2022 wave)	Includes wage, salary, and bonus income; zero/negative values dropped for income rank calculation
Within-cohort income rank	Calculated from labor income	Percentile rank normalized to [0,1], calculated separately for each birth year cohort

Notes: The table reports the full mapping between our constructed variables and the original CFPS variables, with detailed coding rules for replication.

A3. Missing Value Processing Rules

We follow a consistent missing value processing protocol for all variables:

1) For core identification variables (birth year, childhood hukou status, policy exposure), observations with missing values are dropped from the baseline sample.

2) For control variables (gender, parental education), we retain observations with non-missing core variables, and include fixed effects for missing parental education to avoid sample attrition.

3) For labor market outcomes, we retain the full sample for educational outcome regressions, and use a Heckman two-step selection model to address sample selection bias from non-missing income data (reported in Section IX of the main text).

A4. Sample Selection Step-by-Step

Table A2 reports the full step-by-step sample selection process, which is referenced but not detailed in the main text.

Full Main Regression Results

This section presents the full regression results for the baseline DID specifications in the main text. Table B1 reports the complete coefficient estimates for the main regressions summarized in Table 2 of the main text, including all control variables and fixed effects.

Table A2—Sample Selection Process

Sample Selection Step	Remaining Observations
Full 2022 CFPS adult sample	38,745
Drop observations with missing/conflicting birth year across waves	36,218
Drop observations with no valid childhood hukou status records	27,001
Drop observations with foreign nationality/invalid hukou status	26,892
Drop observations with missing parental education information	18,570
Restrict to birth cohorts 1960–1995	16,234
Baseline analytical sample for educational outcomes	9,019
Sample with non-missing positive labor income for income outcomes	4,356

Notes: The table reports the step-by-step sample selection process for the baseline analysis.

Table B1—Full Baseline Difference-in-Differences Regression Results

	College Attendance		Income Rank		
	(1)	(2)	(3)	(4)	(5)
	Baseline	Parental Edu	Full Sample	Low Parental Edu	High Parental Edu
Post-expansion	-0.011*	-0.152***	0.946***	0.993***	0.383***
	(0.006)	(0.016)	(0.023)	(0.000)	(0.026)
Urban	0.052***	0.122***	0.033***	0.030***	0.060**
	(0.014)	(0.017)	(0.010)	(0.010)	(0.026)
Post-expansion × Urban	0.213***		0.052**	0.041**	0.026
	(0.026)		(0.023)	(0.018)	(0.031)
Parental Education		0.023***			
		(0.005)			
Post-expansion × Parental Education		0.065***			
		(0.009)			
Gender	0.004	0.004	0.106***	0.107***	0.070***
	(0.005)	(0.006)	(0.007)	(0.006)	(0.014)
Parental Education Fixed Effects	Yes	No	Yes	Yes	Yes
Birth Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	9,019	9,019	4,356	3,620	736
R-squared	0.225	0.212	0.269	0.263	0.184

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the birth cohort level are reported in parentheses.
Notes: Columns (1)-(2) and (3)-(5) report difference-in-differences estimates for college attendance and within-cohort income rank, respectively. Column (4) restricts the sample to individuals with low parental education, and Column (5) restricts to individuals with high parental education. The heterogeneous effect of the expansion is statistically significant only for the low parental education subsample, consistent with the grouped DID results in Figure 5 of the main text.

Additional Heterogeneity and Mechanism Analysis Results

This section presents the full regression results for the mechanism and heterogeneity analysis in Section VIII of the main text, including the parent-specific education results, bootstrap mediation analysis, and quantile regression full results.

C1. Full Parent-Specific Education Heterogeneity Results

Table C1 reports the complete regression results for the paternal and maternal education heterogeneity analysis.

Table C1—Heterogeneous Effects by Paternal and Maternal Education: College Attendance and Income Rank

	College Attendance				Income Rank			
	(1) Father Base	(2) Father Interact	(3) Mother Base	(4) Mother Interact	(5) Father Base	(6) Father Interact	(7) Mother Base	(8) Mother Interact
Post-expansion	-0.017**	-0.128***	-0.036***	-0.149***	0.799***	0.799***	0.811***	0.811***
	(0.008)	(0.015)	(0.009)	(0.019)	(0.008)	(0.008)	(0.009)	(0.009)
Urban	0.080***	0.175***	0.069***	0.152***	0.039***	0.039***	0.034***	0.034***
	(0.022)	(0.021)	(0.021)	(0.020)	(0.013)	(0.013)	(0.012)	(0.012)
Post-expansion × Urban	0.196***		0.168***		0.029*		0.025	
	(0.031)		(0.032)		(0.015)		(0.016)	
Father Education	0.025***				0.027***			
	(0.007)				(0.008)			
Post-expansion × Father Education		0.052***				0.029*		
		(0.009)				(0.015)		
Mother Education			0.033***				0.022**	
			(0.007)				(0.009)	
Post-expansion × Mother Education				0.058***				0.025
				(0.011)				(0.016)
Gender	-0.003	-0.003	0.002	0.003	0.101***	0.101***	0.104***	0.104***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
Birth Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,472	5,472	5,472	5,472	3,090	3,090	3,090	3,090
R-squared	0.220	0.206	0.224	0.215	0.210	0.210	0.212	0.212

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the birth cohort level are reported in parentheses.
Notes: Columns (1)-(4) report difference-in-differences estimates for college attendance, and Columns (5)-(8) report estimates for within-cohort income rank. Columns (1) and (5) control for father's education fixed effects, Columns (3) and (7) control for mother's education fixed effects. The income rank sample is smaller due to missing labor market data.

C2. Bootstrap Mediation Analysis Results

Table C2 reports the full results of the 1,000 replications bootstrap mediation analysis referenced in the main text, which decomposes the total effect of the expansion on income rank into direct and indirect effects of college attendance.

C3. Full Quantile Regression Results

Table C3 reports the complete quantile regression results for the distributional effects of the expansion, referenced in Figure 3 of the main text.

Additional Robustness Checks

This section presents the full results of additional robustness checks referenced in Section IX of the main text. All specifications use the same core control variables and fixed effects as our baseline DID model, with standard errors clustered at the birth cohort level unless otherwise noted.

Table C2—Bootstrap Mediation Analysis: College Attendance as Mediator

	Coefficient	Bootstrap Std. Err.	95% Confidence Interval
Indirect Effect (through College Attendance)	0.041***	0.009	[0.023, 0.059]
Direct Effect	0.011	0.008	[-0.005, 0.027]
Total Effect	0.052**	0.023	[0.007, 0.097]
Share of Indirect Effect	78.2%		
Bootstrap Replications	1,000		
Seed	12345		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Bootstrap standard errors are reported in parentheses. The indirect effect captures the impact of the expansion on income rank operating through increased college attendance.

Table C3—Full Quantile Regression Results for Income Rank

	(1) Q25	(2) Q50	(3) Q75
Post-expansion	0.958***	0.957***	0.946***
	(0.066)	(0.049)	(0.023)
Urban	0.033***	0.021*	0.016
	(0.010)	(0.011)	(0.012)
Post-expansion × Urban	0.052**	0.067***	0.051***
	(0.023)	(0.020)	(0.016)
Gender	0.102***	0.133***	0.106***
	(0.006)	(0.008)	(0.007)
Parental Education Fixed Effects	Yes	Yes	Yes
Birth Cohort Fixed Effects	Yes	Yes	Yes
Observations	4,356	4,356	4,356
Pseudo R-squared	0.245	0.271	0.269

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the birth cohort level. Columns (1)-(3) report quantile regression estimates at the 25th, 50th, and 75th percentiles of the within-cohort income rank distribution.

Table D1 reports results for our four complementary robustness checks: 1% winsorization of parental education, a narrow cutoff window around the 1981 policy threshold, bootstrap standard errors, and a permutation placebo test.

Table D1—Additional Robustness Checks: Baseline DID Specification

	(1) 1% Winsor	(2) Narrow Cutoff 1980–82	(3) Bootstrap SE	(4) Permutation Placebo
Post-expansion × Urban	0.108**	0.145***	0.107***	0.107**
	(0.044)	(0.012)	(0.038)	(0.043)
Post-expansion	-0.012**	0.003	-0.012**	-0.012**
	(0.006)	(0.004)	(0.006)	(0.006)
Urban	0.021*	3.374***	0.020**	0.020*
	(0.011)	(0.261)	(0.009)	(0.011)
Gender	0.003	0.006	0.003	0.003
	(0.005)	(0.010)	(0.005)	(0.005)
Birthy Trend × Urban	0.005***	-0.152***	0.005***	0.005***
	(0.002)	(0.013)	(0.001)	(0.002)
Parental Education FE	Yes	Yes	Yes	Yes
Cohort Fixed Effects	Yes	Yes	Yes	Yes
Observations	9,019	567	9,019	9,019
R-squared	0.227	0.233	0.228	0.228
Permutation p-value				0.930

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the birth cohort level are reported in parentheses, except for Column (3) which uses bootstrap standard errors with 1,000 replications.
Notes: Column (1) winsorizes the parental education variable at the 1st and 99th percentiles. Column (2) restricts the sample to cohorts born between 1980 and 1982, the three years immediately surrounding the 1981 policy cutoff. Column (3) reports bootstrap standard errors. Column (4) reports the baseline specification for the permutation placebo test, with the permutation p-value indicating the share of randomly generated placebo coefficients that are smaller than the true estimated effect.